

Introduction to Persistent Homology

Žiga Virk

2022

Univerza v Ljubljani
Fakulteta za računalništvo in informatiko

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani

[COBISS.SI-ID=99057667](#)

ISBN 978-961-7059-10-6 (PDF)

Digitalna izdaja je prosto dostopna
This digital publication is freely available
<http://zalozba.fri.uni-lj.si/virk2022.pdf>
DOI: [10.51939/0002](https://doi.org/10.51939/0002)

Recenzenta / Reviewers:
prof. dr. Nežka Mramor Kosta, doc. dr. Boštjan Gabrovšek

Založnik: Založba UL FRI, Ljubljana
Izdajatelj: UL Fakulteta za računalništvo in informatiko, Ljubljana
Urednik: prof. dr. Franc Solina

Copyright © 2022 Založba UL FRI. All rights reserved.

Preface

In the past few decades persistent homology has emerged as one of the focal points of modern topology. While its beginnings have been motivated by practical demand in computational geometry, the scope of persistence soon expanded. Results focused on topological and algebraic fundamentals have been complemented by algorithmic and later statistical development leading to successful applications in sciences. Today, persistent homology is a wide ranging topic engaging a diverse community of mathematicians, computer scientists, data analysts and scientists in general.

This textbook does not aim to encompass the endless variety of topics related to persistence. The main goal is to provide students with a manageable, geometrically intuitive and self contained introduction to persistent homology. The writing arose from the lecture notes on the course that has been taught at the University of Ljubljana for more than a decade. It is intended for a mixed audience of mathematics and computer science students at the masters level. However, any motivated student with scientific or technical inclination should (hopefully) find it accessible.

As a prerequisite only basic linear algebra (including Gaussian elimination) and basic Euclidean geometry are assumed. The textbook is structured so that it gradually introduces fundamental structures. Simplicial complexes and Euler characteristic are first defined in the plane, before generalizing them to the abstract setting. Orientation is first introduced on surfaces, before utilizing it for homology computations. Required algebra is summarized in Chapter 6 to ensure a familiar algebraic footing.

Side notes throughout the text are intended to clarify certain ideas, provide some explanation, or to convey a geometric idea without interrupting the flow of the main text. At the end of each chapter, a reader may find a short comment on the background, along with a few references, keywords, and appendices with further topics or additional material.

Contents

1	<i>Metric spaces</i>	11
	1.1 <i>Definition of metric spaces and basic examples</i>	11
	1.2 <i>Maps and equivalence types</i>	13
	<i>Connectedness</i>	16
2	<i>Planar triangulations</i>	19
	2.1 <i>Definition of planar triangulations</i>	19
	<i>Modifications of triangulations</i>	20
	2.2 <i>Recap on convexity</i>	20
	2.3 <i>Euler characteristic</i>	22
	2.4 <i>Constructing planar triangulations with line sweep</i>	23
	2.5 <i>Voronoi diagram and Delaunay triangulation</i>	24
	<i>Local Delaunay condition</i>	25
	<i>Construction of $\mathcal{D}(S)$</i>	27
	2.6 <i>Concluding remarks</i>	29
	<i>Appendix: Proof of Proposition 2.5.6</i>	30
3	<i>Simplicial complexes</i>	33
	3.1 <i>Affine independence</i>	33
	3.2 <i>Geometric simplicial complex</i>	35
	3.3 <i>Abstract simplicial complex</i>	37
	<i>Two invariants</i>	40

3.4	<i>Simplicial maps</i>	42
	<i>Elementary collapses</i>	43
3.5	<i>Concluding remarks</i>	45
	<i>Appendix: Proof of Theorem 3.3.4</i>	45
4	<i>Surfaces</i>	47
4.1	<i>Surfaces as manifolds</i>	47
	<i>Combinatorial manifolds</i>	49
4.2	<i>Orientability</i>	50
4.3	<i>Connected sum of surfaces</i>	52
4.4	<i>Classification of surfaces</i>	53
	<i>General surfaces</i>	54
4.5	<i>Concluding remarks</i>	55
	<i>Appendix: imagining S^3</i>	56
5	<i>Constructions of simplicial complexes</i>	59
5.1	<i>Rips complexes</i>	59
5.2	<i>Čech complexes</i>	61
5.3	<i>Nerve complexes</i>	62
	<i>Alpha complexes</i>	64
	<i>Mapper</i>	65
5.4	<i>Interleaving properties</i>	67
	<i>Rips-Čech correlation</i>	69
5.5	<i>Concluding remarks</i>	69
	<i>Appendix: the MiniBall algorithm</i>	70
	<i>Appendix: a sketch of a proof of the nerve theorem 5.3.2</i>	71
	<i>Appendix: Dowker duality</i>	73
6	<i>Fields and vector spaces</i>	75
6.1	<i>Fields</i>	75
	<i>The fields of remainders \mathbb{Z}_p</i>	76

6.2	<i>Vector spaces</i>	78
6.3	<i>Concluding remarks</i>	81
	<i>Appendix: A very short introduction to Abelian groups</i>	81
7	<i>Homology: definition and computation</i>	85
7.1	<i>Definition</i>	85
	<i>Chains</i>	85
	<i>Boundary</i>	86
	<i>Homology</i>	88
	<i>Zero-dimensional homology</i>	90
	<i>Homology of a graph</i>	91
7.2	<i>Computing homology</i>	92
	<i>Echelon forms</i>	92
	<i>Smith normal form and representatives</i>	94
	<i>Incremental expansion and elementary collapse</i>	96
7.3	<i>Examples of homology</i>	97
	<i>Disjoint unions</i>	97
	<i>Euler characteristic</i>	97
	<i>Spheres</i>	98
	<i>Surfaces</i>	98
	<i>Impact of coefficients: the Klein bottle</i>	100
	<i>Alexander duality</i>	101
7.4	<i>Concluding remarks</i>	102
	<i>Appendix: Homology with coefficients in Abelian groups</i>	103
	<i>Appendix: cubical homology</i>	106
8	<i>Homology: impact and computation by parts</i>	109
8.1	<i>Impact</i>	109
	<i>Functionality of homology</i>	110
	<i>Brouwer fixed point</i>	111
	<i>Hairy ball</i>	112
	<i>Invariance of domain</i>	113

<i>8.2 Homology by parts</i>	114
<i>Exact sequences</i>	114
<i>Mayer-Vietoris exact sequence</i>	115
<i>8.3 Concluding remarks</i>	116
<i>Appendix: zig-zag lemma</i>	117
<i>Appendix: Relative homology</i>	119
<i>9 Persistent homology: definition and computation</i>	121
<i>9.1 Definition</i>	121
<i>Formal definition</i>	123
<i>9.2 Visualization</i>	125
<i>Barcodes</i>	125
<i>Persistence diagrams</i>	126
<i>The fundamental lemma of persistent homology</i>	127
<i>9.3 Computation</i>	127
<i>Matrix reduction</i>	128
<i>Extracting persistence</i>	128
<i>Representatives</i>	129
<i>Example</i>	130
<i>Computational tricks</i>	132
<i>9.4 Concluding remarks</i>	133
<i>Appendix: zig-zag persistence and multi-parameter persistence</i>	134
<i>10 Persistent homology: stability theorem</i>	137
<i>10.1 Continuous filtrations</i>	137
<i>Interleaving distance for filtrations</i>	140
<i>10.2 Persistence modules</i>	141
<i>Persistence modules</i>	141
<i>Decomposition</i>	142
<i>Interleaving distance for persistence modules</i>	143

<i>10.3 Bottleneck distance and stability theorem</i>	145
<i>Bottleneck distance</i>	145
<i>Stability theorem</i>	147
<i>10.4 Interpretations and examples</i>	148
<i>1-dimensional persistence of geodesic spaces</i>	148
<i>Stability demonstrated</i>	150
<i>Spheres</i>	151
<i>De-noising a function</i>	152
<i>10.5 Concluding remarks</i>	153
<i>Appendix: From the interleaving distance to the bottleneck distance</i>	154
<i>11 Discrete Morse theory</i>	157
<i>11.1 Motivation</i>	157
<i>11.2 Discrete Morse functions and discrete vector fields</i>	159
<i>Gradient vector fields</i>	160
<i>11.3 Morse homology</i>	162
<i>Morse chain complex</i>	162
<i>Morse homology</i>	163
<i>Generating DMFs and gradient vector fields</i>	165
<i>11.4 Concluding remarks</i>	166
<i>A proof of Theorem 11.2.6</i>	166
<i>Index</i>	169
<i>Bibliography</i>	173

1

Metric spaces

TOPOLOGY AND GEOMETRY study the shapes of spaces. In this book we will look at the modelling, computation, and representation of shapes and their properties. Our starting point will be metric spaces. These are sets with a meaningful notion of a distance (metric). In this chapter, we will focus on an intuitive understanding of three equivalence types of metric spaces: the isometry type, the homeomorphism type, and the homotopy type of spaces. These types will play a crucial role in later sections.

1.1 Definition of metric spaces and basic examples

Definition 1.1.1. A *metric space* (X, d) is a pair consisting of a set X and a function $d: X \times X \rightarrow [0, \infty)$, such that for any $x, y, z \in X$ the following hold:

- $d(x, y) = 0$ iff $x = y$,
- symmetry: $d(x, y) = d(y, x)$, and
- triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$.

☞ Word “iff” stands for “if and only if”.

Function d is referred to as a *distance* or a *metric*.

If X is introduced as a metric space, we implicitly assume d or d_X is the metric on X , unless stated otherwise.

Example 1.1.2. The following are some of metrics d_* on \mathbb{R}^n . For $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ we define:

- $d_1(x, y) = \sum_{i=1}^n |x_i - y_i|$
- $d_2(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
- $d_p(x, y) = \sqrt[p]{\sum_{i=1}^n (x_i - y_i)^p}$ for $p > 1$

- $d_\infty(x, y) = \max_{i \in \{1, 2, \dots, n\}} |x_i - y_i|$

From now on, \mathbb{R}^n is always considered to be equipped with the Euclidean d_2 metric, unless stated otherwise.

Example 1.1.3. The underlying space can be different than \mathbb{R}^n . Here are some examples:

- Suppose X is a finite graph with a length associated with each edge. The geodesic distance d_g between two vertices in X is the length of the shortest path between these vertices in X .
- Suppose X is a surface. We can think of it as a sphere or the surface of the earth. Similarly as above, the geodesic distance d_g between two points on X is the length of the shortest path between these points on X . For example, consider the distance between London and Sydney (see Figure 1.2). The distance usually thought of in this case is the geodesic distance on Earth, that is, the length of the shortest path between the two cities. The actual Euclidean distance in space between the two cities is shorter, but usually not of interest, since the path that realizes it passes fairly close to the center of the Earth.
- Suppose A is a finite set which we call alphabet. Let X denote a set of finite sequences (words) consisting of the elements of A (letters). The Levenshtein distance between two words in X is defined as the minimum number of edits required to transform one word into another, where the allowed edits are:
 - an insertion of a letter at any position;
 - a deletion of a letter anywhere;
 - a substitution of a letter in any place by another letter.

See Figure 1.3 for example.

- Let X be a finite set and let 2^X be the collection of all subsets of X . The Jaccard distance on 2^X is defined as

$$d_J(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}.$$

For a metric space (X, d) , $x \in X$ and $r > 0$ we define the closed¹ r -ball around x as

$$B_d(x, r) = \{y \in X \mid d(x, y) \leq r\}.$$

When the metric is apparent from the context we omit it and use $B(x, r)$.

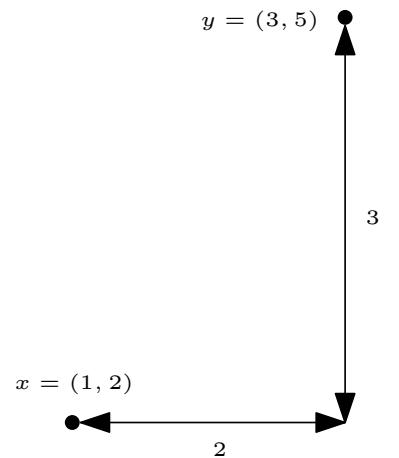


Figure 1.1: A few distances:
 $d_1(x, y) = 5$, $d_2(x, y) = \sqrt{13}$,
 $d_\infty(x, y) = 3$.

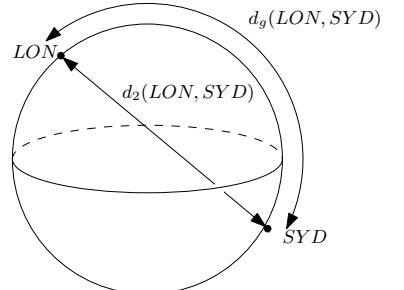


Figure 1.2: Geodesic vs d_2 distance between London and Sydney.

DOG DOL DOLF WOLF
 Figure 1.3: Levenshtein distance between DOG and WOLF is 3 by the following argument. The sequence above demonstrates that the distance is at most 3. As WOLF has three letters that do not appear in DOG, the distance is at least 3.

¹ As we will only consider closed balls, the phrase will be simplified to just “balls”.



Figure 1.4: Balls in d_1, d_2 and d_∞ metric in the plane.

1.2 Maps and equivalence types

When transforming or mapping spaces, we will always be using continuous maps.

Definition 1.2.1. A map $f: X \rightarrow Y$ between metric spaces is **continuous** if for each $x \in X$ and for each $\varepsilon > 0$ there exists $\delta > 0$ so that the following holds for all $y \in X$:

$$d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \varepsilon.$$

The notion of continuity between metric spaces includes the classical continuity from calculus, i.e., all continuous elementary functions $\mathbb{R} \rightarrow \mathbb{R}$ are continuous in the sense of Definition 1.2.1 on (\mathbb{R}, d_1) .

An equivalent definition of continuity could be stated in terms of convergent sequences. A sequence of points $\{z_i\}_{i \in \mathbb{N}}$ in Z converges to $w \in Z$ (notation $\lim_{i \rightarrow \infty} z_i = w$) iff $d_Z(z_i, w)$ converges to zero. It turns out that a map $f: X \rightarrow Y$ between metric spaces is continuous if the following implication holds: If $\{x_i\}_{i \in \mathbb{N}}$ is any sequence in X with $\lim_{i \rightarrow \infty} x_i = u \in X$, then $\lim_{i \rightarrow \infty} f(x_i) = f(u) \in Y$. A practical interpretation of continuity would be the following: if we improve our measurements x_i in the sense that we get a better approximation for the desired state w , then the values over a continuous map $f(x_i)$ also converge to the value $f(w)$. For example, suppose we want to estimate the area of Madagascar from a .bmp image representing a map of the island. We expect that as the resolution increases, we should get a better estimate for the total area.

A continuous map $g: [0, 1] \rightarrow X$ is called a **path** from $g(0)$ to $g(1)$.

Next, we give three different equivalence relations on the class of metric spaces, each of which preserves a different level of geometric information. We start with the strictest equivalence, which preserves the most structure.

Definition 1.2.2. A map $f: X \rightarrow Y$ between metric spaces is an **isometry**, if it is bijective and preserves distances, i.e., for every $x_1, x_2 \in X$, $d_X(x_1, x_2) = d_Y(f(x_1), f(x_2))$. Two metric spaces are **isometric**, if there exists an isometry between them.

Isometries of the plane are combinations of translations, rotations and reflections. In \mathbb{R}^n , isometries are combinations of a translation and a linear map. Linear isometries in \mathbb{R}^n are represented by orthogonal matrices.

It turns out that no patch of a sphere (equipped with the geodesic metric) is² isometric to a subset of a plane. A practical consequence of

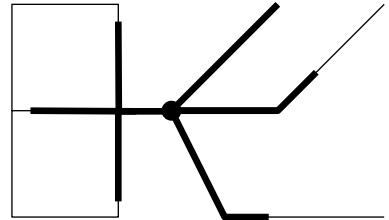
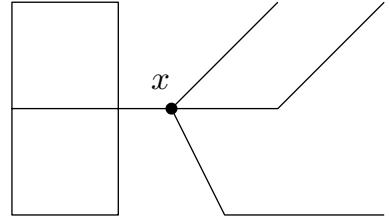


Figure 1.5: A graph (above) and a ball around x in the geodesic metric on the graph (below).

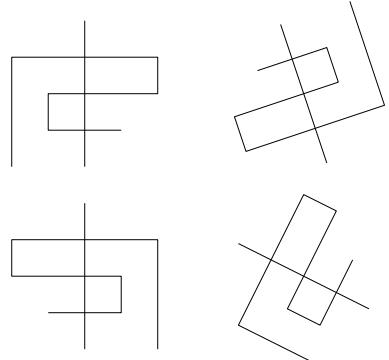


Figure 1.6: Four isometric planar sets.

² This is a consequence of Gauss' Theorem Egregium.

this fact is that all topographic maps are distorted.

Although isometries are convenient in many situations, they are essentially a geometric notion that is too rigid for topological treatment. We next introduce a topological counterpart.

Definition 1.2.3. A map $f: X \rightarrow Y$ between metric spaces is a **homeomorphism**, if it is bijective, continuous, and f^{-1} is continuous. Two metric spaces are **homeomorphic** (or of the same topological type; notation: $X \cong Y$), if there exists a homeomorphism between them.

It is not hard to see that every isometry is a homeomorphism.

While homeomorphisms are much more flexible and preserve a number of invariants of a space (later we will mention dimension, number of components and holes, etc.), they do not preserve some of the geometric properties, e.g. diameter (the supremum of pairwise distances in a space), radii of the smallest enclosing balls, etc.

We will often be referring to the following two spaces:

- For $n \in \mathbb{N}$ an **n -sphere** S^n is any space homeomorphic to the n -dimensional sphere

$$\{x \in \mathbb{R}^{n+1} \mid d_2(x, 0) = 1\} = \left\{(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} \mid \sum_{i=1}^{n+1} x_i^2 = 1\right\},$$

where $0 = (0, 0, \dots, n) \in \mathbb{R}^n$. Observe that S^0 consists of two points, S^1 is homeomorphic to a circle³, S^2 to the usual sphere, etc.

- For $n \in \mathbb{N}$ an **n -disc** D^n is any space homeomorphic to the ball

$$B(0, 1) = \{x \in \mathbb{R}^n \mid d_2(x, 0) \leq 1\},$$

where 0 is the n -tuple of zeros⁴. Observe that D^1 is a closed interval, whose endpoints are S^0 . Similarly, D^2 can be thought of as the unit disc in the plane. Note that its boundary in the plane is S^1 .

Example 1.2.4. Here we provide some examples of homeomorphisms.

- Two finite metric spaces are homeomorphic iff they consist of the same number of points. Each map between finite metric spaces is continuous.
- Any two closed intervals are homeomorphic. In particular, a homeomorphism $f: [0, 1] \rightarrow [a, b]$ for $a < b$ is given by $f(t) = a + t(b - a)$.
- A square $[-1, 1]^2$ in the plane is homeomorphic to the ball $B((0, 0), 1)$ in the plane. One of the homeomorphisms is given by the radial map $B((0, 0), 1) \rightarrow [-1, 1]^2$ mapping:
 - $(0, 0) \mapsto (0, 0)$ and

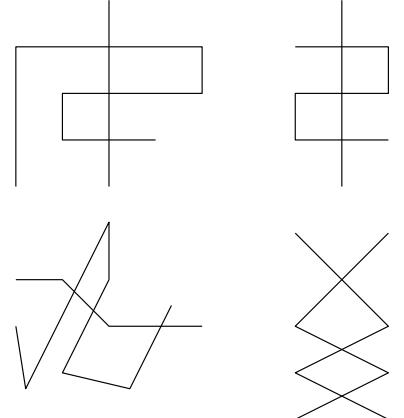


Figure 1.7: Four homeomorphic sets in the plane.

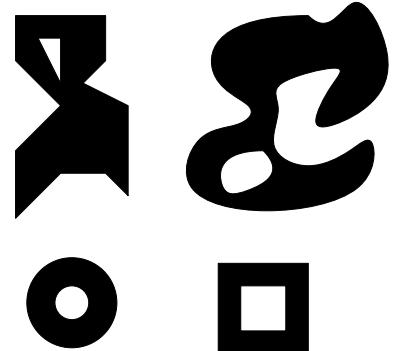


Figure 1.8: Four homeomorphic sets in the plane.

³ A circle is a 1-dimensional subset of \mathbb{R}^2 defined by $(x - a)^2 + (y - b)^2 = r^2$, i.e., it is “empty inside”.

⁴ A clarification on terminology: A ball (a metric concept) in a metric space is a particular specific subspace of that metric space. An n -disc (a topological concept) is any space homeomorphic to the standard unit ball in \mathbb{R}^n , and thus defined up to homeomorphism. A square in the plane is a 2-disc, but is not a ball in the Euclidean metric. Any unit ball of radius at least 1 on a circle of circumference 1 is the entire circle and so is not a 1-disc.

- $(x, y) \mapsto \rho(x, y) \cdot (x, y)$ for

$$\rho(x, y) = \frac{\sqrt{x^2 + y^2}}{\max\{|x|, |y|\}} = \frac{d_2((0, 0), (x, y))}{d_\infty((0, 0), (x, y))}.$$

- All three balls in Figure 1.4 are homeomorphic via radial maps.
- For each $n \in \mathbb{N}$, $S^n \cong S^m$ iff $n = m$. We will prove this result using homology in a later chapter.
- For each $n \in \mathbb{N}$, $D^n \cong D^m$ iff $n = m$. We will prove this result in Theorem 8.1.7.
- No n -disc is homeomorphic to any k -sphere. Each n -sphere can be obtained as a union of two n -discs acting as hemispheres.

While homeomorphism is the focal equivalence in the field of topology, it turns out that many computable invariants are in fact invariant with respect to a continuous deformation of spaces. These deformations are formalized by the concept of homotopy.

Definition 1.2.5. Continuous maps $f, g: X \rightarrow Y$ between metric spaces X and Y are **homotopic** [$f \simeq g$] if there exists a continuous deformation of f into g , i.e., if there exists a map $H: X \times [0, 1] \rightarrow Y$, such that $H|_{X \times \{0\}} = f$ and $H|_{X \times \{1\}} = g$. Map H is called a **homotopy**.

Another way to think about homotopy between f and g would be as a continuous collection of paths from $f(x)$ to $g(x)$ in X . Homotopies induce an equivalence relation on continuous maps between X and Y . Two maps belong to the same homotopy class iff they are homotopic.

Example 1.2.6. Some examples concerning homotopies:

1. For each metric space X , any two maps $f, g: X \rightarrow \mathbb{R}^n$ are homotopic. A homotopy consists of line segments between $f(x)$ and $g(x)$. In particular,

$$H(x, t) = (1 - t)f(x) + tg(x).$$

2. Let $w \in S^1$. Then the identity map $\text{id}: S^1 \rightarrow S^1$ is not homotopic to the constant map $c_w: S^1 \rightarrow S^1$, which maps each point to w . Later we will be able to prove this fact using homology. Note that by the previous example both maps are homotopic in \mathbb{R}^2 , hence the relation of being homotopic depends on the target space of the maps.

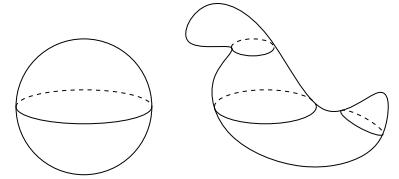


Figure 1.9: Two homeomorphic surfaces.

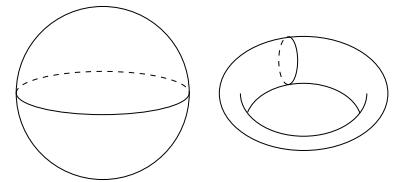


Figure 1.10: Two non-homeomorphic surfaces.

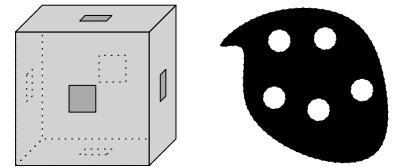


Figure 1.11: The surface of a cube with a puncture in each of the six sides is homeomorphic to a planar set with five holes.

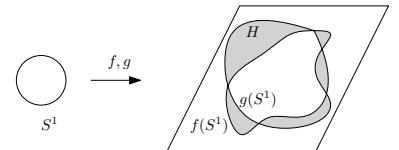


Figure 1.12: Two maps in the plane are homotopic.

3. Consider the two spaces in Figure 1.13. Space X is a single point, space Y consists of a point, an empty triangle (S^1), a square (D^2) and a disc with a tail. Observe that there are four homotopy classes of maps from X to Y , one for each component of Y .

We are now ready to introduce homotopy equivalence.

Definition 1.2.7. Metric spaces X and Y are **homotopy equivalent** $[X \simeq Y]$ if there exist maps $f: X \rightarrow Y$ and $g: Y \rightarrow X$, such that $f \circ g \simeq id_Y$ and $g \circ f \simeq id_X$. Maps f and g are called **homotopy equivalences**.

Homeomorphic spaces are homotopy equivalent. A metric space X is **contractible**, if it is homotopy equivalent to the one-point space.

Example 1.2.8. Some examples concerning homotopy equivalences:

- Let $X = [0, 1]$ and $Y = \{0\}$. Then $X \simeq Y$, i.e., $[0, 1]$ is contractible. Map $f: X \rightarrow Y$ is the constant map and map $g: Y \rightarrow X$ can be chosen to be any map, say $g(0) = 0$. Composition $f \circ g$ is identity. It remains to show that $h = g \circ f: [0, 1] \rightarrow [0, 1]$, which is the constant map at 0, is homotopic to the identity. Such a homotopy is, for example, the linear homotopy from 1. of Example 1.2.6. In the same way we can prove that D^n is contractible for each $n \in \mathbb{N}$.
- Convex sets and trees are contractible.
- It turns out that no S^n is contractible. The case $n = 1$ follows from 2. of Example 1.2.6.
- $\mathbb{R}^n \setminus \{(0, 0, \dots, 0)\} \simeq S^{n-1}$. Thinking of S^{n-1} as the standard unit sphere, this equivalence can be proved using the inclusion map $S^{n-1} \hookrightarrow \mathbb{R}^n \setminus \{(0, 0, \dots, 0)\}$, the radial map (see Figure 1.18) $\mathbb{R}^n \setminus \{(0, 0, \dots, 0)\} \rightarrow S^{n-1}$ defined by $x \mapsto x/\|x\|$, and linear homotopy.

Homotopy equivalence does not preserve all topological properties (for example, dimension), but it does preserve many of those that we can compute: the number of components, holes, etc.

Connectedness

The first homotopy invariant we will mention is connectedness. There are a few versions of it in topology. We will focus on the one generated by paths.

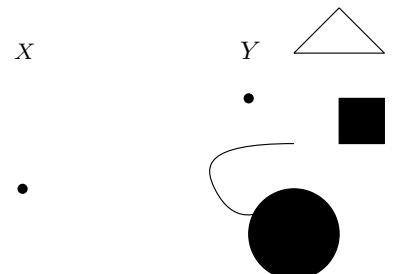


Figure 1.13: There are four homotopy classes of maps from a single point space X to Y .

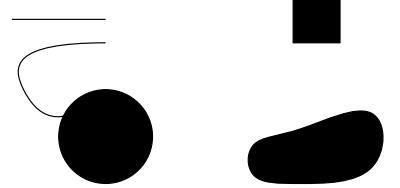


Figure 1.14: Four contractible spaces.

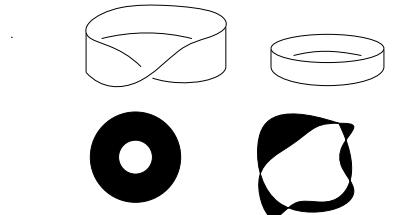


Figure 1.15: Four spaces homotopy equivalent to S^1 : Moebius band (top left), usual band $S^1 \times [0, 1]$ (top right) and two planar sets below. Only two of them are homeomorphic.

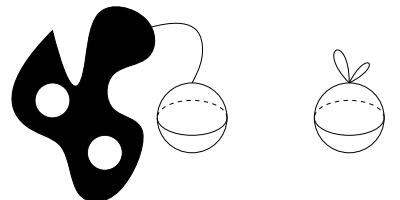


Figure 1.16: Two more homotopy equivalent spaces.



Figure 1.17: A sequence of steps deforming O to P . While the figure demonstrated a continuous deformation (homotopy equivalence), the spaces presented in this case are actually homeomorphic.

Definition 1.2.9. Space X is **path connected**, if for each $x, y \in X$ there exists a path from x to y in X .

Subset $A \subseteq Y$ of a metric space Y is a **path component**, if it is a maximal path connected subset.

A space is path connected iff it is itself a path component. As was mentioned above, path connectedness is a homotopy invariant: if X is path connected and $Y \simeq X$, then Y is also path connected. Similarly, the number of path components of a metric space is a homotopy invariant. Space Y on Figure 1.13 has four components.

1.2 Concluding remarks

Recap (highlights) of this chapter

- Metric spaces;
- Isometry;
- Homeomorphism;
- Homotopy equivalence;
- Connectedness.

Background and applications

Mathematics is the language of science and scientific concepts are modelled by mathematical objects. These objects can range from simple to sophisticated: a simple Boolean value (0 or 1, i.e., TRUE or FALSE), a numeric value (integer, real, complex, etc.), a collection of numeric values (e.g., a point in \mathbb{R}^n), a collection of points in \mathbb{R}^n , a function, a vector space, a probability distribution, a graph, a matrix, a metric space, etc. For most of these notions, there is a useful notion of a metric that transfers the possible outputs into a metric space and thus into the realm of geometry and topology, some of which we have explored here.

The notions introduced in this chapter are covered in standard books on topology⁵.

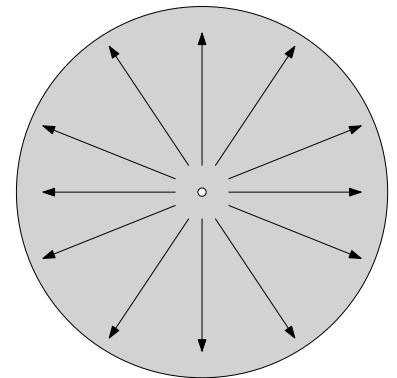


Figure 1.18: Radial map of a punctured ball $B_2((0,0),1) \setminus \{(0,0)\} \subset \mathbb{R}^2$ to the standard unit sphere (S) in the plane. Using the argument used in last part of Example 1.2.8, the induced homotopy equivalence demonstrates $B_2((0,0),1) \setminus \{(0,0)\} \simeq S$.

⚠ From now on we will be dropping adjective “path” and only refer to “connectedness”, and “components”.

⁵ James R. Munkres. Topology. Prentice Hall, Inc, 2nd ed edition, 2000

2

Planar triangulations

IN THE PREVIOUS chapter we learned about metric spaces along with the homeomorphism and homotopy type. However, the descriptions we used are not of combinatorial nature, and one would have difficulties using them for computations. In this chapter we will introduce one of the simplest combinatorial descriptions of planar spaces: triangulations in the plane. Essentially, we would like to describe a planar region as a “nice” union of triangles. Triangles are used primarily because they are easy to describe: we only have to provide three points. In later sections we will use these triangulations to compute various invariants of the space: components, homology, etc.

It turns out that not every planar subset can be triangulated. However, finite triangulations (i.e., triangulations with finitely many triangles) can be obtained for most planar subsets of interest to us .

2.1 Definition of planar triangulations

A triangle in the plane has three edges and three vertices.

Definition 2.1.1. A *triangulation* of a closed region $D \subset \mathbb{R}^2$ is a decomposition of D into triangles, so that:

1. no triangle is degenerate (i.e., a point or just a line segment),
2. interiors of triangles are disjoint, and
3. intersection of any pair of triangles is either a common edge, a common vertex, or empty.

Geometric description of Definition 2.1.1 is provided by Figure 2.3.

The idea of a triangulation may be generalized in various ways. One could use different shapes of pieces to decompose a planar region or the entire plane. Such decompositions are called tessellations. General-

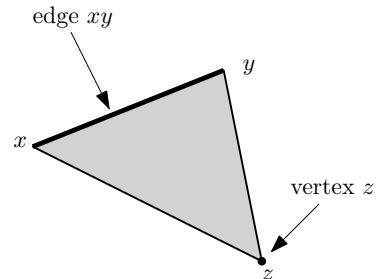


Figure 2.1: Triangle xyz .

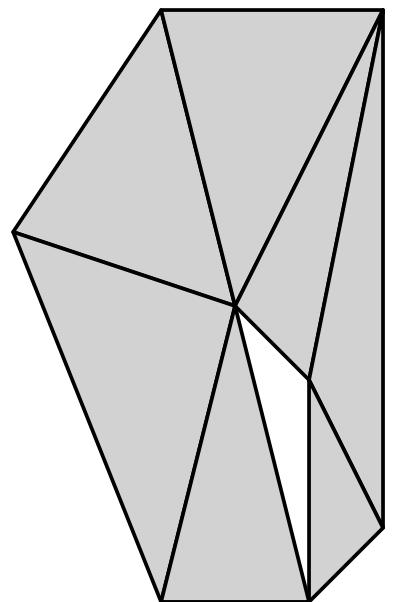


Figure 2.2: A planar triangulation.

☞ A planar region admitting a triangulation is called a *polygonal region*.

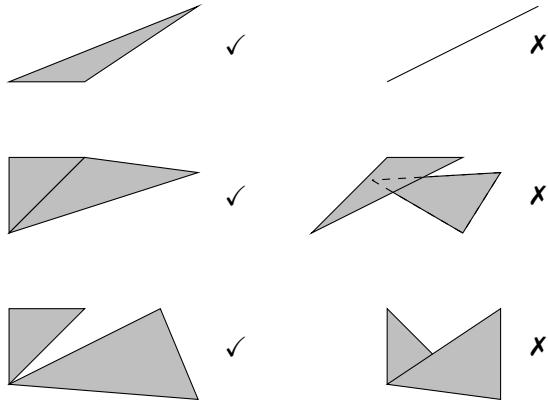


Figure 2.3: Conditions of Definition 2.1.1.

izing by dimension, one could use “higher dimensional triangles”, such as tetrahedra, to decompose a higher dimensional space. This idea will be formalized as simplicial complex in the next chapter.

Modifications of triangulations

Occasionally we will want to modify a triangulation. Here are some of the most used modifications:

- add a triangle;
- remove a triangle;
- flip a common edge;
- refine using a subdivision. An example of a subdivision is the barycentric subdivision: for each edge and each triangle consider its geometric center (centroid) as a new vertex in our triangulation, and then decompose each triangle as demonstrated by Figure 2.4. This subdivision is convenient when we want to refine a triangulation, i.e., systematically decompose the triangles into smaller triangles.

We will often focus on triangulations of convex polyhedra, i.e., convex hulls of finitely many points in the plane, as defined below. Given a finite $S \subset \mathbb{R}^2$ we say a **triangulation on S** is any triangulation of the convex hull of S , whose vertex set is S .

2.2 Recap on convexity

Given points $x, y \in \mathbb{R}^n$, the line segment between them is parameterized as

$$\gamma(t) = tx + (1-t)y, \quad t \in [0, 1].$$

Note that $\gamma(0) = y$, $\gamma(1) = x$, and $\gamma(1/2)$ corresponds to the midpoint of the line segment.

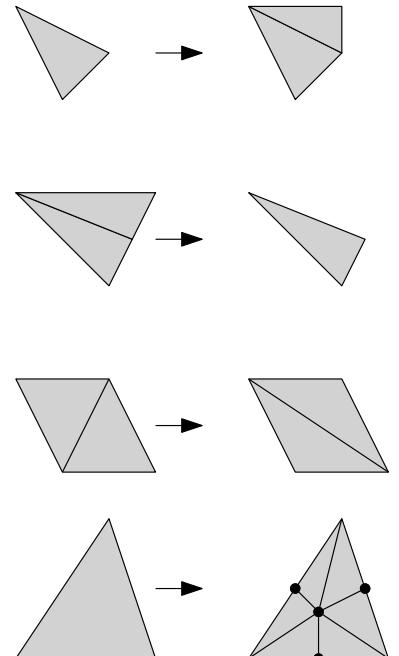


Figure 2.4: Modifications of triangulations.

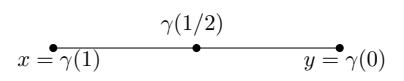


Figure 2.5: Line segment.

Definition 2.2.1. A subset $A \subset \mathbb{R}^n$ is **convex**, if for each $a, b \in A$ the entire line segment between a and b lies in A , i.e., if $\forall t \in [0, 1]$ we have $ta + (1 - t)b \in A$. See Figure 2.6.

Given a subset $B \subset \mathbb{R}^n$, its **convex hull** $\text{Conv}(B)$ is the smallest convex set containing B .

The closed region on Figure 2.2 is not convex, while the ones on Figure 2.4 are convex.

A triangle is the convex hull of the set of its vertices, which provides a convenient description of a triangle: the triangle with affinely independent vertices $x, y, z \in \mathbb{R}^2$ can be parameterized by all possible convex combinations of these vertices:

$$\{\alpha_1x + \alpha_2y + \alpha_3z, \quad | \quad \forall i \in \{1, 2, 3\} : \alpha_i \in [0, 1], \quad \sum_{i=1}^3 \alpha_i = 1\}.$$

The term “convex combination” (as opposed to “linear combination”) refers to the fact that the coefficients α_i are from $[0, 1]$ and add up to 1. These coefficients are called **barycentric coordinates** in a triangle. The point with $\alpha_1 = \alpha_2 = \alpha_3 = 1/3$ is the centroid of the triangle, while points with two barycentric coordinates¹ 1/2 are the midpoints of the corresponding edges; all these points are vertices in the barycentric subdivision shown in Figure 2.4.

Convex hull can be constructed by iteratively adding all feasible line segments. It is important to note that for $B \subset \mathbb{R}^n$ the set obtained by adding all line segments

$$\{\alpha_1x + \alpha_2y \quad | \quad x, y \in B, \quad \forall i : \alpha_i \in [0, 1], \quad \sum_{i=1}^2 \alpha_i = 1\}$$

is typically not the convex hull. For example, starting with three vertices and adding the line segments between all three pairs we would obtain the set consisting of the edges but not the interior of the triangle (which constitutes the convex hull of three points). Instead, we have to add all possible line segments, or alternatively, add all convex combinations in one step:

$$\text{Conv } B = \bigcup_{m \in \mathbb{N}} \left\{ \sum_{i=1}^m \alpha_i x_i \quad | \quad \forall i : x_i \in B, \quad \alpha_i \in [0, 1], \quad \sum_{i=1}^m \alpha_i = 1 \right\}.$$

By the Carathéodory theorem (see original reference² or any modern book treating convexity) we can bound the number of summands³ by the dimension of the ambient space plus one:

$$\text{Conv } B = \left\{ \sum_{i=1}^{n+1} \alpha_i x_i \quad | \quad \forall i : x_i \in B, \quad \alpha_i \in [0, 1], \quad \sum_{i=1}^{n+1} \alpha_i = 1 \right\}.$$

In particular: for a finite subset $F \subset \mathbb{R}^2$, each point of $\text{Conv}(F)$ is contained in the convex hull of some triple of points from F .

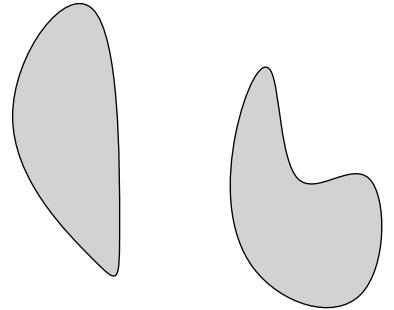


Figure 2.6: A convex (left) and a non-convex (right) subset of the plane.

¹ ...and the third coordinate equal to 0.

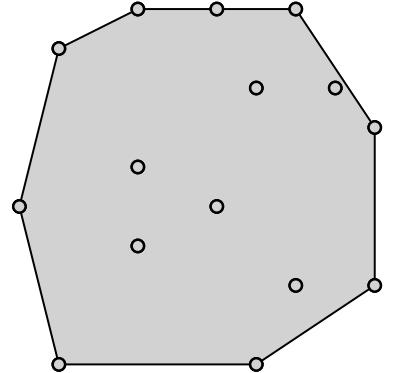


Figure 2.7: A collection of points and its convex hull.

² Constantin Caratheodory. Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen. *Math. Ann.* 64, no. 1, 95–115, 1907. doi: [10.1007/BF01449883](https://doi.org/10.1007/BF01449883)

³ ...and also the number of iterative steps in the procedure above...

2.3 Euler characteristic

Along with the number of components of a space, the Euler characteristic is one of the first real topological invariants we come across. In particular, while there are many triangulations of $\text{Conv}(S)$ on a finite subset $S \subset \mathbb{R}^2$, the Euler characteristic is the same for all of them.

For a given triangulation let:

- V be the number of its vertices,
- E be the number of its edges,
- F be the number of its triangles.

Definition 2.3.1. *The Euler characteristic χ of a triangulation is defined as*

$$\chi = F - E + V.$$

Theorem 2.3.2 (A simple version of the Euler-Poincaré formula). *Assume $S \subset \mathbb{R}^2$ is finite. For each triangulation of $\text{Conv}(S)$ we have*

$$\chi = 1.$$

Proof. Let us assume our triangulation has no vertical edge: if necessary this can be achieved by a small rotation. Assign to each triangle value $+1$ and to each edge value -1 . Position each of these values towards the unique rightmost vertex of the corresponding triangle/edge as suggested by Figure 2.9. Assign to each vertex value $+1$. The total sum of all assigned values is χ .

For each single vertex add: the value at the vertex and all the values of the triangles and edges, that gathered at that vertex. We can see that for each vertex the total sum is zero (arising from a sequence edge-triangle-edge-...-edge-triangle-edge on the left from the vertex plus the vertex itself) except for the leftmost vertex, where the value equals one. □

Remark 2.3.3. *It turns out that $\chi = 1$ for each triangulation of a contractible set⁴ in \mathbb{R}^2 . In fact, for a triangulation of $D \subset \mathbb{R}^2$, χ equals the number of components minus the number of holes of D .*

More technical details on this fact will be provided in later sections.

Let us just mention that the number of holes of a bounded set $D \subset \mathbb{R}^2$ equals the number of components of $\mathbb{R}^2 \setminus D$ minus one (see Figure 2.10).

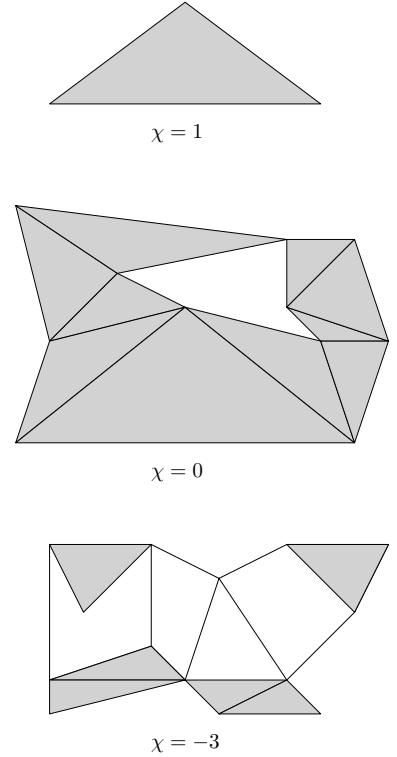


Figure 2.8: A few planar triangulations along with their Euler characteristics.

⁴ Even more: χ is a homotopy invariant.

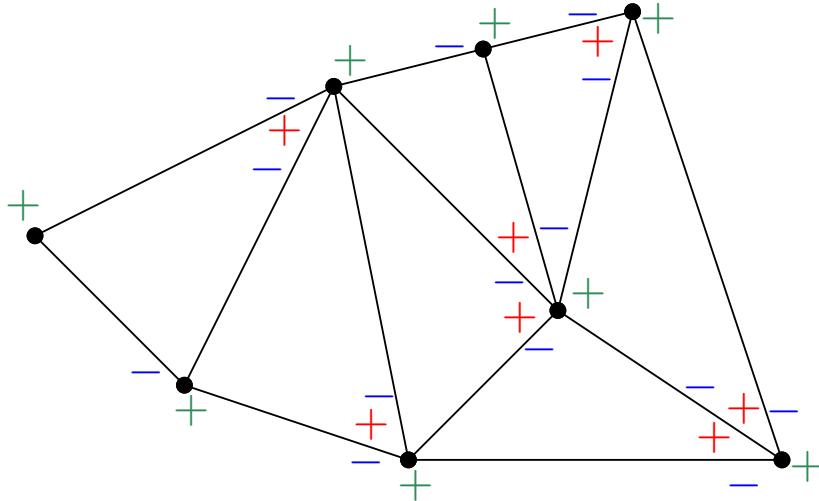
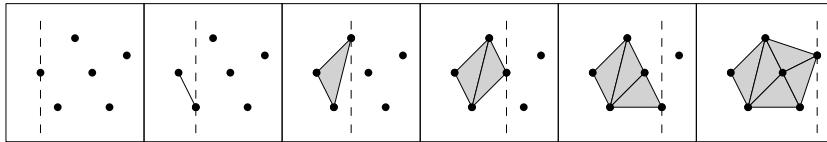


Figure 2.9: The assignment of values on triangles (red +), edges (blue -) and vertices (green +) from the proof of Theorem 2.3.2. Vertices also hold additional +1 value. The triangles are present but not shaded.

2.4 Constructing planar triangulations with line sweep

Let $S \subset \mathbb{R}^2$ be finite. Perhaps the simplest way to construct a triangulation on S is using a **line sweep**, which we now describe. Assume no two points of S have the same horizontal coordinate (this can be achieved by a small rotation if necessary). Now imagine a vertical line sweeping $\text{Conv}(S)$ from left to right. Each time the line reaches a point of S (a vertex in our triangulation), add all possible edges towards left without creating intersections. Furthermore, for each new bounded region add the corresponding triangle. As the line sweeps S we thus obtain a triangulation on S .



The condition that no two points have the same horizontal coordinate was added for reasons of simplicity only. If more points, say a_1, a_2, \dots, a_k have the same horizontal coordinate then, instead of adding all edges for all points a_i at once, proceed point by point: add all possible edges for a_1 , then for a_2 , etc. Depending on the order of points a_i we typically get different triangulations.

It should also be obvious that the line sweep does not need to proceed from left to right, but can proceed along any direction by sweep-

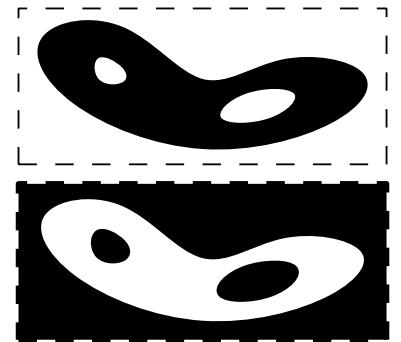


Figure 2.10: The top set has 2 holes. Equivalently, its complement on the bottom has $2 + 1$ components.

Figure 2.11: A line sweep using the vertical dashed line. Each time the vertical line reaches a point, we add all possible edges from that point to a point with smaller horizontal coordinate.

ing a line perpendicular to that direction.

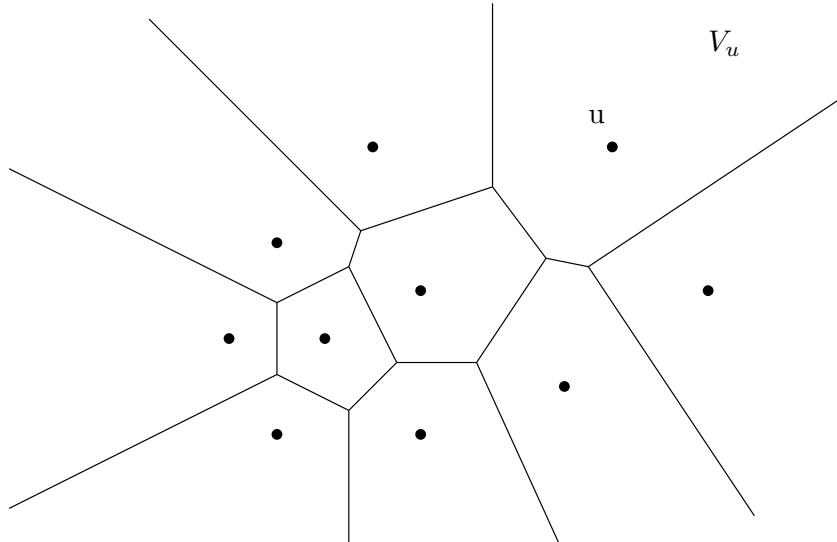
While the line sweep is conceptually simple, it does tend to construct triangulations with very thin triangles, which may be undesirable in applications. The triangulation that avoids thin triangles as much as possible is the Delaunay triangulation.

2.5 Voronoi diagram and Delaunay triangulation

Throughout this section let $S \subset \mathbb{R}^2$ be a finite subset satisfying a general position property: no four points of S lie on the same circle. We will first present the Voronoi diagram of S , which is a decomposition of the plane into specific regions.

For each $s \in S$ define the **Voronoi region** of s :

$$V_s = \{x \in \mathbb{R}^2 \mid \forall u \in S \setminus \{s\} : d(x, s) \leq d(x, u)\}.$$



If a pair of Voronoi regions V_{s_1}, V_{s_2} has a non-empty intersection, then (due to the general position condition above) this intersection is a bounded or unbounded line segment called a **Voronoi edge** and lies on the bisector between s_1 and s_2 .

If a triple of Voronoi regions $V_{s_1}, V_{s_2}, V_{s_3}$ has a non-empty intersection, then this intersection is a point called a **Voronoi vertex**. As this point lies on all three pairwise bisectors, it is the center of the circle containing s_1, s_2 and s_3 .

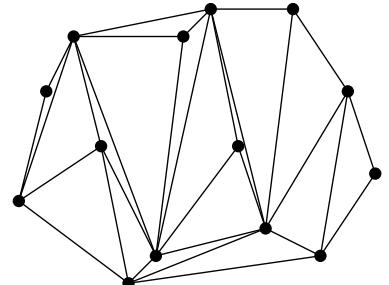


Figure 2.12: A line sweep triangulation resulting in thin triangles.

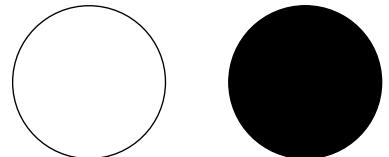


Figure 2.13: Circle on left, and a ball on right. The boundary of the ball is the circle.

Figure 2.14: An example of a Voronoi decomposition.

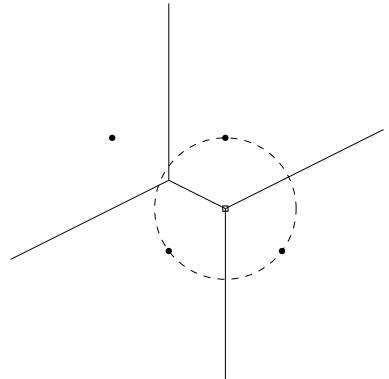


Figure 2.15: A Voronoi vertex \square is the center of the circle containing the corresponding points \bullet of S . Voronoi edges lie on the bisector lines between the corresponding points of S .

Definition 2.5.1. *The Voronoi diagram (or decomposition) of S is the collection of Voronoi regions, edges and vertices.*

A Voronoi region V_s consists of points, whose closest point of S is s . If for some point w there are two such closest points in S , then w is on the corresponding edge. If for some point w there are three such closest points in S , then w is a Voronoi vertex. The general position criterion above states that for each point in the plane, there can be no four closest points in S .

Voronoi diagram can be thought of as a result of a uniform expansion from the points of S . Suppose that in the initial stage we start with a finite set of locations S . Then, as time goes by, each point of S is being expanded into a region by growing at the same speed in all directions. At the beginning all these regions are balls centered at the points of S . As the growing regions collide, the growth towards the regions (edges) of contact stops, but continues along all other directions. The Voronoi decomposition is the final result of such growth, with each Voronoi region V_s containing the points that were reached from s first.

Definition 2.5.2. *The Delaunay triangulation on S , denoted by $\mathcal{D}(S)$, is the triangulation on S , such that:*

- its vertices are all points of S ,
- xy is an edge iff $V_x \cap V_y \neq \emptyset$, and
- xyz is a triangle iff $V_x \cap V_y \cap V_z \neq \emptyset$.

It turns out that Definition 2.5.2 indeed defines a planar triangulation on S . As a curiosity we mention that an edge xy of a Delaunay triangulation may partially lie outside of the union $V_x \cup V_y$.

Note that the edge xy of a Delaunay triangulation is a boundary edge (meaning it is contained in precisely one triangle) iff $V_x \cap V_y$ is unbounded. Similarly, x is a boundary vertex of a Delaunay triangulation (meaning it is an endpoint of some boundary edge) iff V_x is unbounded.

Local Delaunay condition

For a triple of non-collinear points $x, y, z \in \mathbb{R}^2$ in the plane define $\mathcal{C}(x, y, z)$ to be the circle containing x, y, z , and let $\mathcal{B}(x, y, z)$ be the ball whose boundary is $\mathcal{C}(x, y, z)$.

 *Voronoi regions and Voronoi diagrams can be defined for subsets of \mathbb{R}^n in the same way. In this case the general position property is: no collection of $n + 2$ points from S lies on the same sphere.*

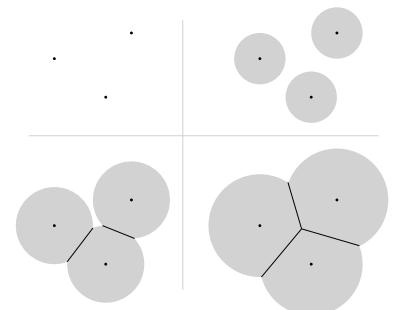


Figure 2.16: Voronoi diagram arising from expansion around points.

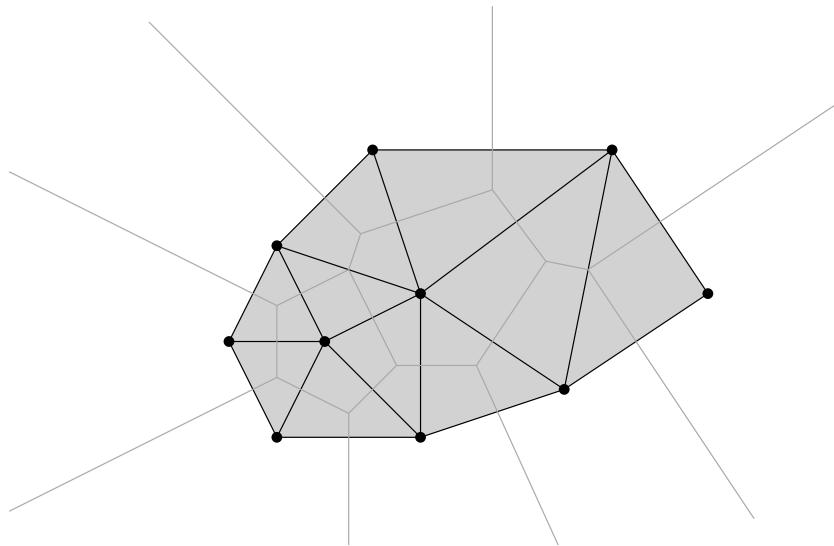


Figure 2.17: An example of a Delaunay triangulation with its underlying Voronoi decomposition.

Definition 2.5.3. Suppose an edge xy is shared by two different triangles xyz and xyw of a triangulation. The edge xy is **locally Delaunay** [abbreviation: LD], if $w \notin \mathcal{B}(x, y, z)$.

Proposition 2.5.4. Suppose edge xy is shared by two different triangles xyz and xyw from a triangulation.

1. Definition 2.5.3 is symmetric, i.e., $w \notin \mathcal{B}(x, y, z)$ iff $z \notin \mathcal{B}(x, y, w)$.
2. Each edge in a Delaunay triangulation is LD.

Proof. Part (1) is apparent from Figure 2.19.

(2): Since abc is a triangle in $\mathcal{D}(S)$, there exists the corresponding Voronoi vertex $q = V_x \cap V_y \cap V_z$, which is the center of $\mathcal{C}(x, y, z)$. As $q \notin V_w$ (recall that no four Voronoi regions have a nonempty intersection by the general position property), $d(q, x) = d(q, y) = d(q, z) < d(q, w)$ by the definition of Voronoi regions, hence $w \notin \mathcal{B}(x, y, z)$. \square

The property of being LD is a local property, shared by all edges of a Delaunay triangulation. It turns out that the converse of Proposition 2.5.4(2) is also true.

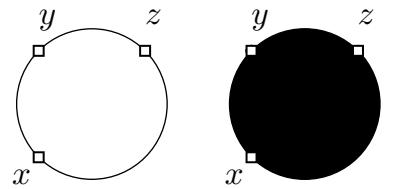


Figure 2.18: $\mathcal{C}(x, y, z)$ on the left and $\mathcal{B}(x, y, z)$ on the right.

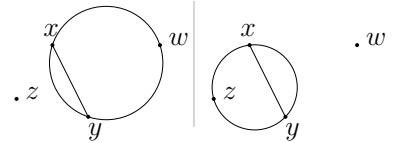


Figure 2.19: Proof of Proposition 2.5.4 (1).

Theorem 2.5.5. Suppose K is a triangulation on S . Then K is the Delaunay triangulation iff each edge is locally Delaunay.

A proof can be found in a textbook⁵.

Construction of $\mathcal{D}(S)$

Theorem 2.5.5 motivates the edge-flipping construction of Delaunay triangulations: starting with any triangulation on S (say, one obtained by a line sweep), keep flipping the non-LD edges. In order to algorithmically implement this construction we have to clarify two issues:

1. How do we verify the LD condition?
2. Does the procedure stop?

We address 1. first. It turns out it is not hard to verify the condition of LD using the incircle test.

Proposition 2.5.6. [Incircle test] Suppose $x = (x_1, x_2), y = (y_1, y_2), z = (z_1, z_2)$ and $w = (w_1, w_2)$ are four points in \mathbb{R}^2 . Assume x, y, z are not collinear and form a positively oriented triple, i.e.:

$$\begin{vmatrix} 1 & x_1 & x_2 \\ 1 & y_1 & y_2 \\ 1 & z_1 & z_2 \end{vmatrix} > 0$$

Then $w \notin \mathcal{B}(x, y, z)$ iff

$$\begin{vmatrix} 1 & x_1 & x_2 & x_1^2 + x_2^2 \\ 1 & y_1 & y_2 & y_1^2 + y_2^2 \\ 1 & z_1 & z_2 & z_1^2 + z_2^2 \\ 1 & w_1 & w_2 & w_1^2 + w_2^2 \end{vmatrix} > 0.$$

A proof and technical details of Proposition 2.5.6 are provided in the Appendix. While Proposition 2.5.6 provides a convenient way to verify LD property (and answer 1.), it does not suggest whether the edge flip algorithm actually stops (2.). In order to address this question we provide a couple more equivalent conditions to LD.

Suppose edge xy is shared by two different triangles xyz and xyw from a triangulation K on S . We say that edge xy is a **MaxMin** edge, if the minimal angle appearing in triangles xyz and xyw is larger than the minimal angle appearing in triangles xzw and yzw .

⁵ Mark de Berg, Marc van Kreveld, Mark Overmars, and Otfried Schwarzkopf. Computational Geometry: Algorithms and Applications. Springer-Verlag, second edition, 2000. doi: 10.1007/978-3-540-77974-2

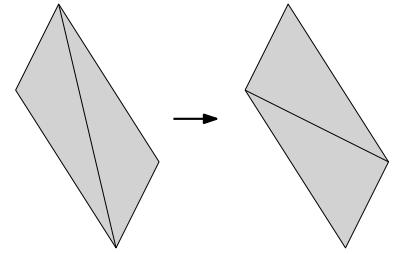


Figure 2.20: Edge flip.

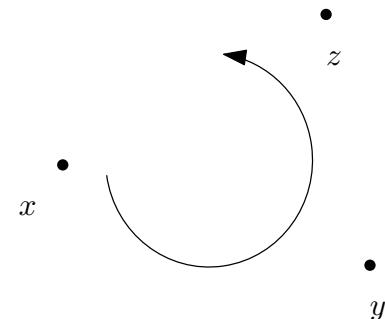


Figure 2.21: Positively oriented triple (x, y, z) .

Proposition 2.5.7. Suppose edge xy is shared by two different triangles xyz and xyw from a triangulation K on S . Then the following conditions are equivalent:

- (i) xy is LD.
- (ii) xy is a MaxMin edge.
- (iii) $\angle xzy + \angle xwy < \pi$.

Proof. Let us prove the equivalence (i) \Leftrightarrow (iii) first using the inscribed angle theorem (see Figure 2.22).

$$xy \text{ is LD} \stackrel{\text{definition}}{\Leftrightarrow} w \notin \mathcal{B}(x, y, z) \stackrel{\text{inscribed angle}}{\Leftrightarrow} \pi - \angle xzy > \angle xwy \Leftrightarrow \angle xzy + \angle xwy < \pi.$$

We now turn our attention to (i) \Leftrightarrow (ii). Let α be the minimal angle appearing in triangles xyz , xyw , xzw and yzw . It is easy to see that α has to lie either along xy or zw as all the other angles get dissected (and hence decreased) by either xy or zw in one of the configurations.

Assume xy is LD. According to the inscribed angle theorem, $\angle xyz > \angle xwz$, hence $\angle xyz$ does not equal α . In the same way we can prove that no angle along xy equals α , hence xy is the MaxMin edge.

Assume now that xy is not LD. Using the identical argument as in the previous paragraph we can prove that each angle along zw is larger than the corresponding angle along xy . Hence α has to lie along xy and therefore xy is not a MaxMin edge. \square

Proposition 2.5.7 implies that each edge flip, which makes an edge in a triangulation LD, increases the minimal angles in the triangulation. Let us explain this statement in more detail. For each triangle T_i in a triangulation K on S let t_i denote the size of its minimal angle. Construct a lexicographically ordered list of these minimal angles, i.e., $t_{i_0} \leq t_{i_1} \leq \dots \leq t_{i_m}$. Proposition 2.5.7 implies that every time we execute an edge flip making an edge in a triangulation LD, the new lexicographically ordered list of the minimal angles $t'_{i_0} \leq t'_{i_1} \leq \dots \leq t'_{i_m}$ is lexicographical larger than the previous list, i.e., $t_{i_j} \leq t'_{i_j}, \forall j$ with strict inequality holding for at least one index j . Hence by making the required edge flips that keep turning edges into LD edges we can't return to the initial or any already visited triangulation. Since there are only finitely many triangulations on S , and therefore finitely many possible ordered lists of minimal angles, the edge flipping algorithm terminates, answering 2. above affirmatively.

Conclusion: the edge flipping algorithm terminates with $\mathcal{D}(S)$.

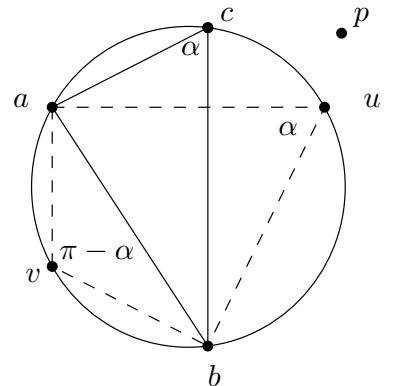


Figure 2.22: Inscribed angle theorem. Suppose $u, v \in \mathcal{C}(a, b, c)$, as the figure demonstrates. Then $\angle acb = \angle aub = \pi - \angle avb$. This obviously implies $\angle acb > \angle apb$.

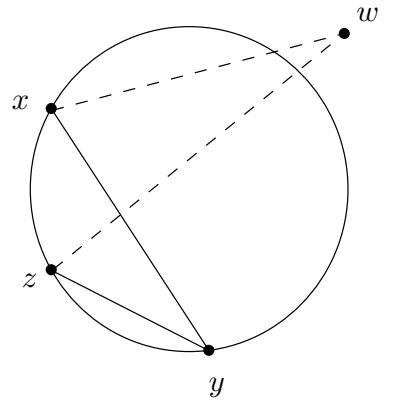


Figure 2.23: Proof of Proposition 2.5.7, (i) \Leftrightarrow (ii).

A triangulation, for which the lexicographically ordered list of the minimal angles is maximal in the lexicographical order, is called a MaxMin triangulation.

Theorem 2.5.8. *A MaxMin triangulation on S coincides with $\mathcal{D}(S)$.*

In particular, there exists only one MaxMin triangulation on S .

2.6 Concluding remarks

Recap (highlights) of this chapter

- Planar triangulations;
- Convexity;
- Euler characteristic;
- Line sweep;
- Voronoi diagram and Delaunay triangulation;
- Constructing the Delaunay triangulation using the locally Delaunay condition and the incircle test.

Background and applications

The Euler characteristic was introduced by Leonhard Euler in the 18th century. The line sweep algorithm, Voronoi diagram and Delaunay triangulation are basic notions studied especially in computational geometry⁶. Applications of the Euler characteristic include image analysis⁷, target enumeration⁸, etc. The edge flip algorithm we mentioned requires $O(n^2)$ edge flips, where n is the number of vertices of S . There are known algorithms to construct the Delaunay triangulation in $O(n \log n)$.

The above mentioned properties of the Delaunay triangulation make it one of the favorite choices for a triangulation on a finite planar set S . For example, assume you are given a collection of points modelling a geographic profile in a small region. The points consist of coordinates and elevations at these coordinates. The task is to model the surface approximating the geographic profile. A standard solution would be to construct the Delaunay triangulation on the set of coordinate points, and then lift these points and triangles according to the given elevations. Triangles lifted this way provide a good approximation of the geographic profile on the sampled region.

⁶ Mark de Berg, Marc van Kreveld, Mark Overmars, and Otfried Schwarzkopf. Computational Geometry: Algorithms and Applications. Springer-Verlag, second edition, 2000. doi: [10.1007/978-3-540-77974-2](https://doi.org/10.1007/978-3-540-77974-2)

⁷ A. Roy, R. A. I. Haque, A. J. Mitra, M. Dutta Choudhury, S. Tarafdar, and T. Dutta. Understanding flow features in drying droplets via Euler characteristic surfaces—a topological tool. *Physics of Fluids*, 32(12):123310, 2020. doi: [10.1063/5.0026807](https://doi.org/10.1063/5.0026807)

⁸ Yuli Baryshnikov and Robert Ghrist. Target enumeration via Euler characteristic integrals. *SIAM Journal on Applied Mathematics*, 70(3):825–844, 2009. doi: [10.1137/070687293](https://doi.org/10.1137/070687293)

Appendix: Proof of Proposition 2.5.6

Proof. Let us explain the positively oriented criterion first, see Figure 2.24. Points x, y, z form a positively oriented triple iff vectors \vec{xy}, \vec{yz} are positively oriented, meaning that the third component of the cross product $(y_1 - x_1, y_2 - x_2, 0) \times (z_1 - x_1, z_2 - x_2)$ is positive. This third component equals the 3×3 determinant

$$\begin{vmatrix} 1 & x_1 & x_2 \\ 1 & y_1 & y_2 \\ 1 & z_1 & z_2 \end{vmatrix}.$$

We now turn our attention to the proof of the proposition. Surprisingly enough, we need to use three-dimensional geometry, see Figure 2.25 throughout the proof. Embed \mathbb{R}^2 (and points x, y, z, w) into \mathbb{R}^3 by assigning the third coordinate to be 0, i.e., $x = (x_1, x_2, 0)$, etc. Consider the graph of the function $f(x, y) = x^2 + y^2$. Lift points x, y, z, w to the graph of f and let x', y', z', w' denote the lifted points, i.e., $x' = (x_1, x_2, x_1^2 + x_2^2)$, etc. Let Π denote the plane containing x', y', z' .

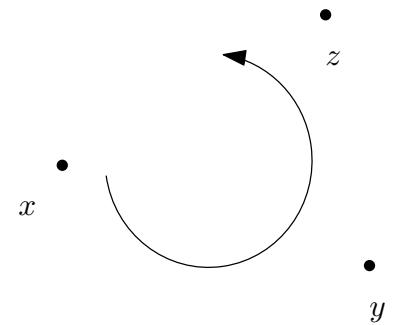


Figure 2.24: Positively oriented triple x, y, z .

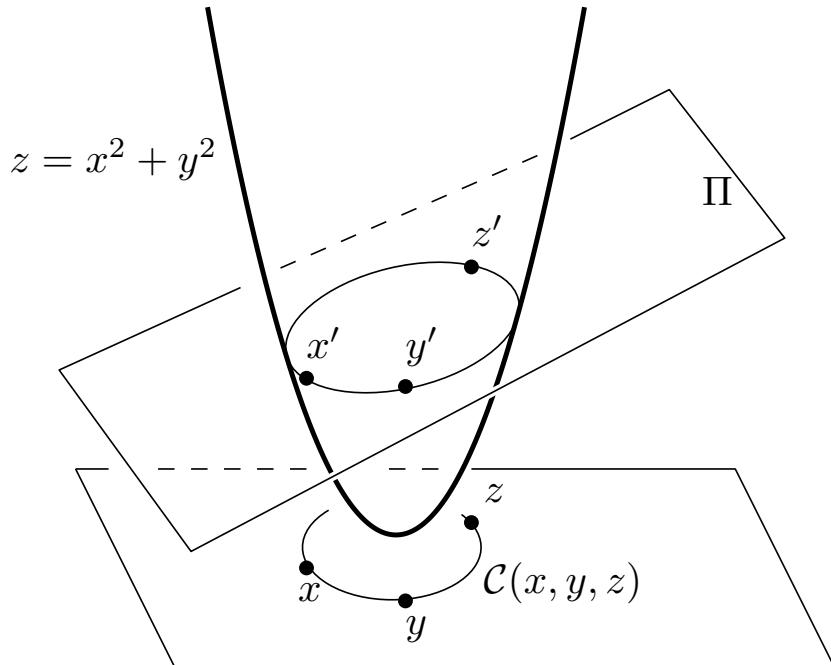


Figure 2.25: Proof of Proposition 2.5.6.

Let C be the intersection of the graph of f and Π . Note that the vertical projection of C onto $\mathbb{R}^2 \times \{0\}$ is a circle: substituting z in $z = x^2 + y^2$ by an equation of a plane $z = ax + by + c$ we obtain an equation of a circle in the plane of the form $x^2 + y^2 - ax - by - c = 0$. As this circle contains x, y, z , it coincides with $C(x, y, z)$. It is

geometrically apparent that $w \notin \mathcal{B}(x, y, z)$ iff w' lies above Π (the region where the graph of f is below Π is $\mathcal{B}(x, y, z)$). Since x, y, z are positively oriented, a normal of Π with a positive third component is $\vec{n} = \overrightarrow{x'y'} \times \overrightarrow{x'z'}$. Point w' lies above Π iff $\vec{n} \cdot \overrightarrow{x'w'}$ is positive. It is elementary to verify that $\vec{n} \cdot \overrightarrow{x'w'}$ equals

$$\begin{vmatrix} 1 & x_1 & x_2 & x_1^2 + x_2^2 \\ 0 & y_1 - x_1 & y_2 - x_2 & y_1^2 + y_2^2 - (x_1^2 + x_2^2) \\ 0 & z_1 - x_1 & z_2 - x_2 & z_1^2 + z_2^2 - (x_1^2 + x_2^2) \\ 0 & w_1 - x_1 & w_2 - x_2 & w_1^2 + w_2^2 - (x_1^2 + x_2^2) \end{vmatrix} = \begin{vmatrix} 1 & x_1 & x_2 & x_1^2 + x_2^2 \\ 1 & y_1 & y_2 & y_1^2 + y_2^2 \\ 1 & z_1 & z_2 & z_1^2 + z_2^2 \\ 1 & w_1 & w_2 & w_1^2 + w_2^2 \end{vmatrix}.$$

□

3

Simplicial complexes

TOPOLOGICAL AND COMPUTATIONAL treatment of metric spaces relies on their convenient description. Given a metric space, we would like to have a finite combinatorial description, that can be used for computations. In the previous chapters we introduced planar triangulations as an example of such a description for planar subsets. In this chapter we will introduce simplicial complexes, which will form the basic structure upon which all our later computations will depend.

Simplicial complexes are higher-dimensional analogues of planar triangulations. While the latter are collections of triangles that fit together nicely, simplicial complexes are collections of higher dimensional simplices (generalizations of triangles) that fit together nicely. Essentially we will be building spaces from simple building blocks (simplices) given a rule describing how these blocks fit together... just like building a castle from LEGO cubes.

3.1 Affine independence

A point, a line segment, a triangle, a tetrahedron, etc. These are some of the geometric simplices. They are basic building blocks of geometric simplicial complexes. A geometric simplex is a convex hull of a finite collection of points. Before we state their formal definition we need to clarify a general position property required of a set of points spanning such a simplex. Under this property we want a pair of points to span a line segment, a triple of points to span a triangle (and not just a line segment), etc.

Choose $d, k \in \mathbb{N}$ and let $V = \{v_0, v_1, \dots, v_k\} \subset \mathbb{R}^d$ be a collection of points. Their **affine combination** is any sum of the form

$$\sum_{i=0}^k \alpha_i v_i, \quad \text{with} \quad \sum_{i=0}^k \alpha_i = 1.$$

The **affine hull** of V is the collection of all affine combinations of ele-

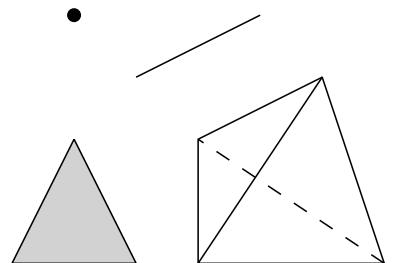


Figure 3.1: Some geometric simplices: a point, a line segment, a triangle, a tetrahedron.

ments of V . An affine hull is always an affine linear subspace in \mathbb{R}^d , meaning it is obtained from a linear subspace of \mathbb{R}^d by a translation.

Points $\{v_0, v_1, \dots, v_k\}$ are **affinely independent** if no v_i can be expressed as an affine combination of points $V \setminus \{v_i\}$. Proposition 3.1.1 explains how to test points for affine independence using linear independence, and why each affine hull is a translated linear subspace.

Proposition 3.1.1. *Points of $V = \{v_0, v_1, \dots, v_k\} \subset \mathbb{R}^d$ are affinely independent iff $\{v_1 - v_0, v_2 - v_0, \dots, v_k - v_0\}$ are linearly independent.*

Proof. Assume points of V are not affinely independent. Then, without the loss of generality, $v_0 = \sum_{i=1}^k \alpha_i v_i$ and $\sum_{i=1}^k \alpha_i = 1$, which implies equality $\sum_{i=1}^k \alpha_i(v_i - v_0) = 0$ and not all α_i are zero. We conclude that the points of V are not linearly independent.

On the other hand assume $\sum_{i=1}^k \beta_i(v_i - v_0) = 0$ with not all β_i being zero. We define $\beta_0 = -\sum_{i=1}^k \beta_i$ and observe that

$$\sum_{i=1}^k \beta_i v_i + \beta_0 v_0 = 0 \quad \text{and} \quad \sum_{i=0}^k \beta_i = 0.$$

Choose $K \in \{0, 1, \dots, k\}$ so that $\beta_K \neq 0$. Then

$$v_K = \sum_{i=0, i \neq K}^k -\frac{\beta_i}{\beta_K} v_i \quad \text{and} \quad \sum_{i=0, i \neq K}^k -\frac{\beta_i}{\beta_K} = 1.$$

Hence points of V are not affinely independent. □

Proposition 3.1.2. *Suppose points of $V = \{v_0, v_1, \dots, v_k\} \subset \mathbb{R}^d$ are affinely independent. Then for each point $x \in \text{Conv}(V)$ there exist unique coefficients $\alpha_i \in [0, 1], i \in \{0, 1, \dots, k\}$, such that*

$$x = \sum_{i=0}^k \alpha_i v_i \quad \text{and} \quad \sum_{i=0}^k \alpha_i = 1.$$

Coefficients α_i in are called **barycentric coordinates** of point x in $\text{Conv}(V)$.

Proof. The existence of such coefficients α_i follows from $x \in \text{Conv}(V)$. In order to prove the coefficients are unique assume the statement holds for two different sets of coefficients α_i and α'_i , i.e.,

$$x = \sum_{i=0}^k \alpha_i v_i = \sum_{i=0}^k \alpha'_i v_i \quad \text{and} \quad \sum_{i=0}^k \alpha_i = \sum_{i=0}^k \alpha'_i = 1.$$

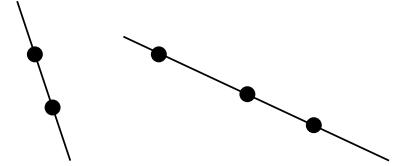


Figure 3.2: The affine hull of the two points on the left is a line. The affine hull of the three colinear points on the right is also a line, implying these three points are not affinely independent.

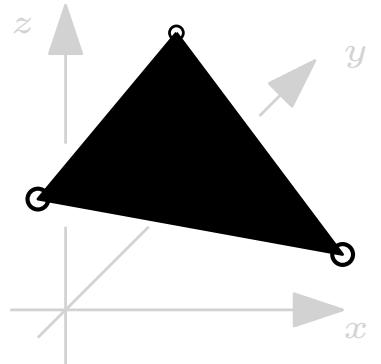


Figure 3.3: The convex hull of three affinely independent points is a triangle. The affine hull is the supporting plane of the triangle.

☞ A linearly independent collection of vectors in \mathbb{R}^d can have at most d elements. An affinely independent collection of points in \mathbb{R}^d can have at most $d+1$ elements.

At some index i the coefficients α_i and α'_i differ. Without loss of generality we can assume that index is zero, i.e., $\alpha_0 - \alpha'_0 \neq 0$. Then

$$(\alpha_0 - \alpha'_0)v_0 = \sum_{i=1}^k (\alpha'_i - \alpha_i)v_i$$

and

$$v_0 = \sum_{i=1}^k \frac{\alpha'_i - \alpha_i}{\alpha_0 - \alpha'_0} v_i,$$

which contradicts the assumption that the points of V are affinely independent. \square

3.2 Geometric simplicial complex

We are now ready to define our basic building blocks.

Definition 3.2.1. Let $k, d \in \{0, 1, \dots\}$ with $k \leq d$. A **geometric k -simplex** σ in \mathbb{R}^d is the convex hull of an affinely independent family $V = \{v_0, v_1, \dots, v_k\} \subset \mathbb{R}^d$, i.e., $\sigma = \text{Conv}(V)$.

The following is some terminology related to a geometric simplex $\sigma = \text{Conv}(\{v_0, v_1, \dots, v_k\})$:

- **Dimension:** $\dim(\sigma) = k$. We sometimes express it by writing it as superscript: $\sigma = \sigma^k$.
- **Vertices** of σ : v_0, v_1, \dots, v_k .
- **Edges** of σ : convex hulls of pairs of vertices.
- We say that σ is **spanned** by the set of its vertices.
- If simplex τ is spanned by a subset of the vertices of σ , we say that:
 - τ is a **face** of σ .
 - σ is a **coface** of τ .
 - τ is a **facet** of σ if $\dim(\tau) = \dim(\sigma) - 1$.

Note that $\sigma^k \cong D^k$. By Proposition 3.1.2 each point of σ is uniquely described by its barycentric coordinates using the vertices of σ .

We can now use these building blocks to assemble more complicated spaces.

Definition 3.2.2. Let $d \in \{0, 1, \dots\}$. A (finite) **geometric simplicial complex** $K \subset \mathbb{R}^d$ is a (finite) collection of geometric simplices, such that:

- a: If $\sigma \in K$ and τ is a face of σ , then $\tau \in K$.

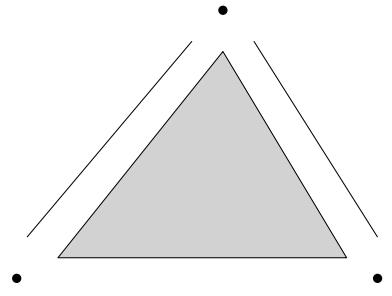


Figure 3.4: A two-dimensional simplex has six simplices as faces (three edges and three vertices), three of which are facets (edges).

☞ All our simplicial complexes will be finite. For that reason we will be dropping the word “finite”. There also exist simplicial complexes with infinitely many simplices. However, a proper definition of infinite simplicial complexes brings along additional technicalities which we want to avoid in our context.

b: If $\sigma, \tau \in K$, then $\sigma \cap \tau$ is either empty or a common face of both.

Each planar triangulation has a corresponding simplicial complex consisting of all triangles, edges and vertices of the triangulation.

Let K be a simplicial complex. We define:

- **Dimension** $\dim(K) = \max_{\sigma \in K} \dim(\sigma)$. A one-dimensional simplicial complex is a graph.
- **Vertices** of K as the collection of all vertices of all simplices of K .
- **Edges** of K as the collection of all edges of all simplices of K .
- A geometric simplicial complex L is a **subcomplex** of K [notation: $L \leq K$], if $L \subseteq K$.
- For $n \in \{0, 1, \dots\}$ the n -**skeleton** of K [notation: $K^{(n)}$] is the simplicial subcomplex of K consisting of all simplices of K of dimension at most n . For example, $K^{(0)}$ is the set of vertices of K .
- The **body** of K [notation: $|K|$] is the union of all simplices of K .

Formally speaking, a geometric simplicial complex in \mathbb{R}^d is a collection of simplices and its body is a subset in \mathbb{R}^d . In practice however we will be often identifying the two objects in geometric discussions. From now on we will be visualizing simplicial complexes by drawing their body and assuming the underlying structure of a simplicial complex.

We are now ready to describe a connection between a metric subspace of a Euclidean space and its combinatorial description.

Definition 3.2.3. Let $d \in \{0, 1, \dots\}$. A **triangulation** of a subspace $X \subset \mathbb{R}^d$ is a simplicial complex K in \mathbb{R}^d , such that $|K| \cong X$.

Not every subspace of \mathbb{R}^d admits a triangulation. However, all the subsets that will arise in our context will admit it. Triangulations of $B_{\mathbb{R}^2}((0, 0), 1)$ include examples in Figure 3.8 and Delaunay triangulations. A geometric simplicial complex is a triangulation of its body.

Occasionally we will want to refine the triangulation of a space, meaning we will want to decrease the size of simplices in order to improve visualisation, level of details, etc. Such refinements are called subdivisions. Given a geometric simplicial complex K , a geometric simplicial complex L is its **subdivision**, if each simplex of K is the union of a collection of simplices from L . As an example we already mentioned the barycentric subdivision of planar triangulations, which also exists for simplicial complexes. At this point we will refrain from introducing its formal definition. The lower right part of Figure 3.8 depicts a subdivision of a single 2-simplex, see also Figure 3.10.

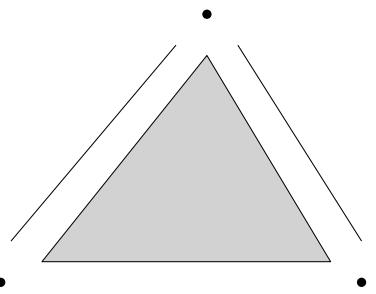


Figure 3.5: The smallest two-dimensional simplicial complex consists of a triangle and all its faces: three edges and three vertices.

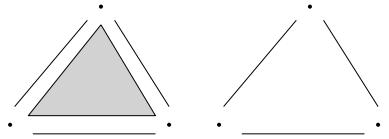


Figure 3.6: The simplicial complex from Figure 3.5 (left) and its 1-skeleton (right) consisting of three edges and three vertices.

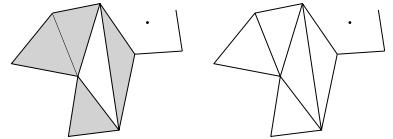


Figure 3.7: On the left there is a geometric simplicial complex presented by drawing its body. Each edge of a sketched triangle and each vertex of a sketched edge is assumed to be in the complex. On the right is the 1-skeleton of the simplicial complex on the left.

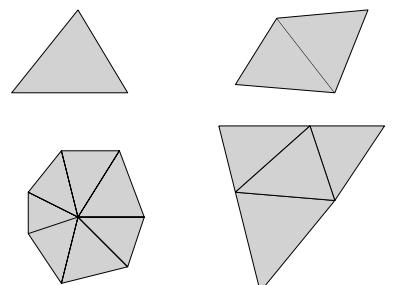


Figure 3.8: A few triangulations of D^2 .

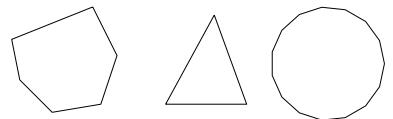


Figure 3.9: A few triangulations of S^1 .

3.3 Abstract simplicial complex

When buying a commercial object to be assembled, be it a piece of furniture, a toy or a model made of cubes, or a picture made of puzzles, the package usually arrives in a big box. On the box is a picture of the object, which in our context represents the body of a geometric simplicial complex. On the picture we can often determine pieces, which in our setting would be geometric simplicial simplices. Pieces on that picture have specific locations and just like geometric simplices, could be described by specific coordinates. However, the assembly instructions contain no coordinates. There is a good reason for that¹. In order to assemble the object, the instructions only provide a list of pieces and instructions about how to put them together. That information is sufficient to reconstruct the object. Abstract simplicial complexes play the role of such instructions.

Assume we want to describe a geometric simplicial complex. That means we have to provide a list of all simplices. A simplex could be provided by a list of coordinates of its vertices, but then we also have to make sure the simplices intersect appropriately. It would be much easier to just list the simplices and describe how they fit together in a coordinate free way. Here is a way to do it.

Definition 3.3.1. Let V be a finite set. An **abstract simplicial complex** L on V is a family of non-empty subsets of V , such that if $\sigma \in L$ and $\tau \subseteq \sigma$ is non-empty, then $\tau \in L$.

A few more accompanying definitions using the notation of Definition 3.3.1:

- An **abstract simplex** σ is an element of L . Its **dimension** is $\dim(\sigma) = |\sigma| - 1$.
- If $\tau \subseteq \sigma \in L$, then:
 - τ is a **face** of σ .
 - σ is a **coface** of τ .
 - τ is a **facet** of σ if $\dim(\tau) + 1 = \dim(\sigma)$.
- **Dimension** $\dim(L)$ of L is the maximal dimension of a simplex in L .
- The (closed) **star** of a vertex $v \in K$ is $\text{St}_K(v) = \text{St}(v) = \{\sigma \in K \mid \sigma \cup \{v\} \in K\} \leq K$.
- The **link** of a vertex $v \in K$ is $\text{Lk}_K(v) = \text{Lk}(v) = \{\tau \in \text{St}(v) \mid v \notin \tau\} \leq \text{St}(v)$.

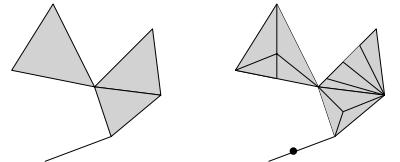


Figure 3.10: A geometric simplicial complex and its subdivision.

¹ Besides the fact that nobody would purchase such an item.

In some sources the non-empty condition in the definition of an abstract simplicial complex is omitted and the empty set is always included as an abstract simplex of dimension -1 .

A geometric simplex is a subset of an Euclidean space, given as the convex hull of the collection of its vertices. Each vertex is given as a point in space, usually in terms of coordinates. A geometric simplicial complex is a set of such simplices, contains all faces and has to satisfy the intersection properties of Definition 3.2.2.

An abstract simplex is just a collection of vertices. No coordinates are needed. An abstract simplicial complex is a set of such collections which contains all faces (all subsets of its elements). There are no intersections to be checked. It is a complete and convenient combinatorial description.

Example 3.3.2. Let K be a geometric simplicial complex provided by Figure 3.12. As a geometric simplicial complex, K contains specific geometric simplices described by the coordinates of their vertices. We can construct a corresponding abstract simplicial complex L . Label the vertices as demonstrated by the figure. Then

$$L = \{\{a, c, d\}, \{a, b\}, \{b, c\}, \{c, d\}, \{d, a\}, \{a, c\}, \{a\}, \{b\}, \{c\}, \{d\}\}.$$

No coordinates are involved. We could also only list the inclusion-maximal simplices, which completely determine the simplicial complex: $\{\{a, c, d\}, \{a, b\}, \{b, c\}\}$.

A simpler structure of an abstract simplicial complex will suffice for most of our topological analysis of spaces and the corresponding computations. Indeed, it will simplify them. A geometric simplicial complex however is still useful when we want to visualise a complex. For example, outputs of various scans come in the form of geometric simplicial complexes modelling the scanned shape. While geometric simplicial complexes describe geometric information about the space (various sizes, lengths, etc.), abstract simplicial complexes contain only topological information (homeomorphism type).

It is easy to turn a geometric simplicial complex into an abstract simplicial complex: replace each coordinate given vertex by a unique label. The opposite is a bit harder. Turning an abstract simplicial complex into a geometric simplicial complex requires us to choose coordinates of vertices in line with the requirements for a geometric simplicial complex. If it can be done, such a geometric simplicial complex is called a **geometric realization** (or just realization) of the original abstract simplicial complex. It turns out that geometric realizations always exists, although obtaining them in a low-dimensional space is typically hard. The following are two special cases of such realizations.

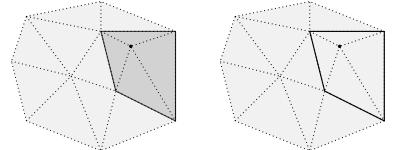


Figure 3.11: A star (left) and a link (right) of a vertex.

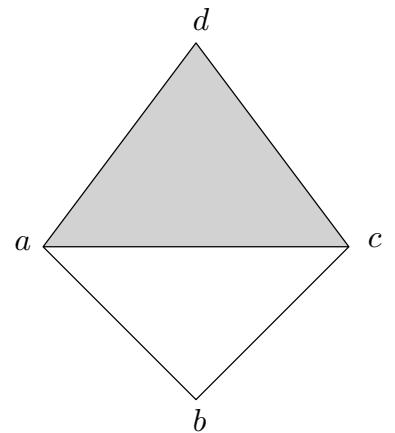


Figure 3.12: A picture accompanying Example 3.3.2.

Theorem 3.3.3. Every abstract simplicial complex K with n vertices admits a geometric realization in \mathbb{R}^{n-1} .

Proof. Simplicial complex K is a subcomplex of the full simplicial complex L on n vertices, i.e., the simplicial complex, whose simplices are all subsets of vertices of K . As L admits a realization in \mathbb{R}^{n-1} as an $(n-1)$ -simplex (i.e., the convex hull of a collection of n affinely independent points), so does K as its subcomplex. \square

Theorem 3.3.4. Every abstract simplicial complex of dimension d admits a geometric realization in \mathbb{R}^{2d+1} .

A proof of Theorem 3.3.4 is provided in Appendix.

As an example consider graphs, i.e., one-dimensional simplicial complexes. It is well known that some graphs are planar, which means they admit a geometric realization in the plane. However, there are graphs, that are not planar. These graphs can only be realized in $\mathbb{R}^3 = \mathbb{R}^{2+1}$ (and, of course, in \mathbb{R}^m for $m > 3$). See Figure 3.13.

One of the goals of this course is the following: given an abstract simplicial complex, extract topological properties of its geometric realization. We will study and analyze spaces by working with their triangulations.

Remark 3.3.5. It is important to understand the differences between a metric space, its triangulation and a corresponding abstract simplicial complex. In practice however, we will be frequently vague in our expression for the sake of simplicity, often referring to just simplicial complex. Given a picture of a space like the one in Figure 3.12, we will keep in mind the three possible interpretations and use the one that fits the context at the moment.

Also note that the terminology is essentially the same for the abstract and geometric simplicial complexes. We declare this to be the case for all further² definitions as well. For example, an abstract simplicial complex L is a triangulation of a metric space X if the corresponding geometric simplicial complex is, i.e., if the body of a geometric realization of L is homeomorphic to X .

Example 3.3.6. One of our standard examples of a metric space will be the torus T . It is a two-dimensional metric space, actually a surface, depicted in Figure 3.15. A triangulation of T in terms of an abstract simplicial complex is provided by Figure 3.17.

Topologically speaking, the torus can be obtained from a square by identifying the opposite sides along the same direction. This construction is depicted in Figure 3.16.

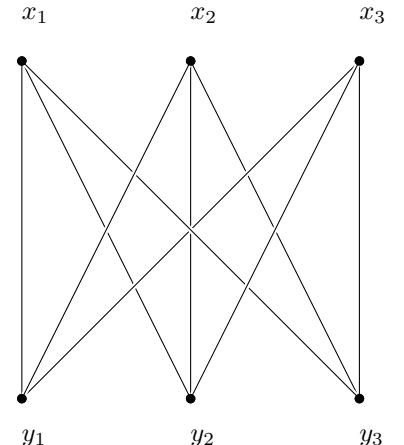


Figure 3.13: A sketch of a one-dimensional abstract simplicial complex (graph) with no realization in \mathbb{R}^2 . The complex consists of all edges between x_i and y_j .

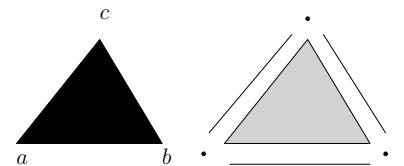


Figure 3.14: A metric space (left) and the corresponding geometric simplicial complex (right). The corresponding abstract simplicial complex is $\{\{a,b,c\}, \{a,b\}, \{b,c\}, \{a,c\}, \{a\}, \{b\}, \{c\}\}$.

² Including the concepts of link and star of a complex at a point.

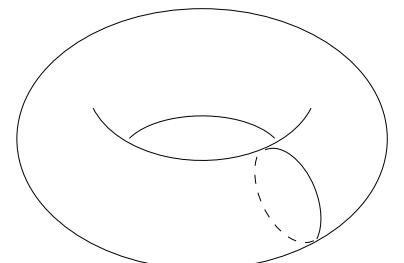


Figure 3.15: Torus.

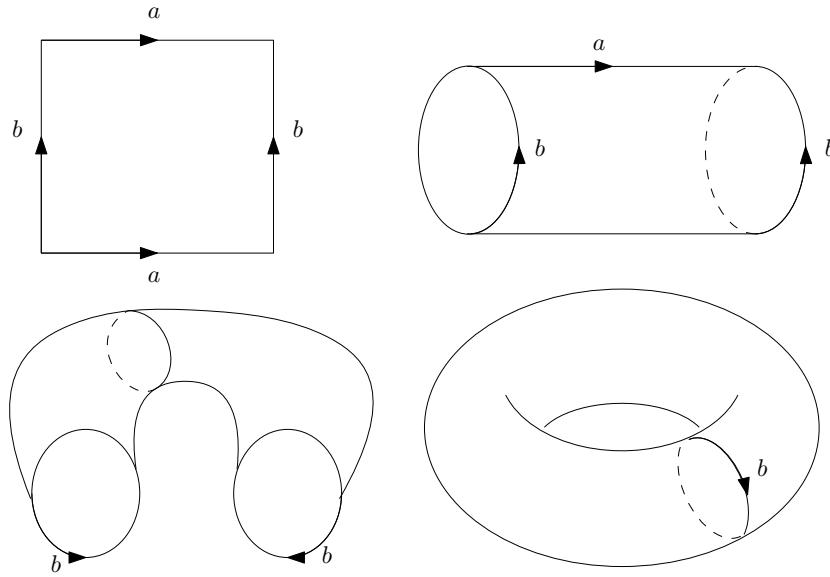


Figure 3.16: The torus arising from a square. Starting with a square (top left) we identify its pairs of opposite sides along the same direction as the labels and arrows suggest. Identifying the sides a we obtain a cylinder (top right). Identifying the other pair of sides, which represent loops b in the cylinder, we obtain the torus (bottom right).

As a result we can obtain a structure of an abstract simplicial complex by triangulating a square and respecting the mentioned identifications. This provides a convenient topological visualisation of a torus. Observe that a triangulation of T in terms of a geometric simplicial complex would be more complicated and not presentable in the plane.

The mentioned abstract simplicial complex is provided by Figure 3.17. We divide the square into 18 triangles and keep in mind the identifications suggested by the arrows. The two sides of the square along the single arrows get identified along the direction of the arrows, and the same holds for the two sides along the double arrows. For the sake of clarity we also labeled the outer vertices of the triangles as each label appears at least twice due to the identifications.

Example 3.3.7. Choose $n \in \{1, 2, \dots\}$. In this example we provide the simplest triangulations of discs and spheres.

Let σ^n be an n -simplex and define K to be the simplicial complex whose only maximal simplex is σ^n , i.e., K contains σ^n and all of its faces. Simplicial complex K is a triangulation of D^n .

To obtain a triangulation of S^{n-1} remove from K the maximal simplex, i.e., $K' = K \setminus \{\sigma^n\}$. Simplicial complex K consists of all faces of σ^n but does not contain σ^n itself. Simplicial complex K' is a triangulation of S^{n-1} .

Two invariants

Here we provide two invariants (of a space) that can be extracted from a triangulation. Both are homotopy invariants (and hence also

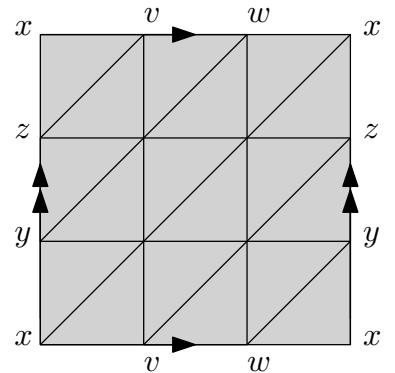


Figure 3.17: A triangulation of a torus in terms of an abstract simplicial complex.

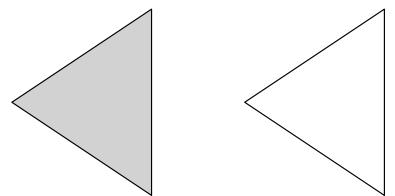


Figure 3.18: A triangle as a triangulation of D^2 and its boundary as a triangulation of S^1 . In a similar way a solid tetrahedron is a triangulation of D^3 while its boundary is a triangulation of S^2 .

topological invariants), meaning they coincide for homotopically equivalent spaces. A space typically has infinitely many possible triangulations. Imagine all possible Delaunay triangulations in \mathbb{R}^2 : they are all triangulations of D^2 . We conclude that the numbers of vertices, edges, or higher dimensional simplices in a triangulation cannot be topological invariants.

The first invariant is the number of components. Given a triangulation K , it is easy to extract that number from $K^{(1)}$ (which is a graph) using standard approaches of graph theory. Later we will explain in detail how to obtain this number in terms of homology. For example, the simplicial complex in Figure 3.19 has two components.

The second invariant is the Euler characteristic, which can be defined for simplicial complexes.

Definition 3.3.8. Suppose K is a simplicial complex and let n_i denote the number of i -simplices in K . The **Euler characteristic** $\chi(K) \in \mathbb{Z}$ is defined as $\chi(K) = n_0 - n_1 + n_2 - n_3 + \dots$

The Euler characteristic of a metric space is the Euler characteristic of any of its triangulations.

As was mentioned above, the Euler characteristic is homotopy invariant. Using this fact we can compute the following cases:

- Let X be a one-point space. Then $\chi(X) = 1$. Since each Delaunay triangulation K in \mathbb{R}^2 is homotopic to a point (meaning that $|K| \simeq X$), we also conclude $\chi(K) = 1$, a statement which we have already proved directly. In fact, the homotopy invariance implies that each triangulation of a contractible space is of Euler characteristic 1.
- The Euler characteristic of a torus is 0. It can be computed directly from the triangulation presented by Figure 3.17, which has 18 triangles, 27 edges and 9 vertices (keep in mind the identifications).
- Let $n \in \{0, 1, 2, \dots\}$. Then $\chi(S^n) = 1 + (-1)^n$.

Proof. Let σ^{n+1} be an $(n+1)$ -simplex and define K to be the simplicial complex whose only maximal simplex is σ^{n+1} . As we know K is a triangulation of D^{n+1} . As D^{n+1} is contractible, $\chi(K) = 1$. We also mentioned that $K' = K \setminus \sigma^{n+1}$ is a triangulation of S^n . As K' is obtained from K by removing an $(n+1)$ -simplex, $\chi(K')$ is obtained from $\chi(K)$ by removing a contribution of that simplex, which is $(-1)^{n+1}$. Hence $\chi(S^n) = 1 - (-1)^{n+1} = 1 + (-1)^n$. \square

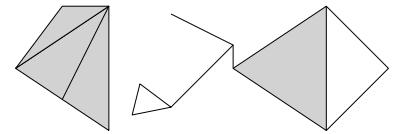


Figure 3.19: A complex with two components.

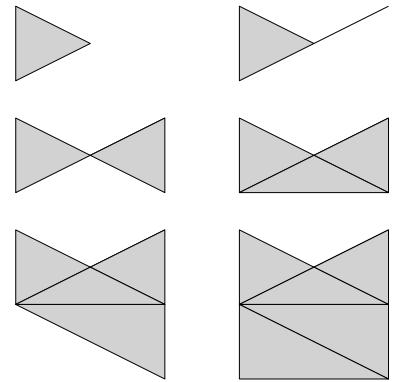


Figure 3.20: A series of homotopic simplicial complexes, each differing from the previous one by a small modification. Note that the modifications preserve the Euler characteristic. In the first step we add a vertex and an edge; in the second step we add a vertex, two edges and a triangle; and so on. Each new addition contributes a total of 0 to the Euler characteristic.

☞ $\chi(S^n)$ could also be computed from the triangulation K' directly using the binomial formula.

3.4 Simplicial maps

Just as simplicial complexes provide a convenient combinatorial description of metric spaces, simplicial maps provide a combinatorial description of continuous maps. We first define them in the abstract setting.

Definition 3.4.1. Suppose K and L are abstract simplicial complexes.

A **simplicial map** between K and L is an assignment $f: K^{(0)} \rightarrow L^{(0)}$ on vertices, such that for each abstract simplex $\{v_0, v_1, \dots, v_k\} \in K$ its image $\{f(v_0), f(v_1), \dots, f(v_k)\}$ is an abstract simplex in L .

Remark 3.4.2. A simplicial map will be usually denoted by $f: K \rightarrow L$. However, since K and L are collections of sets, such a notation would formally include maps that map, say, a vertex to an edge, a highly unfavourable occurrence. When talking about simplicial maps $K \rightarrow L$ we thus always consider only maps in the sense of Definition 3.4.1, i.e., maps that map a vertex to a vertex, while images on simplices are always determined by the values on the vertices:

- For each vertex $v \in K$ we define a corresponding vertex $f(v)$.
- For each abstract simplex $\sigma = \{v_0, v_1, \dots, v_k\} \in K$ its image $f(\sigma) = \{f(v_0), f(v_1), \dots, f(v_k)\}$ is determined by the values on its vertices.
- Note that $f(\sigma)$ is a set, meaning there are no repetitions of elements. In particular this means that each vertex appears at most once in $f(\sigma)$, even if it appears multiple times as $f(v_i)$. As a result the image of an n -dimensional simplex can be of dimension less than or equal to n , but never more than n .

Example 3.4.3. Let K be the simplicial complex in Figure 3.22. Assignment $a \mapsto a$; $b \mapsto c$; $c \mapsto c$; $d \mapsto d$; $e \mapsto b$ can be verified to induce a simplicial map $K \rightarrow K$. Note that triangle $\{a, b, c\}$ gets mapped to edge $\{a, c\}$.

We are now ready to define simplicial maps in the geometric setting.

Definition 3.4.4. Suppose K and L are geometric simplicial complexes. A map $f: K \rightarrow L$ is a **simplicial map**, if:

1. For each vertex v of K its image $f(v)$ is a vertex of L .
2. The corresponding map between the corresponding abstract simplicial complexes is simplicial, i.e., if $\{v_0, v_1, \dots, v_k\}$ span a geo-

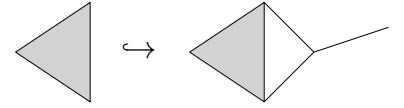


Figure 3.21: An embedding of a simplicial subcomplex $L \leq K$ into K is identity on the vertices and always a simplicial map.

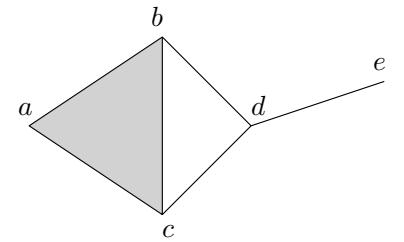


Figure 3.22: Simplicial complex K of example 3.4.3.

metric simplex in K then $\{f(v_0), f(v_1), \dots, f(v_k)\}$ span a geometric simplex in L .

3. Map f is linear on simplices (in terms of barycentric coordinates), i.e.,

$$\forall t_i \in [0, 1], \quad \sum_{i=0}^k t_i = 1, \quad \forall v_i \in K^{(0)} : \quad f\left(\sum_{i=0}^k t_i v_i\right) = \sum_{i=0}^k t_i f(v_i).$$

Given a simplicial map between geometric simplicial complexes the induced map (i.e., the restriction to vertices) between abstract simplicial complexes is simplicial. Conversely, each simplicial map between abstract simplicial complexes corresponds to the unique simplicial map between the corresponding geometric simplicial complexes: the extension from vertices to geometric simplices is defined using the formula of item 3 of Definition 3.4.4. In accordance with our declarations simplicial maps will be used to denote maps either in geometric or abstract setting: in case of a preferred interpretation it will be stated explicitly or should be obvious from the context.

Simplicial maps between geometric simplicial complexes are continuous maps as they are linear (hence) continuous on each simplex. Surprisingly enough, each continuous map can be (up to homotopy) represented by a simplicial map, which means that as long as we are interested in homotopical properties, we can restrict ourselves to simplicial maps.

Theorem 3.4.5. Suppose $f: K \rightarrow L$ is a continuous map between geometric simplicial complexes. Then there exist sufficiently fine subdivisions K' of K and L' of L , and a simplicial map $f': K' \rightarrow L'$, such that $f \simeq f'$. We call f' a **simplicial approximation** of f .

The subdivisions above can be taken to be sufficiently fine barycentric subdivision. A continuous map between simplicial complexes is formally a map between the bodies of the simplicial complexes. In this sense both f and f' map $|K| = |K'|$ to $|L| = |L'|$ hence $f \simeq f'$ makes sense.

Elementary collapses

Elementary collapses are minor local modifications of simplicial complexes, which preserve its homotopy type. Conveniently enough, they can be described in purely combinatorial terms. Their importance stems from the following lemma.

Lemma 3.4.6. Let K be a geometric simplicial complex containing simplex $\sigma = \{v_0, v_1, \dots, v_k\}$ and let $\tau = \{v_1, \dots, v_k\}$ be its facet. If

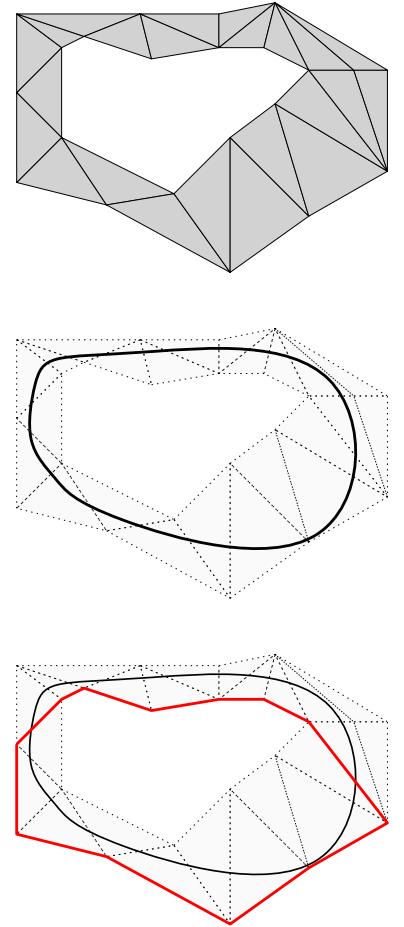


Figure 3.23: The upper part of the figure presents a simplicial complex K . The middle part consists of a smooth curve (bold) given by a continuous map $f: S^1 \rightarrow K$, superposed on the simplicial complex. The bottom part consists of a simplicial approximation (bold red) of the curve superimposed on the simplicial complex and the curve. The indicated simplicial approximation requires a triangulation of the domain S^1 with at least 12 edges.

σ is the only coface³ of τ , then the inclusion $i: K \setminus \{\tau, \sigma\} \hookrightarrow K$ is a homotopy equivalence.

³ i.e., a simplex which contains τ .

Proof. The proof is sketched in Figure 3.24.

We first need to subdivide σ and τ . Choose a point a in the middle of τ and connect it to all vertices of σ . This induces a subdivision of σ and τ and in fact of K as no other simplex contains⁴ σ or τ .

In order to obtain a continuous deformation⁵ from K to $K \setminus \{\tau, \sigma\}$, slide a towards v_0 . This sliding is a linear homotopy and can be easily described in the barycentric coordinates of the new subdivision of K . \square

⁴ If there was another simplex containing σ , then that simplex would have been a coface of τ , which contradicts our assumptions. In particular, If σ is the only coface, then τ is a facet of σ . This fact also refers to conditions of Definition 3.4.7.

⁵ Homotopy equivalence.

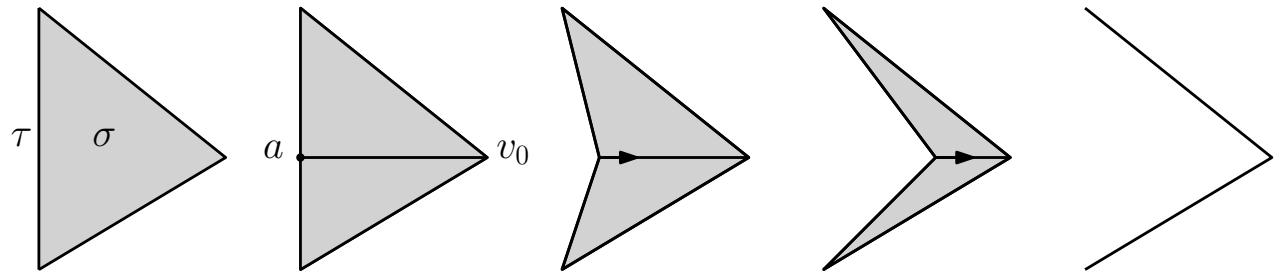


Figure 3.24: Elementary collapse from Lemma 3.4.6 with σ being the triangle and τ its left side.

Definition 3.4.7. Let K be a simplicial complex, $\tau^{(k-1)} \subset \sigma^{(k)} \in K$, and assume σ is the only coface of τ . A removal $K \rightarrow K \setminus \{\tau, \sigma\}$ is called an **elementary collapse**.

By Lemma 3.4.6 each elementary collapse preserves the homotopy type of a complex. Note that the collapsing map $K \rightarrow K \setminus \{\tau, \sigma\}$ is not⁶ a simplicial map on K . It is, however, a simplicial map on the subdivision of K employed in Lemma 3.4.6, defined by mapping $a \mapsto v_0$ and keeping all the other vertices intact. Its homotopy inverse is the inclusion $K \setminus \{\tau, \sigma\} \hookrightarrow K$, which is a simplicial map.

Elementary collapses are convenient because they provide us with a simple combinatorial condition that can be used to induce homotopy equivalence on abstract simplicial complexes. This idea will be further expanded later within the context of discrete Morse theory.

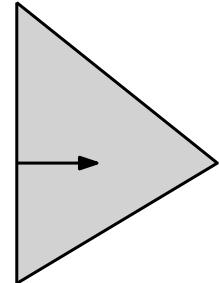


Figure 3.25: The elementary collapse of Figure 3.24 is usually indicated by an arrow from τ into σ .

⁶ Except if τ is a vertex.

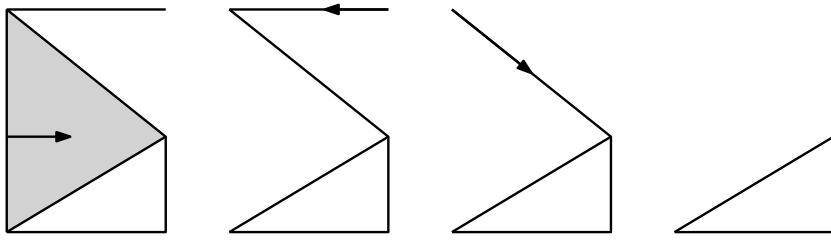


Figure 3.26: An example of a simplification of (the homotopy type of) a simplicial complex using elementary collapses.

3.5 Concluding remarks

Recap (highlights) of this chapter

- Geometric simplex;
- Geometric simplicial complex;
- Abstract simplicial complex;
- Geometric realization;
- Simplicial map;
- Simplicial approximation and elementary collapse.

Background and applications

Simplicial complexes model spaces in a wide spectrum of theory and applications. Great portions of topology and geometry are based on them due to their simple structure and amenability to combinatorial treatment. On the applied side simplicial complexes are typically used to model shapes.

The notions introduced in this chapter are covered in standard books on topology⁷.

Appendix: Proof of Theorem 3.3.4

Before we begin with the proof we clarify a fact that will be used. A generic⁸ (random) collection of $n + 1$ points in \mathbb{R}^n is affinely independent. Geometrically this is easy to believe:

- Generic two points in \mathbb{R} will be different;
- Generic three points in \mathbb{R}^2 will not be colinear;
- Generic four points in \mathbb{R}^3 will not be coplanar.

Similarly, even for $k > n + 1$ a generic set V of k points in \mathbb{R}^n has the same property: each collection of $n + 1$ points in V is affinely

⁷ James Munkres. Elements of Algebraic Topology. Perseus Books, 1984. doi: [10.1201/9780429493911](https://doi.org/10.1201/9780429493911)

⁸ Notion “generic” as used in this appendix is usually referred to as “general linear position” in the literature.

independent. For example, in a generic collection of points in the plane no triple of points will be collinear. We will only use the fact that generic collections exist. This fact can be proved using linear algebra.

Proof of Theorem 3.3.4. Let K be an abstract simplicial complex of dimension d , whose vertices are v_0, v_1, \dots, v_k . Choose a generic collection of points $V = \{x_0, x_1, \dots, x_k\} \subset \mathbb{R}^{2d+1}$, meaning that each collection of $2d + 2$ points from V is affinely independent. We will prove that the correspondence $v_i \leftrightarrow x_i$ for all $i \in \{0, 1, \dots, k\}$ provides a geometric realization of K .

For each abstract simplex $\sigma \in K$ spanned by $v_{j_0}, v_{j_1}, \dots, v_{j_m}$, the corresponding geometric simplex σ' is spanned by the affinely independent collection $x_{j_0}, x_{j_1}, \dots, x_{j_m}$. It remains to prove that if $\sigma, \tau \in K$, then $\sigma' \cap \tau' = (\sigma \cap \tau)'$.

As $(\sigma \cap \tau)' \subseteq \sigma', \tau'$ by the definition, we have $\sigma' \cap \tau' \supseteq (\sigma \cap \tau)'$.

To prove the other inclusion we will make use of the dimension assumption. Let $z \in \sigma' \cap \tau'$, which means that z can be expressed as a convex combination of vertices in σ' and also as a convex combination of vertices in τ' . As the total number of vertices in τ' and σ' is at most $2n + 2$ (by the dimension assumption), the generic condition implies these are affinely independent, and thus the convex (affine) combinations above coincide as they have to be unique by Proposition 3.1.2. In particular, this means that only the barycentric coordinates corresponding to the vertices that lie in both simplices (i.e., $\sigma \cap \tau$) can be non-zero, which implies $z \in (\sigma \cap \tau)'$ and hence $\sigma' \cap \tau' \subseteq (\sigma \cap \tau)'$. \square

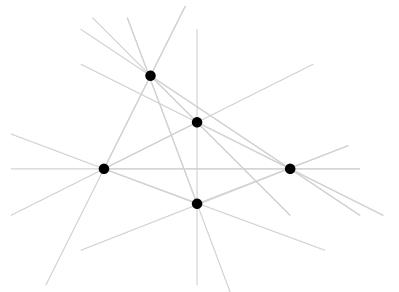


Figure 3.27: A generic collection of 5 points in the plane, meaning that no three are collinear. Only by adding a point on any grey line does the generic condition brake. If we add any other point to this collection, the obtained collection of 6 points is still generic.

4

Surfaces

SURFACES ARE SOME of the simplest topological spaces appearing frequently in science and data analysis. From each local perspective appearing as a part of the plane, the global shape of a surface may take many forms. Think of the surface of the earth: because it appears to be “planar” at each point, it had long been believed that Earth is actually a part of a plane or maybe a disc instead of a sphere.

Surprisingly enough, most surfaces of interest can be recognized up to homeomorphism fairly easily. In this chapter we explain this recognition process and the accompanying theory, both of which will come handy in the chapters to come.

4.1 Surfaces as manifolds

Ever since the ancient times people were wondering about the shape of the world. They agreed that from the perspective of a human being, the world looked like a plane, a part of a surface. What was much harder to figure out was the global picture. The first and most obvious idea was that the world was a flat disc. Later came indications, such as deviations in the angle of the shadow depending on the latitude, that the world might be curved. Magellan’s first circumnavigation of Earth does not constitute a rigorous proof that the world is a sphere by modern mathematical standards, but at the time it was a momentous achievement which confirmed that the Earth is indeed round.

While we will not be sailing around the world in this course, we will be interested in the moral of this story: things that locally look like a plane may globally not be a plane. We will want to determine the global structure from local information.

Spaces that locally look like a plane are called surfaces and their generalizations to other dimensions are called manifolds. Here is a formal definition.

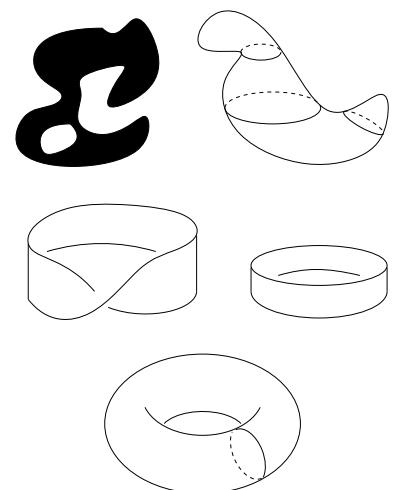


Figure 4.1: Some of the surfaces we mentioned before: a planar set (top left), a space homeomorphic to S^2 (top right), Moebius band (center left), cylinder (center right), and torus (bottom).

Definition 4.1.1. Let $n \in \{0, 1, 2, \dots\}$. A metric space X is an n -manifold, if for each $x \in X$ there exists $r > 0$, such that $B_X(x, r)$ is homeomorphic to the n -dimensional disc D^n .

2-manifolds are called **surfaces**.

Point x on an n -manifold X is:

- a **boundary point** if a homeomorphism $B_X(x, r) \rightarrow D^n$ from Definition 4.1.1 maps x to a point on the boundary of D^n .
- an **interior point** if a homeomorphism $B_X(x, r) \rightarrow D^n$ from Definition 4.1.1 maps x to a point in the interior of D^n .

These two notions are independent of the choice of homeomorphism $B_X(x, r) \rightarrow D^n$. Each point of X is either a boundary point or an interior point. The **boundary** of X consists of all the boundary points, and the **interior** of X consists of all the interior points.

We say that a manifold X is **without boundary**, if it has no boundary points. For an n -manifold Y its boundary is an $(n - 1)$ -manifold without boundary, as can be seen from the examples below. For our purposes a **closed manifold** (closed surface) will be a manifold (surface) without boundary admitting a (finite) triangulation.

Example 4.1.2. We provide some examples of connected n -manifolds listed by dimension n .

- $n = 0$: This one is fairly unimpressive: a single point.
- $n = 1$: Circle and intervals $(0, 1), [0, 1], (0, 1]$. Each connected 1-manifold is homeomorphic to one of these. A circle and an open interval have no boundary, while the boundary of $[0, 1]$ consists¹ of 0 and 1.
- $n = 2$: We will provide a list of all surfaces by the end of this chapter. Here we list some of the more prominent ones. The already mentioned ones are recapped in Figure 4.1: note that the boundary of the band consists of two copies of S^1 , while the Moebius band has a single boundary component. A closed disc D^2 is also a surface, whose boundary is S^1 .

Closely related to the torus are the **Klein bottle** (see Figure 4.3) and the **projective plane**, neither of them has a boundary and neither can be obtained as a subset of \mathbb{R}^3 . However, they can be obtained as subsets of \mathbb{R}^4 . While these two spaces are challenging to imagine geometrically, it is fairly easy to provide their (abstract) triangulations (see Figure 4.3) and compute some of their topological invariants, such as the Euler characteristic. The Torus and the

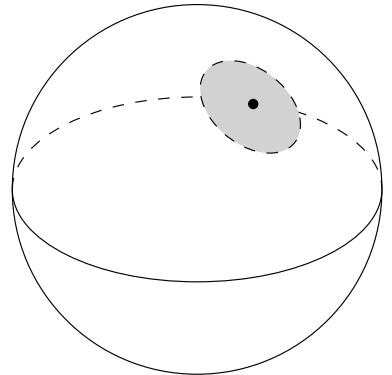


Figure 4.2: A neighborhood of a point on S^2 homeomorphic to D^2

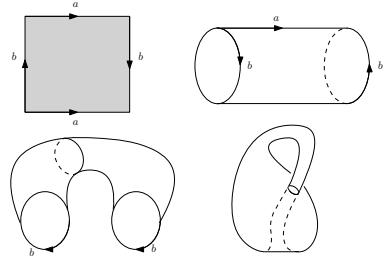


Figure 4.3: Klein bottle obtained by identifying the edges of a square: two along the same direction and two along the opposite direction. The resulting space (bottom right) is not realizable in \mathbb{R}^3 due to the self-intersection. However, Klein bottle can be embedded into \mathbb{R}^4 without self-intersections.

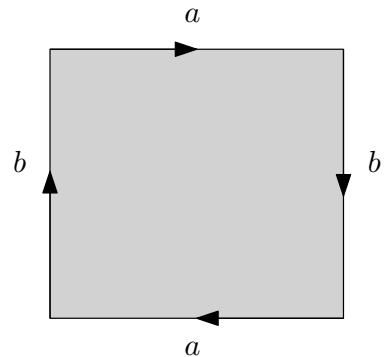


Figure 4.4: Projective plane obtained by identifying the edges of a square: both pairs along the opposite direction. The resulting space is not realizable in \mathbb{R}^3 . However, it can be embedded into \mathbb{R}^4 without self-intersections.

¹ Also, the boundary of $[0, 1]$ is 0.

Klein bottle have Euler characteristic 0, while the Euler characteristic of the projective plane is 1.

The projective space is homeomorphic to the space of all 1-dimensional subspaces in \mathbb{R}^3 .

- General n : D^n and S^n are both n -manifolds. The boundary of D^n is S^{n-1} , while S^n has no boundary. There are many other n -manifolds.

Combinatorial manifolds

We will mostly be working with triangulated manifolds. A natural question that arises in this context is how to recognize whether a given simplicial complex is a triangulation of a manifold. Tackling this task we first introduce nice combinatorial descriptions of manifolds.

Definition 4.1.3. Suppose K is a simplicial complex and $n \in \mathbb{N}$. We say that K is a **combinatorial n -manifold**, if for each vertex $v \in K$ its link $\text{Lk}(v)$ is homeomorphic either to S^{n-1} or D^{n-1} .

Properties and notation:

- Each combinatorial n -manifold is a triangulation of an n -manifold.
- For $n < 4$, each n -manifold admits a triangulation as a combinatorial n -manifold².
- Vertices of a combinatorial manifold K satisfying $\text{Lk}(v) \cong D^{n-1}$ are called **boundary vertices**.
- Vertices of a combinatorial manifold K satisfying $\text{Lk}(v) \cong S^{n-1}$ are called **interior vertices**.
- Edges of a combinatorial surface K that are contained in only one triangle are called **boundary edges**. The union of the boundary edges corresponds to the boundary of the manifold.
- Edges of a combinatorial surface K that are contained in two triangles are called **interior edges**. No edge in a combinatorial surface is contained in more than two triangles.

Using these properties it is fairly easy to recognize whether a given simplicial complex K is a combinatorial surface and thus a triangulation of a surface: for each vertex $v \in K$ we verify whether $\text{Lk}(v)$ is homeomorphic to S^1 or D^1 . It is easy to see that a connected 1-dimensional simplicial complex is homeomorphic to:

- S^1 iff each of its vertices is contained in two edges.

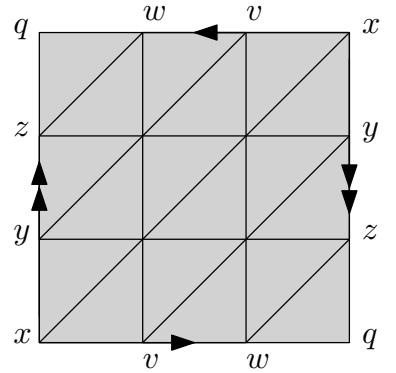
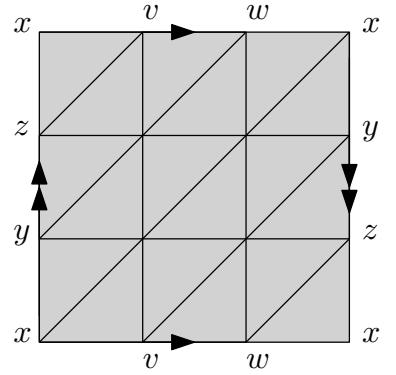


Figure 4.5: Triangulations of the Klein bottle (top) and the projective plane (bottom).

²Surprisingly enough, this does not hold for $n \geq 4$

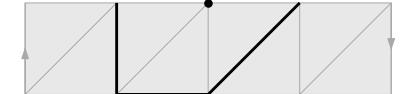
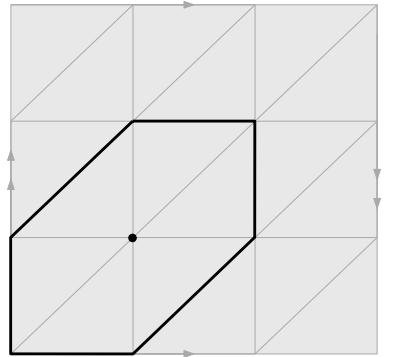


Figure 4.6: Triangulation of the Klein bottle (top) and of the Möbius band (bottom). In the top triangulation each vertex is an interior vertex as each link (bold) is homeomorphic to S^1 . In the bottom case each vertex is a boundary vertex as each link (bold) is homeomorphic to B^1 .

- D^1 , i.e., the line segment, iff two of its vertices are contained in one edge, and all other vertices are contained in two edges.

We will leave the elementary proofs of these two facts to the reader.

4.2 Orientability

Orientability is about defining “up and down”. It is quite easy to agree on the two directions on the surface of the earth. However, in general that may not be the case for all surfaces. Consider the cylinder from Figure 4.1. It is orientable because we can define two different sides of it. To put it into a more colorful language, we can color one side of the band in red and the other side in blue, without the colors ever touching each other. The story is different on the Moebius band: it has only one side. We could start coloring it in red at some spot and keep expanding the color along the surface (but not across the boundary): eventually we will color the whole band, i.e., there is no “other” side.

Orientability is an important property of surfaces. It will be required for our classification result. In order to fully understand it we have to define orientation for simplices first. Besides its application in this section orientation on simplices will feature prominently later within the context of homology computation.

Up to now a simplex was given by a set of its vertices. An oriented simplex is a simplex with a choice of orientation. For an edge that means direction, for a triangle that means “a normal” (see Figures 4.7 and 4.8). This direction/orientation will be described by a choice of an order on vertices.

Definition 4.2.1. An *oriented simplex* on vertices v_0, v_1, \dots, v_k is an ordered $(k+1)$ -tuple $\sigma = \langle v_0, v_1, \dots, v_k \rangle$. For a permutation π on $\{0, 1, \dots, k\}$ we identify:

$$\sigma = (-1)^{\text{sgn}(\pi)} \langle v_{\pi(0)}, v_{\pi(1)}, \dots, v_{\pi(k)} \rangle,$$

where $\text{sgn}(\pi)$ is the signature of permutation π , i.e., value 0 if π is even and value 1 if π is odd.

A 0-dimensional simplex with vertex v can also be oriented in two ways: as $\langle v \rangle$ and as $-\langle v \rangle$.

Figures 4.7 and 4.8 provide examples of descriptions of oriented edges and triangles, and their geometric interpretations. Here are some properties that follow from Definition 4.2.1:

- Each simplex on vertices v_0, v_1, \dots, v_k can be oriented in two different ways: $\sigma = \langle v_0, v_1, \dots, v_k \rangle$ and $-\sigma$.

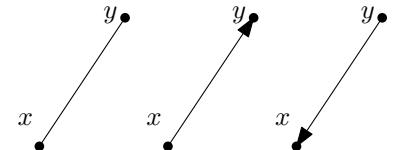


Figure 4.7: The edge $\{x, y\}$ (left) and the oriented edges $\langle x, y \rangle$ (center) and $\langle y, x \rangle = -\langle x, y \rangle$ (right).

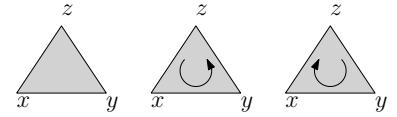


Figure 4.8: The triangle $\{x, y, z\}$ (left) and the oriented triangles $\langle x, y, z \rangle = \langle y, z, x \rangle = \langle z, x, y \rangle$ (center) and $\langle y, x, z \rangle = \langle x, z, y \rangle = \langle z, y, x \rangle = -\langle x, y, z \rangle$ (right).

- An oriented simplex has a sign + (usually omitted) or – prepended.
- Exchanging two vertices in an oriented simplex τ changes the orientation of τ by changing the prefixed sign.

An important property of an oriented simplex is that it induces an orientation on each of its facets.

Definition 4.2.2. Suppose $\sigma = \langle v_0, v_1, \dots, v_k \rangle$ is an oriented simplex and $p \in \{0, 1, \dots, k\}$. Then **induced orientation** of the facet of σ obtained by dropping v_p is

$$(-1)^p \langle v_0, v_1, \dots, v_{p-1}, v_{p+1}, \dots, v_k \rangle.$$

Oriented edge $\langle x, y \rangle$ induces orientations $\langle y \rangle$ and $-\langle x \rangle$ on its facets (vertices). Oriented triangle $\langle x, y, z \rangle$ induces orientations $\langle y, z \rangle$, $-\langle x, z \rangle$ and $\langle x, y \rangle$ on its facets (edges), see Figure 4.9.

Now that we established a way to orient a single simplex, we turn our attention to orienting the whole surface.

Definition 4.2.3. Suppose oriented 2-simplices σ and σ' share a common edge. Simplices σ and σ' are **oriented consistently**, if they induce the opposite orientation on the common edge. (see Figure 4.10)

Definition 4.2.4. Let K be a triangulation of a surface $|K|$. We say that $|K|$ is **oriented**, if all triangles of K are oriented (as simplices) so that the following holds: each pair of oriented triangles with a common edge is oriented consistently.

A surface is **orientable** if it can be oriented.

Orientability of a surface does not depend on a triangulation but on the topological type of the surface only. The following are two basic examples that demonstrate the underlying geometric idea.

Example 4.2.5. The cylinder $S^1 \times [0, 1]$ is orientable as Figure 4.11 demonstrates. To the contrary, the Moebius band is not orientable as Figure 4.12 demonstrates. Since the Klein bottle and the projective plane both contain a copy of the Moebius band (any of the three horizontal strips of triangulations in Figure 4.5), neither of them is orientable.

As Example 4.2.5 and Figure 4.12 suggest it is fairly easy to check whether a connected triangulated surface is orientable. This can be done directly by orienting one triangle and then inductively orienting all neighboring triangles with shared edges, while checking that each newly oriented triangle is oriented consistently with respect to the already oriented triangles.

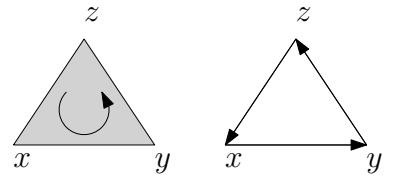


Figure 4.9: Oriented triangle $\langle x, y, z \rangle$ (left) and induced orientation on the edges (right): $\langle x, y \rangle$, $\langle y, z \rangle$, and $\langle z, x \rangle$. Note that the edges are oriented along the direction of the circular arrow indicating the orientation of the triangle.

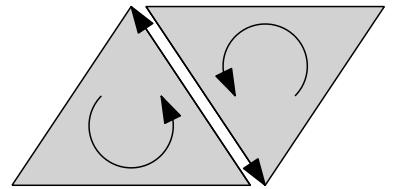


Figure 4.10: Consistent orientation: note that the orientations of the triangles agree (both directed circular arrows point counter-clockwise). This implies that the induced orientations on the common edge are opposite to each other.

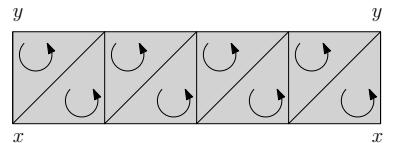


Figure 4.11: Orientable triangulation of a usual band. The oriented simplices induce the opposite orientation on all the edges, including the edge $\{x, y\}$, along which the glueing occurs.

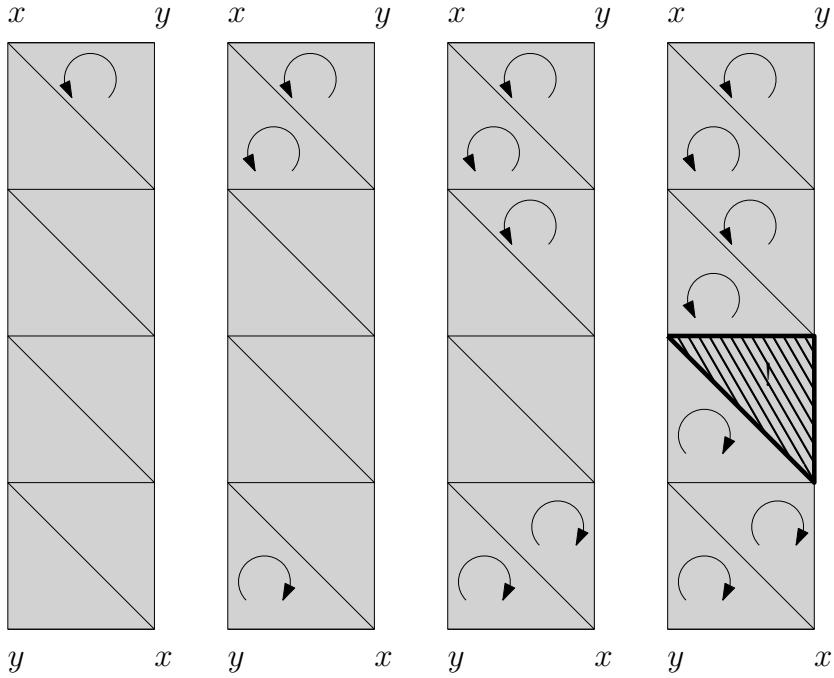


Figure 4.12: A proof that the Moebius band is not orientable. Assume we want to orient a triangulation of the Moebius band on the left. We first choose an orientation of one triangle (far left) and then inductively induce consistent orientation on the neighboring triangles. In the end (far right) we obtain conflicting requirements on the orientation on the last (bold) triangle, which means there is no consistent way to orient all the triangles in this triangulation.

4.3 Connected sum of surfaces

One of the ways to make new surfaces (and actually manifolds in general) out of known ones is the connected sum.

Definition 4.3.1. Suppose X and Y are connected surfaces. Choose topological 2-discs $D_X \subset X$ and $D_Y \subset Y$, neither of which contains any boundary point of the surfaces. The corresponding boundaries of these discs are topological 1-spheres (circles) $S_X \subset X$ and $S_Y \subset Y$ respectively. The **connected sum** $X \# Y$ is obtained by removing the interiors of discs D_X and D_Y from X and Y , and gluing the resulting spaces by identifying S_X with S_Y .

See Figure 4.13 for a sketch of this construction. A few technical remarks about connected sums as defined above:

- It turns out that the topological type of $X \# Y$ does not depend on the choice of discs D_X, D_Y .
- A connected sum is a surface, whose boundary components correspond to the union of the boundary components of X and the boundary components of Y .

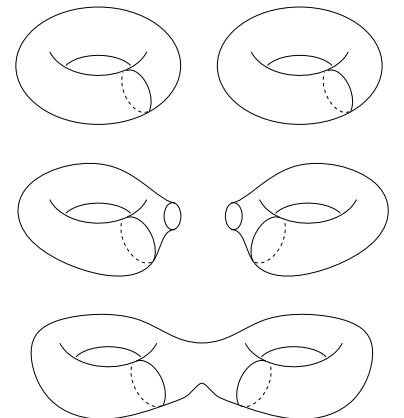


Figure 4.13: Two tori (top) and their connected sum (bottom) obtained by identifying the boundaries of the two removed discs (center).

- Surfaces X and Y are both orientable iff $X \# Y$ is orientable.
- For each surface X , the following holds: $X \# S^2 \cong X$.

If abstract simplicial complexes K and L are triangulations of surfaces X and Y respectively and $X \cap Y = \emptyset$, we can obtain a triangulation M of $X \# Y$ in the following way:

1. Choose triangles Δ_X and Δ_Y in K and L respectively, so that no point of these two triangles lies on the boundary of X or Y .
2. Define

$$M = (K \setminus \{\Delta_X\}) \cup (L \setminus \{\Delta_Y\}) / \sim,$$

where \sim stands for the identification of each of the boundary edges of Δ_X with an appropriate boundary edge of Δ_Y .

In short, M is obtained by removing Δ_X and Δ_Y from the union of K and L , and then identifying the boundaries of the removed triangles. This procedure is a discrete version of the one in Definition 4.3.1.

Proposition 4.3.2. $\chi(X \# Y) = \chi(X) + \chi(Y) - 2$.

Proof. Assume abstract simplicial complexes K and L are triangulations of surfaces X and Y respectively and $K \cap L = \emptyset$. It is obvious that $\chi(K \cup L) = \chi(K) + \chi(L)$. In order to obtain a triangulation of $X \# Y$ from $K \cup L$, we:

- Remove two triangles (change -2 to the Euler characteristic);
- Identify three pairs of vertices, meaning we have three vertices less (change -3 to the Euler characteristic);
- Identify three pairs of edges, meaning we have three edges less (change $+3$ to the Euler characteristic);

The total change to the Euler characteristic after these steps is -2 . □

4.4 Classification of surfaces

We can now describe the classification of surfaces. Let \mathbb{T} denote the torus and let \mathbb{P} denote the projective plane.

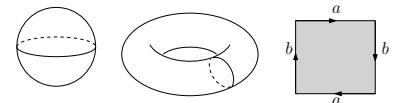


Figure 4.14: The three closed connected surfaces (the sphere S^2 , the torus \mathbb{T} and the projective plane \mathbb{P}), that are used to construct any other closed connected surface using the connected sum operation.

Theorem 4.4.1. [Classification Theorem for closed connected surfaces]

Suppose X is a closed connected surface. Then X is homeomorphic to one of the following:

1. S^2 .
2. n -torus $n\mathbb{T} = \underbrace{\mathbb{T} \# \mathbb{T} \# \dots \# \mathbb{T}}_n$ for some $n \in \mathbb{N}$.
3. $n\mathbb{P} = \underbrace{\mathbb{P} \# \mathbb{P} \# \dots \# \mathbb{P}}_n$ for some $n \in \mathbb{N}$.

It turns out that the surfaces appearing in Theorem 4.4.1 can be distinguished using orientability and the Euler characteristic. From the properties of the connected sum recall that (for each $n \in \mathbb{N}$) S^2 and $n\mathbb{T}$ are orientable while $n\mathbb{P}$ are not. Furthermore, using Proposition 4.3.2 and we can inductively deduce³:

- $\chi(n\mathbb{T}) = 2 - 2n$ as $\chi(\mathbb{T}) = 0$.
- $\chi(n\mathbb{P}) = 2 - n$ as $\chi(\mathbb{P}) = 1$.

Consequently we obtain the following table of connected closed surfaces.

Surface	χ	
S^2	2	
T	0	orientable
nT	$2 - 2n$	
P	1	not orientable
nP	$2 - n$	

³ Using Proposition 4.3.2 we can deduce $\chi(2\mathbb{T}) = \chi(\mathbb{T} \# \mathbb{T}) = \chi(\mathbb{T}) + \chi(\mathbb{T}) - 2 = 0 + 0 - 2 = -2$, $\chi(3\mathbb{T}) = \chi(\mathbb{T}) + \chi(\mathbb{T} \# \mathbb{T}) - 2 = 0 + -2 - 2 = -4$, and proceed inductively. The proof for $n\mathbb{P}$ is analogous.

Table 4.1: A list of closed connected surfaces along with their Euler characteristic and orientability.

Theorem 4.4.1 motivates the following **classification algorithm** for a closed connected surface given as an abstract simplicial complex⁴ K :

1. Check for orientability of K .
2. Compute the Euler characteristic.
3. Consult Table 4.1.

Example 4.4.2. Which of the surfaces in Theorem 4.4.1 is the Klein bottle? We have already discovered that it is not orientable and that its Euler characteristic is 0. By the Classification Theorem the Klein bottle is homeomorphic to $\mathbb{P} \# \mathbb{P}$.

General surfaces

Theorem 4.4.1 can also be used to classify general surfaces admitting a finite triangulation. Suppose X is a surface:

⁴ I.e., we assume the triangulation K is a connected combinatorial 2-manifold and has no boundary components.

⁵ A shortcut to computing χ : while the Euler characteristic is formally defined on a triangulation, it turns out it can also be obtained from the representation of a surface in terms of a polygon with identified sides. For example, the representations of the torus in Figure 4.14 and Klein bottle of Figure 4.3 have one 2-dimensional square, two 1-dimensional edges, and one vertex, yielding $\chi = 2$. The representation of the projective plane in Figure 4.4 has one 2-dimensional square, two 1-dimensional edges, and two vertices, yielding $\chi = 1$. This trick could assist with Figure 4.15. A justification will be provided in the chapter on discrete Morse theory.

1. If X is not connected, it is a disjoint union of connected surfaces and it obviously suffices to recognize each of its components.
2. If X is connected and has a boundary Y , then Y is a 1-manifold without boundary, meaning Y is a disjoint union of k copies of S^1 for some $k \in \mathbb{N}$. By glueing a disc along each component of Y we obtain a closed connected surface X' , which we can recognize⁵. We conclude that X is homeomorphic to X' with k discs removed⁶.

Example 4.4.3. It is easy to see that the cylinder $S^1 \times [0, 1]$ is obtained from S^2 by removing two discs. It is a bit harder to see how to get the Moebius band M this way. It is easy to see that M has one boundary component and has Euler characteristic⁷ 0. Gluing a disc along the boundary component we obtain a closed connected non-orientable surface of Euler characteristic 1, which is \mathbb{P} . Hence the Moebius band is obtained by removing a disc from the projective plane.

We are now ready to state a **classification algorithm** for a surface given as an abstract simplicial complex K :

1. Partition K into its connected components and classify each of them.
2. For each component K' :
 - (a) Count the number $n(K')$ of boundary components of K' .
 - (b) Check for orientability of K' .
 - (c) Compute the Euler characteristic of K' .
 - (d) Let Y be the surface matching the orientability of K' and of Euler characteristic⁸ $\chi(K') + n(K')$ by Table 4.1.
 - (e) Surface K' is homeomorphic to Y with $n(K')$ many discs removed.

With this classification algorithm we can always determine whether two surfaces are homeomorphic or not.

4.5 Concluding remarks

Recap (highlights) of this chapter

- Surfaces, combinatorial surfaces;
- Orientation and orientability;
- Connected sum of surfaces;
- Classification of surfaces;

⁵ Such a gluing of discs does not change the orientation, i.e., X is orientable iff X' is. However, an addition of each disc increases the Euler characteristic by 1.

⁶ It turns out that the homeomorphic type does not depend on the discs we remove from X' , only on their number.

⁷ We could count the simplices in Figure 4.12.

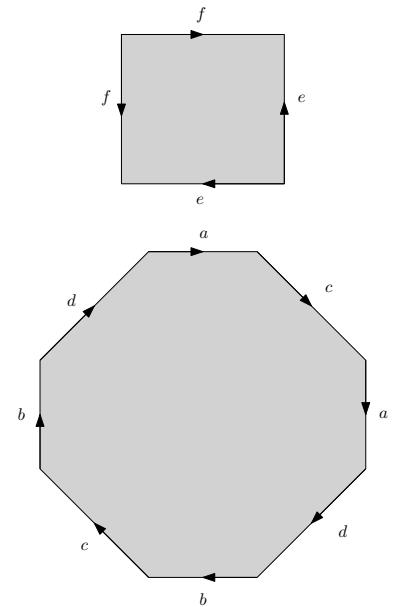


Figure 4.15: Which surfaces are these?

⁸ $\chi(K') + n(K')$ is the Euler characteristic of a surface obtained from K by glueing $n(K')$ discs along the boundary components of K' .

Background and applications

For most of the practical purposes, we live in a three-dimensional space. Objects in our everyday life are often modelled by surfaces enclosing the objects. Outputs of many 3-D scans are given in terms of triangulated surfaces (for example, as .stl files).

Surfaces and other higher-dimensional manifolds are also often assumed to be the underlying spaces in specific settings. A randomly generated bitmap image will seldom represent something reasonable, and yet there is a huge number of images that convey an information to the human eye. A space of “recognizable” images is a huge subspace (perhaps a manifold) in the space of all bitmap images. Manifold recognition approaches the attempt to detect the underlying manifolds from sample data.

The third source of surfaces and manifolds are spaces described with two or more degrees of freedom: configuration spaces of molecules, robotic arms, etc. For example, the configuration space of a robotic arm (i.e., the space of all possible positions of the arm) with two independent joins, each of which allows a full rotational motion, is the torus $S^1 \times S^1$. On a similar note, given two annotated⁹ points on S^1 , the configuration space of all possible positions of the two points is again a torus $S^1 \times S^1$, since the degree of freedom of each point is S^1 . It is interesting to observe the difference that appears if the points are not annotated¹⁰: in such a case all possible configurations actually form the projective plane.

On a more theoretical note, the question of whether each manifold admits a triangulation had been one of the focal points of topology in the previous century. It turns out that every manifold in dimension 3 or less admits a triangulation. Surprisingly enough, there are manifolds in higher dimensions that do not admit any triangulation. A detailed proof of the classification theorem of surfaces can be found in many textbooks^{11,12}.

Appendix: imagining S^3

In this appendix we will try to explain two ways of thinking about the three-dimensional sphere S^3 and spheres in general.

1. The first observation has to do with the relationship between discs and spheres. We have already mentioned that S^{n-1} appears as the boundary of D^n . It should also be apparent (see Figure 4.18) that gluing two copies of a disc D^n along their boundaries (copies of S^{n-1}) results in S^n . In particular, we obtain S^3 by taking two 3-discs (solid balls) and gluing them along the boundary.
2. As for the second observation we will refer to Figure 4.19. It turns

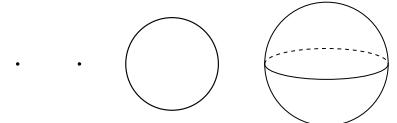


Figure 4.16: Spheres S^0 (two points), S^1 (a circle), and S^2 (a sphere).

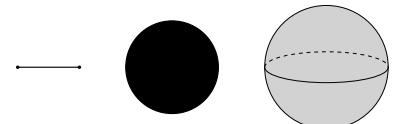


Figure 4.17: Discs D^1 (a line segment), D^2 , and D^3 .

⁹ The annotation refers to the fact that each point has a name in the sense that if the two points lie at different positions, then exchanging them changes the configuration. The situation is sometimes also described using “ordered pairs” of points (x, y) for which $x, y \in S^1$.

¹⁰ In particular, if two non-annotated points lie at different positions, then exchanging them does not change the configuration as the pair represents the same collection of points on S^1 . The situation is sometimes also described using “unordered pairs” of points (x, y) by additionally identifying $(x, y) \sim (x', y')$ if $x = y'$ and $y = x'$.

¹¹ L. Christine Kinsey. Topology of Surfaces. Springer New York, 1993. doi: [10.1007/978-1-4612-0899-0](https://doi.org/10.1007/978-1-4612-0899-0)

¹² Jean Gallier and Dianna Xu. A Guide to the Classification Theorem for Compact Surfaces. Springer Berlin Heidelberg, 2013. doi: [10.1007/978-3-642-34364-3](https://doi.org/10.1007/978-3-642-34364-3)

out that S^n can be obtained in the following way: pick two opposite points (the north pole and the south pole) and span an interval's worth of copies of the sphere S^{n-1} between them, so that the spheres are shrinking as they are approaching the poles.

Note that if we only take one point and have the spheres shrinking only as they approach that point (and have them increase otherwise) we obtain D^n (see Figure 4.20 for some low-dimensional examples).

It may come as a surprise that these points of view can be observed in Dante's *Divine Comedy*, written about seven centuries ago. Dante's description of the universe coincides with the topology of S^3 : on one extreme are the depths of Hell (part of Inferno), from which Dante is guided by Beatrice through the spheres of Inferno, Purgatory, and Paradise, until he reaches Empyrean, the place which contains the essence of God.

This description coincides with 2. above in terms of spheres. Equivalently, we may consider Inferno and Purgatory together as one 3-disc with center at Hell, and Paradise as another 3-disc with center at Empyrean: in this setup the Universe consists of both 3-discs that intersect along the surface of the Earth (which coincides with the boundary S^2).

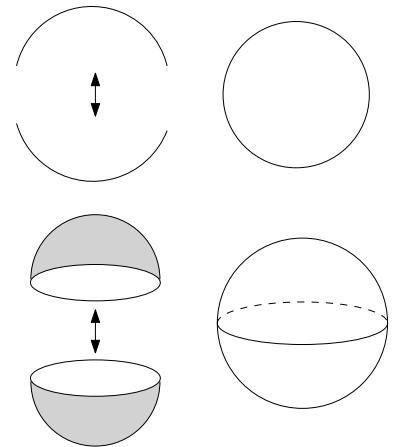


Figure 4.18: Gluing two copies of a disc together results in a sphere.

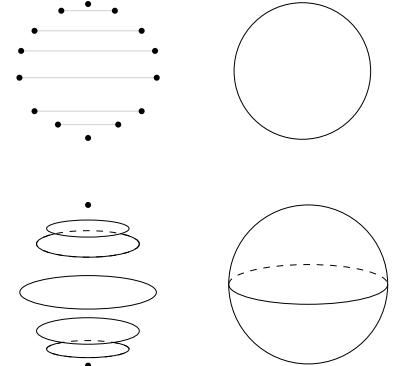


Figure 4.19: Obtaining S^n as a collection of spheres S^{n-1} between two points.

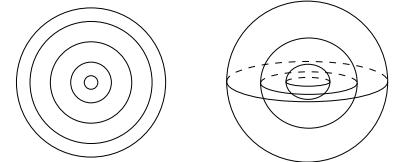


Figure 4.20: Obtaining \mathbb{R}^2 (left) and \mathbb{R}^3 (right) as a collection of concentric spheres of all positive radii. In the same way we can decompose \mathbb{R}^k for any $k \in \{1, 2, \dots\}$.

5

Constructions of simplicial complexes

TOPOLOGICAL METHODS TYPICALLY take a simplicial complex as input. However, objects of interest are often not provided in this form. The first step in a topological treatment is thus frequently a creation of simplicial complexes. In this chapter we will present various constructions of complexes from a point cloud, i.e., from a finite collection of points in a metric space. These points may represent a sample of our shape, a collection of numerical data (a subset of \mathbb{R}^n), etc.

Our discussion will include two properties we expect from such a construction. The first one is a relationship with the underlying shape, which is often guaranteed by the nerve theorem. The second one describes stability to perturbations and a way to measure distance between different constructions, as formalized by the interleaving property.

5.1 Rips complexes

Rips complexes represent perhaps the simplest construction of a complex from a finite collection of points.

Definition 5.1.1. Let X be a metric space and let a sample $S \subset X$ be a finite subset. Choose a scale $r \geq 0$. The **Rips** complex $\text{Rips}(S, r)$ is an abstract simplicial complex defined by the following rules:

1. The vertex set is S .
2. A subset $\sigma \subseteq S$ is a simplex iff $\text{Diam}(\sigma) \leq r$.

Remark 5.1.2. A few comments:

- Rips complexes are sometimes also called Vietoris-Rips complexes.
- $\text{Rips}(S, r)$ represents a combinatorial snapshot of S at scale r .

☞ Rips complexes are a special case of clique complexes. Suppose G is a graph with vertices V and edges E . The clique complex of G is the abstract simplicial complex with the vertex set V , whose simplices satisfy the following condition: a subset $\sigma \subseteq S$ is a simplex iff each pair of vertices of σ is an edge in G . A Rips complex is the clique complex of its 1-skeleton.

☞ Diameter of a finite subset $A \subset X$ of a metric space X is defined as

$$\text{Diam}(A) = \max_{x,y \in A} d(x,y).$$

Diameter of X is defined as

$$\text{Diam}(X) = \sup_{x,y \in X} d(x,y).$$

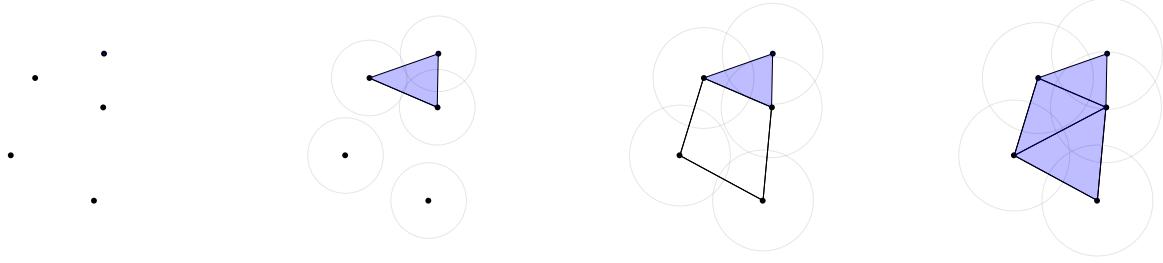


Figure 5.1: Five points in the plane and three corresponding Rips complexes $\text{Rips}(S, r)$. Visualisation is assisted by circles of radius $r/2$ around each point. For much larger scales the Rips complex is not planar and eventually becomes 4-dimensional.

- $\text{Diam}(\sigma)$ is the diameter of σ . Condition $\text{Diam}(\sigma) \leq r$ means that the distance between any two vertices of σ is at most r .
- It is easy to verify that Rips complexes are indeed abstract simplicial complexes: if σ is a simplex then so is each of its subsets.

Remark 5.1.3. Some properties of the Rips complexes:

1. Rips complexes are often the preferred construction in TDA due to their simplicity in terms of the definition and computation.
2. $\text{Rips}(S, r)$ is an abstract simplicial complex, typically not embeddable in X .
3. For r smaller than the smallest pairwise distance between the points in S , $\text{Rips}(S, r)$ is a discrete set, i.e., a complex with no edges or higher-dimensional simplices.
4. For r at least as large as $\text{Diam}(S)$, i.e., the largest pairwise distance between the points in S , $\text{Rips}(S, r)$ is the $(|S| - 1)$ -simplex, i.e., the simplicial complex on S containing all subsets of S .
5. If $r_1 \leq r_2$, then $\text{Rips}(S, r_1) \subseteq \text{Rips}(S, r_2)$.

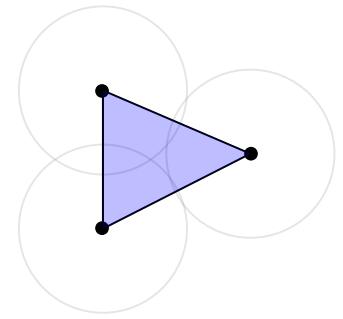


Figure 5.2: Suppose we are given three points in the plane. These three points span a triangle in the Rips complex for r greater or equal to the maximal pairwise distance between these points. Equivalently, the three balls of radius $r/2$ should pairwise intersect.

Definition 5.1.4. Let X be a metric space and let a sample $S \subset X$ be a finite subset. The **Rips filtration** on S is the collection of abstract simplicial complexes $\{\text{Rips}(S, r)\}_{r \geq 0}$ along with inclusions

$$i_{r_1, r_2}: \text{Rips}(S, r_1) \hookrightarrow \text{Rips}(S, r_2) \quad \text{for all } r_1 \leq r_2.$$

A Rips filtration provides the collection of all Rips complexes on S . While a single Rips complex depends on the choice of the scale, the filtration does not. Filtrations will play a fundamental role later in the definition of persistent homology.

☞ More generally, a **filtration** of a simplicial complex K is a family of subcomplexes $\{K_r\}_{r \geq 0}$ indexed by a parameter r , such that $K_r \leq K_{r'}$ for all $r < r'$. The Rips filtration on a finite set S is a filtration of the $(|S| - 1)$ -simplex.

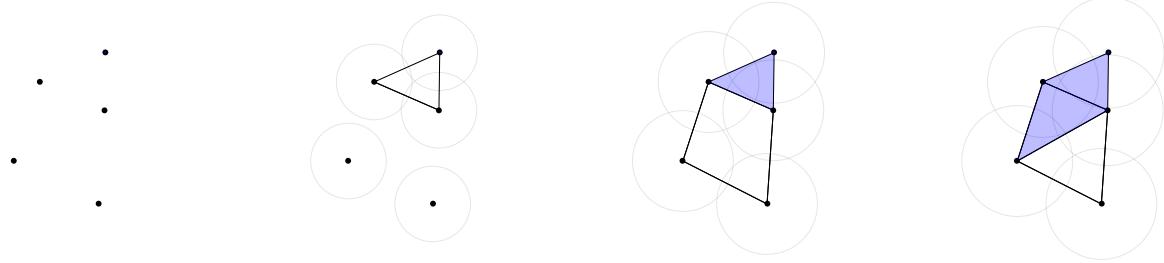
5.2 Čech complexes

Definition 5.2.1. Let X be a metric space and let a sample $S \subset X$ be a finite subset. Choose a scale $r \geq 0$. The **Čech complex** $\text{Cech}(S, r)$ is an abstract simplicial complex defined by the following rules:

1. The vertex set is S .
2. A subset $\sigma \subseteq S$ is a simplex iff $\bigcap_{x \in \sigma} B(x, r) \neq \emptyset$.

☞ Recall $B(x, r) = \{y \in X \mid d(x, y) \leq r\}$ is a closed ball.

☞ Čech complexes are a special case of nerve complexes, a connection that will be explained below.



Remark 5.2.2. A few comments about the definition:

- Čech complexes are a classical topological construction used in many contexts throughout topology.
- $\text{Cech}(S, r)$ represents a combinatorial snapshot of S at scale r .
- It is easy to verify that Čech complexes are indeed abstract simplicial complexes.

Remark 5.2.3. Some properties of the Čech complexes:

1. While harder to compute¹, Čech complexes are attractive due to a well understood geometric interpretation, which will be explained in the next section within the context of nerve complexes.
2. $\text{Cech}(S, r)$ is an abstract simplicial complex, typically not embeddable in X , although it is often homotopy equivalent² to a subset of X .
3. For r smaller than one half of the smallest pairwise distance between the points in S , $\text{Cech}(S, r)$ is a discrete set.
4. For r at least as large as twice the largest pairwise distance between points in S , $\text{Cech}(S, r)$ is the $(|S| - 1)$ -simplex.
5. If $r_1 \leq r_2$, then $\text{Cech}(S, r_1) \subseteq \text{Cech}(S, r_2)$.

¹ See the MiniBall algorithm in the Appendix.

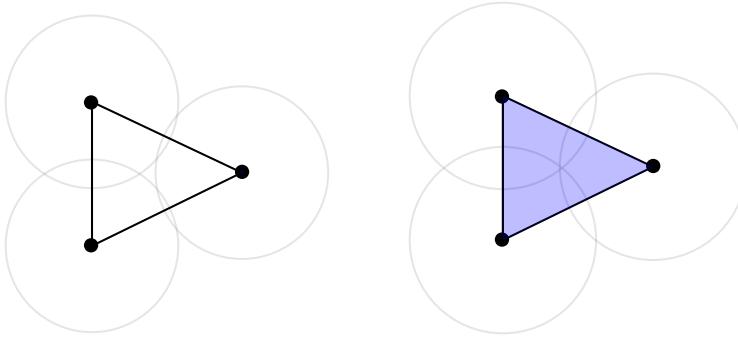
² See the nerve theorem below for details.

6. It is easy to verify³ that $\text{Cech}(S, r) \subseteq \text{Rips}(S, 2r)$.
7. It is also easy to see that $\text{Rips}(S, r) \subseteq \text{Cech}(S, r)$. A non-trivial inclusion $\text{Rips}(S, r\sqrt{2}) \subseteq \text{Cech}(S, r)$ holds in Euclidean spaces by Jung's theorem⁴.

Definition 5.2.5. Let X be a metric space and let a sample $S \subset X$ be a finite subset. The **Čech filtration** of S is the collection of abstract simplicial complexes $\{\text{Cech}(S, r)\}_{r \geq 0}$ along with inclusions

$$i_{r_1, r_2}: \text{Cech}(S, r_1) \hookrightarrow \text{Cech}(S, r_2) \quad \text{for all } r_1 \leq r_2.$$

As was the case with Rips filtrations, Čech filtrations also provide a scale-free approximation of the underlying set by simplicial complexes. The smallest example demonstrating a difference between the two constructions is given in Figure 5.4.



³ If balls of radius r intersect then the pairwise distances between the centers are at most $2r$.

⁴

Theorem 5.2.4 (Jung's theorem). If D is the diameter of a finite subset $F \subset \mathbb{R}^n$, then F is contained in a ball of radius at most $D\sqrt{\frac{n}{2(n+1)}}$.

For $X = (\mathbb{R}^n, d_2)$ we actually obtain

$$\text{Rips}\left(S, r\sqrt{\frac{2(n+1)}{n}}\right) \subseteq \text{Cech}(S, r).$$

The factor $r\sqrt{2}$ in 7 of Remark 5.2.3 is thus only the smallest upper bound that holds for all n .

Figure 5.4: Suppose we are given three points in the plane. These three points span a triangle in the Čech complex iff the three balls of radius r intersect. The left complex consisting of three edges and no triangle does not appear as a Rips complex of any triple of points.

5.3 Nerve complexes

Čech complexes are a special case of a classical topological construction called the nerve.

Definition 5.3.1. For $k \in \mathbb{N}$ let $\mathcal{U} = \{U_1, U_2, \dots, U_k\}$ be a collection of subsets of X . The **nerve** of \mathcal{U} is the abstract simplicial complex $\mathcal{N}(\mathcal{U})$ defined by the following rules:

1. The vertex set is $\mathcal{U} = \{U_1, U_2, \dots, U_k\}$, consisting of k elements.
2. A subset $\sigma \subseteq \mathcal{U}$ is a simplex iff $\bigcap_{i \in \sigma} U_i \neq \emptyset$.

A Čech complex is the nerve of the corresponding collection of r -balls, i.e., $\text{Cech}(S, r) = \mathcal{N}(\{B(s, r)\}_{s \in S})$. Another example is the Delaunay triangulation, which is the nerve of the Voronoi diagram.

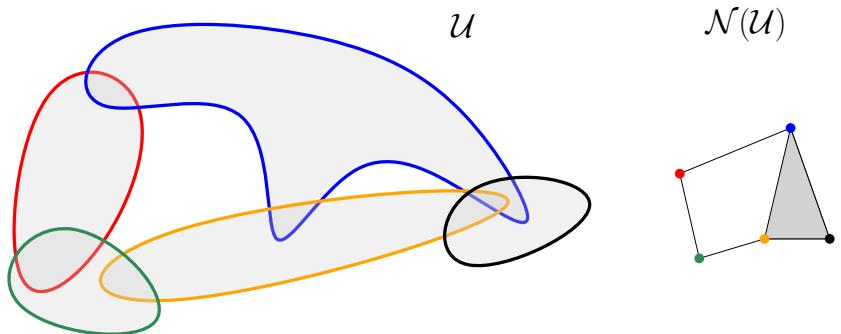
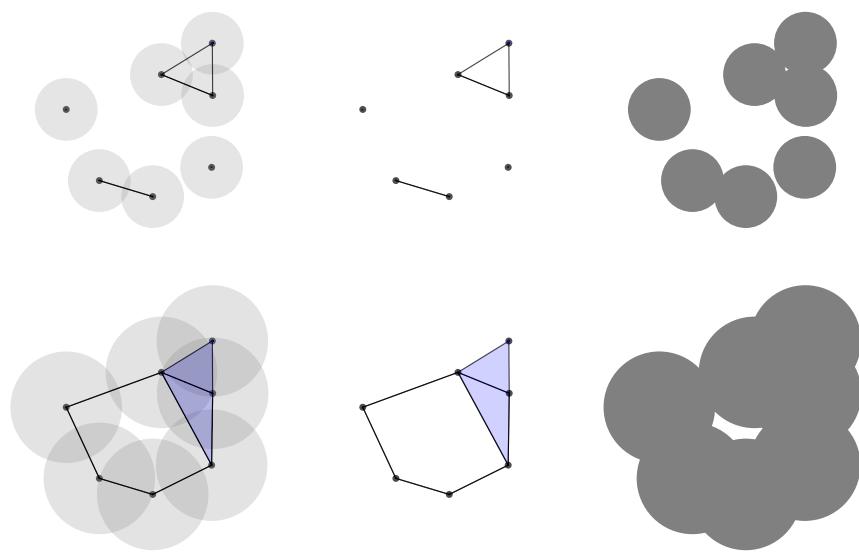


Figure 5.5: An example of a nerve.



One of the main advantages of nerve complexes is that their homotopy type in some cases represents the union of the elements of \mathcal{U} . This is formalized within the context of the nerve theorem, of which we now state a special case.

Theorem 5.3.2. [Nerve theorem] Let $n \in \mathbb{N}$ and assume a collection $\mathcal{U} = \{U_1, U_2, \dots, U_k\}$ consists of closed convex subsets of \mathbb{R}^n . Then $U_1 \cup U_2 \cup \dots \cup U_k \simeq \mathcal{N}(\mathcal{U})$.

An idea of a proof is given in Appendix. The nerve theorem does not hold for an arbitrary collection of subsets as Figure 5.5 demonstrates.

For Delaunay triangulations the nerve theorem provides no additional information. As the Voronoi cells are convex, the nerve theorem implies that the Delaunay triangulation is contractible, a fact we already know as it triangulates a convex (hence contractible by Lemma 5.3.3) planar region.

On the other hand, the nerve theorem provides a homotopical description of the Čech complex. As Euclidean balls are convex, we obtain

$$\text{Cech}(S, r) = \mathcal{N}(\{B(s, r)\}_{s \in S}) \simeq \bigcup_{s \in S} B(s, r),$$

i.e., the Čech complex $\text{Cech}(S, r)$ has the homotopy type of the r -neighborhood of S . This fact is the foremost reason for the use of Čech complexes: while they are harder to compute than Rips complexes, we know that the obtained homotopy type represents the r -neighborhood of S . In this spirit we can interpret Figure 5.6. Furthermore, this observation can be used to prove reconstruction results: given a closed connected surface X in an Euclidean space, for each sufficiently small scale parameter $r \geq 0$ and for each sufficiently dense finite subset $S \subset X$ we have $X \simeq \text{Cech}(S, r)$, i.e., the homotopy type of a space X can be reconstructed using Čech complexes (in Euclidean or geodesic metric)⁵. In the Euclidean metric this holds as⁶ $X \simeq N(X, r)$ for small r .

Alpha complexes

Alpha complexes are a fusion between planar Čech complexes and Delaunay triangulations.

☞ Nerve theorem actually holds much more generally. For example, assume each finite intersection of sets of \mathcal{U} (including each member $U \in \mathcal{U}$, since it appears as the intersection of $\{U\} \subseteq \mathcal{U}$) is either empty or contractible. If \mathcal{U} is a finite collection of closed subsets in \mathbb{R}^n , or an arbitrary collection of open sets in a metric space, then

$$\bigcup_{U \in \mathcal{U}} U \simeq \mathcal{N}(\mathcal{U}).$$

This is a stronger statement than Theorem 5.3.2 by Lemma 5.3.3.

Lemma 5.3.3. Let $n \in \mathbb{N}$. Each convex subset of \mathbb{R}^n is contractible.

Proof. Assume $A \subset \mathbb{R}^n$ is convex and fix $x_0 \in A$. We can slide each $a \in A$ into x_0 along the line segment from a to x_0 . This results in a homotopy $H(a, t) = (1 - t)a + t x_0$ between the identity map on A and the constant map at x_0 , hence A is contractible. □

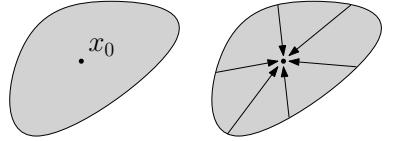


Figure 5.7: A sketch of Lemma 5.3.3.

⁵ The same result holds for more general spaces and under appropriate conditions also for Rips complexes. However, almost all of the proofs are based on the application of the nerve theorem to Čech complexes.

⁶ Imagine a circle in the plane, a knot in \mathbb{R}^3 or a surface in \mathbb{R}^3 : its small thickening is homotopy equivalent to the space itself.

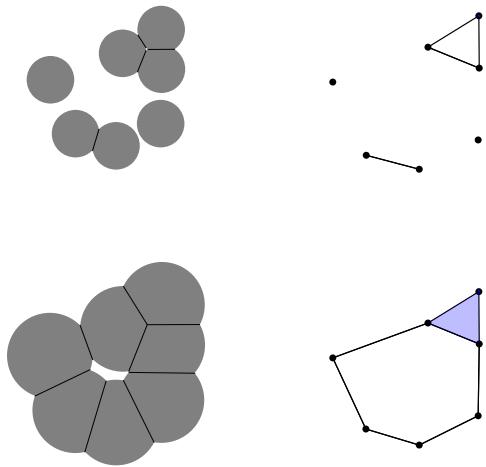
Definition 5.3.4. Let $r \geq 0$ and assume $S \subset \mathbb{R}^n$ is a finite collection of points satisfying a general position property: no $n+2$ points of S lie on the same $(n-1)$ -sphere. For each $s \in S$ let V_s denote the corresponding Voronoi cell. The **alpha complex** of S at scale r is the following nerve: $\mathcal{N}\left(\{V_s \cap B(s, r)\}_{s \in S}\right)$.

Assume S is as in Definition 5.3.4. While $\text{Čech}(S, r)$ may be of arbitrarily high dimension⁷, the nerve theorem guarantees it is homotopy equivalent to a subset of \mathbb{R}^n . The alpha complex of S at scale r is a complex⁸, that is homotopy equivalent to $\text{Čech}(S, r)$. To see this note that by the nerve theorem⁹ both are homotopy equivalent to

$$\bigcup_{s \in S} B(s, r) = \bigcup_{s \in S} (V_s \cap B(s, r)).$$

Thus alpha complexes may be seen as an efficient way of obtaining the homotopy type of a Čech complex in \mathbb{R}^n .

Another way of thinking of alpha complexes is as a model for molecules. Each atom in a molecule has a radius¹⁰ and touches (rather than intersects) other atoms within the range.



Mapper

Another example of a construction based on the idea of the nerve is Mapper. In contrast to the constructions above it is typically¹¹ a one-dimensional simplicial complex, i.e., a graph. Mapper can be thought of as a one-dimensional sketch of a space X as detected through the lens of a single map on X .

We first describe a **theoretical setup**. Assume:

⁷ This implies it may be computationally inefficient.

⁸ Note that it is a subcomplex of the Delaunay triangulation on S . It is planar if S is.

⁹ Sets $V_s \cap B(s, r)$ are intersections of closed convex sets thus closed and convex themselves.

¹⁰ To see $\bigcup_{s \in S} B(s, r) = \bigcup_{s \in S} (V_s \cap B(s, r))$, take any $x \in \bigcup_{s \in S} B(s, r)$ and note that if $s \in S$ is a closest point to x in S , then $x \in V_s \cap B(x, r)$.

¹¹ Assume all the radii are the same. For different radii there is a well studied concept of a weighted alpha complex.

Figure 5.8: Alpha complexes corresponding to the situation in Figure 5.6. Note that the alpha complexes are smaller (or equal) yet still homotopy equivalent to the corresponding Čech complexes. For larger r the Čech complexes become higher-dimensional while the alpha complexes of planar subsets maintain the dimensionality bound 2.

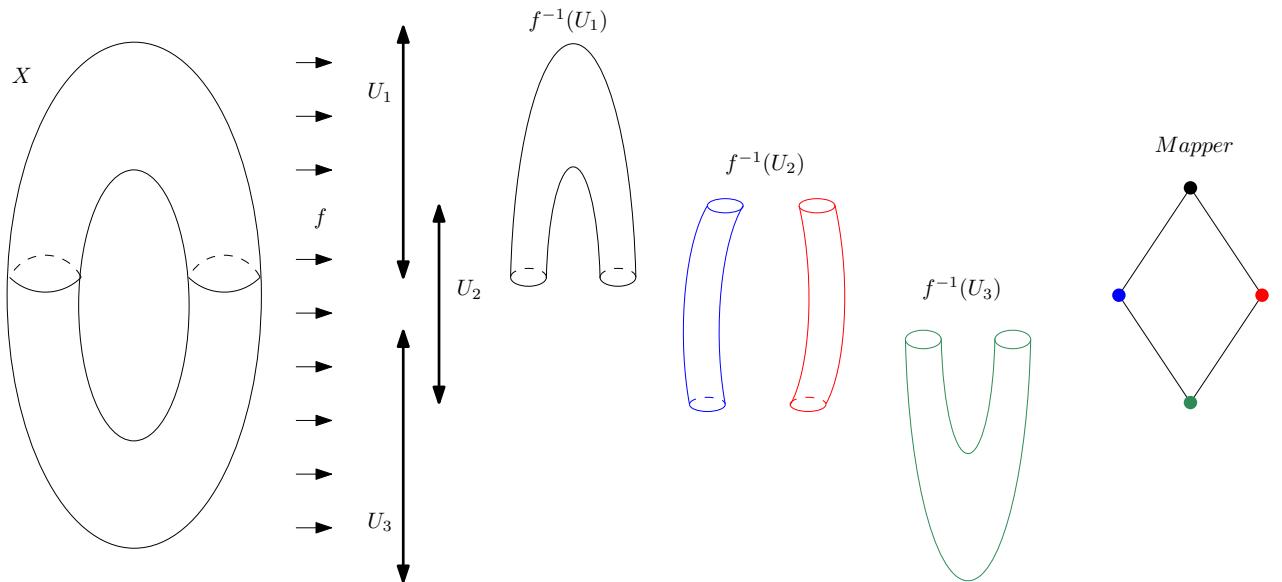
A decomposition into regions of the form $V_s \cap B(s, r)$ (on the left) mimics the decomposition of molecules into atoms.

¹¹ By the definition that will be provided, a Mapper is a simplicial complex of arbitrary dimension. However, our discussion and examples will focus on one-dimensional case, as do the practical applications in which Mapper is used.

- X is a metric space;
- $f: X \rightarrow [0, 1]$ is a (continuous) map¹²;
- \mathcal{U} is a collection¹³ of subsets of $[0, 1]$, whose union is $[0, 1]$.

Definition 5.3.5. For each U let \mathcal{V}_U denote the collection of all components of $f^{-1}(U)$ and define $\mathcal{V} = \bigcup_{U \in \mathcal{U}} \mathcal{V}_U$ as the collection of all subsets of X that appear as a component of a preimage $f^{-1}(U)$ for some $U \in \mathcal{U}$. **Mapper** is defined as $\mathcal{M}(X, f, \mathcal{U}) = \mathcal{N}(\mathcal{V})$.

An example is provided by Figure 5.9.



In practice data is often given as a finite set of points along with certain measurements. For example, we may have a collection of patients along with their heart rate and blood pressure, or a collection of basketball players along with their statistics, etc. In this case a modified **practical setup** comes into play. Assume:

- X is a finite set;
- $f: X \rightarrow I$ is a map (measurement);
- \mathcal{U} is a chosen partition of $[0, 1]$ into intervals, typically of fixed length $\varepsilon > 0$.
- Π is a chosen clustering scheme¹⁴ on X .

¹² We restrict ourselves to the cases when the target space is $[0, 1]$. However, there is no theoretical reason for doing so and the construction is well defined even if we replace $[0, 1]$ by some more complicated space.

¹³ Typically we restrict to cases when no three subsets of \mathcal{U} intersect. In such cases Mapper is a one-dimensional simplicial complex.

Figure 5.9: The construction of Mapper: a space (the torus on the left), a continuous map (projection f onto the vertical axis), cover $\mathcal{U} = \{U_1, U_2, U_3\}$ of the interval $[0, 1]$, a decomposition of the preimages into the four components and the resulting graph (right).

¹⁴ This step possibly includes additional choices of clustering algorithms.

Definition 5.3.6. For each $U \in \mathcal{U}$ let \mathcal{V}_U denote the collection of all clusters of $f^{-1}(U)$ with respect to Π and define $\mathcal{V} = \bigcup_{U \in \mathcal{U}} \mathcal{V}_U$ as the collection of all subsets of X that appear as a cluster of a preimage $f^{-1}(U)$ for some $U \in \mathcal{U}$. **Mapper** is defined as $\mathcal{M}(X, f, \mathcal{U}) = \mathcal{N}(\mathcal{V})$.

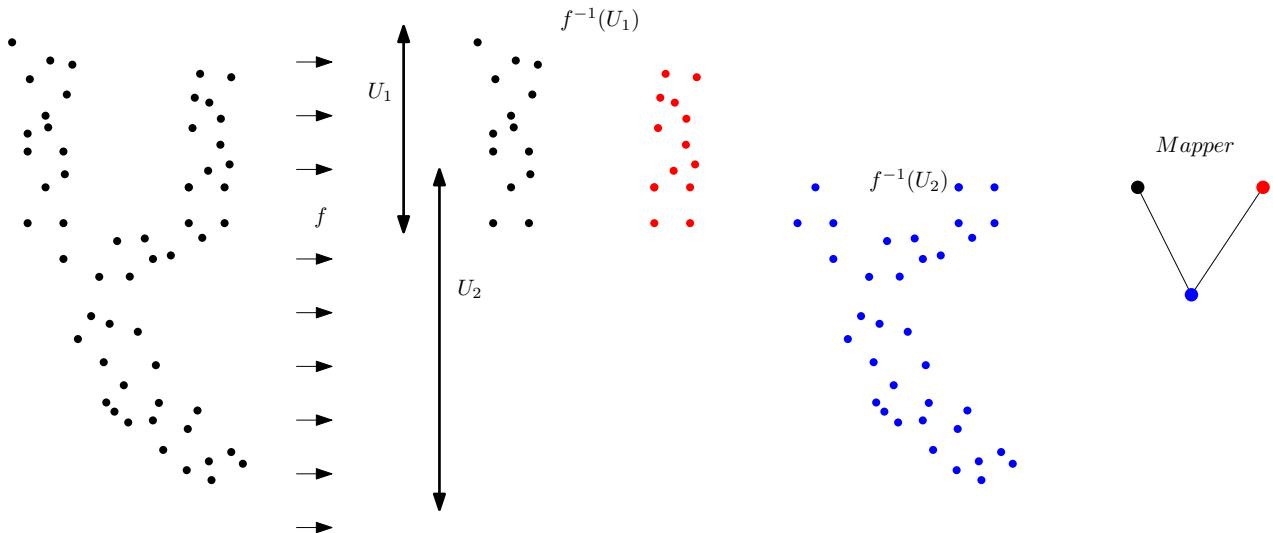


Figure 5.10: The construction of a Mapper when X is a point cloud.

While a point cloud and a number of measurements on it are often given, one has to construct a single function f and a partition \mathcal{U} , and choose other parameters very carefully to extract the desired information. Mapper is usually not analyzed further with topological tools but rather visualized, which is why the one-dimensionality is preferable.

5.4 Interleaving properties

Given a finite subset of a metric space we have described how to associate various complexes with that set. If we think for a moment about finite abstract complexes we note that these objects are discrete: it would be hard to define an obvious distance between abstract simplicial complexes. On the other hand, we have a continuous selection of inputs and input parameters: scale r is typically positive and there are reasonable notions of a distance between finite subsets of a metric space. As a result any assignment of a single complex is bound to have discontinuities¹⁵ (instabilities) of some sort, see Example 5.4.1 for a demonstration.

However, it turns out we can define a distance on filtrations, for

Example 5.4.1. Let $X = \{0, 1\} \subset \mathbb{R}$. Note that $\text{Rips}(X, r)$ changes discontinuously at $r = 1$: while $\text{Rips}(X, 1)$ is a single edge (along with the two boundary points), for each $r < 1$ the complex $\text{Rips}(X, 1)$ consists of only two vertices.

¹⁵ Unless it assigns a constant complex, of course.

which the assignment of a filtration¹⁶ becomes a continuous function of the input set and the scale parameter.¹⁷

Definition 5.4.2. Choose $\varepsilon > 0$. Filtrations $\{A_r\}_{r \geq 0}$ and $\{B_r\}_{r \geq 0}$ (obtained by the Rips or the Čech construction) are ε -interleaved if there exist simplicial maps $\varphi_r: A_r \rightarrow B_{r+\varepsilon}$ and $\psi_r: B_r \rightarrow A_{r+\varepsilon}$ such that $\varphi_{r+\varepsilon} \circ \psi_r: B_r \rightarrow B_{r+2\varepsilon}$ and $\psi_{r+\varepsilon} \circ \varphi_r: A_r \rightarrow A_{r+2\varepsilon}$ are equal to the corresponding inclusions.

Maps of Definition 5.4.2 can be visualised by drawing the following commutative¹⁸ “ladder” diagram.

$$\begin{array}{ccccccc} \cdots & \longrightarrow & A_r & \longrightarrow & A_{r+\varepsilon} & \longrightarrow & A_{r+2\varepsilon} \longrightarrow \cdots \\ & & \downarrow \varphi_r & & \downarrow & & \downarrow \\ \cdots & \longrightarrow & B_r & \longrightarrow & B_{r+\varepsilon} & \longrightarrow & B_{r+2\varepsilon} \longrightarrow \cdots \end{array}$$

Definition 5.4.3. Given two filtrations their **interleaving distance** is defined as the infimum of all values $\varepsilon > 0$, for which the filtrations are ε -interleaved.

It turns out that in our context (Rips and Čech filtrations on finite collections of points) the interleaving distance is a metric on the set of filtrations. In contrast, recall that there seems to be no geometrically meaningful metric on the set of single finite simplicial complexes.

The concept of interleaving will play an important role later in the context of the stability of persistent homology. At this point we can use it to phrase two proximity results.

Theorem 5.4.5 (Stability with respect to spaces). Choose $\varepsilon > 0$ and assume $X = \{x_1, x_2, \dots, x_k\}$ and $Y = \{y_1, y_2, \dots, y_k\}$ with $d(x_i, y_i) \leq \varepsilon, \forall i$, i.e., X and Y each consist of k points, such that the corresponding distances are at most ε . Then:

- The Rips filtrations of X and Y are 2ε -interleaved.
- The Čech filtrations of X and Y are ε -interleaved.

Proof. It follows directly from the triangle inequality (see Figure 5.11) that if a subset $\sigma \subset X$ is of diameter r , then the corresponding¹⁹ subset $\tau \subset Y$ is of diameter at most $r + 2\varepsilon$. Hence if σ is a simplex in $\text{Rips}(X, r)$, then τ is a simplex in $\text{Rips}(Y, r + 2\varepsilon)$. Consequently we may deduce that:

- maps $\text{Rips}(X, r) \rightarrow \text{Rips}(Y, r + 2\varepsilon)$ defined by $x_i \mapsto y_i$ are simplicial;

¹⁶ Of course, this eliminates the dependency on the scale parameter r .

¹⁷ For the sake of simplicity we will restrict ourselves to the mentioned Rips and Čech filtrations although the concept can be defined more generally.

¹⁸ Adjective “commutative” refers to the fact that all maps commute, i.e., going from one complex to another through any viable sequence of maps gives the same result.

Example 5.4.4. Let $X = \{0, 1\} \subset \mathbb{R}$ and $Y = \{0.1, 1.2\} \subset \mathbb{R}$. $\text{Rips}(X, r)$ consists of:

- two points for $r < 1$;
- one edge for $r \geq 1$.

$\text{Rips}(Y, r)$ consists of:

- two points for $r < 1.1$;
- one edge for $r \geq 1.1$.

The filtrations are 0.1-interleaved.

¹⁹ Subset τ is formed by taking the points of Y with the same indices as appear in the points of σ .

- maps $\text{Rips}(Y, r) \rightarrow \text{Rips}(X, r + 2\epsilon)$ defined by $y_i \mapsto x_i$ are simplicial;
- as the above two maps obviously commute with the inclusions we conclude that the Rips filtrations of X and Y are 2ϵ -interleaved;
- in a similar fashion we may conclude that the Čech filtrations of X and Y are ϵ -interleaved.

□

These conclusions tell us that if we perturb our point set slightly, the resulting filtration does not change much in terms of the interleaving distance, i.e., the construction of a filtration is stable. In a similar fashion we can express the relationship between Rips and Čech filtrations.

Rips-Čech correlation

Recall that $\text{Cech}(S, r) \subseteq \text{Rips}(S, 2r)$ and $\text{Rips}(S, r) \subseteq \text{Cech}(S, r) \subseteq \text{Cech}(S, 2r)$. This implies that the Rips and Čech filtrations, when constructed with logarithmic scales²⁰, are $(\log 2)$ -interleaved, i.e.,

$$\{\text{Rips}(S, e^r)\}_{r \geq 0} \quad \text{and} \quad \{\text{Cech}(S, e^r)\}_{r \geq 0}$$

are $(\log 2)$ -interleaved.

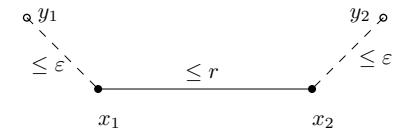


Figure 5.11: If $d(x_1, x_2) \leq r$ and $d(x_i, y_i) \leq \epsilon$ then it is apparent that $d(y_1, y_2) \leq r + 2\epsilon$.

²⁰ Note that $\text{Cech}(S, e^r) \subseteq \text{Rips}(S, 2e^r) = \text{Rips}(S, e^{\log 2+r})$ and similarly $\text{Rips}(S, e^r) \subseteq \text{Cech}(S, e^r) \subseteq \text{Cech}(S, e^{\log 2+r})$. These inclusions are the interleaving maps.

5.5 Concluding remarks

Recap (highlights) of this chapter

- Complexes: Rips, Čech, nerve, alpha, Mapper;
- Nerve theorem;
- Interleaving;

Background and applications

Constructions of simplicial complexes greatly depend on the intended use. Rips and Čech complexes along with the nerve construction are relatively well understood and have been originally introduced for theoretical purposes in the first half of the twentieth century. Čech complexes and particularly nerves were instrumental in development of Čech homology and cohomology theories, which later led to shape theory. Rips complexes have seen various independent introductions, including in geometric group theory. Their use has recently been extended to the applied setting. They are the complexes most exposed to the curse of dimensionality. Alpha complexes arose decades later

 *The curse of dimensionality: an inconvenient fact that the number of simplices typically grows fast with the dimension of a simplicial complex. This presents challenges for their computational applications, which are partially addressed by alternative constructions of complexes.*

within the realm of computational geometry and are intended for computationally intense applications. For more details and references on historical background on these complexes see a textbook²¹. There is also a modern perspective²² on nerve theorem, Dowker duality, and connections to Rips complexes. Mapper is a more recent construction²³. It is often thought of as a low-dimensional projection method and has turned out to be a commercial success.

At about the same time the interleaving distance emerged²⁴ as a measure of stability of filtrations and persistent homology, although equivalent concepts have been known in pure topology for a long time.

Appendix: the MiniBall algorithm

Given a finite subset $\sigma \subset X \subset \mathbb{R}^n$ the MiniBall algorithm²⁵ is a recursive algorithm that returns the miniball of σ , i.e., the minimal²⁶ ball in \mathbb{R}^n containing σ . As such, the algorithm provides a computational verification of the containment of σ in a Čech complex: $\sigma \in \text{Cech}(X, r)$ iff²⁷ the radius of the miniball is at most r . As the radius of the ball is also provided, the algorithm actually provides the exact lower bound for the scales r at which σ is a simplex in $\text{Cech}(X, r)$, hence a single execution of the algorithm suffices for the entire filtration.

- Input: disjoint finite sets $\tau, \nu \subset \mathbb{R}^n$.
- Output: the minimal ball with:
 - τ in the ball;
 - ν on the boundary of the ball.

Algorithm 1: Miniball(τ, ν).

```

if  $\tau = \emptyset$  then
  | compute miniball  $B$  directly;
else
  | choose  $u \in \tau$ ;
  |  $B = \text{miniball}(\tau - \{u\}, \nu)$ ;
  | if  $u \notin B$  then
    |   |  $B = \text{miniball}(\tau - \{u\}, \nu \cup \{u\})$ ;
return  $B$ 
```

The algorithm is initiated by calling $\text{Miniball}(\sigma, \emptyset)$ ²⁸ and terminates with the miniball B when $\tau = \emptyset$. It inductively scans through the points of τ . At each step it either removes a point (if removing it from the set does not change the miniball of the set) or puts a point

²¹ Herbert Edelsbrunner and John Harer. Computational Topology: An Introduction. *Applied Mathematics*. American Mathematical Society, 2010. doi: [10.1090/mhk/069](https://doi.org/10.1090/mhk/069)

²² Žiga Virk. Rips complexes as nerves and a functorial Dowker-nerve diagram. *Mediterranean Journal of Mathematics*, 18(2):58, 2021. doi: [10.1007/s00009-021-01699-4](https://doi.org/10.1007/s00009-021-01699-4)

²³ Gurjeet Singh, Facundo Memoli, and Gunnar Carlsson. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In M. Botsch, R. Palearola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007. doi: [10.2312/SPBG/SPBG07/091-100](https://doi.org/10.2312/SPBG/SPBG07/091-100)

²⁴ Steve Y. Oudot. Persistence Theory: From Quiver Representations to Data Analysis. *Number 209 in Mathematical Surveys and Monographs*. American Mathematical Society, 2015. doi: [10.1090/surv/209](https://doi.org/10.1090/surv/209)

²⁵ Emo Welzl. Smallest enclosing disks (balls and ellipsoids). In Hermann Maurer, editor, *New Results and New Trends in Computer Science*, pages 359–370, Berlin, Heidelberg, 1991. Springer Berlin Heidelberg

²⁶ Minimality is considered with respect to the radius. Such a ball is unique.

²⁷ $z \in \bigcap_{x \in \sigma} B(x, r) \Leftrightarrow \sigma \subset B(z, r)$.

\triangleleft Given random finite $\tau, \nu \subset \mathbb{R}^n$, there typically exists no ball containing τ and having ν on the boundary. The algorithm is designed so that only the pairs (τ, ν) , for which this condition is satisfied are called.

²⁸ I.e., $\tau = \sigma, \nu = \emptyset$

into ν (if removing the point decreases the miniball). When $\tau = \emptyset$ the set ν consists of at most $n + 1$ points that lie on the boundary of the miniball of σ and determine it. In this case we can use the standard circumsphere and circumradius formulas in terms of determinants to get the miniball.

Appendix: a sketch of a proof of the nerve theorem 5.3.2

A special case of the proof is illustrated by Figures 5.12, 5.13, and 5.14.

Proof. For the sake of simplicity²⁹ let us assume the nerve is of dimension 1, i.e., all triple intersections of sets of \mathcal{U} are empty. Define $X = U_1 \cup U_2 \cup \dots \cup U_k$ and $Z \subset X \times \mathcal{N}(\mathcal{U})$ as:

$$Z = \bigcup_{\sigma \in \mathcal{N}(\mathcal{U})} \left(\bigcap_{s \in \sigma} U_s \times \sigma \right).$$

We will prove that $Z \simeq X$ and $Z \simeq \mathcal{N}(\mathcal{U})$.

In order to prove $Z \simeq X$ note that for each $x \in X$ the section $(\{x\} \times \mathcal{N}(\mathcal{U})) \cap Z$ is a simplex³⁰ in the nerve spanned by all $s \in S$, for which $x \in U_s$. Contracting each such simplex to a point in a synchronized manner for each $x \in X$ we obtain a deformation of Z to X , hence $Z \simeq X$.

In order to prove $Z \simeq \mathcal{N}(\mathcal{U})$ note that for each $y \in \mathcal{N}(\mathcal{U})$ the section $(X \times \{y\}) \cap Z$ is a contractible set by assumptions. Contract first the sections of this form for all non-vertices y , and then conclude by contracting all the sections for vertices. We obtain a deformation of Z to $\mathcal{N}(\mathcal{U})$, hence $Z \simeq \mathcal{N}(\mathcal{U})$. \square

²⁹ A complete general proof is much more technical but broadly follows the same steps as are presented here.

³⁰ In our case, either an edge or a vertex.

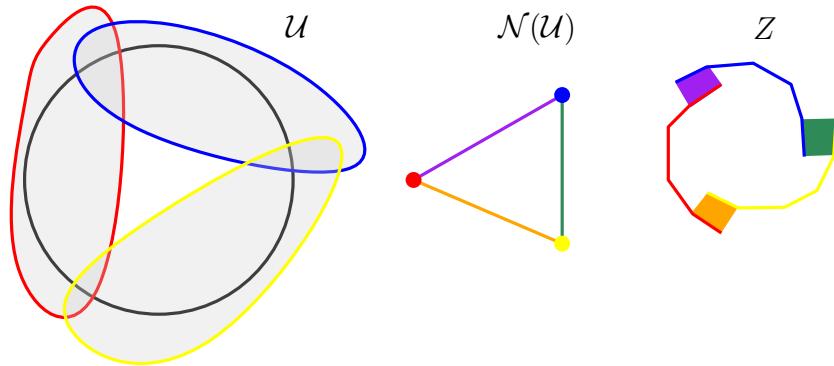


Figure 5.12: A collection \mathcal{U} of subsets of a circle $X = S^1$ (left), the corresponding nerve (center) and space Z constructed in the proof (right). The sets of \mathcal{U} are illustrated as subsets of the plane for greater clarity, while formally \mathcal{U} consists of their intersections with X .

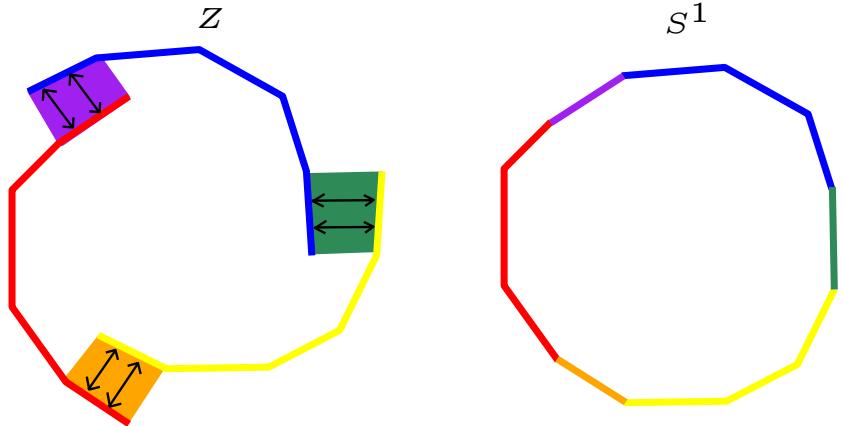


Figure 5.13: Proving $Z \simeq S^1$ we contract the sections above points of $x \in X$ in Z corresponding to edges in the nerve complex (contract along the indicated arrows on the left) to obtain S^1 .

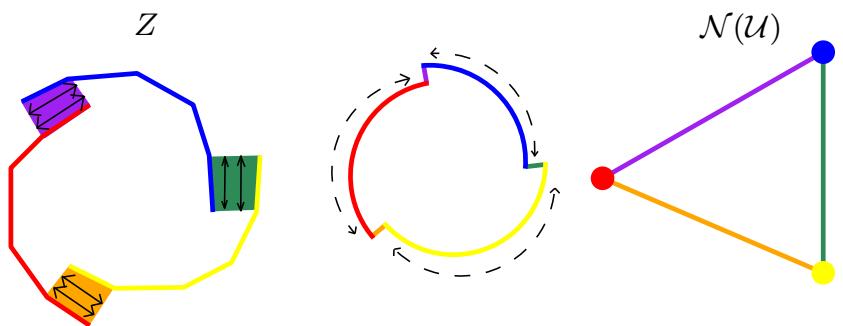


Figure 5.14: Proving $Z \simeq \mathcal{N}(\mathcal{U})$ we first contract the sections above non-vertices $y \in \mathcal{N}(\mathcal{U})$ in Z (contract along the indicated arrows on the left) to obtain the space in the center. Conclude by contracting the sections above vertices $y \in \mathcal{N}(\mathcal{U})$ in Z (contract along the indicated dashed arrows in the center) to obtain $\mathcal{N}(\mathcal{U})$ on the right.

Appendix: Dowker duality

Nerve complexes are natural complexes arising from a collection of subsets. There is another similar construction, called the Vietoris complexes, that is in a way dual to the nerve construction.

Definition 5.5.1. For $k \in \mathbb{N}$ let $\mathcal{U} = \{U_1, U_2, \dots, U_k\}$ be a collection of subsets of a finite space X , whose union is X . The **Vietoris complex** of \mathcal{U} is the abstract simplicial complex $\mathcal{V}(\mathcal{U})$ defined by the following rules:

1. The vertex set is X .
2. A set $\sigma \subseteq X$ is a simplex iff there exists $U \in \mathcal{U}$ containing σ .

We see that maximal simplices of $\mathcal{V}(\mathcal{U})$ are determined by (inclusion-wise) maximal sets of \mathcal{U} . There is a surprising connection³¹ between the nerve complexes and Vietoris complexes.

Theorem 5.5.2 (Dowker duality). For $k \in \mathbb{N}$ let $\mathcal{U} = \{U_1, U_2, \dots, U_k\}$ be a collection of subsets of a finite space X , whose union is X . Then $\mathcal{N}(\mathcal{U}) \simeq \mathcal{V}(\mathcal{U})$.

Proof. Consider $\mathcal{V}(\mathcal{U})$ as a subspace of a Euclidean space. Each U_i determines a simplex Δ_{U_i} spanned by all points of U_i . Note that $\{\Delta_{U_1}, \Delta_{U_2}, \dots, \Delta_{U_k}\}$ are closed convex sets whose union is $\mathcal{V}(\mathcal{U})$. By the nerve theorem 5.3.2 $\mathcal{V}(\mathcal{U})$ is homotopy equivalent to the nerve of $\{\Delta_{U_1}, \Delta_{U_2}, \dots, \Delta_{U_k}\}$. This nerve, on the other hand, is actually $\mathcal{N}(\mathcal{U})$ via the correspondence $\Delta_{U_i} \mapsto U_i$, see Figure 5.15 for a visual sketch of the proof. \square

 Dowker duality actually holds for an arbitrary collection of subsets \mathcal{U} of an arbitrary set X , no additional structure is necessary. Even in such generality it can be proved with ease using a general form of the nerve theorem.

³¹ Clifford H. Dowker. Homology groups of relations. *Annals of Mathematics*, 56(1):84–95, 1952. doi: [10.2307/1969768](https://doi.org/10.2307/1969768)

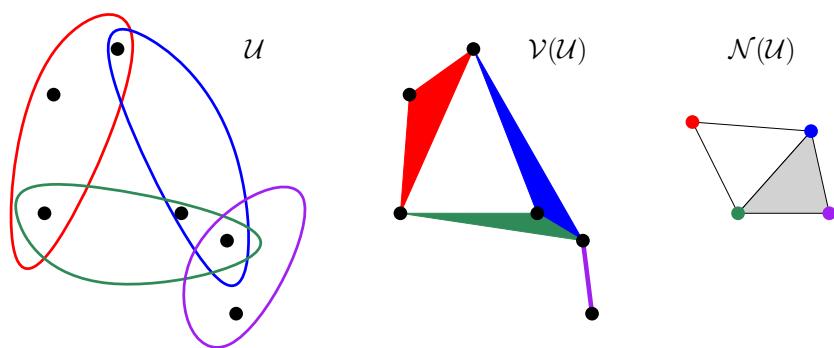


Figure 5.15: A cover of six points by four colored sets (left), its Vietoris complex (center) and nerve (right). Colored simplices on the central picture provide a collection of subsets satisfying the conditions of the nerve theorem. On the other hand, their nerve is actually $\mathcal{N}(\mathcal{U})$ on the right. By the nerve theorem we conclude the Dowker duality.

While Čech complexes are nerves associated to a collection of balls of radius r , Rips complexes are Vietoris complexes associated to a collection of sets of diameter at most r .

6

Fields and vector spaces

THE MATERIAL PRESENTED up to this point mostly falls into the premise of geometric topology and combinatorics: we introduced metric spaces and their combinatorial descriptions, simplicial complexes. Our eventual goal however is to compute meaningful topological invariants from these combinatorial descriptions. Within mathematics the field dealing with operations is called algebra and our milestone on the path to computational implementation is an algebraic formulation based on simplicial complexes. With that intention in mind we first review and introduce some algebraic concepts.

In this lecture we will present fields and vector spaces. Specific cases of the first two notions are probably familiar to the reader: real numbers and vectors in Euclidean space. We will introduce a few more fields and vector space constructions, which will provide us with enough structure to introduce homology in the next chapter.

6.1 Fields

Within the context of algebra, a field is a set with two operations satisfying a number of properties. For our purposes we will deflect a formal introduction and rather introduce specific fields which will be of our interest.

We will think of a field as our number system. We will want to be able to **add, subtract, multiply and divide (except by zero)** in our field. The fields a reader is most familiar¹ with are probably \mathbb{Q} , \mathbb{R} , and \mathbb{C} . However, there is also a family of finite fields (consisting of finitely many numbers) which often provides convenient examples: fields of remainders.

¹ \mathbb{N} is not a field as it does not contain all results of subtractions, for example, $3 - 5 \notin \mathbb{N}$. \mathbb{Z} is not a field as it does not contain all quotients by non-zero numbers, for example $3/5 \notin \mathbb{Z}$.

The fields of remainders \mathbb{Z}_p

Definition 6.1.1. Let $p \in \{2, 3, 5, \dots\}$ be a prime number. Define:

- (a) $p\mathbb{Z} = \{p \cdot n \mid n \in \mathbb{Z}\} = \{\dots, -2p, -p, 0, p, 2p, 3p, \dots\};$
- (b) $\mathbb{Z}_p = \mathbb{Z}/(p\mathbb{Z})$ as the quotient consisting of **remainders when dividing by p** .

Let us discuss (b)² in detail. The **quotient** $\mathbb{Z}/(p\mathbb{Z})$ consists of classes, each of which can be represented by a number from the “numerator” \mathbb{Z} . If $a \in \mathbb{Z}$ then the corresponding class is represented by $[a]$. Two such numbers represent the same class in the quotient iff their difference is³ in the “denominator” \mathbb{Z}_p . To phrase it differently⁴,

$$[a] = [b] \Leftrightarrow b - a \in p\mathbb{Z}.$$

Example 6.1.2. Let $p = 5$. In \mathbb{Z}_5 two numbers represent the same class iff their difference is divisible by 5. Classes $[0], [1], [2], [3], [4]$ are all distinct but⁵:

- $[5] = [0]$ as $5 - 0 = 1 \cdot 5$.
- $[6] = [1]$ as $6 - 1 = 1 \cdot 5$.
- $[-1] = [4]$ as $-1 - 4 = -1 \cdot 5$.
- $[126] = [1]$ as $126 - 1 = 25 \cdot 5$.

In particular, two positive numbers represent the same class iff their remainder when dividing by 5 is the same.

We draw another conclusion from Example 6.1.2: the most convenient representation⁶ of \mathbb{Z}_p is given by p classes $[0], [1], \dots, [p-1]$. These classes are all distinct⁷ and together form \mathbb{Z}_p .

Example 6.1.3. The structure of \mathbb{Z}_2 encodes parity: for $a \in \mathbb{Z}$ we observe that $[a] = 0$ iff a is even, and $[a] = 1$ iff a is odd.

Defining addition, subtraction and multiplication in \mathbb{Z}_p These three operations are defined in the obvious way:

$$[a] + [b] = [a + b], \quad [a] - [b] = [a - b] \quad \text{and} \quad [a] \cdot [b] = [a \cdot b].$$

It turns out that the operations are well defined⁸ in the following sense:

$$[a] = [a'], [b] = [b'] \implies [a + b] = [a' + b']$$

and the same holds for subtraction and multiplication.

² I.e., the fields of remainders and the quotient construction that defines it.

³ I.e., iff their difference is a multiple of p . In particular this means $[a] = [b]$ if the remainder after dividing by p is the same for both a and b .

⁴ What we just described is a general construction of an algebraic quotient structure. We will come across it again in the context of vector spaces.

⁵ A few more examples in \mathbb{Z}_5 : $[4] = [1], [8] = [3], [17] = [2], [1346134523451] = [1], [3457] = [2], [-23513252] = [3]$.

⁶ A few examples in \mathbb{Z}_7 : $[8] = [1] = [15], [-5] = [2] = [72]$.

⁷ They form all possible remainders after division by p .

⁸ A proof that addition is well defined:

$[a] = [a'], [b] = [b'] \implies$
 $\exists k_a, k_b \in \mathbb{Z} : a' = a + k_a p, b' = b + k_b p.$
 Thus

$$\begin{aligned} [a' + b'] &= [a + k_a p + b + k_b p] = \\ &= [(a + b) + (k_a + k_b)p] = [a + b]. \end{aligned}$$

Example 6.1.4. *Addition:*

In \mathbb{Z}_5 : $[3] + [4] = [2]$, $[3] - [4] = [4]$, $[1] + [2] = [3]$.

In \mathbb{Z}_7 : $[3] + [4] = [0]$, $[3] - [4] = [6]$, $[1] + [2] = [3]$.

Multiplication:

In \mathbb{Z}_5 : $[3] \cdot [4] = [2]$, $[2] \cdot [4] = [3]$, $[2] \cdot [3] = [1]$.

In \mathbb{Z}_7 : $[3] \cdot [4] = [5]$, $[2] \cdot [4] = [1]$, $[2] \cdot [3] = [6]$.

Example 6.1.5. Note that in $\mathbb{Z}_2 = \{[0], [1]\}$ we have $[a] = [-a]$ hence addition is the same as subtraction. In fact, if we identify $[0]$ and $[1]$ with their Boolean values, addition and multiplication encode⁹ logical operations “Exclusive or” (*XOR*) and “Conjunction” (*AND*):

$$[a] + [b] = [a \text{ XOR } b], \quad [a] \cdot [b] = [a \text{ AND } b].$$

Defining division in \mathbb{Z}_p

Up to this point the described structure of \mathbb{Z}_p did not require p to be prime. This assumption, however, is required¹⁰ if we want to define division. From number theory we know that if p is a prime, then for each number $a \in \mathbb{Z}$ with $[a] \neq [0]$, the classes $[a], [2a], \dots, [pa] = [0]$ represent the entire \mathbb{Z}_p . In particular, we can choose a coefficient k representing $[1] = [kp]$ and define the inverse of p by $[p]^{-1} = [k]$. We can consequently define the division by

$$[a] / [b] = [a] \cdot [b]^{-1},$$

which turns out to be well defined if p is prime and $[b] \neq [0]$.

Example 6.1.6. In \mathbb{Z}_5 we have $[1 \cdot 3] = [3]$, $[2 \cdot 3] = [1]$, $[3 \cdot 3] = [4]$, $[4 \cdot 3] = [2]$, $[5 \cdot 3] = [0 \cdot 3] = [0]$. The products $[k \cdot 3]$ exhaust entire \mathbb{Z}_5 and $[3]^{-1} = [2]$. Similarly, $[2]^{-1} = [3]$.

In \mathbb{Z}_7 we have $[2]^{-1} = [4]$, $[3]^{-1} = [5]$, ...

We are now able to add, subtract, multiply and divide (except by zero) in \mathbb{Z}_p , which makes \mathbb{Z}_p a field.

Remark 6.1.7. Counting and computing in \mathbb{Z}_p is surprisingly common in everyday life. It appears whenever we have a periodic behaviour.

- \mathbb{Z}_2 is a model for true/false in logic, odd/even numbers, and binary numbers.
- We use \mathbb{Z}_4 when thinking about seasons of the year.
- We use \mathbb{Z}_7 when thinking about days of the week (if today is the a^{th} day of the week then b days from today it will be $[a+b]^{\text{th}}$ day of the week).
- We use \mathbb{Z}_{10} whenever we are computing in decimal numbers. Given $a, b \in \mathbb{N}$, the first digit of $a+b$ equals $[a+b]$ in \mathbb{Z}_{10} and the same goes for multiplication.

⁹ Which means, amongst others, that these operations in \mathbb{Z}_2 are fairly natural in computer implementations, exact, and fast. In fact, computations in topological data analysis are often performed using \mathbb{Z}_2 .

¹⁰ If q is not a prime then \mathbb{Z}_q contains *divisors of zero*, i.e., non-zero classes, whose product is the zero class. For example, $[2] \in \mathbb{Z}_4$ is a non-zero class, but $[2] \cdot [2] = [4] = [0] \in \mathbb{Z}_4$ is the zero class. If we wanted to find an inverse of $[2]$ in \mathbb{Z}_4 we would need to find an integer $k \in \mathbb{Z}$, so that $[2k] = [1] \in \mathbb{Z}_4$, an unattainable feat as $2k$ is always even. A divisor of zero has no inverse.

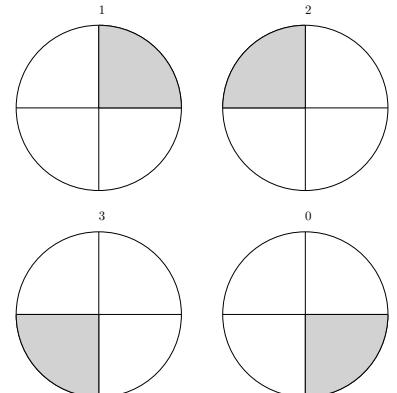


Figure 6.1: Quotient \mathbb{Z}_p models rotations by $2\pi/p$. Adding p such rotations we arrive at the original situation $0 \in \mathbb{Z}_p$. The Figure represents \mathbb{Z}_4 . Given any situation the addition of 1 is represented by a rotation by $\pi/2$ in the positive direction.

- We use \mathbb{Z}_{10} whenever we are converting units in the metric system¹¹.
- We use \mathbb{Z}_{24} when thinking about hours in a day. When thinking about hours coupled with the am/pm prefixes we actually do a combination of \mathbb{Z}_2 and \mathbb{Z}_{12} .
- We use \mathbb{Z}_{60} when thinking about minutes and seconds.

As a summary let us recall all the fields we mentioned: $\mathbb{Q}, \mathbb{R}, \mathbb{C}$, and \mathbb{Z}_p for any prime number p . These are the only fields we will be considering.

6.2 Vector spaces

Let \mathbb{F} be a field. A prototype of a vector space over field \mathbb{R} a reader is familiar with is \mathbb{R}^n for any $n \in \mathbb{N}$. It consists of n -tuples (vectors) of real numbers, which we can add, subtract, and multiply by any element of our field \mathbb{R} . In a similar way \mathbb{F}^n is a vector field over \mathbb{F} : it consists of n -tuples (vectors) of numbers from \mathbb{F} , which we can add, subtract, and multiply by any element of our field \mathbb{F} . While all our vector spaces will essentially¹² be of the form \mathbb{F}^n , some of our constructions will require us to use a more formal definition.

Definition 6.2.1. Let \mathbb{F} be a field. A **vector space** V over field \mathbb{F} is a set of elements (vectors) equipped with two operations,

1. addition $+: V \times V \rightarrow V$ and
2. scalar multiplication $\cdot: \mathbb{F} \times V \rightarrow V$

satisfying the following properties:

- addition is associative, commutative, contains the identity (zero) vector 0 , and V contains the opposite element of each vector;
- scalar multiplication is compatible, distributive, and normalized.

Roughly speaking, if we have a set of vectors we can reasonably add, subtract, and multiply by elements of some field, then this set forms a vector space.

Example 6.2.2. Let $n \in \mathbb{N}$. Given symbols v_1, \dots, v_n and a field \mathbb{F} , all formal¹³ sums $\sum_{i=1}^n a_i v_i$ where $a_i \in \mathbb{F}$ form a vector space. Operations are defined in the obvious way:

$$\sum_{i=1}^n a_i v_i + \sum_{i=1}^n a'_i v_i = \sum_{i=1}^n (a_i + a'_i) v_i, \quad \text{and} \quad b \sum_{i=1}^n a_i v_i = \sum_{i=1}^n (ba_i) v_i.$$

¹¹ As the reader might imagine, there is no reasonable algebraic explanation for the imperial system.

¹² I.e., up to isomorphism, which will be defined later.

 *Glossary of algebraic properties mentioned in Definition 6.2.1:*

- **associativity:** $(u + v) + w = u + (v + w), \forall u, v, w \in V$
- **commutativity:** $u + v = v + u, \forall u, v \in V$
- **zero vector:** $0 + v = v, \forall v \in V$ [it should always be clear from the context whether 0 denotes a number in \mathbb{F} or the zero vector in V]
- **the opposite element of $v \in V$ is denoted by $-v \in V$ and satisfies $v - v = 0$**
- **compatibility:** $(ab)v = a(bv), \forall a, b \in \mathbb{F}, \forall v \in V$
- **distributivity:** $(a+b)v = av + bv$ and $a(v+w) = av + aw, \forall a, b \in \mathbb{F}, \forall v, w \in V$
- **normalization:** $1 \cdot v = v, \forall v \in V$.

¹³ A “formal sum” in this setting means that $v_i + v_j$ is not defined as a single element v_k (as a result of a summation) in a vector space, but is rather thought of as an abstract element in itself. For example, if we want to shop for an apple and a pear, our shopping list should be apple + pear, which does not equal any other single fruit.

When $\mathbb{F} = \mathbb{Z}_2$ the corresponding vector space models the power set of v_1, \dots, v_n . A subset $\{v_{i+1}, \dots, v_{i_k}\}$ corresponds to $v_{i+1} + \dots + v_{i_k}$. The sum of two formal sums in this setting models the symmetric difference¹⁴ between the corresponding sets.

For a prime number p and $n \in \mathbb{N}$ the vector space $(\mathbb{Z}_p)^n = \mathbb{Z}_p^n$ is a finite vector space consisting of p^n elements. While this vector space appears different from \mathbb{R}^n , the formal theory, concepts, and proofs are the same in both cases. We next recast the familiar notions from \mathbb{R}^n in the setting of vector spaces over \mathbb{F} .

Let V, W be a vector space over field \mathbb{F} .

1. A linear combination of vectors in V is any expression of the form

$$\sum_{i=1}^k a_i v_i, \quad a_i \in \mathbb{F}, v_i \in V$$

2. A set of vectors $\{v_1, v_2, \dots, v_k\} \subset V$ is linearly independent¹⁵ if the only coefficients $a_i \in \mathbb{F}$ satisfying $\sum_{i=1}^k a_i v_i = 0 \in V$ are the zero coefficients, i.e., $a_i = 0, \forall i$.
3. A basis of V is a maximal¹⁶ linearly independent set in V . A vector space typically has many different bases. However, if V is finite dimensional¹⁷, then the cardinality of each basis is the same. This number is called the dimension of V .
4. A subset $U \subseteq V$ is a vector subspace [notation: $U \leq V$] of V if it is itself a vector space over \mathbb{F} .
5. A map $f: V \rightarrow W$ is linear if it is additive¹⁸ and multiplicative¹⁹. A linear map is completely determined²⁰ by the images of its basis.
6. A bijective linear map is called an isomorphism [notation: \cong]. Every vector space over \mathbb{F} of dimension $d \in \mathbb{N}$ is isomorphic to \mathbb{F}^d .
7. Let $f: V \rightarrow W$ be a linear map.

- (a) The kernel of f is defined as

$$\ker(f) = \{v \in V; f(v) = 0\} \leq V$$

- (b) The image of f is defined as

$$\text{Im}(f) = \{f(v); v \in V\} \leq W.$$

The dimension of $\text{Im}(f)$ is called the rank of f .

- (c) Given bases $\{v_1, v_2, \dots, v_k\}$ of V and $\{w_1, w_2, \dots, w_l\}$ of W , map f may be represented by an $l \times k$ matrix with entries in \mathbb{F} . If $f(v_i) = \sum_{j=1}^l a_{i,j} w_j$, then the entry at (j, i) equals $a_{i,j}$.

¹⁴ The symmetric difference of sets A, B equals $A \cup B \setminus A \cap B$.

¹⁵ Let X be a metric space and $m, n \in \mathbb{N}$. The following are vector spaces over \mathbb{F} : the set of all $m \times n$ matrices with entries in \mathbb{F} , the set of all functions $X \rightarrow \mathbb{F}$, the set of all continuous functions $X \rightarrow \mathbb{F}$, the set of all differentiable functions $X \rightarrow \mathbb{F}$ if $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$, ... Operations on functions in these examples are defined pointwise.

¹⁶ For example, vectors $(1, 3)$ and $(2, 1)$ are linearly independent in $\mathbb{R}^2, \mathbb{Q}^2, \mathbb{Z}_{13}^2$, but not in \mathbb{Z}_5^2 .

¹⁷ In particular, each element of V can be expressed uniquely as a linear combination of the basis vectors.

¹⁸ I.e., if it admits a finite basis.

¹⁹ $f(v+w) = f(v) + f(w), \forall v, w \in V$

²⁰ $f(av) = af(v), \forall a \in \mathbb{F}, v \in V$

²⁰ Consequently, a linear map can be represented by a matrix M with coefficients in \mathbb{F} if we chose bases of V and W , with the matrix-vector product $M \cdot v$ representing $f(v)$.

8. Given a matrix with coefficients in \mathbb{F} , we can still perform Gauss reduction to, for example, compute the rank of a linear map, solve systems of linear equations, ... The procedure is the same as in \mathbb{R}^n .
9. Given $U \leq V$, the quotient V/U is defined as the vector space over \mathbb{F} consisting of classes $[v]$ for $v \in V$ under the following identification²¹:

$$[u] = [v] \Leftrightarrow u - v \in U.$$

In particular, $[v] = [0]$ iff $v \in U$.

Our forthcoming descriptions of holes in simplicial complexes will be expressed in terms of dimensions and bases of vector spaces, for which the following proposition will turn out to be very handy.

Proposition 6.2.3. *Assume U, V, W are vector spaces over a field \mathbb{F} .*

1. *Let $f: V \rightarrow W$ be a linear map. Then $\text{Im}(f) \cong V/\ker(f)$.*
2. *Let $U \leq V$ be a subspace. Then $\dim(V/U) = \dim(U) - \dim(V)$.*

Proof. 1. Consider the map $g: V/\ker(f) \rightarrow \text{Im}(f)$ defined by $[u] \mapsto f(u)$. The map is:

- well defined because $[u] = [v] \implies u - v \in \ker(f) \implies f(u - v) = 0 \implies f(u) = f(v) \implies g([u]) = g([v]);$
- surjective by the definitions of $\text{Im } f$ and g ;
- injective as $g([u]) = f(u) = 0$ implies $u \in \ker(f)$ and thus $[u] = [0]$.

We conclude that g is an isomorphism.

2. Let $\{w_1, \dots, w_k\}$ be a basis²² of U . Complete it by a set $B_1 = \{v_1, \dots, v_l\}$ to a basis²³ of V . Observe that $B_2 = \{[v_1], \dots, [v_l]\}$ is a basis²⁴ of U/V :

- B_2 is linearly independent: if a linear combination of B_2 was the zero vector in V/U then the corresponding combination of the elements of B_1 was in U . This can only happen if the later combination equals 0 by the choice of B_1 and thus all the coefficients equal 0 by the linear independence of B_1
- B_2 spans the whole V/U because²⁵ B_1 and U span the whole V .

□

²¹ The operations of addition $[u] + [v] = [u + v]$ and multiplication by a scalar $a[u] = [au]$ for $a \in \mathbb{F}, u, v \in V$ are well defined by the same argument that was provided in the previous section for the fields.

²² Take the following system in \mathbb{Z}_5 :

$$\begin{aligned} (1) \quad & 2x + 3y = 2 \\ (2) \quad & 3x - y = 1. \end{aligned}$$

Multiply (1) by $2^{-1} = 3$ to obtain

$$(3) \quad x + 4y = 1.$$

To match with the leading coefficient of (2) multiply (3) by 3 to obtain

$$(4) \quad 3x + 2y = 3.$$

Now subtract (4) – (2) to obtain

$$3y = 2$$

and thus $y = 4$ and $x = 2$.

²² This implies $\dim(U) = k$.

²³ This implies $\dim(V) = k + l$.

²⁴ This implies $\dim(V) = l$ and thus proves our claim. Furthermore, it demonstrates a way to obtain a basis of U/V .

²⁵ Take any $v \in V$ and express it as $v = v' + v''$, where $v' \in U$ and v'' is a linear combination of B_1 . Then $[v] = [v'']$.

6.3 Concluding remarks

Recap (highlights) of this chapter

- fields, vector spaces
- quotients and dimension

Background and applications

Fields and abstract vector spaces have a long presence in mathematics going back centuries. Finite fields are attractive for computational implementation due to their simplicity. Computations in them are typically faster than in real numbers. Furthermore they are resistant to some numerical issues present in reals and floating point computations. On the other hand, there is a potential issue of overflowing with computer stored numbers, say integers. Algebraically it stems from the fact that counting with integers in a computer is typically performed in \mathbb{Z}_p where $\log_2 p$ is the number of bits assigned to a variable. Some of related issues are described in a non-technical book²⁶, including an interesting rule in Swiss train regulations²⁷.

Algebraic predecessor of fields and vector spaces are (algebraic) groups, another classical subject of algebra, which is now present in virtually every corner of mathematics. For further algebraic background see a textbook²⁸.

Homology of the forthcoming section is typically introduced through groups in the theoretical setting, while in practice fields are used almost exclusively. A short recap of groups is given in the appendix. Persistent homology, on the other hand, is almost exclusively introduced through coefficients in a field and the resulting persistence modules due to the accessible description in terms of a persistence diagram.

Appendix: A very short introduction to Abelian groups

Definition 6.3.1. An **Abelian group** $(G, +)$ is a set G with an associative commutative operation $+ : G \times G \rightarrow G$, such that:

1. there exists the zero element $0 \in G$ satisfying $0 + g = g + 0, \forall g \in G$;
2. for each $g \in G$ there exists its converse $-g \in G$ satisfying $g + (-g) = 0$.

Examples of Abelian groups include $(\mathbb{R}, +)$, $(\mathbb{C}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{Z}, +)$,

²⁶ Matt Parker. Humble pi: a comedy of maths errors. *Allen Lane*, 2019

²⁷ Apparently trains in Switzerland are not allowed to have an effective total number of axles equal to 256. Note that $256 = 2^8$.

²⁸ David S. Dummit and Richard M. Foote. Abstract algebra. *Wiley*, 3rd edition, 2004

☞ The term "Abelian" refers to commutativity. If the commutativity condition is not satisfied, the structure is called a (non-Abelian) group. These include the groups of permutations (with the operation being the composition) on n elements, the groups of isometries of a metric space (with the operation being the composition), the group of invertible matrices (with the operation being the product), etc.

☞ We will typically shorten $a + (-b)$ to $a - b$.

$(\mathbb{Z}_q, +)$ for any q , $(\mathbb{R} \setminus \{0\}, \cdot)$, $(\mathbb{C} \setminus \{0\}, \cdot)$, $(\mathbb{Q} \setminus \{0\}, \cdot)$, $(\mathbb{Z}_q \setminus \{0\}, \cdot)$ for any prime p , etc.

Many definitions concerning groups are the same as those of fields and vector spaces.

Definition 6.3.2. Suppose G, H are Abelian groups. A map $f: G \rightarrow H$ is a **homomorphism** iff $f(a+b) = f(a) + f(b), \forall a, b \in A$. A bijective homomorphism is called an **isomorphism** [notation: \cong].

Suppose G, H are Abelian groups and map $f: G \rightarrow H$ is a homomorphism.

1. A subset $G' \subseteq G$ is a subgroup [notation: $G' \leq G$] of G' if it is itself a group for the same operation.
2. The kernel of f is defined as

$$\ker(f) = \{a \in G; f(a) = 0\} \leq G.$$

3. The image of f is defined as

$$\text{Im}(f) = \{f(a); a \in G\} \leq H.$$

4. A set of elements $a_1, a_2, \dots, a_k \in G$ is called a generating set²⁹ of G , if each element of G can be expressed³⁰ as a sum³¹ $\sum_{i=1}^k n_i a_i$ for some $n_i \in \mathbb{Z}$.
5. Group G is finitely generated if there exists a finite generating set.
6. If $G' \leq G$, the quotient G/G' is defined as the group consisting of classes $[a]$ for $a \in G$ under the following identification:

$$[a] = [b] \Leftrightarrow a - b \in G'.$$

7. The direct sum of groups G and H is the group denoted by $G \oplus H$ and defined as

$$G \oplus H = \{(a, b); a \in G, b \in H\}$$

and the operation being defined coordinate-wise.

A remarkable fact about finitely generated Abelian groups is that they can be classified in a wonderful way.

²⁹ Or just “generators”.

³⁰ As opposed to vector spaces, such expressions in groups are often not unique, which is why the expression “generating set” is used instead of “basis”.

³¹ For $n \in \mathbb{N}$ and $a \in G$ we define

$$n \cdot a = \underbrace{a \cdot a \cdot \dots \cdot a}_{n-\text{times}}$$

and $(-n) \cdot a = -(n \cdot a)$.

Theorem 6.3.3. [Classification theorem for finitely generated Abelian groups] Let G be a finitely generated Abelian group. Then there exist :

- $k, r \in \{0, 1, \dots\}$,
- $q_1, q_2, \dots, q_k \in \mathbb{N}$, and
- prime numbers $p_1, p_2, \dots, p_k \in \mathbb{N}$,

such that G is isomorphic to

$$\underbrace{\mathbb{Z}^r}_{\text{free part of } G} \oplus \underbrace{\mathbb{Z}_{p_1^{q_1}} \oplus \mathbb{Z}_{p_2^{q_2}} \oplus \dots \oplus \mathbb{Z}_{p_k^{q_k}}}_{\text{torsion of } G}.$$

Number $r = \text{rank}(G)$ is called the rank of G .

Example 6.3.4. $\mathbb{Z}_{12} \cong \mathbb{Z}_3 \oplus \mathbb{Z}_4$, while $\mathbb{Z}_4 \not\cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$: for each element $a \in \mathbb{Z}_2 \oplus \mathbb{Z}_2$ we have $a + a = 0$, while the same does not hold in \mathbb{Z}_4 .

Proposition 6.3.5. Suppose G, H are Abelian groups, a map $f: G \rightarrow H$ is a homomorphism, and $G' \leq G$. Then:

1. $\text{Im}(f) \cong G / \ker(f)$.
2. $\text{rank}(G/G') = \text{rank}(G) - \text{rank}(G')$.

7

Homology: definition and computation

NOW THAT WE PRESENTED combinatorial and algebraic prerequisites, we are ready to define homology. The notion of homology arose from the need to detect the holes in a simplicial complex or a more general space. Its definition is not as straight forward as one might hope, but nonetheless results in a notion amenable to practical computations and consistent with the geometric intuition we presented in the first chapter.

In this chapter we will journey through a geometric introduction and definition of homology, and study the basic methods of computation. We will provide examples of homologies, which should build up our understanding and detection of holes of all dimension not only for subspaces of Euclidean spaces, but also within the combinatorial context of abstract simplicial complexes.

7.1 Definition

Homology measures holes in a simplicial complex. As the latter is provided by a collection of simplices, we need to devise a computational framework based on the simplices that will result in a meaningful result. The formal treatment of this section will be provided in parallel to a simple example on the right.

Let K be an abstract simplicial complex of dimension n and choose a field of coefficients \mathbb{F} .

Chains

Chains are formal sums of simplices along with coefficients from \mathbb{F} . They are an algebraic model of collections of simplices as demonstrated in Figure 7.3.

For each $p \in \{0, 1, \dots, n\}$ let n_p denote the number of simplices of dimension p in K .

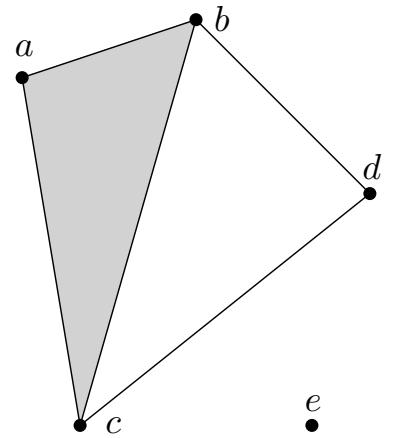


Figure 7.1: Abstract simplicial complex L .

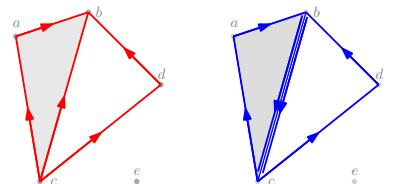


Figure 7.2: Two 1-chains in L : the red chain on the left $\langle c, a \rangle + \langle a, b \rangle + \langle c, d \rangle + \langle d, b \rangle + \langle c, b \rangle$ coincides with the blue chain on the right $\langle c, a \rangle + \langle a, b \rangle + \langle c, d \rangle + \langle d, b \rangle - 2\langle c, b \rangle = \langle c, a \rangle + \langle a, b \rangle + \langle c, d \rangle + \langle d, b \rangle + 2\langle b, c \rangle$ iff the coefficients are from \mathbb{Z}_3 .

Definition 7.1.1. A p -chain is a formal sum $\sum_{i=1}^{n_p} \lambda_i \sigma_i^p$ with $\lambda_i \in \mathbb{F}$ and σ_i^p being an oriented simplex of dimension p in K for each i .

This formalism incorporates orientation: if σ is an oriented simplex then $(-1) \cdot \sigma = -\sigma$ is the simplex σ with the opposite orientation.

We assume that $\{\sigma_1^p, \sigma_2^p, \dots, \sigma_{n_p}^p\}$ is the collection of all p -simplices of K . The p -simplices that are "absent" in a p -chain have coefficient 0. p -chains can be added/subtracted and multiplied by any scalar:

$$\sum_{i=1}^{n_p} \lambda_i \sigma_i^p + \sum_{i=1}^{n_p} \lambda'_i \sigma_i^p = \sum_{i=1}^{n_p} (\lambda_i + \lambda'_i) \sigma_i^p \quad \forall \lambda_i, \lambda'_i \in \mathbb{F}.$$

$$k \sum_{i=1}^{n_p} \lambda_i \sigma_i^p = \sum_{i=1}^{n_p} (k \lambda_i) \sigma_i^p, \quad \forall k, \lambda_i \in \mathbb{F}.$$

Example 7.1.2. Consider the simplicial complex L from Figure 7.1. Two examples of 1-chains and their additions are presented in Figure 7.3.

- Working in \mathbb{Z}_2 (top of Figure 7.3) the 1-chains are merely subsets of the collection of edges as the orientation does not matter ($+1 = -1$ in \mathbb{Z}_2). Adding the red chain $\{a, c\} + \{b, c\}$ and the blue chain $\{b, c\} + \{b, d\}$ results in the purple chain $\{a, c\} + \{b, d\}$.
- Computing in any other field (bottom of Figure 7.3) the orientation does matter. Adding the red chain $\langle b, a \rangle + \langle a, c \rangle + \langle c, d \rangle$ and the blue chain $\langle b, a \rangle + \langle d, c \rangle$ results in the purple chain $\langle a, c \rangle + 2\langle b, a \rangle$.

As a result the collection of all chains forms a vector space¹.

Definition 7.1.3. The **chain group** $C_p(K; \mathbb{F})$ is the vector space of all p -chains.

Thinking of p -simplices of K as an abstract collection of linearly independent vectors, the resulting linear space (with coefficients in \mathbb{F}) spanned by them is the chain group. If n_p is the number of p -simplices of K then $C_p(K; \mathbb{F}) \cong \mathbb{F}^{n_p}$.

Boundary

With the definition of chain groups in place, we can now express the boundary relation as a linear map. The boundary map encodes the assembly instruction for a simplicial complex.

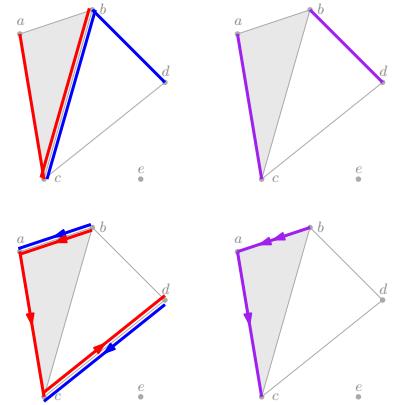


Figure 7.3: Top row: addition of chains in \mathbb{Z}_2 . Bottom row: addition of chains in any other field.

¹ For historical and practical reasons we will match the established terminology in the literature and call this vector space a chain group, the reason being that if the coefficients are in a group (as is standard in classical theoretical approaches, see also Appendix), the resulting chains form only a group. In our case the chains still form a group for addition, but the overall structure along with multiplication by a scalar is that of a vector space.

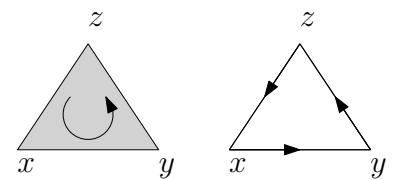


Figure 7.4: Oriented triangle $\langle x, y, z \rangle$ and its boundary $\partial_2(\langle x, y, z \rangle) = \langle x, y \rangle + \langle y, z \rangle + \langle z, x \rangle$.

Definition 7.1.4. Let $p \in \mathbb{N}$. The **boundary map**

$$\partial_p: C_p(K; \mathbb{F}) \rightarrow C_{p-1}(K; \mathbb{F})$$

is the linear map defined by the following rule on the basis of $C_p(G; \mathbb{F})$: for each oriented p -simplex $\sigma = \langle v_0, v_1, \dots, v_p \rangle$ the image $\partial_p \sigma$ is the sum of facets of σ equipped with the induced orientation from σ , i.e.,:

$$\partial_p \sigma = \sum_{i=0}^p (-1)^i \langle v_0, v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_p \rangle.$$

For technical reasons we additionally define $\partial_0: C_0(K; \mathbb{F}) \rightarrow 0$ to be the trivial map (actually, the only map) into the trivial vector space (the space only containing the 0 vector).

A crucial fact for the algebraic formulation of a homology is that the composition of two boundary maps is the trivial map. In particular, this implies that the image of a boundary map is contained in the kernel of the subsequent boundary map. See the note on the right concerning the notation in the following statement.

Theorem 7.1.5. $\partial^2 = 0$.

Proof. It suffices to prove that $\partial^2 \sigma = 0$ for an oriented p -simplex $\sigma = \langle v_0, v_1, \dots, v_p \rangle$. Note that $\partial^2 \sigma$ is a formal sum of faces of σ of dimension $p - 2$. Choose indices $i < j$ from $\{0, 1, \dots, p\}$ and consider how does the face²

$$\sigma' = \langle v_0, v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_{j-1}, v_{j+1}, \dots, v_p \rangle$$

appear in $\partial^2 \sigma$. Such a face appears from two terms:

- By first removing vertex v_j from σ in the expression of ∂_p and then removing vertex v_i from the resulting simplex in the expression of ∂_{p-1} . The indices of removed vertices are j and i hence the sign in front of σ' is $(-1)^i(-1)^j$.
- By first removing vertex v_i from σ in the expression of ∂_p and then removing vertex v_j from the resulting simplex in the expression of ∂_{p-1} . The indices of removed vertices are i and³ $(j-1)$ hence the sign in front of σ' is $(-1)^i(-1)^{j-1}$.

As the signs are opposite, the total sum equals zero. \square

Corollary 7.1.6. $\text{Im}(\partial) \subset \ker(\partial)$.

☞ The 0 vector is also called the trivial vector. The trivial map between vector spaces is the map whose image is the 0 vector.

⚠ We will typically be dropping the index of the boundary map ∂ whenever it will be evident either that the statement relating to the use of ∂ refers to all indices p or to a specific p . For example, when talking about $\partial \sigma^p$, it is apparent that the map in question is ∂_p . On the other hand, notation $\partial^2 = 0$ means that for each $p \in \mathbb{N}$, $\partial_p \circ \partial_{p-1}$ is the trivial map whose image is the zero vector.

² The following face is obtained from σ by dropping vertices v_i and v_j .

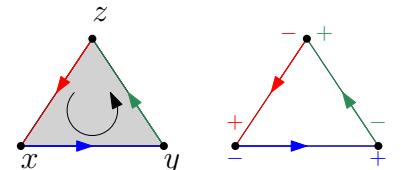


Figure 7.5: An example of Theorem 7.1.5: Oriented triangle $\langle x, y, z \rangle$ on the left, its boundary $\partial(\langle x, y, z \rangle) = \langle x, y \rangle + \langle y, z \rangle + \langle z, x \rangle$, and $\partial^2(\langle x, y, z \rangle) = \langle x \rangle - \langle x \rangle + \langle y \rangle - \langle y \rangle + \langle z \rangle - \langle z \rangle = 0$ as indicated by the signs at the vertices on the right.

³ As vertex v_i has already been removed and $i < j$, the vertex v_j is now on position $j-1$.

Definition 7.1.7. The collection of chain groups bound together by the boundary maps is called the **chain complex**:

$$\cdots \xrightarrow{\partial} C_n(K; \mathbb{F}) \xrightarrow{\partial} C_{n-1}(K; \mathbb{F}) \xrightarrow{\partial} \cdots \xrightarrow{\partial} C_1(K; \mathbb{F}) \xrightarrow{\partial} C_0(K; \mathbb{F}) \xrightarrow{\partial} 0$$

For computational purposes the boundary maps are typically represented as matrices with entries in \mathbb{F} . For each $p \in \mathbb{N}$ a matrix M_p corresponding to ∂_p is obtained as follows:

- Columns are enumerated by oriented p -simplices of K .
- Rows are enumerated by oriented $(p-1)$ -simplices of K .
- Entry at position (i,j) equals $+1$ or -1 if the i -th row appears with orientation $+1$ or -1 correspondingly in the boundary of the j -th column. All other entries are zero.

Example 7.1.8. In particular, the boundary $\partial\alpha$ of a chain α is obtained by multiplying the boundary matrix with the natural representation of α in the chosen⁴ basis.

Labeled boundary matrices⁵ for complex L of Figure 7.6:

$$M_2 = \begin{pmatrix} \langle a, b, c \rangle \\ \langle a, b \rangle & 1 \\ \langle b, c \rangle & 1 \\ \langle a, c \rangle & -1 \\ \langle b, d \rangle & 0 \\ \langle c, d \rangle & 0 \end{pmatrix}, M_1 = \begin{pmatrix} \langle a \rangle & \langle a, b \rangle & \langle b, c \rangle & \langle a, c \rangle & \langle b, d \rangle & \langle c, d \rangle \\ \langle b \rangle & -1 & 1 & -1 & -1 & \\ \langle c \rangle & 1 & -1 & 1 & -1 & \\ \langle d \rangle & & & & 1 & 1 \\ \langle e \rangle & & & & & \end{pmatrix}$$

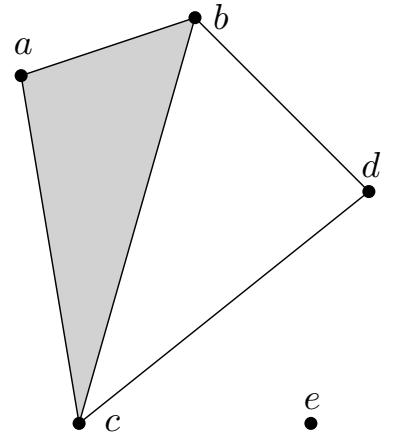


Figure 7.6: Abstract simplicial complex L .

⁴ The same basis that is used to enumerate rows.

⁵ Only non-zero entries are provided. Matrix M_0 has no formal rows as it represents the zero-map into the one-element vector space 0.

Homology

We are now finally ready to define homology as a measure of holes. Let us first build an intuition on the simplicial complex L from Figure 7.6. This will be followed up by a formal introduction in Definition 7.1.9.

Our task is to compute that L has one hole. In the figure the hole seems to be enclosed by edges cd, db and bc . Following this observation we decide that holes will be represented by a special kind of chains called cycles, see Figure 7.7. These are the chains that model closed simplicial loops in our simplicial complex, just as the one describing the hole in L above. Formally, we define cycles to be those chains, whose boundary is zero. These are our candidates for the representatives of holes.

However, not all cycles represent loops. For example, the top right cycle in Figure 7.7 is the boundary of a triangle and thus does not enclose any hole. Such cycles thus do not represent a hole and should

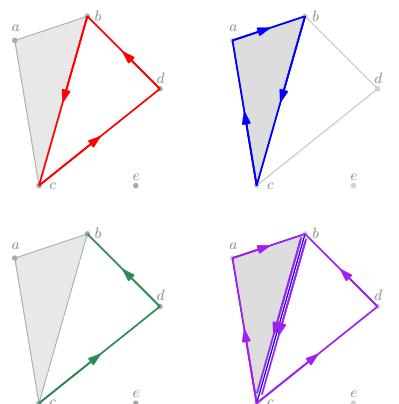


Figure 7.7: Top row: Two cycles. Bottom row: a chain that is not a cycle (left) and the cycle, that is the sum of the cycles of the top row.

be treated as trivial. Similarly, if a cycle is obtained as the boundary of a 2-chain, then it should be treated as trivial. Such cycles are called boundaries⁶ and the structure formalizing the triviality of boundaries is the quotient space.

Summing up the idea, the holes are represented by the quotient group cycles/boundaries.

Recall that for each $p \in \{0, 1, \dots\}$ we have $\text{Im } \partial_{p+1} \leq \ker \partial_p$.

Definition 7.1.9. Let K be an abstract simplicial complex. Choose a field \mathbb{F} and $q \in \{0, 1, \dots\}$. We define

- the group of q -cycles as $Z_q(K; \mathbb{F}) = \ker \partial_q \leq C_q(K; \mathbb{F})$.
- the group of q -boundaries as $B_q(K; \mathbb{F}) = \text{Im } \partial_{q+1} \leq Z_q(K; \mathbb{F}) \leq C_q(K; \mathbb{F})$.
- q -homology group as the quotient $H_q(K; \mathbb{F}) = Z_q(K; \mathbb{F}) / B_q(K; \mathbb{F})$.

The dimension of $H_q(K; \mathbb{F})$ is called the **q -Betti number** (of K with coefficients in \mathbb{F}) and is denoted by $b_q = b_q(K; \mathbb{F})$.

In particular, each element of a homology group is an equivalence class⁷ of cycles. The homology group of example L from Figure 7.6 will depend on \mathbb{F} . Defining $\alpha = \langle b, c \rangle + \langle c, d \rangle + \langle d, b \rangle$ as the top left cycle in Figure 7.7, we see that $H_1(L; \mathbb{F})$ is $\{k[\alpha] \mid k \in \mathbb{F}\}$. Even though we have only one hole, the homology group typically has more elements. However, the entire $H_1(L; \mathbb{F})$ is spanned by $[\alpha]$ and thus the number of holes should be interpreted⁸ as the dimension of the homology group, in this case 1.

More generally, each homology group with coefficients in a field \mathbb{F} is a vector space and thus isomorphic to \mathbb{F}^r for some dimension r . The main goal of our computations is thus to compute $r = b_q$, which represents the number of q -dimensional holes:

- b_0 is the same for all fields \mathbb{F} and equals the number of components of K (0-dimensional holes).
- b_1 is the number of holes in the usual geometric sense (1-dimensional holes), although various fields detect different⁹ holes in this setting. For planar graphs however, b_1 is always the number of the holes.
- b_2 is the number of caves/enclosures.

These interpretations will be explored, demonstrated and partially proved throughout the rest of this chapter. Before we do that let us mention that homology groups are homotopy invariants even though cycles and boundaries are not.

⁶ At this point, the term “boundary” can refer to a geometric boundary of a simplex, a boundary map, or a chain, that is the image of a boundary map.

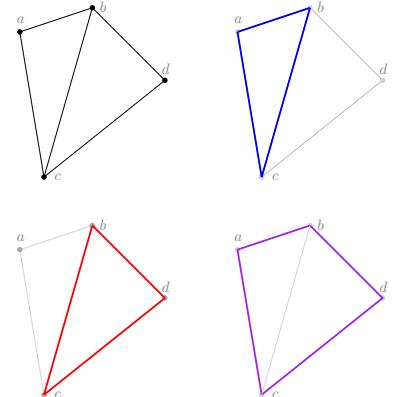


Figure 7.8: Top left: a simplicial complex with two holes. Its first homology group H_1 with coefficients in \mathbb{Z}_2 has three non-trivial elements, depicted as the blue, the red, and the purple chain. However, that does not mean that the number of holes equals 3. Along with the trivial homology class, the homology groups consists of 4 elements. This means that its dimension over \mathbb{Z}_2 equals 2, which is the number of holes. We also observe that any two of the three non-trivial chains above could form the basis of H_1 . In fact, each of the three non-trivial chains is the sum of the other two.

⁷ Given a cycle β , the corresponding class in homology will be denoted by $[\beta]$.

⁸ At this point we observe that it is crucial to preserve the algebraic structure (of a vector space) of the homology group in order to compute the dimension as the number of holes.

⁹ See the example of the Klein bottle later in this section.

Theorem 7.1.10. Assume K and K' are simplicial complexes. Then a homotopy equivalence $K \simeq K'$ implies $H_p(K; \mathbb{F}) \cong H_p(K'; \mathbb{F})$, for each field \mathbb{F} and for each $p \in \{0, 1, \dots\}$.

Zero-dimensional homology

In this subsection we prove that b_0 is the number of components¹⁰ of the underlying simplicial complex. Let K be a simplicial complex and \mathbb{F} any field. The homology group $H_0(K; \mathbb{F})$ is computed from the following piece of information:

$$C_1(K; \mathbb{F}) \xrightarrow{\partial_1} C_0(K; \mathbb{F}) \xrightarrow{\partial_0} 0.$$

In order to compute $H_0(K; \mathbb{F})$ we need to determine $\ker \partial_0$ and $\text{Im } \partial_1$. Since ∂_0 is trivial we have¹¹ $\ker \partial_0 = C_0(K; \mathbb{F})$. In order to determine $\text{Im } \partial_1$ we prove the following proposition.

Proposition 7.1.11. Let K be a simplicial complex, \mathbb{F} any field and assume $x, y \in K^{(0)}$ are vertices. Then $\langle y \rangle - \langle x \rangle \in \text{Im } \partial_1$ iff x and y lie in the same component of K .

Proof. Assume x and y lie in the same component of K . Then there exists¹² a path from x to y tracing edges. Let $x = x_0, x_1, \dots, x_k = y$ denote the sequence of vertices traced by one such path. Then the chain $\langle y \rangle - \langle x \rangle$ is the boundary of the 1-chain $\sum_{i=0}^{k-1} \langle x_i, x_{i+1} \rangle$. See Figure 7.9 for an example.

In order to prove the other direction assume $\langle y \rangle - \langle x \rangle = \partial\alpha$ for some 1-chain α . Let $K' \leq K$ be the component of K containing vertex x and define α' to be the part of α contained in K' . i.e., α' contains all those terms of α whose edge is in K' . No vertex of $K'^{(0)} \setminus \{x, y\}$ appears in $\partial\alpha'$ as none appears in $\partial\alpha$ either and the terms containing edges with such a vertex as an endpoint are the same in both α and α' . Hence $\partial\alpha'$ is either $\langle y \rangle - \langle x \rangle$ in case $y \in K'$ or $-\langle x \rangle$ otherwise. Since the coefficients in front of vertices of any boundary add¹³ up to zero, only the first of these two options is possible. \square

Assume K_1, K_2, \dots, K_n are the components of K with $x_i \in K_i, \forall i$. We now combine the following information that allows us to describe $H_0(K; \mathbb{F})$:

1. Equality $\ker \partial_0 = C_0(K; \mathbb{F})$ means $\ker \partial_0 = Z_0(K; \mathbb{F})$ has a basis $\{\langle v \rangle\}_{v \in K^{(0)}}$.
2. For each edge $\langle x, y \rangle \in K$ we have $\partial\langle x, y \rangle = \langle y \rangle - \langle x \rangle$, meaning that $\langle x \rangle$ and $\langle y \rangle$ get identified in the homology group, i.e., $[\langle x \rangle] = [\langle y \rangle]$.

¹⁰ While there are alternative ways to obtain the number of components employing a smaller amount of algebra, there are no alternatives to homological constructions when it comes to 1- and higher-dimensional holes.

¹¹ Dimension 0 is the only case where a single simplex forms a cycle.

¹² ...by the simplicial approximation theorem.

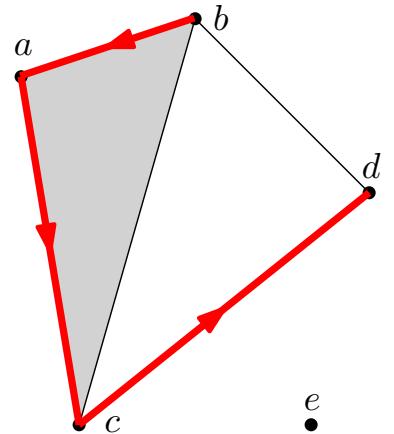


Figure 7.9: The boundary of the depicted chain is $\langle d \rangle - \langle b \rangle$, which is also the boundary of $\langle b, d \rangle$. As a consequence, the column corresponding to $\langle b, d \rangle$ in the matrix of ∂_1 is the sum of the columns corresponding to the edges of the chain.

¹³ As $\partial(k\langle z, w \rangle) = k\langle w \rangle - k\langle z \rangle$ this property holds for boundaries of single terms. By linearity of ∂ the same also holds for chains.

3. By Proposition 7.1.11 the equivalence classes of two vertices are identified in homology iff the vertices lie in the same components.
4. By 1. $\{[\langle v \rangle]\}_{v \in K^{(0)}}$ span $H_0(K; \mathbb{F})$ and by 2. and 3. so do $\{[\langle x_i \rangle]\}_{i=1}^n$.
5. The collection $\{[\langle x_i \rangle]\}_{i=1}^n$ is linearly independent, the proof of this claim being similar to the second part of the proof of Proposition 7.1.11.

As a result $\{[\langle x_i \rangle]\}_{i=1}^n$ is a basis of $H_0(X; \mathbb{F})$ and thus the dimension of $H_0(X; \mathbb{F})$ equals the number of components of K , i.e., $b_0 = n$. For example see Figure 7.10.

Homology of a graph

Let K be a simplicial complex which is a connected planar graph, and let \mathbb{F} be any field. In this subsection we prove that b_1 is the number of holes K generates in the plane.

The homology group $H_1(K; \mathbb{F})$ is computed from the following piece of information:

$$C_2(K; \mathbb{F}) \xrightarrow{\partial_2} C_1(K; \mathbb{F}) \xrightarrow{\partial_1} C_0(K; \mathbb{F}).$$

As $C_2(K; \mathbb{F}) = 0$ we have $H_1(K; \mathbb{F}) = \ker \partial_1$ so it suffices to determine the kernel of ∂_1 .

1. Let $K' \leq K$ be a maximal tree with edges e_1, e_2, \dots, e_n .
2. The collection $\partial e_1, \partial e_2, \dots, \partial e_n$ is linearly independent by the following argument¹⁴. As $K' \simeq 0$ its first homology is trivial by Theorem 7.1.10 and as K' contains no triangles, $H_1(K'; \mathbb{F}) = \ker \partial_1|_{C_1(K'; \mathbb{F})}$. In particular, $\partial_1|_{C_1(K'; \mathbb{F})}$ is injective. Its matrix contains $\partial e_1, \partial e_2, \dots, \partial e_n$ as columns and injectivity implies the columns are linearly independent.
3. Let W denote the span of $\partial e_1, \partial e_2, \dots, \partial e_n$.
4. Let $e_{n+1}, e_{n+2}, \dots, e_m$ be the edges of K that are not contained in K' , with each e_j being the edge from vertex x_j to vertex y_j .
5. Adding edges $e_{n+1}, e_{n+2}, \dots, e_m$ to K' inductively, each addition of an edge increases the number of holes generated by the resulting graph by one.
6. In a parallel fashion, each addition of an edge increases the dimension of the kernel of the first boundary map by 1 as $\partial e_j \in W, \forall j \in \{n+1, n+2, \dots, m\}$ by Proposition 7.1.11.
7. In the end of this process of adding edges we have generated $m - n$ holes and the dimension of $\ker \partial_1$ (and b_1) turns out to be $m - n$.

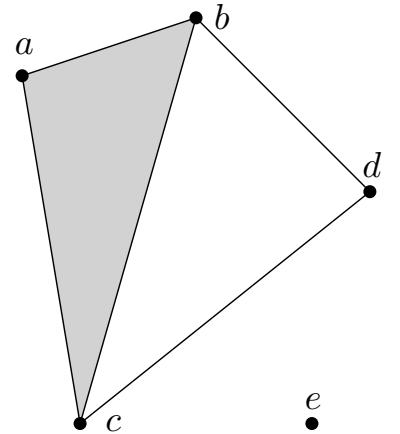


Figure 7.10: Abstract simplicial complex L . $H_0(L; \mathbb{F})$ is of dimension two (representing two components) with a basis $[< a >] = [< b >] = [< c >] = [< d >]$ and $[< e >]$. $H_1(L; \mathbb{F})$ is of dimension one representing one hole, with a basis $[\langle c, d \rangle + \langle d, b \rangle + \langle b, c \rangle]$.

¹⁴ For an alternative geometric argument see Figure 7.11.

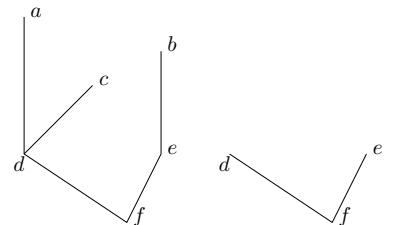


Figure 7.11: In this figure we demonstrate a geometric reason why the collection of the boundaries of all edges of a tree is linearly independent. Given a tree (on the left side of the figure) assume a linear combination of the boundaries of its edges is the zero vector. Since vertex a only appears in edge $\langle a, d \rangle$, the coefficient in front of that edge in the mentioned linear combination equals 0. The same argument holds for b and c and thus the mentioned linear combination only contains edges from the subtree on the right. Repeating the argument above, now for vertices d and e , we conclude that the mentioned linear combination is trivial and thus the claim holds. The same argument works for any tree.

8. For each $j \in \{n+1, n+1, \dots, m\}$ let c_j denote the (simplicial) path in K' from x_j to y_j represented as a 1-chain. The following form a basis of $H_1(K; \mathbb{F})$: $[e_j - c_j]$ for $j \in \{n+1, n+1, \dots, m\}$.

An example is displayed in Figure 7.12.

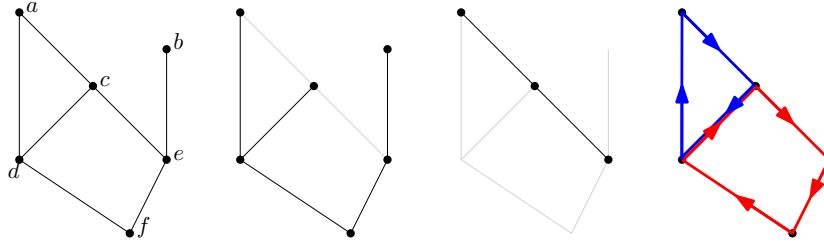


Figure 7.12: From left to right, the pictures represent a planar graph, a maximal tree, edges not contained in the chosen maximal tree and two cycles representing a basis of the first homology. Note that the graph induces two holes and thus $b_1 = 2$.

7.2 Computing homology

A systematic way to compute homology groups is through matrix reduction which allows us to obtain the rank¹⁵ of a linear map. Before we provide the details on the rank computation, let us explain how to use it in order to compute the Betti numbers.

Let K be an abstract simplicial complex of dimension n and choose a field of coefficients \mathbb{F} . For each $p \in \{0, 1, \dots, n\}$ let n_p denote the number of simplices of dimension p in K .

- Proposition 7.2.2.**
1. $\dim \ker \partial_p = n_p - \text{rank } \partial_p$
 2. $b_p = n_p - \text{rank } \partial_p - \text{rank } \partial_{p+1}$

Proof. Part 1. follows from 1. of Proposition 7.2.1 as $n_p = \dim C_p(K; \mathbb{F})$. Part 2. follows from 2. of Proposition 7.2.1 and 1. \square

Thus the Betti numbers can be expressed only using the number of simplices of a given dimension and the ranks of the corresponding boundary maps. Now let us turn our attention to rank computations. Given a boundary matrix, its rank¹⁶ is easily obtained¹⁷ from the row or column echelon form.

Echelon forms

In order to obtain a **row echelon form** of a matrix we can use the following operations¹⁸:

R1: exchange two rows;

R2: multiply a row by a non-zero element of \mathbb{F} ;

¹⁵ Given a linear map of vector spaces, its rank is the dimension of its image.

Proposition 7.2.1. Let $f: A \rightarrow B$ be a linear map of vector spaces. Then:

1. $\dim A = \dim(\ker f) + \text{rank } f$
2. $\dim(B / \text{Im } f) = \dim B - \text{rank } f$

Part 1. of Proposition 7.2.1 is a standard statement of linear algebra. Part 2. was proved in the previous chapter.

¹⁶ And thus also the rank of the boundary map. Equivalent definitions of the rank of a matrix include: the maximal number of linearly independent columns; the maximal number of linearly independent rows.

¹⁷ When using coefficients in \mathbb{R} or \mathbb{Q} the numerical procedure to obtain rank might in some cases result in certain instabilities. When using coefficients in \mathbb{Z}_p however such issues do not arise, at least not for reasonably small p .

¹⁸ Operations are considered in \mathbb{F} .

R3: add a multiple of one row to a different row.

C1: exchange two columns.

In the end we are aiming for the following transformation¹⁹

$$\left(\begin{array}{cccccc} a_{1,1} & \cdots & \cdots & \cdots & a_{1,n} \\ \vdots & & & & \vdots \\ a_{m,1} & \cdots & \cdots & \cdots & a_{m,n} \end{array} \right) \rightsquigarrow \left(\begin{array}{cccccc} 1 & * & \cdots & \cdots & * \\ 0 & \ddots & & & \vdots \\ \vdots & & 1 & * & \cdots & * \\ 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & & 0 \end{array} \right)$$

The number of the non-trivial diagonal entries, r , equals the rank of the matrix. In practice we will sometimes refrain from using C1 and only reduce to the classical row echelon form that is typically obtained through Gaussian elimination.

In a similar way we can also compute the **column echelon form** using the corresponding column operations C1, C2, C3 and (possibly) R1.

Example 7.2.3. Let us compute the homology of simplicial complex L from Figure 7.13. The boundary matrices are

$$M_2 = \begin{pmatrix} \langle a, b, c \rangle & & & & & \\ \langle a, b \rangle & 1 & & & & \\ \langle b, c \rangle & 1 & & & & \\ \langle a, c \rangle & -1 & & & & \\ \langle b, d \rangle & 0 & & & & \\ \langle c, d \rangle & 0 & & & & \end{pmatrix}, M_1 = \begin{pmatrix} \langle a, b \rangle & \langle b, c \rangle & \langle a, c \rangle & \langle b, d \rangle & \langle c, d \rangle \\ \langle a \rangle & -1 & -1 & -1 & \\ \langle b \rangle & 1 & -1 & & \\ \langle c \rangle & & 1 & 1 & -1 \\ \langle d \rangle & & & 1 & 1 \\ \langle e \rangle & & & & \end{pmatrix}.$$

Performing only row operations we obtain²⁰

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{ and } \left(\begin{array}{c|c|c|c|c} 1 & & 1 & 1 & \\ \hline & 1 & 1 & 1 & \\ \hline & & & 1 & 1 \end{array} \right)$$

These are the classical row echelon forms typically obtained through Gaussian reduction²¹ and the rank of such a matrix is the number of pivots²². The corresponding ranks of the matrices are 1 and 3. We thus have $\text{rank } \partial_2 = 1$, $\text{rank } \partial_1 = 3$, $n_2 = 1$, $n_1 = 5$, $n_0 = 5$ and we conclude:

- $b_2 = n_2 - \text{rank } \partial_2 = 0$, the complex encloses no “void”.
- $b_1 = n_1 - \text{rank } \partial_1 - \text{rank } \partial_2 = 1$, which is the number of holes.
- $b_0 = n_0 - \text{rank } \partial_1 = 2$, which is the number of components.

¹⁹ Symbols $*$ denote arbitrary elements of \mathbb{F} . The first r elements of the diagonal are declared to be 1 in our case. This is one version of the row echelon form and can always be achieved. However, there is a variant of row echelon form in which these diagonal entries are non-zero but not necessarily 1, with the other $*$ entries still being arbitrary (possibly zero) elements of \mathbb{F} . Using this variant the rank is still obtained in the same way and as a benefit, the number of row operations required to reach it is typically smaller.

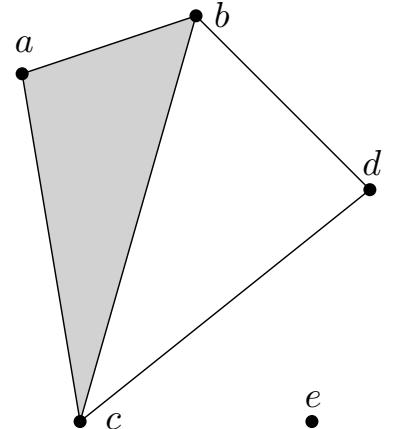


Figure 7.13: Simplicial complex L .

²⁰ The reduced form in this case coincides for all fields \mathbb{F} . Later we will see, for example with the Klein bottle, that the reduced forms and ranks in general depend on \mathbb{F} .

²¹ In order to obtain pivots only on the diagonal, as the row echelon form as we defined it requires, we would need to exchange columns 3 and 4.

²² Equivalently, the number of non-zero rows.

Row and column operations amount to changes in the bases of the domain and target vector spaces. These changes can be encoded in transformation matrices and in fact, most special forms or reductions of matrices are often expressed in terms of matrix factorizations. For our illustrative purposes though we will stick with the annotations.

Smith normal form and representatives

While the computation of the echelon forms suffices to compute the Betti numbers, we are often interested in the representing cycles²³ of homology groups as well. To that end we employ a different canonical form of a matrix: the **Smith normal form**. It is obtained from the row echelon form by eliminating the * entries to zero using the row and column operations $R1, R2, R3, C1, C2, C3$.

$$\left(\begin{array}{cccccc} a_{1,1} & \cdots & \cdots & \cdots & a_{1,n} \\ \vdots & & & & \vdots \\ a_{m,1} & \cdots & \cdots & \cdots & a_{m,n} \end{array} \right) \rightsquigarrow \left(\begin{array}{cccccc} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 \end{array} \right)$$

In order to obtain **representing cycles** though, we need to use annotated rows and columns:

- The annotations of columns from index $r + 1$ on form the basis of the kernel.
- The boundaries of annotations of columns up to r on form the basis of the image.

$$\left(\begin{array}{cccccc} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 \end{array} \right) \xrightarrow{\text{ker}}$$

²³ There are also other ways to compute the representing cycles although, at the end of the day, most of them use a similar amount of linear algebra. A high-level approach would be the following. First compute the basis of $\text{Im } \partial_{p+1}$, which is the column space of the corresponding boundary matrix. Then complete it to the basis of $\ker \partial_p$. The vectors forming the completion represent the basis of p -homology. As mentioned, there are many ways to practically formalize these steps, including the presented one through the Smith normal form.

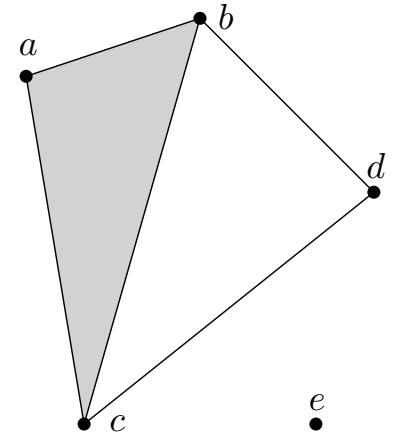


Figure 7.14: Abstract simplicial complex L .

Example 7.2.4. Let us compute the representatives of the homology groups of simplicial complex L from Figure 7.14. The annotated boundary matrices are

$$M_2 = \langle a, b, c \rangle \begin{pmatrix} \langle a, b \rangle & 1 \\ \langle b, c \rangle & 1 \\ \langle a, c \rangle & -1 \\ \langle b, d \rangle & 0 \\ \langle c, d \rangle & 0 \end{pmatrix}, M_1 = \langle a \rangle \begin{pmatrix} \langle a, b \rangle & \langle b, c \rangle & \langle a, c \rangle & \langle b, d \rangle & \langle c, d \rangle \\ \langle b \rangle & -1 & 1 & -1 & \\ \langle c \rangle & 1 & 1 & -1 & \\ \langle d \rangle & & & 1 & 1 \\ \langle e \rangle & & & & \end{pmatrix}$$

with the annotated row echelon forms being²⁴

²⁴ Only the column annotations will be displayed as the row annotations are not required.

$$\begin{pmatrix} \langle a, b, c \rangle \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} \langle a, b \rangle & \langle b, c \rangle & \langle b, d \rangle & \langle a, c \rangle & \langle c, d \rangle \\ 1 & 1 & 1 & 1 & \\ \hline & & 1 & & \\ & & & & 1 \end{pmatrix}.$$

The first of these two matrices is already in the Smith normal form.

The Smith normal form of the second matrix is:

$$\begin{pmatrix} \langle a, b \rangle & \langle b, c \rangle & \langle b, d \rangle - \langle b, c \rangle & \langle a, c \rangle - \langle b, c \rangle - \langle a, b \rangle & \langle c, d \rangle - \langle b, d \rangle + \langle b, c \rangle \\ 1 & 1 & & & \\ \hline & & 1 & & \\ & & & & \\ & & & & \end{pmatrix}.$$

We now construct the homology representatives by dimension:

Dimension 0:

1. $\ker \partial_0$ has a basis $\langle a \rangle, \langle b \rangle, \langle c \rangle, \langle d \rangle, \langle e \rangle$.
2. $\text{Im } \partial_1$ has a basis formed by the images of the first three annotated columns of the Smith normal form, i.e., $\langle a, b \rangle, \langle b, c \rangle$, and $\langle b, d \rangle - \langle b, c \rangle$. The basis obtained in this way is $\langle b \rangle - \langle a \rangle, \langle c \rangle - \langle b \rangle$, and $\langle d \rangle - \langle b \rangle + \langle c \rangle - \langle b \rangle$.
3. We may complete the basis from 2. to the basis of $\ker \partial_0$ by, for example, adding $\langle a \rangle$ and $\langle e \rangle$ and thus $\langle a \rangle$ and $\langle e \rangle$ represent the two 0-holes²⁵ spanning $H_0(L; \mathbb{F})$.

Dimension 1:

1. $\ker \partial_1$ has a basis $\langle a, c \rangle - \langle b, c \rangle - \langle a, b \rangle$ and $\langle c, d \rangle - \langle b, d \rangle + \langle b, c \rangle$.
2. $\text{Im } \partial_1$ has a basis formed by the images (boundaries) of the first annotated column of the Smith normal form, i.e., $\langle a, b, c \rangle$. The basis obtained in this way is $\langle a, b \rangle + \langle b, c \rangle - \langle a, c \rangle$.

3. We may complete²⁶ the basis from 2. to the basis of $\ker \partial_0$ by, for example, adding $\langle c, d \rangle - \langle b, d \rangle + \langle b, c \rangle$ and thus $\langle c, d \rangle - \langle b, d \rangle + \langle b, c \rangle$ represents a 1-holes spanning $H_1(L; \mathbb{F})$.

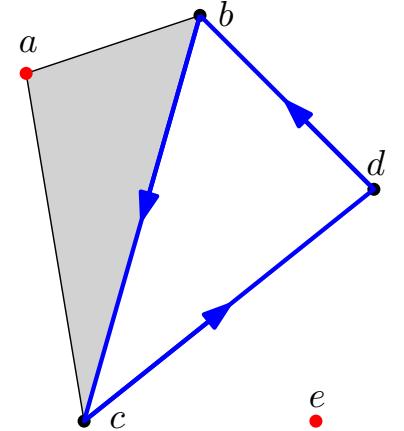


Figure 7.15: Obtained representatives of bases of the homology groups of L . Representatives $\langle a \rangle$ and $\langle e \rangle$ in red spanning $H_0(L; \mathbb{F})$, and representative $\langle c, d \rangle - \langle b, d \rangle + \langle b, c \rangle$ in blue spanning $H_0(L; \mathbb{F})$.

²⁵I.e., components.

²⁶The fact that the basis element from 2. is a member from the basis of 1. helps us to see this completion immediately. However, such a situation is an exception and a completion of basis typically involves some work with linear algebra.

Incremental expansion and elementary collapse

We conclude the section by analysing how a minimal change to a simplicial complex, an addition of one or two simplices, affects the homology.

We first discuss the **incremental expansion**, or how an addition²⁷ of a simplex to a simplicial complex changes the homology. Let K be a simplicial complex and let $\sigma^{(n)} \notin K$ be an n -simplex on vertices of K such that $K \cup \{\sigma\}$ is²⁸ also a simplicial complex. The addition of σ to K has the following effect to the homology computation scheme:

1. The number of n -simplices increases²⁹ by 1.
2. If chain $\partial\sigma$ is already contained in $\ker \partial_n$, then the addition of σ to the boundary matrix of ∂_n adds a column, which is linearly dependent on other columns and in effect, the dimension of the kernel is increased by 1.
3. If chain $\partial\sigma$ is not in $\ker \partial_n$, then the addition of σ to the boundary matrix of ∂_n adds a column, which is linearly independent on other columns and in effect, the rank of the matrix is increased by 1.

As a result (see Figure 7.16), an incremental expansion either increases b_n by 1 (case 2.), or decreases b_{n-1} by 1 (case 3.).

We next discuss an **elementary collapse**. We have already mentioned it in the chapter on simplicial complexes. Let K be a simplicial complex, $\tau^{(k-1)} \subset \sigma^{(k)} \in K$, and assume σ is the only coface of τ . A removal $K \rightarrow K \setminus \{\tau, \sigma\}$ is called an elementary collapse. It is a modification that does not change the homotopy type, and hence the homology is preserved.

Let us see how an elementary collapse effects the computation of homology.

- The boundary of σ is not a linear combination of boundaries of other k -simplices as σ is the only³⁰ coface of τ . Hence removing σ decreases $\text{rank } \partial_k$ by 1.
- The boundary of τ is a linear combination of boundaries of other $(k-1)$ -simplices by the following argument. Simplex τ is contained³¹ in the chain $\partial\sigma$. Since the boundary of this chain equals zero³², we can express $\partial\tau$ as a sum of boundaries of other facets of σ with the appropriate coefficients ± 1 . Hence τ is a linear combination of boundaries of other $(k-1)$ -simplices and thus removing it decreases $\ker \partial_{k-1}$ by 1.

In total, the dimensions of the homology groups do not³³ change.

²⁷ Or a removal, which can be analyzed in a similar fashion.

²⁸ In particular, all faces of σ should be present in K .

²⁹ This means that either the dimension of the kernel of ∂_n or the rank of ∂_n increases by 1.

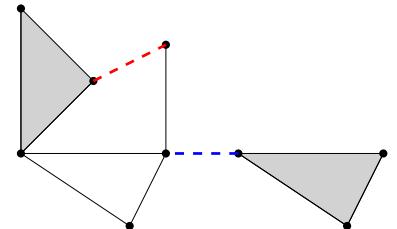


Figure 7.16: A demonstration of incremental expansion. Adding an edge to a simplicial complex may either reduce b_0 (the number of components) by 1 (blue case) or increase b_1 (the number of holes) by 1 (red case).

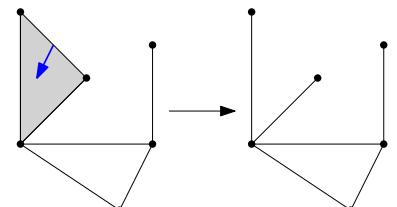


Figure 7.17: An elementary collapse.

³⁰ Meaning that $\partial\sigma$ is the only boundary of a k -simplex containing a term with τ .

³¹ ...with coefficient +1 or -1.

³² $\partial^2\sigma = 0$

³³ Recall that the only homology group that may potentially change is H_{k-1} . It is defined as $\ker \partial_{k-1} / \text{Im } \partial_k$ and since the dimension of both $\ker \partial_{k-1}$ and $\text{Im } \partial_k$ decreases by one, the dimension of the quotient is preserved.

7.3 Examples of homology

In this section we present some further aspects of homology that should aid our understanding of the concept.

Disjoint unions

Two abstract simplicial complexes are said to be **disjoint** if their collections are disjoint³⁴. Two geometric simplicial complexes are said to be **disjoint** if their bodies are disjoint. The union of disjoint simplicial complexes K, L is called a **disjoint union** and is denoted by $K \sqcup L$.

Given two disjoint simplicial complexes K, L , the homology of their disjoint union is the cartesian product³⁵ of the individual homologies: $H_i(K \sqcup L; G) \cong H_i(K; G) \times H_i(L; G)$. Computationally we can see this by observing that the boundary map ∂ has block-diagonal matrices: boundaries of chains from K lie in K and the same holds for L . Since each simplicial complex is the disjoint union of its components, the technical computations and treatments of homology are typically restricted to connected simplicial complexes.

Example 7.3.1. Given a planar graph K and any field \mathbb{F} :

- b_0 is the number of components of K .
- b_1 is the number of holes of K induces in the plane.

This is consistent with our observation for disjoint union in case K is not connected (as in Figure 7.18).

Euler characteristic

Suppose K is a simplicial complex and let n_i denote the number of i -simplices in K . Recall that the Euler characteristic $\chi(K) \in \mathbb{Z}$ is defined as $\chi(K) = n_0 - n_1 + n_2 - n_3 + \dots$

This invariant has an interesting interpretation in terms of homology.

Proposition 7.3.2. $\chi(K) = b_0 - b_1 + b_2 - b_3 + \dots$

Proof. By 2. of Proposition 7.2.2 we have $b_p = n_p - \text{rank } \partial_p - \text{rank } \partial_{p+1}$. Substituting these equality into $b_0 - b_1 + b_2 - b_3 + \dots$ we obtain χ . \square

Example 7.3.3. Given a planar graph K and any field \mathbb{F} , $\chi(K)$ equals the number of components subtracted by the number of holes K generates in the plane.

³⁴I.e., if there is no intersection between the sets of vertices. Formally speaking, if such an intersection existed it would mean that we are treating both collection of vertices as subsets of some larger set.

³⁵In our setting, the term “direct sum” could also be used.

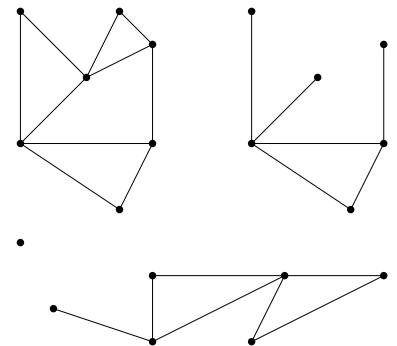


Figure 7.18: A planar graph with four components: $b_0 = 4, b_1 = 7, \chi = -3$.

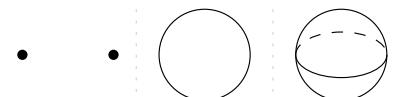


Figure 7.19: S^0 demonstrates non-trivial H_0 , S^1 represents a one-dimensional hole, and S^2 encloses a two-dimensional hole.

Spheres

Holes as measured by homology are represented by cycles and the fundamental examples of holes are provided³⁶ by spheres. In this subsection we prove that given a triangulation of an n -sphere for $n \geq 1$, the consistently oriented collection of n -simplices represents an n -hole. In fact, this is the only hole a sphere has. A convenient triangulation of S^n we will be using will be the one³⁷ consisting of all faces of an $(n+1)$ -simplex.

Proposition 7.3.4. *For each \mathbb{F} and $n \in \{1, 2, \dots\}$ we have:*

- $H_0(S^n; \mathbb{F}) \cong H_n(S^n; \mathbb{F}) \cong \mathbb{F};$
- $H_i(S^n; \mathbb{F}) = 0, \forall i \notin \{0, n\}.$

Proof. The full simplicial complex on $n+2$ points is contractible hence all its homology groups are trivial except for H_0 , which is of rank 1. Removing the only $(n+1)$ -simplex reduces $\text{rank } \partial_{n+1}$ by one and hence increases b_n by 1, as was explained in the context of incremental expansions and removals. \square

Since S^0 is a collection of two points it is easy to see³⁸. that the only non-trivial homology group of S^0 is $H_0(S^0; \mathbb{F}) \cong \mathbb{F}^2$.

Surfaces

A beautiful demonstration of the two-dimensional homology is provided by surfaces.

Proposition 7.3.5. *Let K be a triangulation of a closed (i.e., without boundary) connected orientable surface. For each group \mathbb{F} we have*

$$H_2(K; \mathbb{F}) \cong \mathbb{F}.$$

Proof. Recall that K being orientable means there exists a consistent choice of orientations on all triangles of K . Let us fix such an orientation on them.

1. The structure of a surface implies that each edge of K is a face of at most two triangles.
2. The structure of a *closed* surface implies that each edge of K is a face of precisely two triangles.
3. Consistency of orientations on triangles implies that whenever two triangles intersect in an edge, the induced orientations on the edge are the opposite.

³⁶ The homology of a metric space is, for our purposes, the homology of any triangulation of that space.

³⁷ To be precise: take an $(n+1)$ -simplex, add all of its faces to obtain a simplicial complex called the **full simplex on $n+2$ points** (sometimes also called the **full $(n+1)$ -simplex**), and then remove the $(n+1)$ -simplex to obtain a triangulation of S^n .

³⁸ An observation: The full simplex on n points is contractible hence its Euler characteristic equals 1. On the other hand, computing its Euler characteristic by definition we get n points $- \binom{n}{2}$ edges $+ \binom{n}{3}$ triangles $\dots (-1)^n \cdot 1$ n -simplex. Summing it up we get:

$$\binom{n}{1} - \binom{n}{2} + \binom{n}{3} - \dots (-1)^n \binom{n}{n} = 1,$$

by the binomial formula.

³⁸ ..by a direct computation or by the argument of Proposition 7.3.4

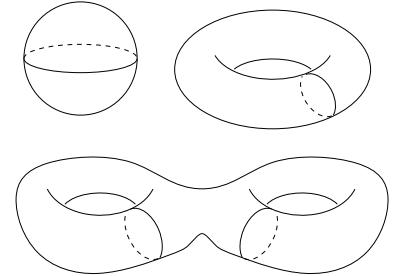


Figure 7.20: Examples of closed connected surfaces: they all enclose one 2-dimensional hole in the form of a “cave”, which is manifested in the fact that $b_2 = 1$.

\triangle The statement of Proposition 7.3.5 does not hold for connected surfaces with boundary. If there was a non-trivial 2-cycle in such a case, the same argument as in the proof of the proposition would imply that the cycle would be the oriented sum of all triangles (possibly multiplied by a single non-trivial factor $\lambda \in \mathbb{F}$). Since a presence of a boundary of a manifold implies the existence of an edge, which is a face of precisely one triangle, such a triangle (multiplied by λ) would thus appear in the boundary of the cycle, a contradiction.

The second homology of a connected manifold with a boundary is thus always trivial.

Let us define chain α as the sum of all oriented triangles. By 1.-3. above each edge appears in $\partial\alpha$ twice, once with each orientation (see Figure 7.21), and thus $\partial\alpha = 0$, meaning that α is a chain. As the image of ∂_3 is trivial, α represents a non-trivial homology class.

On the other hand, whenever a 2-cycle β contains a term³⁹ $+\sigma$, where σ is an oriented triangle, observations 2. and 3. imply that all oriented simplices sharing an edge with σ also appear in α with coefficient 1. Inductively expanding this conclusion to further neighbors we reach all triangles as K is connected and thus deduce that $\beta = \alpha$. The proposition is thus proved. \square

Homology class $[\alpha]$ generating $H_2(K; \mathbb{F})$ as defined⁴⁰ in the proof is called the **fundamental class** of the surface. In the same way we can prove that if K is a closed connected orientable manifold of dimension n , then $H_n(K; \mathbb{F}) \cong \mathbb{F}$ with the generator, which is again called the fundamental class, being the sum of all consistently oriented n -simplices of K .

The case of non-orientable surfaces is the first presented situation in which the choice of coefficients matters.

Proposition 7.3.6. *Let K be a triangulation of a closed connected non-orientable surface. Then $H_2(K; \mathbb{Z}_2) \cong \mathbb{Z}_2$ and $H_2(K; \mathbb{F}) \cong 0$ for each $\mathbb{F} \not\cong \mathbb{Z}_2$.*

Proof. As in the proof of Proposition 7.3.5, the fact that K is a surface means that if a 2-cycle α contains a term $+\sigma$ for some oriented triangle σ , it also contains a term $+\sigma'$ for each oriented triangle σ' sharing⁴¹ an edge with σ . Again, as K is connected, this means that α is the sum of all oriented triangles. However, as K is non-orientable, there is no consistent orientation on triangles and thus⁴² some edges appear with coefficient 2 in the boundary, see Figure 7.21. Thus if⁴³ $0 \neq 2$ the boundary is non-trivial and the assumed 2-cycle does not have an empty boundary, a contradiction. Hence the only 2-cycle is the trivial cycle.

However, if $\mathbb{F} = \mathbb{Z}_2$, the obtained boundary equals zero and thus α is the only non-trivial cycle. As a result, $H_2(K; \mathbb{Z}_2) \cong \mathbb{Z}_2$. \square

We may summarize these two propositions and the corresponding comments as follows:

- A connected surface K is closed iff $H_2(K; \mathbb{Z}_2) \neq 0$.
- Given any field $\mathbb{F} \neq \mathbb{Z}_2$, a closed connected surface is orientable if $H_2(K; \mathbb{F}) \cong \mathbb{F}$.

³⁹ If σ appear in the term $\lambda\sigma$ for some non-zero $\lambda \in \mathbb{F}$, we repeat the same argument for the chain divided by λ .

⁴⁰ Formally speaking, there are two fundamental classes, one for each orientation of triangles...except when $\mathbb{F} = \mathbb{Z}_2$. If $\mathbb{F} = \mathbb{Z}_2$ there is only one non-trivial homology class which is its own converse.

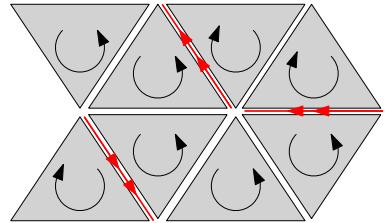
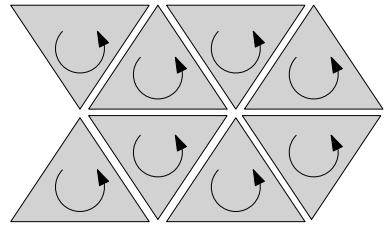


Figure 7.21: Top: The boundary of a chain consisting of all consistently oriented triangles of a surface without boundary is zero, as the induced orientations on edges cancel out. Bottom: The boundary of a chain consisting of a not-consistently oriented collection of all triangles of a surface without boundary contains each edge between two non-consistently oriented triangles twice.

⁴¹ Sharing in the sense of consistent orientation, meaning that the induced orientation on the shared edge are the opposite.

⁴² As each edge appears twice in the boundary of such a chain and not all such appearances may cancel each other out by the non-orientability.

⁴³ Equivalently, if $\mathbb{F} \neq \mathbb{Z}_2$.

Proposition 7.3.6 also generalizes to the n -dimensional homology of closed connected non-orientable n -manifolds.

Impact of coefficients: the Klein bottle

An example where the choice of coefficients makes a difference in 1-dimensional homology computations is the Klein bottle, which will be denoted by \mathcal{K} in this subsection. It is depicted in Figure 7.22. Its triangulation is given by the black portion in Figure 7.24. We already know that $b_0 = 1$ as \mathcal{K} is connected. However, the second Betti number of this closed surface depends on the coefficients due to the non-orientability:

- $b_2(\mathcal{K}; \mathbb{Z}_2) = 1$.
- For $\mathbb{F} \not\cong \mathbb{Z}_2$, $b_2(\mathcal{K}; \mathbb{F}) = 0$.

From this information and the expression of the Euler characteristic as the alternative sum of Betti numbers we conclude:

- $b_1(\mathcal{K}; \mathbb{Z}_2) = 2$, i.e., $H_2(\mathcal{K}; \mathbb{Z}_2) \cong \mathbb{Z}_2^2$.
- For $\mathbb{F} \not\cong \mathbb{Z}_2$, $b_1(\mathcal{K}; \mathbb{F}) = 1$.

These Betti numbers can also be computed through the matrix reduction. Instead of computing them, we will rather demonstrate the geometric reason for the difference in b_1 depending on the coefficients. The explanation will be based on Figure 7.24. On each of the five parts of the figure a triangulation of \mathcal{K} is provided by the black/grey portion. The black arrows indicate the direction in which the identifications are performed. A single red horizontal directed line represents a cycle α generating the extra dimension of $H_2(\mathcal{K}; \mathbb{Z}_2)$. It is also depicted in Figure 7.22. It turns out that $[\alpha]$ is homologically non-zero iff coefficients are \mathbb{Z}_2 . In order to prove this statement we first present a claim.

We claim that $2[\alpha] = 0$ in 1-dimensional homology. In order to prove the claim the leftmost part of Figure 7.24 has two copies of α drawn slightly apart from each other. The corresponding homology class does not change if we move⁴⁴ each of the copies of α separately. So let's move them as on the Figure:

- move the upper copy slightly higher;
- move the lower copy to the bottom of the side. Due to the reversed orientation, the chain then appears on the top of the square with (in the plane seemingly) reversed orientation. Moving this representative lower to the first copy of α we see, that the copies cancel each other out: they consist of the same edges with converse orientations.

As a result, the claim holds, i.e., $2[\alpha]$ is the trivial homology class. Depending on the coefficients of our computation this has the following ramifications:

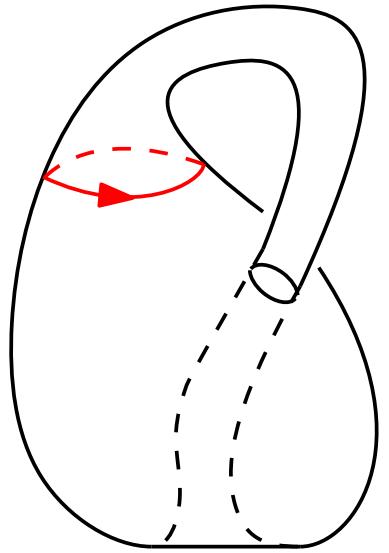


Figure 7.22: The Klein bottle.

⁴⁴ “Moving” in this setting can be thought of as a homotopic change. Formally speaking, a moved chain represents the same homology class if the difference between the original and the new chain is in the boundary group, see Figure 7.23.

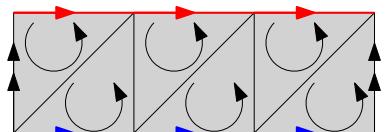


Figure 7.23: Excerpt from the transformation in Figure 7.24. The blue and the red chain represent the same homology class because their difference (blue – red) is the boundary of the 2-chain consisting of the strip of depicted oriented triangles.

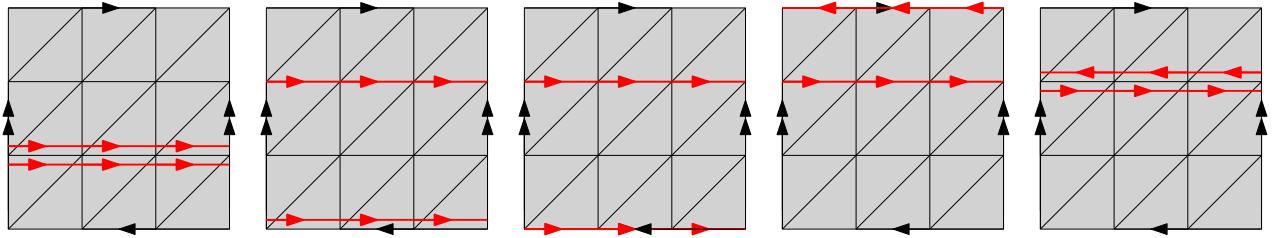


Figure 7.24: The Klein bottle.

- if $\mathbb{F} \not\cong \mathbb{Z}_2$ then we can divide equation $2[\alpha] = 0$ by 2 and obtain that $[\alpha] = 0 \in H_1(\mathcal{K}; \mathbb{F})$.
- if $\mathbb{F} \cong \mathbb{Z}_2$ then we can't divide equation $2[\alpha] = 0$ by 2 as $2 = 0$. It turns out that $[\alpha] \neq 0 \in H_1(\mathcal{K}; \mathbb{Z}_2)$ and thus α provides an extra dimension to $H_1(\mathcal{K}; \mathbb{Z}_2)$.

For an alternative argument proving the claim see Figure 7.25.

Alexander duality

Homology is defined for any abstract simplicial complex. However, if there is an underlying geometric simplicial complex K embedded in a sphere or a Euclidean space, there is a connection between the homology of K and that of its complement. The relationship is formally known as Alexander duality.

Before we state the duality we should explain a few technical details of the complement construction. Let $K \subset \mathbb{R}^2$ be a geometric simplicial complex. In particular, K consists⁴⁵ of finitely many simplices. The complement of K , denoted by $K^C = \mathbb{R}^2 \setminus K$, is unfortunately not homeomorphic to a (finite⁴⁶) simplicial complex. As a proof of this claim observe that K is a closed⁴⁷ subset of the plane, while K^C is usually⁴⁸ not. However, K^C is homotopic to a finite simplicial complex. For example see Figure 7.26. At this point we defer from specifying details of triangulation of K^C or its homotopy type and rather conclude with the declaration: K^C is homotopy equivalent to a finite simplicial complex K' and so whenever we will be talking about the homology of K^C , we will formally be thinking of the homology of K' . The same discussion applies if K is a geometric simplicial complex in any Euclidean space or a sphere.

Alexander duality provides a connection between the homologies of K and its complement.

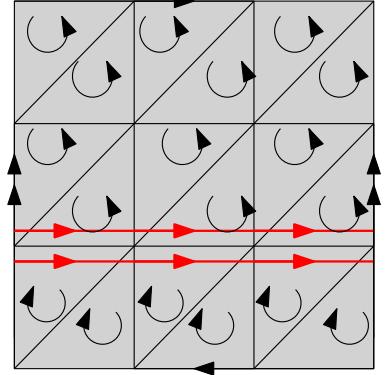


Figure 7.25: Another proof of the fact that $2[\alpha]$ is homologically trivial within the Klein bottle. The chain $2[\alpha]$ is depicted in red and is the boundary of the 2-chain consisting of all depicted oriented triangles.

⁴⁵ Formally, the body of K is the union of finitely many simplices.

⁴⁶ Recall that all simplicial complexes considered here are finite. Within the context of infinite simplicial complexes though, the complement can be triangulated and the treatment of complements presented here is immaterial.

⁴⁷ In particular, this means that the limit of each converging sequence in K lies in K .

⁴⁸ Except if K is empty.

Theorem 7.3.7 (Alexander duality). *Let $n \in \mathbb{N}$ and suppose $K \subset S^n$ is a geometric simplicial complex. Then for any coefficients \mathbb{F} we have:*

1. $b_0(K; \mathbb{F}) - 1 = b_{n-1}(K^C; \mathbb{F})$.
2. $b_{n-1}(K; \mathbb{F}) = b_0(K^C; \mathbb{F}) - 1$.
3. $b_q(K; \mathbb{F}) = b_{n-q-1}(K^C; \mathbb{F})$ for all $q \in \{1, 2, \dots, n-1\}$.

For a proof see a textbook⁴⁹. From the Alexander duality we may draw a similar conclusion for complexes in Euclidean spaces by taking into account that removing a point⁵⁰ from S^n results in a space homeomorphic to \mathbb{R}^n .

Corollary 7.3.8. *Let $n \in \mathbb{N}$ and suppose $K \subset \mathbb{R}^n$ is a geometric simplicial complex. Then for any coefficients \mathbb{F} we have:*

1. $b_0(K; \mathbb{F}) = b_{n-1}(K^C; \mathbb{F})$.
2. $b_{n-1}(K; \mathbb{F}) = b_0(K^C; \mathbb{F}) - 1$.
3. $b_q(K; \mathbb{F}) = b_{n-q-1}(K^C; \mathbb{F})$ for all $q \in \{1, 2, \dots, n-1\}$.

Alexander duality is handy when computing homology groups of simplicial complexes in Euclidean spaces or spheres. For example, instead of computing the one-dimensional homology of a planar simplicial complex, we can⁵¹ compute the number of components of its complement, which is typically much faster.

7.4 Concluding remarks

Recap (highlights) of this chapter

- Cycles, boundaries, homology
- Detecting components and holes with homology
- Computing homology through matrix reduction
- Euler characteristic
- Alexander duality

Background and applications

Homology is one of the focal invariants in topology and geometry. Homological conditions and constructions can be found throughout mathematics. We will present one of them in the appendix (Cubical

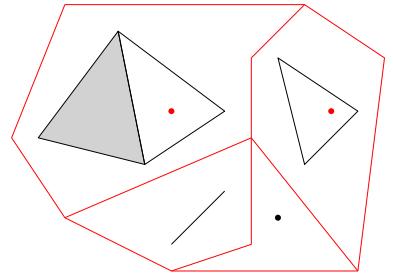


Figure 7.26: Simplicial complex K in the plane in black, and a simplicial complex K' homotopy equivalent to its complement in red. Note the number of holes of K is one less than the number of components of K' , i.e., $b_1(K) = b_0(K') - 1$. Also, number of holes of K' equals the number of components of K , i.e., $b_1(K') = b_0(K)$.

⁴⁹ James Munkres. Elements of Algebraic Topology. Perseus Books, 1984. doi: 10.1201/9780429493911

⁵⁰ A removal of a point from $K^C \subset S^n$ increases b_{n-1} by one.

⁵¹ Provided there is an easy description of a complement. Such examples would include bitmap images.

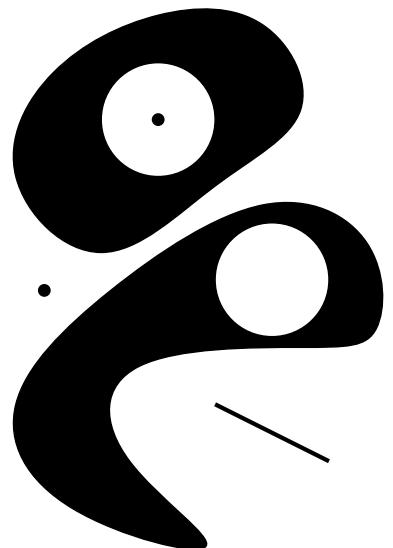


Figure 7.27: A demonstration of Alexander duality: given a bounded subset X of the plane, each component of X corresponds to a hole in X^C , and each hole in X corresponds to bounded component of X^C .

homology). The version presented here is usually called “simplicial homology” as it arises from the structure of a simplicial complex. For non-triangulated spaces a version called “singular homology” can be defined. In general though, any reasonable boundary map ∂ satisfying ∂^2 induces its own homology structure. Examples⁵² include cubical homology (see appendix) and cohomology. For a general reference we mention a few textbooks⁵³.

Amenability to algorithmic computations through matrix reductions and, as we will see later, Discrete Morse Theory makes homology an obvious tool with which we could determine topological properties of data. In practice, though, the usual homology is often superseded by persistent homology, which is a richer, parameterized version of homology described in later chapters.

Appendix: Homology with coefficients in Abelian groups

Classical introductions of homology typically consider coefficients from an Abelian group rather than a field. By far the most popular choice among non-fields is the group of integers \mathbb{Z} . In this subsection we review the construction and properties of homology using coefficients⁵⁴ in a group \mathbb{Z} .

Let K be an abstract simplicial complex of dimension n . For each $q \in \{0, 1, \dots, n\}$ let n_q denote the number of simplices of dimension q in K .

The definition of homology in this case remains the same with the only difference being that the structure of the resulting algebraic invariants is that of Abelian groups, and the boundary operator ∂ is a homomorphism:

1. A q -chain is a formal sum $\sum_{i=1}^{n_q} a_i \sigma_i^q$ where $a_i \in \mathbb{Z}$ and σ_i^q is an oriented simplex of dimension q in K .
2. The chain group $C_q(K; \mathbb{Z}) \cong \mathbb{Z}^{n_q}$ is the group of all q -chains. Its generators are oriented q -simplices of K .
3. For each $p \in \mathbb{N}$ the boundary map

$$\partial_p: C_p(K; \mathbb{Z}) \rightarrow C_{p-1}(K; \mathbb{Z})$$

is the homomorphism defined by

$$\partial_p \langle v_0, v_1, \dots, v_p \rangle = \sum_{i=0}^p (-1)^i \langle v_0, v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_p \rangle.$$

As before, $\partial^2 = 0$. Additionally define $\partial_0 = 0$.

4. The collection of chain groups bound together by the boundary homomorphisms is called the chain complex:

$$\cdots \xrightarrow{\partial} C_n(K; \mathbb{Z}) \xrightarrow{\partial} C_{n-1}(K; \mathbb{Z}) \xrightarrow{\partial} \cdots \xrightarrow{\partial} C_1(K; \mathbb{Z}) \xrightarrow{\partial} C_0(K; \mathbb{Z}) \xrightarrow{\partial} 0$$

⁵² Another example is the De Rham cohomology and exterior derivative. While the theory itself is quite involved, a snapshot of the fact that $\partial^2 = 0$ can be observed in low dimensions via specific derivatives: gradient, divergence, and curl, are specific boundary maps as the composition of a consecutive pair amongst them equals zero.

⁵³ James Munkres. Elements of Algebraic Topology. *Perseus Books*, 1984. doi: [10.1201/9780429493911](https://doi.org/10.1201/9780429493911); Allen Hatcher. Algebraic topology. *Cambridge Univ. Press, Cambridge*, 2000; and Raoul Bott and Loring W. Tu. Differential Forms in Algebraic Topology. *Springer New York, New York, NY*, 1982. doi: [10.1007/978-1-4757-3951-0](https://doi.org/10.1007/978-1-4757-3951-0)

⁵⁴ The presented treatment would be practically identical for any Abelian group as the coefficient group.

5. For each $q \in \{0, 1, \dots\}$. We define groups:

- q -cycles as $Z_q(K; \mathbb{Z}) = \ker \partial_q \leq C_q(K; \mathbb{Z})$.
- q -boundaries as $B_q(K; \mathbb{Z}) = \text{Im } \partial_{q+1} \leq Z_q(K; \mathbb{Z}) \leq C_q(K; \mathbb{Z})$.
- q -homology group as the quotient group

$$H_q(K; \mathbb{Z}) = Z_q(K; \mathbb{Z}) / B_q(K; \mathbb{Z}).$$

Up to this point the introduction has been analogous to the one where coefficients form a field. However, as $H_q(K; \mathbb{Z})$ is an Abelian group, its rank does not completely determine it. In particular,

$$H_q(K; \mathbb{Z}) \cong \underbrace{\mathbb{Z}^r}_{\text{free part of } G} \oplus \underbrace{\mathbb{Z}_{p_1^{q_1}} \oplus \mathbb{Z}_{p_2^{q_2}} \oplus \dots \oplus \mathbb{Z}_{p_k^{q_k}}}_{\text{torsion of } G},$$

where the rank of the group $r = b_q(K; \mathbb{Z})$, referred to as the q -Betti number, only determines⁵⁵ the free part of the group.

Let $\text{rank } \partial_q$ be the rank of the image of ∂_q . By Proposition 7.4.1, numbers b_q can be deduced⁵⁶ from the ranks of $\partial_q, \partial_{q+1}$ and n_q . However, in order to compute torsion we need to delve deeper into the structure of the boundary maps.

For example, suppose the ranks of the two maps in the following diagram are 1:

$$\mathbb{Z} \xrightarrow{\varphi} \mathbb{Z} \xrightarrow{\psi} \mathbb{Z},$$

and assume $\text{Im } \varphi \subseteq \ker \psi$. Defining $H = \ker \psi / \text{Im } \varphi$, we know that $\text{rank } H = 0$. However, depending on maps φ, ψ group H could be any group of the form \mathbb{Z}_m . For example, if $\psi(n) = k \cdot s \cdot n$ for some $k, s \in \mathbb{N}$ and $\varphi(n) = k \cdot n$, then $H \cong \mathbb{Z}_s$.

In order to compute homology with coefficients in \mathbb{Z} we may reduce each boundary matrix to its **Smith normal form**. Given a matrix with entries in \mathbb{Z} , its Smith normal form is:

$$D = \begin{pmatrix} a_1 & 0 & \dots & \dots & \dots & 0 \\ 0 & a_2 & & & & \vdots \\ \vdots & & \ddots & & & \vdots \\ & & & a_r & 0 & \dots & 0 \\ & & & & 0 & \dots & 0 \\ & & & & & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 \end{pmatrix},$$

where each diagonal entry a_i divides⁵⁷ the next one. The diagonal entries a_i are called **elementary divisors** and r is the rank⁵⁸ of the matrix.

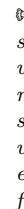
Proposition 7.4.1. Suppose G, H are Abelian groups, a map $f: G \rightarrow H$ is a homomorphism, and $G' \leq G$. Then:

1. $\ker(f) \leq G$.
2. $\text{Im}(f) \leq H$.
3. $\text{Im}(f) \cong G / \ker(f)$.
4. $\text{rank}(G/G') = \text{rank}(G) - \text{rank}(G')$.

⁵⁵ Two cases when the homology group has no torsion:

- For any simplicial complex K we have $H_0(K; \mathbb{Z}) \cong \mathbb{Z}^{b_0}$, where b_0 is the number of components of K .
- If K is a planar graph then $H_1(K; \mathbb{Z}) \cong \mathbb{Z}^{b_1}$, where b_1 is the number of holes K generates in the plane.

⁵⁶ I.e., $b_p = n_p - \text{rank } \partial_p - \text{rank } \partial_{p+1}$.

 It turns out that amongst all possible choices of coefficients, homology with coefficients in \mathbb{Z} contains the most information. Details of this statement are formalized in the universal coefficient theorem, which explains the connection between coefficients \mathbb{Z} and all other coefficients.

⁵⁷ I.e., $a_i | a_{i+1}, \forall i \in \{1, 2, \dots, r-1\}$.

⁵⁸ The rank of the matrix corresponding to a boundary map coincides with the rank of the boundary map.

Each matrix with entries in \mathbb{Z} has⁵⁹ a Smith normal form. Some of its properties are:

1. The form is obtained through a combination⁶⁰ of row reduction and the Euclidean algorithm for computing greatest common divisors.
2. The form is unique up to the signs of the elementary divisors.

Elementary divisors generate the torsion part of homology.

We now describe how to obtain homology groups using the Smith normal form.

- Choose $q \in \{0, 1, \dots\}$.
- Assume matrix D above is the Smith normal form of ∂_{q+1} with all diagonal entries positive.
- Compute the rank of ∂_q , possibly also through its Smith normal form.
- Then:

$$H_q(K; \mathbb{Z}) \cong \mathbb{Z}^{n_q - \text{rank } \partial_q - \text{rank } \partial_{q+1}} \oplus \bigoplus_{i=1}^r \mathbb{Z}_{a_i}.$$

Note that this form may potentially be simplified⁶¹ further.

We conclude by providing analogues of the examples of homology with field coefficients:

- The formula for disjoint union holds as before: $H_i(K \coprod L; \mathbb{Z}) \cong H_i(K; \mathbb{Z}) \times H_i(L; \mathbb{Z})$.
- The expression for the Euler characteristic with integer Betti numbers is the same: $\chi(K) = b_0 - b_1 + b_2 - b_3 + \dots$
- For each $n \in \{1, 2, \dots\}$ we have:
 - $H_0(S^n; \mathbb{Z}) \cong H_n(S^n; \mathbb{Z}) \cong \mathbb{Z}$;
 - $H_i(S^n; \mathbb{Z}) = 0, \forall i \notin \{0, n\}$.
- For each connected manifold K of positive dimension n we have:
 - $H_n(K; \mathbb{Z}) = 0$ if K has boundary.
 - $H_n(K; \mathbb{Z}) \cong \mathbb{Z}$ if K is closed orientable.
 - $H_n(K; \mathbb{Z}) \cong \mathbb{Z}_2$ if K is closed non-orientable.
- If \mathcal{K} is the Klein bottle, then $H_1(\mathcal{K}; \mathbb{Z}) \cong \mathbb{Z} \oplus \mathbb{Z}_2$.

For an extended treatment of these examples see a book⁶².

⁵⁹ Formally, every matrix A with entries in \mathbb{Z} can be factored as $A = UDV$, where D is its Smith normal form, and U and V are matrices with entries in \mathbb{Z} with determinant ± 1 . In particular, the last condition means that U and V are invertible, and that its inverses have entries in \mathbb{Z} .

⁶⁰ At this point the difference of the structure of a group as compared to that of a field becomes prominent. When coefficients were in a field, we could always divide a row by a non-zero entry. When working with coefficients in \mathbb{Z} that is not allowed (except for ± 1 , which doesn't really help). As a result, obtaining the desired form of a matrix requires us to involve greatest common divisors and even then not all non-trivial diagonal entries can be transformed to 1.

⁶¹ If s_1, s_2 are relatively prime, then $\mathbb{Z}_{s_1 \cdot s_2} \cong \mathbb{Z}_{s_1} \oplus \mathbb{Z}_{s_2}$. Also, if some $a_i = 1$, then \mathbb{Z}_{a_i} is the trivial group, i.e., it can be omitted from the expression.

⁶² Allen Hatcher. Algebraic topology. Cambridge Univ. Press, Cambridge, 2000

Appendix: cubical homology

The homology construction we described above is called simplicial homology as it is based on the structure of a simplicial complex: a space assembled using simplices. However, there are settings in which alternative shapes of basic building blocks appear to be more suitable. One such setting is image analysis, where we work with an image or a video consisting of pixels. In this setting it would be natural to consider pixels as the building blocks.

This leads to a new construction⁶³ of complexes and homology: cubical complexes and cubical homology. We will restrict ourselves to the setting of two-dimensional images, meaning the pixels are chosen from a fixed grid. The construction could easily be generalized to three-dimensional (movies of 2-D images or a 3-D image) four-dimensional (movies of 3-D images) or higher-dimensional images with different shapes of grids and cubes, or even without a fixed grid.

Let $n \in \mathbb{N}$ and consider a square grid of size $n \times n$, where n refers to the number of squares along each side, see Figure 7.28. Our image is given by a collection of pixels (grey squares). The first task is to define the building blocks:

- 0-dimensional cubes are the vertices appearing on the grid. There are $(n + 1)^2$ vertices.
- 1-dimensional cubes are the vertical and horizontal edges between vertices appearing on the grid. There are $2n(n + 1)$ edges.
- 2-dimensional cubes are the squares of the grid. There are n^2 squares.

A **cubical complex** K on an $n \times n$ grid is a collection of cubes such that if $\sigma \in K$ and $\tau \subseteq \sigma$, then $\tau \in K$.

Our next task is to determine a convenient systematic labelling for the squares, edges and vertices. In the context of simplicial complexes the labels were just the oriented collections of vertices. While the same approach could⁶⁴ be used here, there is a more elegant enumeration of the cubes.

Instead of thinking about coordinates in terms of the $n \times n$ grid, we systematically imagine all potential cubes of a complex drawn in a table-like pattern as Figure 7.29 demonstrates. Each cube can be assigned coordinates (x, y) where $x, y \in \{0, 1, 2, \dots, 2n\}$ according to this pattern. Drawing the corresponding coordinate axes superimposed over the original $n \times n$ grid (Figure 7.30) we see that a pair of coordinates (x, y) represents the cube⁶⁵, whose center is (x, y) . We additionally define the orientations:

- Each square is oriented with the ordering of its vertices in the positive-rotational order.

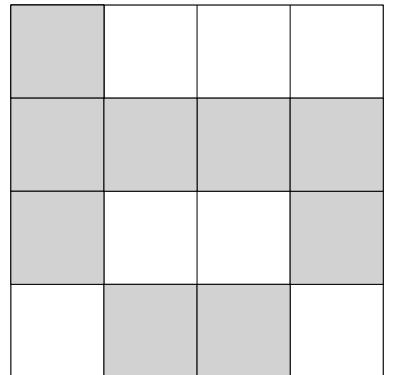


Figure 7.28: A 4×4 image consisting of grey pixels.

⁶³ Actually, we could build complexes and the corresponding theory from many different shapes of basic building blocks.

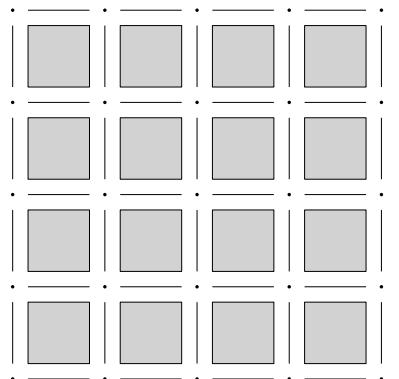


Figure 7.29: The collection of all potential cubical simplices.

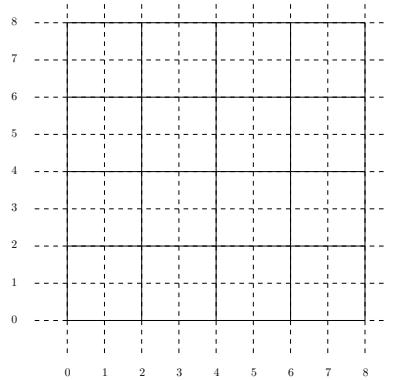


Figure 7.30: The assignment scheme.

⁶⁴ Although, the approach would be cumbersome. We would need 4 vertices to describe a square.

⁶⁵ A square, and edge, or a vertex.

- Vertical edges are oriented upwards, horizontal to the right.

The resulting assignment of coordinates/labels has the following properties (see Figure 7.31):

- If x, y are both odd, then (x, y) is a square.
- If exactly one of x, y is odd, then (x, y) is an edge.
- If x, y are both even, then (x, y) is a vertex.

We are now in a position to define cubical homology. The structure of the definition is the same as for simplicial homology with the only essential difference in the boundary map.

Let K be a cubical complex and choose⁶⁶ a field of coefficients \mathbb{F} . For each $q \in \{0, 1, 2\}$ let n_q denote the number of q -cubes in K .

1. A q -chain is a formal sum $\sum_{i=1}^{n_q} a_i \sigma_i^q$ where $a_i \in \mathbb{F}$ and σ_i^q is an oriented cube of dimension q in K .
2. The chain group $\mathfrak{C}_q(K; \mathbb{F}) \cong \mathbb{F}^{n_q}$ is the vector space of all q -chains. Its generators are oriented q -cubes of K .
3. For each $p \in \mathbb{N}$ the boundary map

$$\partial_p: \mathfrak{C}_p(K; \mathbb{F}) \rightarrow \mathfrak{C}_{p-1}(K; \mathbb{F})$$

is the linear map defined⁶⁷ by the following rules:

If x and y are both even (vertex): $\partial_p(x, y) = 0$.

If x is odd and y is even (horizontal edge):

$$\partial_p(x, y) = (x+1, y) - (x-1, y).$$

If x is even and y is odd (vertical edge):

$$\partial_p(x, y) = (x, y+1) - (x, y-1).$$

If x and y are both odd (square):

$$\partial_p(x, y) = (x+1, y) - (x, y+1) - (x-1, y) + (x, y-1).$$

As before, $\partial^2 = 0$.

4. The collection of chain groups bound together by the boundary homomorphisms is the chain complex:

$$\dots \xrightarrow{\partial} \mathfrak{C}_n(K; \mathbb{F}) \xrightarrow{\partial} \mathfrak{C}_{n-1}(K; \mathbb{F}) \xrightarrow{\partial} \dots \xrightarrow{\partial} \mathfrak{C}_1(K; \mathbb{F}) \xrightarrow{\partial} \mathfrak{C}_0(K; \mathbb{F}) \xrightarrow{\partial} 0$$

5. For each $q \in \{0, 1, \dots\}$. We define groups:

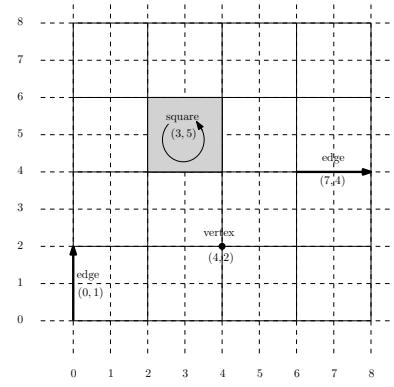


Figure 7.31: The assignment scheme.

⁶⁶ We could also choose the coefficients from an Abelian group, the construction would be analogous.

⁶⁷ The map encodes the geometric boundary.

⚠ The operations between coordinates in the boundary map are formal summations and subtractions in the chain group and should not be considered as operations on pairs. The coordinates (x, y) are only labels and shouldn't be added to or subtracted from each other. For example, $(0, 0) - (2, 0)$ is a formal chain consisting of two vertices with coefficients 1 and -1, while label $(-2, 0)$ is undefined.

- q -cycles as $\mathfrak{Z}_q(K; \mathbb{F}) = \ker \partial_q \leq \mathfrak{C}_q(K; \mathbb{F})$.
- q -boundaries as $\mathfrak{B}_q(K; \mathbb{F}) = \text{Im } \partial_{q+1} \leq \mathfrak{Z}_q(K; \mathbb{F}) \leq \mathfrak{C}_q(K; \mathbb{F})$.
- cubical q -homology group as the quotient

$$\mathfrak{H}_q(K; \mathbb{F}) = \mathfrak{Z}_q(K; \mathbb{Z}) / \mathfrak{B}_q(K; \mathbb{F}).$$

It turns out that the cubical homology of a cubical complex K is isomorphic to the homology of the simplicial complex obtained by subdividing the cubes into simplices. In particular, the homology detects components, holes, and (in the case of higher dimensional cubical complexes) higher-dimensional holes as simplicial homology would.

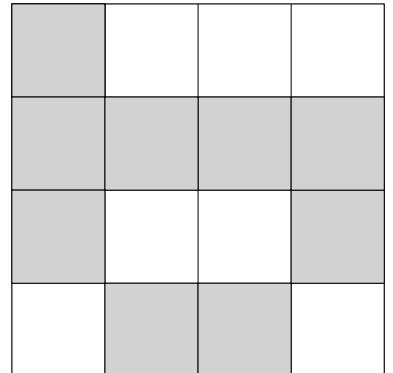


Figure 7.32: The cubical homology of the above image is given by $\mathfrak{H}_0 \cong \mathbb{F}$ (one component) and $\mathfrak{H}_1 \cong \mathbb{F}$ (one hole).

8

Homology: impact and computation by parts

HOMOLOGY AS DEFINED in the previous chapter is an invariant assigned to a simplicial complex. While its homotopy invariance and computational amenability make homology a suitable tool for computational purposes, the structural depth of the underlying theory goes far beyond the presented material.

In this chapter we present some further properties of homology. The first one is functoriality and its impact on significant topological results from the beginning of the twentieth century: Brouwer fixed point theorem, hairy ball theorem, and invariance of domains. The second property is the ability to combine homology computations of two parts of a space in order to deduce the homology of the whole space.

8.1 Impact

One of the fundamental tasks of mathematics is a construction of new objects (invariants) assigned to known objects. For example, given a closed surface we can assign to it a triangulation. In turn, we can assign homology groups to the obtained triangulation.

It turns out to be very beneficial if such an assignment can be extended in a consistent manner to maps between the objects as well. When this is the case, we say the assignment is functorial¹. It turns out that homology is functorial as Proposition 8.1.2 demonstrates.

¹ Functoriality and its formal consequences are studied within the category theory.

Functionality of homology

Definition 8.1.1. Suppose $f: K \rightarrow L$ is a simplicial map between simplicial complexes, $q \in \{0, 1, \dots\}$, and \mathbb{F} is a field. The **induced maps** f_* and f_* are defined as follows:

- $f_*: C_q(K; \mathbb{F}) \rightarrow C_q(L; \mathbb{F})$ is the linear map of chain groups defined as

$$f_* \left(\sum_i a_i \sigma_i \right) = \sum_{\{i \mid \dim(f(\sigma_i))=q\}} a_i f(\sigma_i), \quad a_i \in \mathbb{F}, \sigma_i^q \in K.$$

- $f_*: H_q(K; \mathbb{F}) \rightarrow H_q(L; \mathbb{F})$ is the linear map defined as

$$f_*([\alpha]) = [f_*(\alpha)].$$

Comments on Definition 8.1.1 using the notation established in it:

- Given a simplex $\sigma \in K$ of dimension q , its image $f(\sigma)$ is a simplex of dimension q or less. The condition on dimension in the definition of f_* means that only the images of those simplices σ_i , which are of full dimension q , are taken into account. In particular,

$$f_*(\sigma) = \begin{cases} f(\sigma); & \dim(f(\sigma)) = q \\ 0; & \text{else.} \end{cases}$$

- The induced map f_* turns out to be well defined, i.e., if $[\alpha] = [\beta]$ then $f_*([\alpha]) = f_*([\beta])$.
- Homotopic maps induce the same maps on homology.
- Suppose X and Y are metric spaces with triangulations K and L . By the simplicial approximation theorem there exists for each continuous map $f_1: X \rightarrow Y$ a simplicial map f_2 between some subdivisions of K and L , such that $f_1 \simeq f_2$. Whenever we mention the homology of X , we formally think of the homology of K . In a similar manner, whenever we talk about the maps on homology induced by a continuous map f_1 , we formally think² of maps induced by the simplicial map f_2 .

The induced maps are consistent with respect to compositions³ as the following proposition explains.

\triangle We refrain from specifying q and \mathbb{F} in the notation f_* in order not to overload it with the indices. As such f_* represents the induced map on homology in any dimension or with any coefficients. The relevant choice of the dimension(s) and coefficients should always be apparent from the context.

\bowtie The induced maps in the case of coefficients in a group are homomorphisms and are still well defined.

\bowtie Identity maps between spaces induce identity maps on homology. Constant maps between spaces induce trivial (i.e., zero) maps on homology.

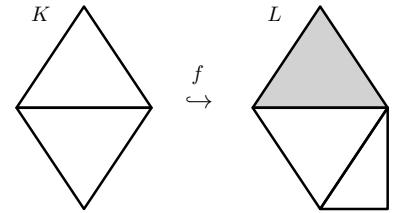


Figure 8.1: An embedding $f: K \rightarrow L$. While the first homology groups of K and L are of dimension 2, the image $f_*(H_1(K; \mathbb{F}))$ is of dimension 1, demonstrating that the embedding preserves only one hole. This interpretation will be significantly expanded within the context of persistent homology.

² With this explanation, the notion of a map on homology induced by a continuous map between spaces X and Y is well defined.

³ Formally speaking we express this property by saying that homology is **functorial**.

Proposition 8.1.2. [Functoriality of the induced maps] Suppose maps $f: K \rightarrow L$ and $g: L \rightarrow M$ between simplicial complexes are simplicial. Then for each $q \in \{0, 1, \dots\}$ and for each \mathbb{F} we have

$$(g \circ f)_{\#} = g_{\#} \circ f_{\#}, \quad \text{and} \quad (g \circ f)_* = g_* \circ f_*.$$

The proof follows straight from the definition.

One of the most natural demonstrations of the power of functoriality concerns the existence of retractions. Given a space X and its subspace⁴ $A \subset X$, a **retraction** of X to A is any continuous map $f: X \rightarrow A$ such that $f(a) = a, \forall a \in A$. For example, the radial map from a two-dimensional Euclidean ball with the center removed to its boundary sphere is a retraction, see Figure 8.7.

Example 8.1.3. For each $n \in \mathbb{N}$ the standard $(n - 1)$ -sphere S^{n-1} is the boundary of the standard n -ball B^n . We claim there is no retraction $B^n \rightarrow S^{n-1}$. As a special case, there is no retraction of the unit interval onto its endpoints.

Proof. Assume such a retraction $f: B^n \rightarrow S^{n-1} = \partial B^n$ exists. Precompose it with the inclusion $g: S^{n-1} \hookrightarrow B^n$, see Figure 8.3. Let $[\alpha] \neq 0$ be a basis (generator) of $H_{n-1}(S^{n-1}; \mathbb{F})$. We combine two observations:

- As $f \circ g: S^{n-1} \rightarrow S^{n-1}$ is identity, $(f \circ g)_*([\alpha]) = [\alpha] \neq 0$.
- As $H_{n-1}(B^n; \mathbb{F}) = 0$, $g_*([\alpha]) = 0$ and thus $f_*(g_*([\alpha])) = 0$.

By Proposition 8.1.2 $(f \circ g)_*([\alpha]) = f_*(g_*([\alpha]))$, a contradiction.

Hence a retraction f does not exist. \square

Brouwer fixed point

Brouwer fixed point theorem is probably one of the most famous early results of topology. It has a surprisingly short proof using the functoriality of homology.

Theorem 8.1.4. Every continuous map $f: B^2 \rightarrow B^2$ has a fixed point, i.e., a point $x_0 \in B^2$ such that $f(x_0) = x_0$.

Proof. Assume map f has no fixed point. Define map $g: B^2 \rightarrow S^1$ by declaring that for each $x \in B^2$, point $g(x) \in S^1 = \partial B^2$ is the intersection of S^1 with the ray starting at $f(x)$ containing x , see Figure 8.4. As f has no fixed point, such a ray always exists. Map g is a continuous retraction, a contradiction according to Example 8.1.3. Hence a fixed point exists. \square

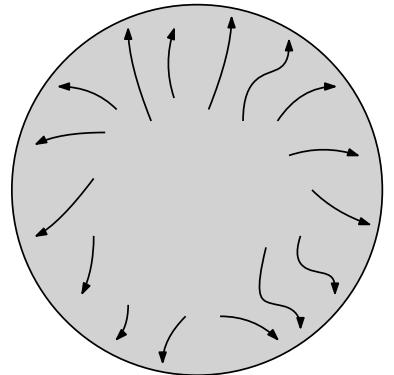


Figure 8.2: Geometric intuition dictates that if we want to retract B^2 onto $S^1 = \partial B^2$, the resulting map would need to have a discontinuity, i.e., at least one point where we “tear” the disc. A fairly simple proof of this fact is given using homology.

⁴ A required condition for the existence of a retraction is for A to be closed in X .

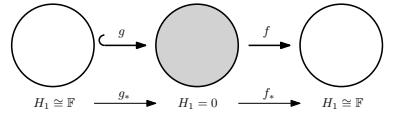


Figure 8.3: The proof of Example 8.1.3. The composition of maps is identity on S^1 , while the composition of induced maps can't be identity as it factors through 0.

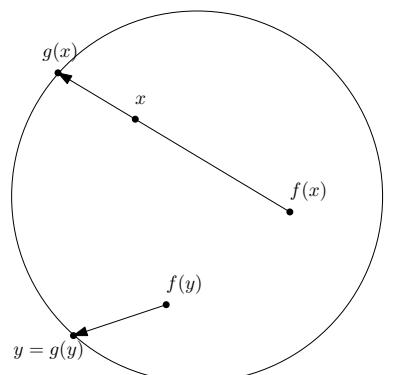


Figure 8.4: Map g from the proof of the Brouwer fixed point theorem.

Hairy ball

Another prominent theorem that can be conveniently proved using homology is the hairy ball theorem. The name comes from a popular adaptation of the result: one can't comb the hair on a hairy ball without creating a hair whorl.

Before we state the theorem we need to clarify a few technical details. Throughout this subsection let S^2 denote the unit two-dimensional sphere in \mathbb{R}^3 .

1. A **tangent vector field** on the sphere S^2 is a continuous map $f: S^2 \rightarrow \mathbb{R}^3$ such that for each $x \in S^2$ we have $x \perp f(x)$. A vector field f is non-vanishing, if it is non-zero at each point.
2. Given a centrally symmetric⁵ triangulation K of S^2 , let $\alpha = \sum_{\sigma^{(2)} \in K} \sigma$ be the cycle defined as the sum of consistently oriented triangles of K . Without loss of generality we may assume the triangles are oriented so that their “upwards” direction is pointing away from the point $(0, 0, 0)$. Recall that $[\alpha]$ is the fundamental class spanning $H_2(K; \mathbb{R}) \cong \mathbb{R}$.
3. For each triangle $\sigma \in K$ the reflection of σ through $(0, 0, 0)$ is again a simplex σ' of K . However, if σ has the chosen orientation from the previous point⁶, then the reflected triangle has the opposite orientation⁷ from the originally chosen orientation on σ' , see the left portion of Figure 8.6. In particular, $[\alpha'] = \sum_{\sigma^{(2)} \in K} \sigma'$ is a non-trivial homology class representing $-[\alpha]$.
4. Let $\rho: K \rightarrow K$ be the reflection map and let $g: K \rightarrow K$ be the identity map. By 2. and 3. maps g and ρ are not homotopic as $g_*([\alpha]) = [\alpha] \neq \rho_*([\alpha]) = [\alpha']$.

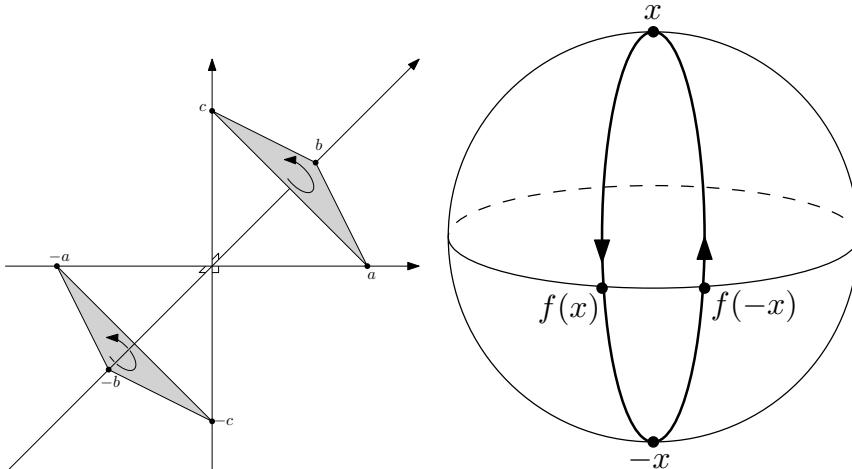


Figure 8.5: A tangent vector field on a sphere induces a flow presented by the streamlines on this figure. Theorem 8.1.5 states that the vector field must have a zero, which can be demonstrated on our example by the source of streamlines. To the contrary, there are non-trivial tangent vector fields in the plane and on the torus.
⁵I.e., the triangulation K has the following property: for each simplex $\tau \in K$ its reflection through the point $(0, 0, 0)$ is also a simplex.

⁶I.e., such that the chosen normal is pointing away from the point $(0, 0, 0)$.

⁷I.e., such that the chosen normal is pointing towards the point $(0, 0, 0)$.

Figure 8.6: Elements of the proof of Theorem 8.1.5.

On the left side is simplex $\langle a, b, c \rangle$ and its (oriented) reflection through $(0, 0, 0)$: $\langle -a, -b, -c \rangle$. Observe that in both cases the normal to the simplex is in the direction $(1, 1, 1)$. The same argument and picture work for any odd dimension, which leads to Theorem 8.1.6.

On the right side is the construction of homotopy from the proof of Theorem 8.1.5. Point x is connected to $-x$ by the geodesic passing through $f(x)$ and vice versa.

Theorem 8.1.5. *Every tangent vector field on S^2 has a zero, i.e., there is no non-vanishing tangent vector field on S^2 .*

Proof. Suppose f is a non-vanishing vector field on S^2 . Without loss of generality⁸ we can assume $\|f(x)\| = 1, \forall x \in S^2$. Using the notation leading to this theorem, we will prove that $g \simeq \rho$, which is a contradiction by 4. above. We will construct an explicit homotopy between g and ρ . Such a homotopy can be thought of as a continuous collection⁹ of paths from x to $-x$ for all¹⁰ $x \in S^2$.

The simplest way to connect two diametrically opposite points on a sphere, for the sake of simplicity let us assume we are connecting the north pole N to the south pole S , is by drawing a meridian between them. Such a meridian is completely determined by the point at which it intersects the equator. We define this intersection point to be¹¹ $f(N)$.

In general, connect x to $-x$ by a geodesic¹² on the sphere passing through $f(x)$. This is a continuous assignment of paths and constitutes the homotopy between g and ρ , which completes the proof. \square

The argument of Theorem 8.1.5 works for any even dimension which leads to a more general result.

Theorem 8.1.6. *The sphere S^n admits a non-vanishing tangent field iff n is even.*

When n is even there is an easy construction of a non-vanishing tangent field:

$$(x_1, y_1, x_2, y_2, \dots, x_m, y_m) \mapsto (y_1, -x_1, y_2, -x_2, \dots, y_m, -x_m).$$

Invariance of domain

The last classical result we mention explains why Euclidean balls of different dimensions are fundamentally different in the sense that they can't be homeomorphic¹³.

Theorem 8.1.7. *For any pair of natural numbers $m \neq n$ the closed balls $B_1 = B_{\mathbb{R}^m}(0, 1)$ and $B_2 = B_{\mathbb{R}^n}(0, 1)$ are not homeomorphic.*

Proof. Assume there exists a homeomorphism $f: B_1 \rightarrow B_2$. Then $f|_{B_1 \setminus \{0\}}: B_1 \setminus \{0\} \rightarrow B_2 \setminus \{f(0)\}$ is also a homeomorphism. Recall that $B_1 \setminus \{0\} \simeq S^{n-1}$ via the radial projection (see Figure 8.7), which means $H_{n-1}(B_1 \setminus \{0\}; \mathbb{F})$ is non-trivial for any \mathbb{F} . On the other hand, $B_2 \setminus \{f(0)\}$ is either:

⁸ I.e., by normalizing each vector in the image of f .

⁹ A homotopy in question is of the form $H: S^2 \times [0, 1] \rightarrow S^2$. For each $y \in S^2$ the restriction $H|_{\{y\} \times [0, 1]}$ is thus a path from y to $-y$. The fact that H is continuous means that the collection of such paths is continuous.

¹⁰ While the homology setup above is performed in the simplicial setting of K , the homology here will be constructed on a “smooth” sphere S^2 .

¹¹ Recall that $\|f(N)\| = 1$ and $f(N) \perp N$, since $f(N)$ lies on the equator.

¹² A **geodesic** on S^2 is the shortest path between two points on S^2 . Geodesics between N and S are meridians.

¹³ This geodesic traces the trail of x as translated by the resulting homotopy. On the other hand, the trail of $-x$ as translated by the resulting homotopy is given by the geodesic from $-x$ to x passing through $f(-x)$. See the right portion of Figure 8.6 for a sketch.

¹⁴ While homology itself is a homotopy invariant, the trick we will use will allow us to use it to differentiate homeomorphic types of spaces.

¹⁵ As a consequence of Theorem 8.1.7, $D^n \not\cong D^m$ if $m \neq n$. The same argument gives $\mathbb{R}^n \not\cong \mathbb{R}^m$ if $m \neq n$.

- homotopy equivalent to S^{m-1} if $f(0) \notin \partial B_2$, or
- contractible if $f(0) \in \partial B_2$.

In both cases $H_{n-1}(B_2 \setminus \{f(0)\}; \mathbb{F}) = 0$, a contradiction. \square

8.2 Homology by parts

Given a decomposition of a simplicial complex $K = A \cup B$ as the union of subcomplexes A and B , can¹⁴ we compute the homology of X from the homology of A and B ?

The answer to this question is unfortunately negative, for example:

- As the sidenote on the right on a similar question demonstrates, the cumulative zero-dimensional homology of K and L may be too large and should possibly be decreased by the zero-dimensional union of the intersection.
- On the other hand, a circle is the union of two 1-discs, i.e., a space with a one-dimensional hole is the union of two subspaces without holes.

These two examples show that A and B can have cumulatively “too much” or “too little” homology to deduce the homology of the union X and that one should probably take into account the homology of the intersections as well. The algebraic structure through which the connection between the homologies of X , A , and B is expressed is that of exact sequences.

Exact sequences

Definition 8.2.1. A sequence of vector spaces V_0, V_1, \dots and linear maps $\varphi_n: V_n \rightarrow V_{n-1}$ is **exact**, if for each n we have $\text{Im } \varphi_{n+1} = \ker \varphi_n$.

It turns out that in an exact sequence, the dimension of each vector space (except for the last one) can be deduced from the ranks of the neighboring maps.

Proposition 8.2.2. Suppose the following sequence is exact:

$$\cdots \rightarrow V_{n+1} \xrightarrow{\varphi_{n+1}} V_n \xrightarrow{\varphi_n} V_{n-1} \rightarrow \cdots \rightarrow V_2 \xrightarrow{\varphi_2} V_1 \xrightarrow{\varphi_1} V_0.$$

Then for each $n > 0$, $\dim V_n = \text{rank } \varphi_{n+1} + \text{rank } \varphi_n$.

Proof. We know that $\dim V_n = \dim \ker \varphi_n + \text{rank } \varphi_n$. Now use exactness: $\text{Im } \varphi_{n+1} = \ker \varphi_n$. \square

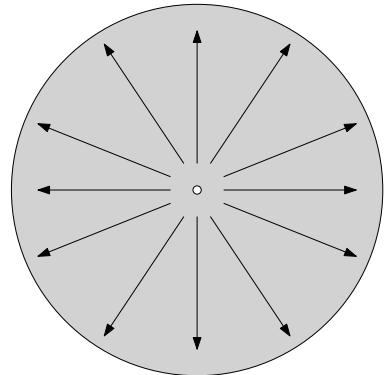


Figure 8.7: Radial projection of a disc with the center removed to the boundary of the disc. The induced homotopy equivalence demonstrates $B_1 \setminus \{0\} \simeq S^{n-1}$ in the proof of Theorem 8.1.7.

¹⁴ A similar question: Given a finite set $Y = C \cup D$, can we determine the cardinality $|Y|$ from $|C|$ and $|D|$? The answer $|Y| = |C| + |D| - |C \cap D|$, which also includes the intersection, is not unlike the answer to our question about homology..especially since, for discrete sets, the cardinality represents the zero-dimensional homology.

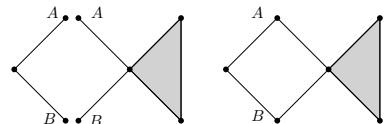


Figure 8.8: Two contractible complexes, whose union is not contractible.

Recall that homology is defined from a sequence of chain groups called the chain complex; it is defined as the quotient $\ker \partial / \text{Im } \partial$. In particular, the homology of a chain complex is zero at all dimensions iff the chain complex forms an exact sequence. Or, to put it locally, $H_q = 0$ iff the chain complex is “exact at C_q ” in the sense that $\text{Im } \partial_{q+1} = \ker \partial_q$. Homology thus measures the extent to which a chain complex is not exact.

Mayer-Vietoris exact sequence

We are now able to express¹⁵ the connection between the homology of X and the homology of its two parts A and B , a connection that also includes the homology of the intersection $A \cap B$.

¹⁵ Standard proofs use zig-zag lemma given in an appendix.

Theorem 8.2.3. Suppose $A, B \leq X$ are subcomplexes of a simplicial complex K such that $A \cup B = X$. Then for each choice of coefficients the following sequence of homology groups is exact:

$$\cdots \rightarrow H_{n+1}(X) \xrightarrow{\delta_{n+1}} H_n(A \cap B) \xrightarrow{(i_n, j_n)} H_n(A) \oplus H_n(B) \xrightarrow{\mu_n} H_n(X) \rightarrow \cdots$$

$$\cdots \rightarrow H_0(A \cap B) \xrightarrow{(i_0, j_0)} H_0(A) \oplus H_0(B) \xrightarrow{\mu_0} H_0(X) \rightarrow 0,$$

with the involved maps defined as follows:

- i_*, j_* are inclusion induced maps, i.e., $i_*[\alpha] = [\alpha]$ and $j_*[\alpha] = [\alpha]$.
- μ is the subtraction map, i.e., $\mu_*([\alpha], [\beta]) = [\alpha - \beta]$.
- δ is a variant of a boundary map defined as follows. Given an n -cycle α in X , decompose it as $\alpha = \alpha_A + \alpha_B$ where α_A is an n -chain in A and α_B is an n -chain in B . Define $\delta[\alpha] = [\partial\alpha_A]$ as the homology class corresponding to the boundary of the chain α_A .

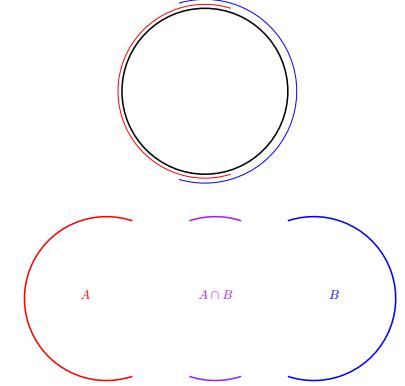


Figure 8.9: A decomposition of S^1 into two 1-discs.

Example 8.2.4. We will compute the homology of S^1 with coefficients in a field \mathbb{F} . Express S^1 as the union of two 1-discs A and B as Figure 8.9 suggests. The only non-trivial part of the corresponding Mayer-Vietoris sequence is the following:

$$H_1(A) \oplus H_1(B) \rightarrow H_1(X) \rightarrow H_0(A \cap B) \rightarrow H_0(A) \oplus H_0(B) \rightarrow H_0(X) \rightarrow 0,$$

which is of the form

$$0 \rightarrow H_1(X) \xrightarrow{\delta_1} \mathbb{F}^2 \xrightarrow{(i_0, j_0)} \mathbb{F}^2 \xrightarrow{\mu_0} H_0(X) \xrightarrow{\delta_0} 0,$$

since $A \cap B$ has two components. We proceed by the following sequence of deductions:

1. $\text{rank } \delta_0 = 0$ as it is the trivial map.
2. Map μ_0 is of rank¹⁶ 1.
3. By Proposition 8.2.2 we get $\dim H_0(S^1) = 1$.
4. By exactness and observation 2. we have $\dim \text{Im}(i_0, j_0) = \dim \ker \mu_0 = 1$, hence $\text{rank}(i_0, j_0) = 1$.
5. By exactness and the previous item we have $\dim \text{Im } \delta_1 = \dim \ker(i_0, j_0) = 1$, hence $\text{rank } \delta_1 = 1$.

¹⁶ Recall that $\mu(u, v) = u - v$. Its rank is either 0, 1, or 2. It can't be 0 as the map is nontrivial. It can't be 2, as it has a non-trivial kernel generated by (u, u) since A and B are in the same component of X .

6. By Proposition 8.2.2 we get $\dim H_1(S^1) = 1$.
7. All higher homotopy groups (for $n > 1$) are trivial as they appear as $\cdots 0 \rightarrow H_n(X) \rightarrow 0 \cdots$ in the exact sequence which, by Proposition 8.2.2, means they are trivial.

Remark 8.2.5. In the same manner we could compute the homology groups of S^m for each m by observing that it can be decomposed as the union of two hemispheres (m -discs) whose intersection is homotopy equivalent to S^{m-1} , see Figure 8.10.

Example 8.2.6. In a similar manner we can compute the homology of the torus X with coefficients in any field \mathbb{F} . We will only mention how to compute its first homology as the homology groups of other dimensions are already known.

We will use the decomposition of Figure 8.11. The relevant part of the Mayer-Vietoris sequence is

$$H_1(A) \oplus H_1(B) \rightarrow H_1(X) \rightarrow H_0(A \cap B) \rightarrow H_0(A) \oplus H_0(B) \rightarrow H_0(X) \rightarrow 0$$

which is of the form

$$\mathbb{F}^2 \xrightarrow{\mu_1} H_1(X) \xrightarrow{\delta_1} \mathbb{F}^2 \xrightarrow{(i_0, j_0)} \mathbb{F}^2 \rightarrow \mathbb{F} \rightarrow 0.$$

We proceed by the following sequence of deductions:

1. By the same argument as in Example 8.2.4 we have $\text{rank } \delta_1 = 1$.
2. The generators of $H_1(A)$ and $H_1(B)$ are cycles/loops α and β respectively. Note that $\alpha \simeq \beta$ in X thus $[\alpha] = [\beta] \in H_1(X)$. Furthermore, as¹⁷ $0 \neq [\alpha] \in H_1(X)$, we have¹⁸ $\text{rank } \mu_1 = 1$.
3. By Proposition 8.2.2 we get $\dim H_1(X) = 2$.

8.3 Concluding remarks

Recap (highlights) of this chapter

- Induced maps on homology and functoriality;
- Brouwer fixed point theorem and hairy ball theorem;
- Exact sequences;
- Mayer-Vietoris exact sequence.

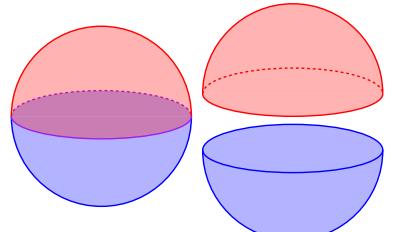


Figure 8.10: A decomposition of S^2 into two discs, whose intersection is S^1 .

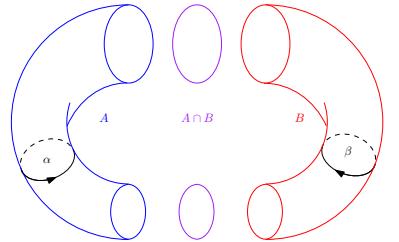


Figure 8.11: A decomposition of the torus into two parts, whose intersection is the disjoint union of two copies of S^1 .

¹⁷ An algebraic way to see that $0 \neq [\alpha] \in H_1(X)$ is through the Mayer-Vietoris sequence:

Proof. If $[\alpha]$ was trivial in $H_1(X)$ then $([\alpha], 0)$ would be in $\ker \mu_1$. By exactness, this would mean that $([\alpha], 0) \in \text{Im}(i_1, j_1)$. However, $\text{Im}(i_1, j_1)$ is generated by the images of the two obvious cycles in $A \cap B$, each of which maps into $(\pm[\alpha], \mp[\beta])$. Space $\text{Im}(i_1, j_1)$ is thus one-dimensional and generated by $(\pm[\alpha], \mp[\beta])$, hence $([\alpha], 0) \notin \text{Im}(i_1, j_1)$ as $[\alpha] \neq 0$ in $H_1(A)$ and $[\beta] \neq 0$ in $H_1(B)$. \square

¹⁸ The map μ_1 is defined as $\mu_1(u, v) = u - v$ in the basis $([\alpha], 0), (0, [\beta])$ of $H_1(A) \oplus H_1(B)$. Its rank is either 0, 1, or 2. Its can't be 0 as the map is nontrivial, since $0 \neq [\alpha] \in H_1(X)$. It can't be 2, as it has a non-trivial kernel generated by (u, u) since $[\alpha] = [\beta] \in H_1(X)$.

Background and applications

Invariants of homological nature appear throughout topology, geometry and other fields of mathematics. The examples of theoretical applications presented here barely scratch the surface. Some of the settings in which such constructions contributed to significant development include knot theory (Khovanov homology), differential geometry (De Rham cohomology, Floer homology), etc. For further background on theoretical foundations see Hatcher's book¹⁹.

The Mayer-Vietoris sequence arises from a decomposition of a space into two pieces. A natural question about a similar result in the context of decompositions into more pieces is treated within the context of spectral sequences, an algebraic formalism far above the reach of our presentation. These theoretical developments allow for a certain level of distributed computation of homology.

Appendix: zig-zag lemma

Lemma 8.3.1. [Zig-zag lemma] Let \mathbb{F} be a field of coefficients. Assume the following diagram of vector spaces over \mathbb{F} and linear maps²⁰ is commutative²¹:

$$\begin{array}{ccccccc}
 & \vdots & \vdots & & \vdots & & \\
 & \downarrow \partial & \downarrow \partial & & \downarrow \partial & & \\
 0 & \longrightarrow & A_{q+1} & \xrightarrow{\alpha} & B_{q+1} & \xrightarrow{\beta} & C_{q+1} \longrightarrow 0 \\
 & \downarrow \partial & \downarrow \partial & & \downarrow \partial & & \\
 0 & \longrightarrow & A_q & \xrightarrow{\alpha} & B_q & \xrightarrow{\beta} & C_q \longrightarrow 0 \\
 & \downarrow \partial & \downarrow \partial & & \downarrow \partial & & \\
 0 & \longrightarrow & A_{q-1} & \xrightarrow{\alpha} & B_{q-1} & \xrightarrow{\beta} & C_{q-1} \longrightarrow 0 \\
 & \downarrow \partial & \downarrow \partial & & \downarrow \partial & & \\
 & \vdots & \vdots & & \vdots & &
 \end{array}$$

If each row is a short exact sequence, and each columns is a chain complex²², then there exists a long exact sequence of homology groups²³

$$\begin{array}{ccccc}
 \cdots & \xrightarrow{\alpha_*} & H_{q+1}(B) & \xrightarrow{\beta^*} & H_{q+1}(C) \\
 & & \searrow \delta & & \\
 & & H_q(A) & \xleftarrow{\alpha_*} & H_q(B) \xrightarrow{\beta^*} H_q(C) \\
 & & & \searrow \delta & \\
 & & & & H_{q-1}(A) \xleftarrow{\alpha_*} H_{q-1}(B) \xrightarrow{\beta^*} \cdots
 \end{array}$$

¹⁹ Allen Hatcher. Algebraic topology. Cambridge Univ. Press, Cambridge, 2000

²⁰ For the sake of simplicity the indices of maps will be omitted. For example, maps $\alpha_q: A_q \rightarrow B_q$ are all denoted by α even though they depend on q . For the same reason we will refrain from mentioning \mathbb{F} again.

²¹ I.e., $\partial \circ \alpha = \alpha \circ \partial$ and $\partial \circ \beta = \beta \circ \partial$.

²² An exact sequence of the form

$$0 \rightarrow A \rightarrow B \rightarrow C \rightarrow 0$$

is called a **short exact sequence**. In such a situation, map $A \rightarrow B$ is injective as its kernel is the trivial image of the map $0 \rightarrow A$. On a similar note, $B \rightarrow C$ is surjective as its image is the kernel of the map $C \rightarrow 0$, which is C . As $\text{Im}(B \rightarrow C) \cong B / \ker(B \rightarrow C)$ we conclude $C \cong B / A$ since equality $\ker(B \rightarrow C) = \text{Im}(A \rightarrow B)$ holds by exactness.

²³ I.e., $\partial^2 = 0$.

²³ I.e., the homology groups arising from the vertical chain complexes. In particular, $H_q(A)$ is the quotient

$$\ker(A_q \rightarrow A_{q-1}) / \text{Im}(A_{q+1} \rightarrow A_q).$$

The idea of a proof. The proof is performed using the “diagram chasing” technique. We will only prove the existence of the δ map.

In order to define δ let us choose a non-trivial cycle $c \in C_{q+1}$. Charted by the diagrams on the right, the chase after $\delta([c])$ begins:

Diagram 1: $\partial(c) = 0$ as c is a cycle.

Diagram 2: By the exactness of the row map β is surjective, thus there exists $b_1 \in \beta^{-1}(c)$. Define $b_2 = \partial(b_1)$. By the commutativity $\beta(b_2) = 0$.

Diagram 3: By the exactness of the row map there exists $a_1 \in \alpha^{-1}(b_2)$. Define $\delta([c]) = [a_1]$.

The rest of the proof goes along the same lines. For example, in order to prove a_1 is a cycle we use diagram 4:

- Define $a_2 = \partial(a_1)$ observe $\partial(b_2) = 0$ as $\partial^2 = 0$.
- By the commutativity $\alpha(a_2) = 0$.
- By the exactness of the row $a_2 = 0$, hence a_1 is a cycle. \square

Remark 8.3.2. *The construction and proof of the Mayer-Vietoris sequence follows from the zig-zag lemma using the following commutative diagram (using notation of Theorem 8.2.3)*

$$\begin{array}{ccccccc}
 & \vdots & & \vdots & & \vdots & \\
 & \downarrow \partial & & \downarrow \partial & & \downarrow \partial & \\
 0 & \longrightarrow & C_{q+1}(A \cap B) & \xrightarrow{\alpha} & C_{q+1}(A) \oplus H_{q+1}(B) & \xrightarrow{\beta} & C_{q+1}(X) \longrightarrow 0 \\
 & \downarrow \partial & & \downarrow \partial & & \downarrow \partial & \\
 0 & \longrightarrow & C_q(A \cap B) & \xrightarrow{\alpha} & C_q(A) \oplus H_q(B) & \xrightarrow{\beta} & C_q(X) \longrightarrow 0 \\
 & \downarrow \partial & & \downarrow \partial & & \downarrow \partial & \\
 0 & \longrightarrow & C_{q-1}(A \cap B) & \xrightarrow{\alpha} & C_{q-1}(A) \oplus H_{q-1}(B) & \xrightarrow{\beta} & C_{q-1}(X) \longrightarrow 0 \\
 & \downarrow \partial & & \downarrow \partial & & \downarrow \partial & \\
 & \vdots & & \vdots & & \vdots &
 \end{array}$$

with maps α being induced by inclusion, and maps β being the subtraction maps²⁴. Observe that the horizontal maps are short exact sequences.

Zig-zag lemma provides a useful template for constructions of exact sequences. Another setting in which it applies is that of relative homology.

☞ *Diagram 1:*

$$\begin{array}{ccc}
 c & \xrightarrow{\quad} & 0 \\
 \downarrow \partial & & \\
 0 & &
 \end{array}$$

☞ *Diagram 2:*

$$\begin{array}{ccccc}
 b_1 & \xrightarrow{\quad} & c & \xrightarrow{\quad} & 0 \\
 \downarrow \partial & & \downarrow & & \\
 b_2 & \xrightarrow{\beta} & 0 & &
 \end{array}$$

☞ *Diagram 3:*

$$\begin{array}{ccccc}
 b_1 & \xrightarrow{\quad} & c & \xrightarrow{\quad} & 0 \\
 \downarrow & & \downarrow & & \\
 a_1 & \xrightarrow{\alpha} & b_2 & \xrightarrow{\quad} & 0
 \end{array}$$

☞ *Diagram 4:*

$$\begin{array}{ccccc}
 b_1 & \xrightarrow{\quad} & c & \xrightarrow{\quad} & 0 \\
 \downarrow & & \downarrow & & \\
 a_1 & \xrightarrow{\quad} & b_2 & \xrightarrow{\quad} & 0 \\
 \downarrow & & \downarrow & & \\
 a_2 & \xrightarrow{\quad} & 0 & &
 \end{array}$$

²⁴I.e., $\beta([\gamma_1], [\gamma_2]) = [\gamma_1 - \gamma_2]$.

Appendix: Relative homology

Let us fix a field \mathbb{F} , a simplicial complex K , and $L \leq K$. Homology construction on K is based on cycles: chains whose boundaries are trivial. The concept of relative homology expands this construction in the following way: given $L \leq K$, the relative homology construction is based on relative cycles, i.e., chains, whose boundaries are contained in L .

Algebraic specifics of the definition. From the chain complexes of K and L we can construct the quotient chain complex:

$$\cdots \xrightarrow{\partial} C_q(K)/C_q(L) \xrightarrow{\partial} C_{q-1}(K)/C_{q-1}(L) \xrightarrow{\partial} \cdots \xrightarrow{\partial} C_0(K)/C_0(L) \xrightarrow{\partial} 0$$

The **relative homology groups** $H_q(K, L)$ are the homology groups arising from this chain complex. In particular:

$$H_q(K, L; \mathbb{F}) = \frac{\ker(C_{q+1}(K)/C_{q+1}(L) \xrightarrow{\partial} C_q(K)/C_q(L))}{\text{Im}(C_q(K)/C_q(L) \xrightarrow{\partial} C_{q-1}(K)/C_{q-1}(L))}.$$

☞ Observe that $H_q(K, \emptyset) = H_q(K)$.

Combining Lemma 8.3.1 and the commutative diagram

$$\begin{array}{ccccccc} & \vdots & \vdots & & \vdots & & \\ & \downarrow \partial & \downarrow \partial & & \downarrow \partial & & \\ 0 \longrightarrow C_{q+1}(L) & \hookrightarrow & C_{q+1}(K) & \longrightarrow & C_{q+1}(K)/C_{q+1}(L) & \longrightarrow & 0 \\ & \downarrow \partial & \downarrow \partial & & \downarrow \partial & & \\ 0 \longrightarrow C_q(L) & \hookrightarrow & C_q(K) & \longrightarrow & C_q(K)/C_q(L) & \longrightarrow & 0 \\ & \downarrow \partial & \downarrow \partial & & \downarrow \partial & & \\ 0 \longrightarrow C_{q-1}(L) & \hookrightarrow & C_{q-1}(K) & \longrightarrow & C_{q-1}(K)/C_{q-1}(L) & \longrightarrow & 0 \\ & \downarrow \partial & \downarrow \partial & & \downarrow \partial & & \\ & \vdots & \vdots & & \vdots & & \end{array}$$

we conclude that relative homology groups fit into the following exact sequence:

$$\cdots \rightarrow H_{q+1}(K, L) \rightarrow H_q(L) \rightarrow H_q(K) \rightarrow H_q(K, L) \rightarrow \cdots$$

$$\cdots \rightarrow H_0(L) \rightarrow H_0(K) \rightarrow H_0(K, L) \rightarrow 0,$$

Relative homology has a geometric meaning, which expands that of the usual homology. Table 8.1 summarizes the relative homology of the pair (K, L) of simplicial complexes from Figure 8.12.

Let us geometrically interpret Table 8.1:

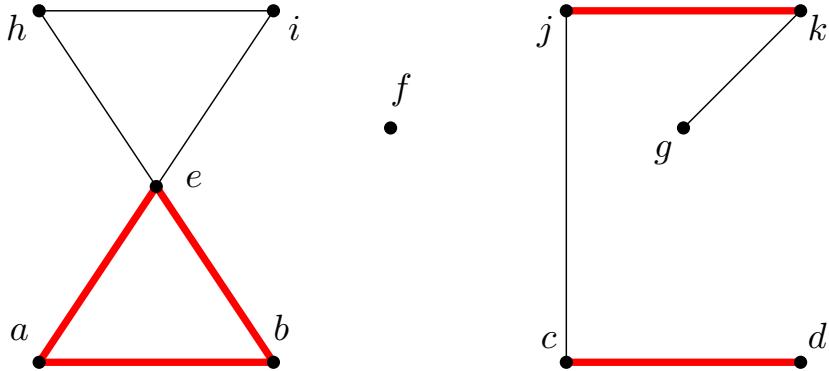


Figure 8.12: Simplicial complex K . Its subcomplex $L \leq K$ contains vertices a, c, b, d, e, j, k and all edges between these vertices. It is depicted by bold red edges.

q	$\dim H_q(K)$	$\dim H_q(K, L)$
0	3	1
1	2	2

Table 8.1: The comparison of the homology of K and the relative homology of the pair (K, L) from Figure 8.12.

Dimension 0: K has three components. However, the relative homology detects only component $[f]$. Homology class $[e]$ is contained in L and is thus trivial by the definition. Homology class $[h]$ is homologous to $[e]$ and thus trivial as well.

Dimension 1: A convenient basis for $H_1(K)$ would consist of $[\langle a, b \rangle + \langle b, e \rangle + \langle e, a \rangle]$ and $[\langle e, i \rangle + \langle i, h \rangle + \langle h, e \rangle]$. A basis for $H_1(K, L)$ however would consist of $[\langle e, i \rangle + \langle i, h \rangle + \langle h, e \rangle]$ and $[\langle j, c \rangle]$. Note that:

- $[\langle a, b \rangle + \langle b, e \rangle + \langle e, a \rangle]$ is a trivial in $H_1(K, L)$ as it is contained in L .
- $\langle j, c \rangle$ is a cycle in the relative homology chain complex as its boundary is contained in L and thus trivial.

Geometrically we can think of the relative homology $H_*(K, L)$ as the homology of the space obtained from K when the subcomplex L is contracted to a point, see Figure 8.13 for an example. The only exception to this rule is $H_0(K, L)$, whose dimension is one less²⁵ than the number of the components of the resulting space²⁶.

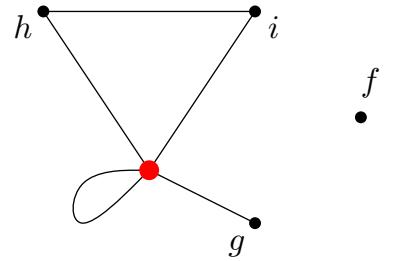


Figure 8.13: The space obtained from simplicial complex K from Figure 8.12 by contracting the subcomplex L to a point. The space has two holes but is not a simplicial complex in general.

²⁵ In the literature this exception is usually encoded in the phrase “reduced homology”.

²⁶ Note that the resulting space does not inherit the structure of a simplicial complex from K . However, it can be triangulated.

9

Persistent homology: definition and computation

THE CONCEPT OF PERSISTENT HOMOLOGY along with its variations is at the forefront of topological data analysis. Mathematically speaking, persistent homology is an obvious extension of homology: the functoriality of homology is applied to a sequence of inclusions. The resulting structure is, somewhat surprisingly, not harder to compute than ordinary homology. When coupled with the standard constructions of complexes, persistent homology contains not only topological but also geometric information.

We will start this chapter by explaining geometric intuition on persistent homology. We will continue by presenting formal and convenient visualisation techniques. We will conclude with a fairly simple algorithm for computation one could call a “labelled matrix reduction”.

9.1 Definition

We first describe the geometric intuition of persistent homology. Given a “growing” simplicial complex, persistent homology describes the evolution of its holes. As an illustrative example we consider four simplicial complexes $K_1 \leq K_2 \leq K_3 \leq K_4$ of Figures 9.1 and 9.2.

Here is how we interpret the corresponding zero-dimensional barcode¹ described by Figure 9.1:

K_1 : There are two components of K_1 . This fact is visualised by the fact that there are two bars (blue and red) starting at that time. The corresponding homology generators² (points) are colored accordingly.

K_2 : There are three components of K_1 : this fact is visualised by the fact that there are three bars (blue, purple and green) passing from that time on. The corresponding homology generators of the new components are colored accordingly. However, the two components of K_1 merge, which we interpret as one of the components of K_1

¹ I.e., the evolution of the components.

² Note that the generator of the red component is unique. On the other hand, we could have chosen any vertex of the other component as a generator and color it blue.

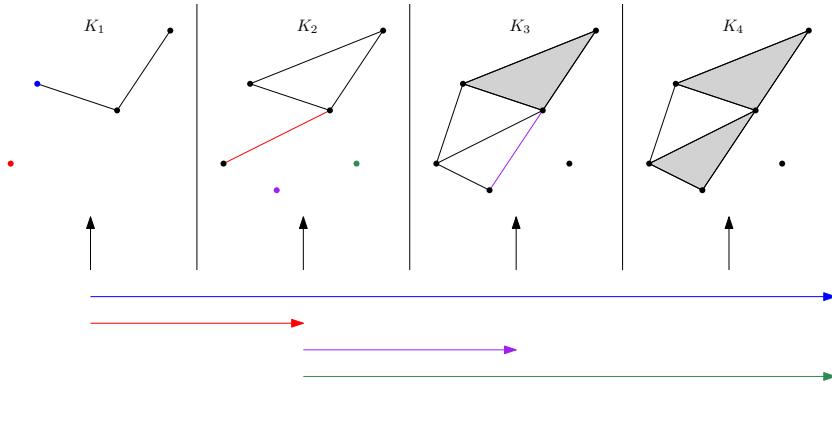
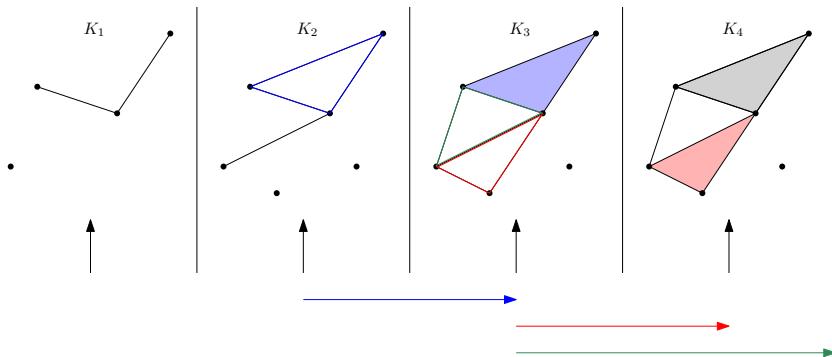


Figure 9.1: Nested simplicial complexes $K_1 \leq K_2 \leq K_3 \leq K_4$ are divided by vertical lines. The horizontal arrows below are called “bars” and form a **barcode**. They indicate the persistence of zero-dimensional homology classes: components. The left endpoint of each bar corresponds to the birth complex of a component. The right endpoint of each bar corresponds to the terminal complex of a component. The color of each bar also appears on one vertex (the representative of the component) and potentially on one edge (the edge, that terminates the component).

disappearing. We declare³ that the component disappearing is the red component, which is visualised by the fact that the red bar terminates just before K_2 . The edge making the connection between the two components is colored in red.

K_3 : The purple component terminates by connecting to the blue component via two edges, one of which is indicated by the purple color.

K_4 : There is no change in components as compared to K_3 , both bars are passing through to infinity.



³ As the components appeared at the same time, we might as well have chosen to have the blue bar terminated and keep the red bar going. The uncolored barcode would have remained the same. However, whenever there is a merger of components with different birth times we act according to the **elder rule**: the older component survives. This will be apparent at K_3 . The reader may rest assured this is not a product of discrimination but rather a rule that is consistent with the mathematical structure of persistence (especially the interleaving and stability) that will be described later.

Figure 9.2: Nested simplicial complexes $K_1 \leq K_2 \leq K_3 \leq K_4$ and the corresponding one-dimensional homology barcode.

In a similar fashion we interpret the corresponding one-dimensional barcode⁴ described by Figure 9.2:

K_1 : There are no holes and hence no bars passing on.

K_2 : A blue hole appears inducing a blue bar.

K_3 : The blue hole becomes trivial by the blue triangle and hence the blue bar terminates. However, two new holes appear, the red one

⁴ I.e., the indicated evolution of the holes.

and the green one. Consequently, there are two bars passing from K_3 on.

K_4 : The red hole becomes trivial while the green hole lives on forever, just as the corresponding bar.

The goal of this chapter is to present the theoretical background formalizing the presented geometric idea of persistent homology, and to introduce the computational procedure to obtain the barcodes.

Formal definition

We first formally introduce a filtration: a nested sequence of ever larger simplicial complexes modelling a growing simplicial complex.

Definition 9.1.1. Let K be a simplicial complex. A (discrete) **filtration** of K is a sequence of subcomplexes

$$K_1 \leq K_2 \leq \dots \leq K_m = K.$$

An example of a filtration is given in Figures 9.1 and 9.2.

Persistent homology measures how homology elements⁵ persist⁶ through steps of a filtration. A filtration of a simplicial complex K in Definition 9.1.1 can be expressed as a sequence of natural inclusion maps denoted by map⁷ $i_{-, -}$:

$$K_1 \xhookrightarrow{i_{1,2}} K_2 \xhookrightarrow{i_{2,3}} \dots \xhookrightarrow{i_{m-1,m}} K_m = K.$$

Given a field \mathbb{F} and $q \in \{0, 1, 2, \dots\}$ we can apply homology $H_q(-; \mathbb{F})$ to obtain a sequence⁸ of homology groups connected by linear maps:

$$H_q(K_1; \mathbb{F}) \xrightarrow{(i_{1,2})_*} H_q(K_2; \mathbb{F}) \xrightarrow{(i_{2,3})_*} \dots \xrightarrow{(i_{m-1,m})_*} H_q(K_m; \mathbb{F}) = H_q(K; \mathbb{F})$$

Definition 9.1.2. Assume K is a simplicial complex, \mathbb{F} is a field, and $q \in \{0, 1, 2, \dots\}$. Given a filtration

$$K_1 \leq K_2 \leq \dots \leq K_m = K$$

of K , the corresponding q -dimensional **persistent homology** groups with coefficients in \mathbb{F} are images of the maps

$$(i_{s,t})_*: H_q(K_s; \mathbb{F}) \rightarrow H_q(K_t; \mathbb{F})$$

for all $0 \leq s \leq t \leq m$. The corresponding ranks $\beta_{s,t}^q = \text{rank}(i_{s,t})_*$ are called **persistent Betti numbers**.

⁵ I.e., components, holes, etc.

⁶ I.e., remain non-trivial

⁷ For example, $i_{s,t}: K_s \hookrightarrow K_t$.

⁸ By the functoriality of the homology we have $(i_{u,t})_* \circ (i_{s,u})_* = (i_{s,t})_*$.

 In each step of a filtration we add simplices. The addition of a single d -dimensional simplex in one step may either “terminate” a non-trivial homological element of dimension $d-1$, create a non-trivial homological element of dimension d , or have no effect on homology.

 Note that $\beta_{s,t}^q$ is a non-increasing function in t and a non-decreasing function in s .

As is the case with ordinary homology with coefficients in a field, each persistent homology group is determined⁹ up to isomorphism by its Betti number. A single filtration results in a table of persistent Betti numbers.

Example 9.1.3. Given any field \mathbb{F} the following are the tables of the zero-dimensional and one-dimensional persistent Betti numbers of the filtration of Figure 9.3:

	$s \ t$	1	2	3	4		$s \ t$	1	2	3	4
$\beta_{s,t}^0 \rightarrow$	1	2	1	1	1		1	0	0	0	0
	2	/	3	2	2		2	/	1	0	0
	3	/	/	2	2		3	/	/	2	1
	4	/	/	/	2		4	/	/	/	1

Let us demonstrate how to interpret these numbers geometrically:

- $\beta_{2,3}^0 = 2$ means that two of the different components of K_2 are still disconnected from each other in K_3 .
- $\beta_{3,4}^1 = 1$ roughly means that only one homologically non-trivial loop of K_3 is still¹⁰ homologically non-trivial in K_4 .
- $\beta_{2,3}^1 = 0$ means all one-dimensional homology elements in $H_1(K_2; \mathbb{F})$ are homologically trivial in K_3 .

⁹ While the rank of $(i_{s,t})_*$ determines the image of the map $i_{s,t}$ up to isomorphism, it does not determine a specific $\beta_{s,t}$ -dimensional subspace of $H_q(K_t; \mathbb{F})$. In this aspect persistent homology as a specific subgroup of $H_q(K_t; \mathbb{F})$ contains more information than persistent Betti numbers, i.e., its basis consists of homology representatives spanning the persistent homology group.

Table 9.1: The table of persistent Betti numbers corresponding to the filtration of Figure 9.3. The diagonal entries coincide with the Betti numbers of the corresponding stages of the filtration. The sub-diagonal entries are undefined.

¹⁰ A mathematically correct statement would be: the space of one-dimensional homology elements in $H_1(K_4; \mathbb{F})$ which have representatives in $C_1(K_3; \mathbb{F})$ is of dimension one.

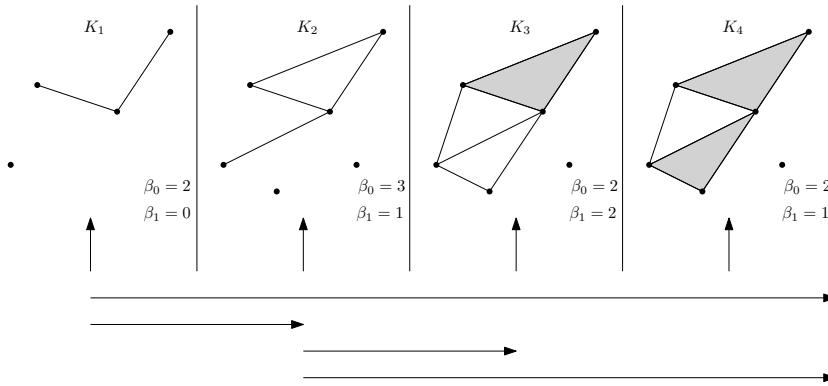


Figure 9.3: A filtration $K_1 \leq K_2 \leq K_3 \leq K_4$ along with the corresponding Betti numbers of each of the stage and the zero-dimensional barcode.

While the tables of persistent Betti numbers are useful, there are other ways to visualize the evolution of homology groups through a filtration. One such visualization we have already presented is the barcode.

9.2 Visualization

Throughout this section we fix a field \mathbb{F} , $q \in \{0, 1, \dots\}$, a filtration

$$K_1 \leq K_2 \leq \dots \leq K_m = K,$$

and $1 \leq s < t \leq m$.

Barcodes

Barcodes have been geometrically introduced above. In this subsection we will provide their formal definition.

Persistent Betti number $\beta_{s,t}^q$ represents the dimension of the subspace of homology elements in K_t that have a representative in K_s . Putting it differently, $\beta_{s,t}^q$ indicates the dimension of the collection¹¹ of homology elements in K_s that are still non-trivial in K_t in the sense that $\beta_{s,t}^q = \dim H_q(K_s; \mathbb{F}) / \ker(i_{s,t})_*$. Barcodes as indicated above, however, have more specific information: a bar $[s, t)$ represents a homology element that is born precisely at s and terminates precisely at t . Let us phrase this formally:

1. The number of bars containing s and passing through t equals¹² $\beta_{s,t}^q$.
2. Homology **born** at s is defined as¹³ $H_q(K_s; \mathbb{F}) / (\text{Im } i_{s-1,s})_*$. Its dimension is $\beta_{s,s} - \beta_{s-1,s}$ and represents the number of bars starting at s .
3. Homology **terminating** at t is defined as $\ker(i_{t-1,t})_*$. Its dimension¹⁴ is $\beta_{t-1,t-1} - \beta_{t-1,t}$ and represents the number of bars terminating at t .
4. The quantity $\beta_{s,t} - \beta_{s-1,t}$ represents¹⁵ the dimension of homology born at s which is still alive at t . It represents the number of bars starting at s which are passing through t .
5. Quantity $\mathbf{n}_{s,t} = \beta_{s,t-1} - \beta_{s-1,t-1} - (\beta_{s,t} - \beta_{s-1,t})$ represents¹⁶ the dimension of homology born at s which terminates at t . It represents the number of bars starting at s and terminating at t .
6. We additionally define $\mathbf{n}_{s,\infty} = \beta_{s,m} - \beta_{s-1,m}$, which represents the dimension of homology born at s which is still alive at the end of the filtration.

The q -dimensional **barcode**¹⁷ consists of intervals¹⁸ of the form

- i. $[s, t)$ for $1 \leq s < t \leq m$, and
- ii. $[s, \infty)$ for $1 \leq s < m$.

¹¹ This collection is formally not a linear subspace. In a formal setting we represent it as the quotient linear subspace appearing at the end of the sentence.

¹² Through the rest of the section we will drop the superscript q indicating the fixed dimension.

¹³ For formal reasons we define $(i_{0,t})_*$ to be the trivial map.

¹⁴ Using the fact that $\ker(i_{t-1,t})_* \cong H_q(K_{t-1}; \mathbb{F}) / \text{Im}(i_{t-1,t})_*$.

¹⁵ Compare to the interpretation of persistent Betti numbers above.

Also note that $\beta_{s,t} - \beta_{s-1,t} = \dim((\text{Im } i_{s,t})_* / \text{Im}(i_{s-1,t})_*)$, i.e., the dimension of the homology elements in $H_q(K_t; \mathbb{F})$ that have a representative in K_s module the ones that have a representative in K_{s-1} .

¹⁶ Observation 4. interprets this formula as [the dimension of homology born at s which is alive at $t-1$] - [the dimension of homology born at s which is still alive at t].

¹⁷ ...of the chosen filtration with coefficients in \mathbb{F} ...

¹⁸ In the setting of a barcode these intervals will be called **bars**.

A barcode can have multiple¹⁹ copies²⁰ of each interval. Fixing $1 \leq s < t \leq m$:

- The number of the intervals $[s, t)$ is denoted by $n_{s,t}$.
- The number of the intervals $[s, \infty)$ is denoted by $n_{s,\infty}$.

Example 9.2.1. We again turn our attention the familiar filtration in Figure 9.3. From the table on the right we can deduce that $n_{2,3} = 1 - 1 - (2 - 3) = 1$ and as a result there is 1 bar of the form $[2, 3)$, as displayed in the figure. In a similar fashion we compute $n_{1,2} = n_{1,\infty} = n_{2,\infty} = 1$ and $n_{1,3} = n_{1,4} = n_{2,4} = n_{3,4} = n_{3,\infty} = n_{4,\infty} = 0$.

A barcode represents the persistences of homology elements. The longer a bar, the longer the corresponding homology element persists. In most settings the longer persistence of a homology element also means higher importance²¹. However, there are also settings in which information is contained in shorter bars, especially when the bars are numerous and specifically distributed.

Persistence diagrams

Another established method of visualisation of persistent homology are persistence diagrams defined as follows. Given a barcode as defined above we can think of an interval $[s, t)$ as a pair of numbers and visualize²² it as a point $(s, t) \in \mathbb{R}^2$. A point of the form (s, ∞) obviously can't be drawn in a plane so we choose a y -coordinate above k , perhaps most conveniently as $k + 1$, to act as a representative of ∞ , i.e., a bar $[s, \infty)$ corresponds to a point $(s, k + 1)$. Each point (s, t) of a persistence diagram has an assigned **multiplicity** $n_{s,t}$, which represents the number of bars of the form $[s, t)$. In the case of $(s, k + 1)$, the multiplicity is $n_{s,\infty}$.

The result is a collection of weighted points in the plane called a **persistence diagram**. An example is provided in Figure 9.4.

A barcode encodes precisely the same information as a persistence diagram. While the persistence of a bar is measured by its length, the persistence of a point on a persistence diagram is measured by its distance from the **diagonal** $\Delta = \{(x, x) \mid x \in \mathbb{R}\}$. All points of a persistence diagram lie above Δ .

Persistence diagrams are often the chosen method of visualization when it comes to representation of persistent homology. Especially when the number of points and bars is large, their distribution seems to be well represented by persistence diagrams. On the other hand, when the number of points and bars is low, a barcode is often more descriptive.

¹⁹ ...or none...

²⁰ Alternatively, we could think of the barcode as the collection of all possible intervals of the forms (i) and (ii), each with an assigned multiplicity from $\{0, 1, 2, \dots\}$.

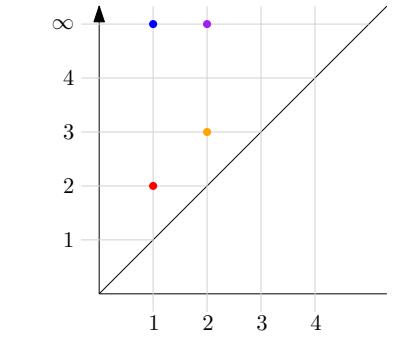
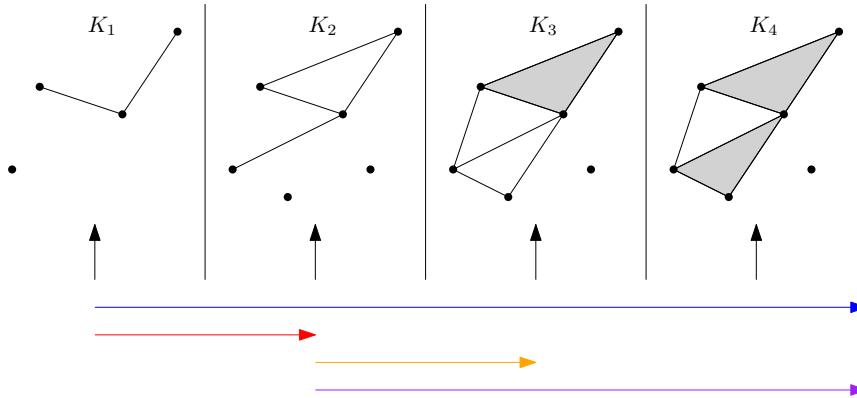
s	t	1	2	3	4
1	2	2	1 ₋	1 ₊	1
2	/	/	3 ₊	2 ₋	2
3	/	/	/	2	2
4	/	/	/	/	2

Computing $n_{2,3}$ of Figure 9.3: the coloring corresponds to the defining formula and the subscripts to the signs within.

²¹ I.e., a more prominent topological feature.

²² Just as there can be more bars with the same endpoints in a barcode, there can be more copies of the same point visualized at the same location in a persistence diagram. While multiple such intervals can be visualized in a vertical stack, the same can not be done with points. For this reason we always consider a point (s, t) in a persistence diagram as a weighted point with weight (multiplicity) $n_{s,t}$.

 Theoretically speaking, if there existed bars $[s, s)$ of length zero, then these would have been the shortest bars. They would have corresponded to diagonal points (s, s) . This point of view will come handy in the next chapter in the context of stability.



The fundamental lemma of persistent homology

Numbers $n_{s,t}$ are defined using persistent Betti numbers $\beta_{s,t}$. It turns out that the reverse expression also exists.

Lemma 9.2.2. [Fundamental lemma of persistent homology]

$$\beta_{s,t} = \sum_{s' \leq s, t' > t} n_{s',t'}$$

The formula in the lemma can be verified explicitly. However, the statement is apparent from the definitions, as

- $\beta_{s,t}$ represents the homology born at s or before and terminating after t ;
- $n_{s,t}$ represents the homology born precisely at s and terminating precisely at t .

Lemma 9.2.2 implies that the information encoded in a barcode or in a persistence diagram is precisely the same as the information encoded by persistent Betti numbers.

9.3 Computation

While the multiplicities $n_{s,t}$ of points of persistence diagrams are formally expressed by persistent Betti numbers, there is an algorithm to obtain them directly without referring to the Betti numbers and the corresponding $k(k+1)/2$ ranks of maps. In this section we will present perhaps the simplest²³ version of the algorithm, which is also the most illustrative. We will proceed in two steps:

- compute the matrix reduction, and
- extract the persistent homology.

Figure 9.4: A filtration along with the corresponding zero-dimensional barcode and persistence diagram. The colors of bars match the colors of the corresponding points in the persistence diagram.

\triangle In this setting, condition $t' > t$ implies that t' as an index in n could also take the value ∞ .
 \bowtie Lemma 9.2.2 has a geometric interpretation in the context of persistence diagrams, see Figure 9.5. It essentially states that $\beta_{s,t}$ is the sum of all multiplicities of points of a persistence diagram, which lie in the upper-left quadrant $[0,s] \times [t,\infty]$ with the apex at (s,t) . In the context of this interpretation, the formula for multiplicity $n_{s,t} = \beta_{s,t-1} - \beta_{s-1,t-1} - \beta_{s,t} + \beta_{s-1,t}$ is the expression of the square $(s-1,s] \times [t-1,t)$ in terms of such quadrants.

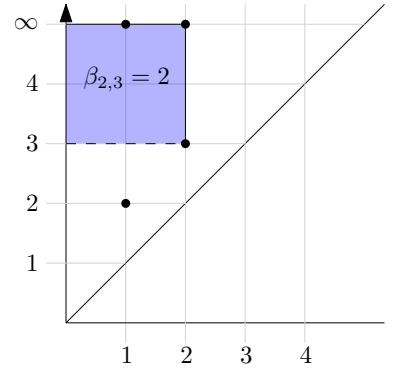


Figure 9.5: The sum of multiplicities of points in the blue quadrant with apex $(2,3)$ is $\beta_{2,3}$ by Lemma 9.2.2. It equals 2 as the point $(2,3)$ is not contained in it, see also the table in Example 9.1.3.

²³ There exist many improvements of this algorithm which may significantly improve the computing time.

We will conclude the section with an example.

Throughout this section we fix a field \mathbb{F} and filtration

$$K_1 \leq K_2 \leq \dots \leq K_m = K.$$

Parameter $q \in \{0, 1, \dots\}$ will denote the dimension of a considered object.

Matrix reduction

This part could be called an annotated matrix reduction using only column operation from the left.

1. **Order simplices consistently with the filtration.** For each q order all q -simplices in an order in which they appear in the filtration. If more simplices appear at the same time their internal order is immaterial²⁴.
2. For each q **construct the boundary matrix M_q using the order** chosen in 1 to label columns and rows.
3. For each q **reduce M_q from the left** using a **single type of column operations**: the addition of an \mathbb{F} -multiple of any of the previous²⁵ columns to a treated column. Specifically, starting with the leftmost column go through all the columns by passing to the right and for each column:
 - (a) Determine the pivot²⁶.
 - (b) If any of the previous columns on the left has a pivot in the same row, subtract the appropriate multiple of that column so that the pivot of the current column either disappears or its location is moved up.
 - (c) Repeat as long as there are matching pivots on the left.

For each q the resulting matrix is denoted by M'_q . Each of its columns is either trivial or has a pivot, whose row is unique amongst all pivots.

Extracting persistence

At this point we have sufficient information to extract homology of K_m from the number of pivots²⁷. However, we can also use the locations of pivots to extract numbers $n_{s,k}$ required to construct the barcode and persistence diagram. In order to explain the extraction process we first recall the incremental expansion.

Given a simplicial complex, an addition of a single q -simplex can change the homology in two ways:

²⁴ It would eventually effect the obtained critical simplices and representatives, but not the persistent homology.

²⁵ In the chosen order from 1.

²⁶ The lowest non-trivial entry in the column

²⁷ Note that the rank of a matrix is the number of its pivots in a reduced form, and the ranks themselves suffice to compute the Betti numbers.

- If its boundary is a linear combination of boundaries of other q -simplices²⁸, then the simplex gives birth to a non-trivial q -dimensional homology element. In this case we call the simplex a **birth simplex**.
- If its boundary is *not* a linear combination of boundaries of other q -simplices²⁹, then the simplex terminates a non-trivial $(q - 1)$ -dimensional homology element. In this case we call the simplex a **terminal simplex**.

A filtration can be considered to be a sequence³⁰ of incremental expansions. At each stage of the filtration we may assume we first add all vertices according to the ordering in 1., then all edges, etc. Combining such an ordering through all stages we get a sequence of incremental expansions inducing boundary matrices M_q and their reduced forms M'_q .

Based on such an ordering each simplex of K is either a terminal simplex or a birth simplex. We are now in a position to **extract persistence diagram and barcode**:

- For each terminal q -simplex τ there exists a paired birth $(q - 1)$ -simplex σ , which is the label of the pivot in the column τ . Such a pair **induces a bar** $[s, t)$ in the corresponding barcode or, equivalently, a **point** (s, t) in the corresponding persistence diagram, where s, t are the stages of the filtration at which σ and τ appear³¹.
- Each birth simplex which is not paired to a terminal simplex **induces a bar** $[s, \infty)$ in the corresponding barcode or, equivalently, a **point** $(s, m + 1)$ in the corresponding persistence diagram, where s is the stage of the filtration at which σ appears.

As a result we obtain a barcode and a persistence diagram as demonstrated in the example in the last subsection.

Representatives

Occasionally we are also interested in homology representatives of the bars and points of persistence diagrams. These can be extracted from the reduction process. In this subsection we present the most direct way of generating representatives. Given a bar with the birth simplex σ and the terminal simplex τ we define:

- The **birth representative** of σ as the chain formulated³² by the reduction of the column corresponding to σ to the zero column in the column reduction scheme. In particular, if the linear combination turning column σ into the zero column in our column reduction scheme is encoded in terms of columns as $\partial\sigma - \sum_i \lambda_i \partial\sigma_i = 0$, then

²⁸I.e., if, after adding the simplex to the boundary matrix as the rightmost column, its column gets reduced to the trivial column by the above reduction.

²⁹I.e., if, after adding the simplex to the boundary matrix as the rightmost column, its column does not reduce to the trivial column.

³⁰A specific sequence should respect the ordering given by a filtration as in 1. above, and also the structure of a simplicial complex, i.e., a simplex cannot be added before all of its faces are present.

³¹As each row contains at most one pivot, each birth simplex is paired to at most one terminal simplex. A terminal simplex cannot appear as the label of a pivot column.

³²Note that if $s = t$ we obtain an empty interval in the barcode and a point on the diagonal in the persistence diagram, both of which we ignore in the visualization as they represent elements of persistence zero. This is consistent with our interpretation of persistent homology, which measures only holes that persist through at least one stage of the filtration.

³³It turns out that the presented definition and computation of the barcode respects the elder rule mentioned at the beginning of the chapter.

³²For example, in the next subsection we provide an example in which the column $\langle c, d \rangle$ is reduced to the zero column by subtracting the column $\langle b, d \rangle$ and adding the column $\langle b, c \rangle$. This means $\partial\langle c, d \rangle - \partial\langle b, d \rangle + \partial\langle b, c \rangle = 0$ and hence $\langle c, d \rangle - \langle b, d \rangle + \langle b, c \rangle$ is the chain that is our birth representative.

the birth representative is $\alpha = \sigma - \sum_i \lambda_i \sigma_i$. The birth representative gives a homology class $[\alpha]$ that is born³³ by the addition of σ .

- The **terminal representative** is encoded by the column corresponding³⁴ to τ in the reduced matrix.

The birth representative and the terminal representative typically do not represent the same homology class. The birth representative may not even be a good representative of the corresponding bar in the sense that it may remain homologically non-trivial beyond³⁵ the appearance of the corresponding terminal simplex. On the other hand, infinite intervals do not have a terminal representative. As a result we define the **representative of a bar** as follows:

1. If the bar is a finite interval, the representative of the bar is the terminal representative.
2. If the bar is an infinite interval, the representative of the bar is the birth representative.

This choice of representatives is algebraically sound in the sense that the representatives form a basis of the elementary intervals of the decomposition described in the structure theorem for persistent homology, a result we discuss in details in the next chapter. This statement includes the fact that the lifespan of each representative matches the lifespan of the corresponding bar, and that the representatives are linearly independent³⁶ at all times.

In practice we sometimes deviate from the algebraically orthodox choice when declaring the representatives of 0-dimensional bars: we choose the birth representative as a bar representative even if the bar is bounded. Let us explain this geometrically motivated exception on the example of the next subsection, where pair $(\langle b \rangle, \langle a, b \rangle)$ induces a 0-dimensional bar. Sometimes we would geometrically like to think of this bar as a representation of the component containing b merging with a larger component, hence the choice of the birth representative $\langle b \rangle$ which fits into this geometric intuition. However, we should be aware that the homological element $[\langle b \rangle]$ does not become trivial³⁷ after adding $\langle a, b \rangle$. In terms of homology the appearance of $\langle a, b \rangle$ identifies³⁸ $[\langle a \rangle] = [\langle b \rangle]$ rather than sets $[\langle b \rangle] = 0$.

Example

Let us compute persistent homology of our standard example, see Figure 9.4. The annotation of simplices we will be using is provided in Figure 9.6. The chosen order is apparent from the following boundary matrices, in which vertical and horizontal lines divide simplices from different stages of the filtration.

³³ \triangle Homology class $[\alpha]$ is not the only homology class born by the addition of σ . If $[\beta]$ is another homology class of the same dimension that has existed before the addition of σ , then $[\alpha + \beta]$ is also a homology class born by adding σ .

³⁴ For example, in the next subsection we provide an example in which the column corresponding to $\langle b, c, d \rangle$ in M'_2 encodes the terminal representative $\langle b, c \rangle + \langle d, b \rangle + \langle c, d \rangle$.

³⁵ See the discussion on the representatives of 0-dimensional bars below for an example.

\bowtie There is no guarantee that these representatives are geometrically the most convenient. There are more involved ways of obtaining representatives that optimize given criterion function. For example, we may want to obtain the shortest 1-dimensional representatives, etc..

\bowtie Let us prove that the lifespan of the homology class of the terminal representative β corresponds to the lifespan of the corresponding bar:

- $[\beta]$ appears by the time σ appears by construction;
- if a representative β' of $[\beta]$ appeared before σ , then the column corresponding to τ could have been reduced further to β' thus eliminating the pivot labeled as σ , a contradiction.
- $[\beta]$ becomes trivial by the time τ emerges by definition;
- if $[\beta]$ became trivial sooner, its expression as a boundary could be used to reduce the τ column to the zero column, a contradiction.

³⁶ ...or trivial beyond their lifespans

³⁷ The terminal representative $\langle b \rangle - \langle a \rangle$ does become trivial. In fact, $[\langle b \rangle]$ never becomes trivial.

³⁸ In this sense, the terminal representative tells us which two components merge.

$$M_1 = \left(\begin{array}{cc|cc|cc|c} \langle b, c \rangle & \langle b, d \rangle & \langle a, b \rangle & \langle c, d \rangle & \langle a, c \rangle & \langle a, e \rangle & \langle b, e \rangle \\ \langle a \rangle & -1 & -1 & -1 & -1 & -1 & -1 \\ \langle b \rangle & -1 & -1 & 1 & 1 & 1 & 1 \\ \langle c \rangle & 1 & 1 & -1 & 1 & 1 & 1 \\ \langle d \rangle & & 1 & 1 & & & \\ \hline \langle e \rangle & & & & & 1 & 1 \\ \langle f \rangle & & & & & & 1 \end{array} \right)$$

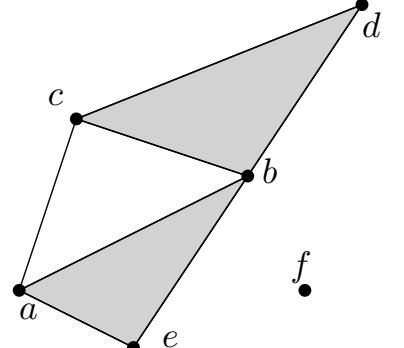


Figure 9.6: The annotation of simplices of K .

$$M_2 = M'_2 = \left(\begin{array}{cc|c} \langle b, c, d \rangle & \langle a, b, e \rangle \\ \langle b, c \rangle & 1 & \\ \langle b, d \rangle & -1 & \\ \hline \langle a, b \rangle & & 1 \\ \langle c, d \rangle & 1 & \\ \hline \langle a, c \rangle & & \\ \langle a, e \rangle & & -1 \\ \langle b, e \rangle & & 1 \end{array} \right)$$

We now perform the labelled matrix reduction as described above.

$$M_1 = \left(\begin{array}{cc|cc|cc|c} \langle b, c \rangle & \langle b, d \rangle & \langle a, b \rangle & \langle c, d \rangle & \langle a, c \rangle & \langle a, e \rangle & \langle b, e \rangle \\ \langle a \rangle & -1 & -1 & -1 & -1 & -1 & -1 \\ \langle b \rangle & 1 & 1 & 1 & 1 & 1 & 1 \\ \langle c \rangle & & 1 & -1 & 1 & 1 & 1 \\ \langle d \rangle & & & 1 & & & \\ \hline \langle e \rangle & & & & & 1 & 1 \\ \langle f \rangle & & & & & & 1 \end{array} \right)$$

In the matrices below a blue column is modified using red columns.

$\partial \langle c, d \rangle = \partial \langle b, d \rangle - \partial \langle b, c \rangle$.

$$\left(\begin{array}{cc|cc|cc|c} \langle b, c \rangle & \langle b, d \rangle & \langle a, b \rangle & \langle c, d \rangle & \langle a, c \rangle & \langle a, e \rangle & \langle b, e \rangle \\ \langle a \rangle & -1 & -1 & -1 & -1 & -1 & -1 \\ \langle b \rangle & 1 & 1 & 1 & 1 & 1 & 1 \\ \langle c \rangle & & 1 & -1 & 1 & 1 & 1 \\ \langle d \rangle & & & 1 & & & \\ \hline \langle e \rangle & & & & & 1 & 1 \\ \langle f \rangle & & & & & & 1 \end{array} \right)$$

$\partial \langle a, c \rangle = \partial \langle a, b \rangle + \partial \langle b, c \rangle$.

$$\left(\begin{array}{cc|cc|cc|c} \langle b, c \rangle & \langle b, d \rangle & \langle a, b \rangle & \langle c, d \rangle & \langle a, c \rangle & \langle a, e \rangle & \langle b, e \rangle \\ \langle a \rangle & -1 & -1 & -1 & -1 & -1 & -1 \\ \langle b \rangle & 1 & 1 & 1 & 1 & 1 & 1 \\ \langle c \rangle & & 1 & -1 & 1 & 1 & 1 \\ \langle d \rangle & & & 1 & & & \\ \hline \langle e \rangle & & & & & 1 & 1 \\ \langle f \rangle & & & & & & 1 \end{array} \right)$$

$\partial \langle b, e \rangle = \partial \langle a, e \rangle - \partial \langle a, b \rangle$.

$$M'_1 = \left(\begin{array}{cccc|cc} & \langle b,c \rangle & \langle b,d \rangle & \langle a,b \rangle & \langle c,d \rangle & \langle a,c \rangle & \langle a,e \rangle & \langle b,e \rangle \\ \langle a \rangle & & & -1 & & & & -1 \\ \langle b \rangle & -1 & & & & & & \\ \langle c \rangle & & 1 & & & & & \\ \langle d \rangle & 1 & & & & & & \\ \langle e \rangle & & & & & & 1 & \\ \langle f \rangle & & & & & & & \end{array} \right)$$

We can now extract the barcode from the birth-terminal pairs and unpaired birth simplices. We start by extracting the zero-dimensional barcode from the pivots of M'_1 and unpaired vertices.

- Pairs $(\langle c \rangle, \langle b, c \rangle)$ and $(\langle d \rangle, \langle b, d \rangle)$ provide no contribution³⁹.
- Pair $(\langle b \rangle, \langle a, b \rangle)$ induces a 0-dimensional (component) bar⁴⁰ $[1, 2)$ represented by $[\langle b \rangle]$.
- Pair $(\langle e \rangle, \langle a, e \rangle)$ induces a 0-dimensional (component) bar $[2, 3)$ represented by $[\langle e \rangle]$.
- Vertices a and f are unpaired and thus induce 0-dimensional bars $[1, \infty)$ (generated by $[\langle a \rangle]$) and $[2, \infty)$ (generated by $[\langle f \rangle]$).

We next extract the one-dimensional barcode from the pivots of M'_2 and unpaired edges.

- Pair $(\langle c, d \rangle, \langle b, c, d \rangle)$ induces a 1-dimensional bar $[2, 3)$ represented⁴¹ by $[\langle c, d \rangle - \langle b, d \rangle + \langle b, c \rangle]$.
- Pair $(\langle b, e \rangle, \langle a, b, e \rangle)$ induces a 1-dimensional bar $[3, 4)$ represented by $[\langle b, e \rangle - \langle a, e \rangle + \langle a, b \rangle]$.
- Edge $\langle a, c \rangle$ is unpaired and thus induces the 1-dimensional bar $[3, \infty)$ generated by $[\langle a, c \rangle - \langle a, b \rangle - \langle b, c \rangle]$.

Computational tricks

We conclude by mentioning a trick that speeds up the computation of persistent homology. It is based on an observation that boundary matrices M_i that are being reduced in the reduction process are not completely independent of each other.

If the reduction process of M_q reduces the column corresponding to q -simplex τ to a non-trivial column, we can extract the following information

1. τ is a terminal simplex and hence the row corresponding to τ in M_{q+1} will have been reduced to the zero-row, which means we can set it to zero immediately.

³⁹ Formally, they contribute the empty interval $[1, 1)$ as all involved simplices appear at K_1 .

⁴⁰ Recall that $\langle b \rangle$ appears at K_1 while $\langle a, b \rangle$ appears at K_2 , hence the values of the endpoints.

⁴¹ In this example the 1-dimensional birth and terminal representatives coincide. This is not generally the case.

⁴¹ Recall that $\langle c, d \rangle$ appears at K_2 while $\langle b, c, d \rangle$ appears at K_3 , hence the values of the endpoints. The linear combination that made the column corresponding to $\langle c, d \rangle$ trivial in M'_1 was $\partial\langle c, d \rangle - \partial\langle b, d \rangle + \partial\langle b, c \rangle$ and hence the representative.

2. The pivot location reveals the corresponding birth simplex σ . As a result the column corresponding to σ in M_{q-1} will have been reduced to the zero-column and can hence be set to zero immediately.

Hence a single reduction of a column in M_q corresponding to a terminal simplex also reveals a zero-row in M_{q+1} and a zero-column in M_{q-1} . Of course, this information can't be of much help if it has already been extracted from previous reductions. For this reasons the matrices M_q can be reduced in the order of decreasing dimension: this way no column in M_{q-1} has been reduced by the time M_q has been reduced and as a result we avoid reducing almost half⁴² the columns resulting in a significant speedup.

9.4 Concluding remarks

Recap (highlights) of this chapter

- Persistent homology;
- Barcode;
- Persistence diagram;
- Computing persistent homology.

Background and applications

Persistent homology^{43,44,45} is perhaps the most popular and fruitful construction of topological data analysis. For the past two decades it has been an inspiration to extensive theoretical and practical treatments, spanning from purely mathematical theoretical foundations⁴⁶ to computable aspects and applications in numerous fields of science and engineering. When coupled with standard constructions of complexes, persistent homology contains information about geometry of data. As such, the method is applied whenever the geometric shape of data is thought to contain significant information.

Applications include de-noising schemes, dimension reduction schemes, feature extraction methods, and specific data analysis of materials, molecular structures, medical images, weather patterns, etc.

The combinatorial treatment of this chapter will be followed by further properties in the following chapter. While the definition of persistent homology could have been expressed using coefficients in an Abelian group, the visualizations⁴⁷ and efficient implementations⁴⁸ crucially depend on the structure of a field.

 *Overview: reducing a τ -column to a non-zero column reveals:*

- τ is a terminal simplex;
- τ -row is trivial;
- pivot label σ is a birth simplex;
- σ -column is trivial.

⁴² This estimate depends on the filtration but seems to hold for most of the practical cases.

⁴³ Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28(4):511–533, 2002. doi: [10.1007/s00454-002-2885-2](https://doi.org/10.1007/s00454-002-2885-2)

⁴⁴ Herbert Edelsbrunner and John Harer. Computational Topology: An Introduction. *Applied Mathematics. American Mathematical Society*, 2010. doi: [10.1090/mhk/069](https://doi.org/10.1090/mhk/069)

⁴⁵ Tamal K. Dey and Yusu Wang. Computational Topology for Data Analysis. *Cambridge: Cambridge University Press*, 2022. doi: [10.1017/9781009099950](https://doi.org/10.1017/9781009099950)

⁴⁶ We will mention two ideas of generalizations of the standard persistent homology in the appendix.

⁴⁷ I.e., barcodes and persistence diagrams.

⁴⁸ I.e., matrix reductions.

Appendix: zig-zag persistence and multi-parameter persistence

In this appendix we will sketch the ideas of two generalizations of the standard persistent homology as presented throughout the chapter. In both cases the generalization refers to the type of filtration used.

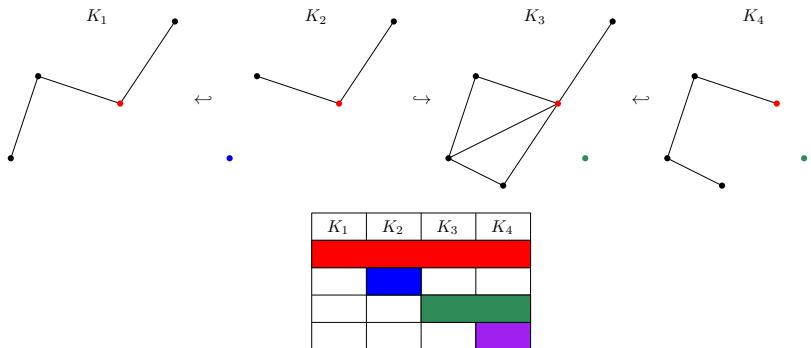
The first generalization is based on zig-zag filtration. While a standard filtration models a growing simplicial complex, a zig-zag filtration models a changing simplicial complex, in which the simplices may be appearing or disappearing.

Definition 9.4.1. Let K be a simplicial complex. A **zig-zag filtration** of K is a sequence of subcomplexes K_1, K_2, \dots, K_m of K , such that for each $i \in \{1, 2, \dots, m-1\}$ either $K_i \leq K_{i+1}$ or $K_{i+1} \leq K_i$.

For example, a zig-zag filtration may be of the following sort:

$$K_1 \hookleftarrow K_2 \hookleftarrow K_3 \hookleftarrow K_4 \hookleftarrow K_5 \hookleftarrow K_6 \hookrightarrow K_7.$$

It turns out that even in this setting there exists an algorithm based on matrix reductions which will produce a well defined barcode⁴⁹ describing what is called **zig-zag persistence**. An example is displayed in Figure 9.7.



☞ In case L_1, L_2, \dots, L_m are subcomplexes of a simplicial complex L not satisfying the condition of a zig-zag filtration, and we still want to compute a meaningful zig-zag homology, a standard way to construct a corresponding zig-zag filtration is to connect them either by unions or intersections:

$$K_1 \hookleftarrow K_1 \cup K_2 \hookleftarrow K_2 \hookleftarrow \dots \hookleftarrow K_m,$$

$$K_1 \hookleftarrow K_1 \cap K_2 \hookleftarrow K_2 \hookleftarrow \dots \hookleftarrow K_m.$$

Interestingly enough, while the two options induce generally different barcodes, they encode precisely the same information.

⁴⁹ Or, equivalently, a persistence diagram.

Figure 9.7: A zig-zag filtration and the corresponding zero-dimensional barcode, visualized as a table for technical reasons (i.e., the absence of a designated direction of all arrows as endpoints of bars). In the same way we could have presented the barcodes of ordinary persistent homology as well.

The second generalization is based on a multi-parameter filtration. While standard filtration models a one-parameter⁵⁰ growth of a simplicial complex, multi-parameter filtration models growth with more degrees of freedom. For our demonstrative purposes it suffices to formally introduce only a 2-parameter filtration.

⁵⁰ I.e., a sequential.

Definition 9.4.2. Let K be a simplicial complex. A **2-parameter filtration** of K is a collection of subcomplexes $K_{j,k} \leq K$ parameterized with $j, k \in \{1, 2, \dots, m\}$, such that for each $j \in \{1, 2, \dots, m-1\}$ and for each k the following containments hold (see Figure 9.8):

- $K_{j,k} \leq K_{j+1,k}$, and
- $K_{k,j} \leq K_{k,j+1}$.

There are theoretical and practical settings in which multi-parameter filtrations arise naturally. A **multi-parameter persistent homology** is the object obtained by applying the homology to spaces and maps of such a filtration. Unfortunately, there exists no convenient⁵¹ visualization⁵² in this setting. As a result, theoretical treatments of multi-parameter persistent homology typically deal with a multi-dimensional grid of interconnected homology groups, while practical applications of the same object use incomplete information about it such as multi-parameter tables of Betti numbers, restrictions to a 1-parameter filtrations yielding a standard barcode, etc. For more details see a book⁵³.

$$\begin{array}{ccccccc} K_{1,m} & \hookrightarrow & K_{2,m} & \hookrightarrow & K_{3,m} & \hookrightarrow & \cdots \hookrightarrow K_{m,m} \\ \uparrow & & \uparrow & & \uparrow & & \uparrow \\ \vdots & & \vdots & & \vdots & & \vdots \\ \uparrow & & \uparrow & & \uparrow & & \uparrow \\ K_{1,2} & \hookrightarrow & K_{2,2} & \hookrightarrow & K_{3,2} & \hookrightarrow & \cdots \hookrightarrow K_{m,2} \\ \uparrow & & \uparrow & & \uparrow & & \uparrow \\ K_{1,1} & \hookrightarrow & K_{2,1} & \hookrightarrow & K_{3,1} & \hookrightarrow & \cdots \hookrightarrow K_{m,1} \end{array}$$

Figure 9.8: A scheme of a 2-parameter filtration.

⁵¹ While a 1-parameter persistent homology “decomposes” into simple pieces called bars (we will explain this statement in detail in the next chapter), the pieces of a multi-parameter persistent homology can be quite complicated and not easily visualized or encoded.

⁵² ...such as multi-dimensional barcode or persistence diagram.

⁵³ Tamal K. Dey and Yusu Wang. Computational Topology for Data Analysis. Cambridge: Cambridge University Press, 2022. doi: [10.1017/9781009099950](https://doi.org/10.1017/9781009099950)

10

Persistent homology: stability theorem

IN THE PREVIOUS CHAPTER we introduced persistent homology and its basic method of computation in the discrete setting. However, it turns out that the concept of persistent homology can be treated on a much deeper theoretical level, through which many of its advantages become apparent.

In this chapter we will delve further into the theoretical machinery of persistent homology. We will introduce continuous filtrations and the underlying algebraic structure of persistence modules. These structures will be crucial in the formulation of the stability theorem, which states that, unlike homology, persistent homology behaves continuously with respect to the underlying filtration. We conclude by mentioning a series of interpretations and examples of our expanded scope of persistence.

10.1 Continuous filtrations

Recall that a discrete filtration of a simplicial complex K is a sequence of subcomplexes

$$K_1 \leq K_2 \leq \dots \leq K_m = K.$$

An example of a filtration is given in Figure 10.1.

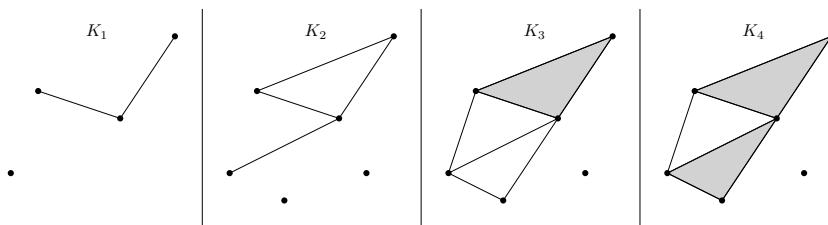


Figure 10.1: A discrete filtration.

Discrete filtrations¹ formalize finite nested sequences of complexes.

¹ I.e., filtrations given by finitely many nested simplicial complexes

While this approach is geometrically intuitive, there is an alternative shorter description of a filtration.

Rather than storing a sequence of separate subcomplexes, we **annotate** each simplex $\sigma \in K$ by the index $f(\sigma)$ at which σ first appears. Given such an annotation it is easy to reconstruct $K_i = \{\sigma \in K \mid f(\sigma) \leq i\}$. See Figure 10.2 for an example. This observation motivates us to expand the scope of filtrations in two ways: by considering continuous filtrations with infinitely many subcomplexes; and by defining filtrations from an appropriate annotation function.

1. A **continuous filtration** of a finite simplicial complex K is a collection of subcomplexes² $\{K_r\}_{r \geq 0}$ of K such that

$$\forall r < q : K_r \leq K_q \leq K.$$

2. Given a simplicial complex K , let f be a **filtration function**, i.e., an annotation of each of the simplices of K by a non-negative number such that $\sigma \leq \tau \implies f(\sigma) \leq f(\tau)$. The **sublevel filtration** associated to f is a continuous filtration consisting of sublevel complexes³ $K_r = \{\sigma \in K \mid f(\sigma) \leq r\} \leq K$ for $r \geq 0$.

There are two motivating reasons for the introduction of continuous filtrations:

- Most of the standard constructions of filtrations actually yield continuous filtrations: Čech filtration⁴, Rips filtration⁵, filtration by alpha complexes, sublevel filtration, etc.
- The interleaving structure and the resulting stability theorem depend⁶ on the continuous choice of parameter.

The definition of persistent homology groups for continuous filtrations is the same, with the only difference being the continuous range of the indices $0 \leq s \leq t$, which results in nominally infinitely many persistent homology groups. The matrix-reduction based computation of persistent homology is unhindered by the expansion to continuous filtrations as the computations depend only on filtration function defined on (finitely many) simplices, and do not require all (infinitely many) complexes of filtration separately. In a nutshell, we can compute the same barcodes using the same procedure for continuous or discrete filtrations.

We next discuss the interplay between discrete and continuous filtrations:

1. Given a discrete filtration, there is an obvious extension of it as the sublevel filtration of the annotation function.

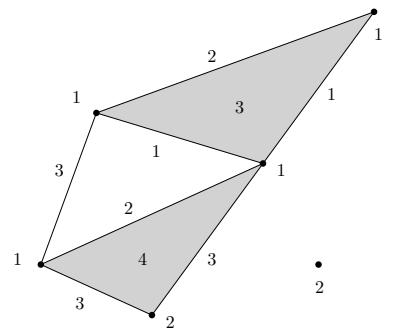


Figure 10.2: The annotation of simplices encodes the filtration of Figures 10.1.

² Throughout the book we will **additionally require** that for each simplex $\sigma \in K$ the minimum $\text{argmin}_r \{\sigma \in K_r\}$ exists, i.e., there exists the smallest scale r at which σ appears. An equivalent condition is the following: for each r there exists $r' > r$ such that $K_r = K_{r'}$, i.e., if a simplex is absent at scale r , it is also absent at slightly larger scales.

Under this condition each continuous filtration is the sublevel filtration of its associated annotation function. In particular, each sublevel filtration is a continuous filtration and vice versa. The Rips and Čech filtrations as defined in Chapter 5 are continuous filtrations of this sort.

³ In this setting parameter r is often referred to as the scale, a notion arising from Rips and Čech filtrations, or the level, a notion arising from the filtration function.

⁴ The corresponding filtration function on simplices being the radius of the smallest enclosing ball of the vertices.

⁵ The corresponding filtration function on simplices being the diameter of the set of vertices.

⁶ To be discussed in detail later. In a nutshell, the continuous choice of the parameter eventually results in the continuity of persistent homology.

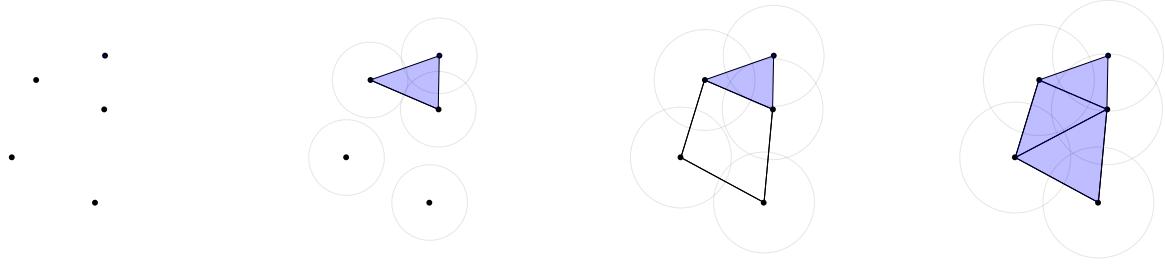


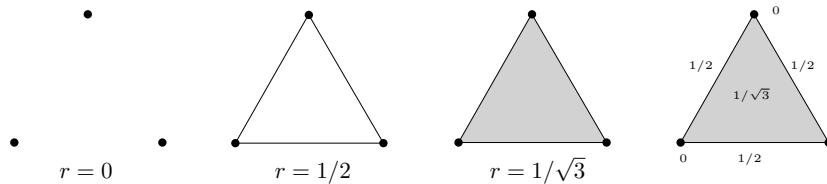
Figure 10.3: An excerpt from the Rips filtration on the five points on the left.

2. Given a continuous sublevel filtration $\{K_r\}_{r \geq 0}$ associated to a filtration function f there are two ways of generating a discrete filtration:
 - (a) By restriction to $K_1 \leq K_2 \leq \dots \leq K_{\lceil \max f \rceil}$. While mathematically convenient, this approach has many drawbacks⁷ and is mostly avoided.
 - (b) A more beneficial way of thinking about the index i of a discrete filtration is not as the scale parameter⁸ but rather as the index of the **critical scale**⁹ of the continuous filtration. Formally speaking we define critical scales $r_1 < r_2 < \dots < r_k$ as the enumeration¹⁰ of the image of f and define

$$K_i = \{\sigma \in K \mid f(\sigma) \leq r_i\}.$$

The corresponding finite filtration contains information about all changes in the original continuous filtration.

Continuous filtrations conveniently model the geometric setup of the standard filtrations. On the other hand, discrete filtrations are a convenient finite description on which we may develop algorithmic approaches.



Example 10.1.1. [Topology of offsets] Given a finite collection of points $S \subset \mathbb{R}^n$ we have already mentioned that the nerve theorem implies that for each $r > 0$ the Čech complex $\text{Čech}(S, r)$ is homotopy equivalent to the r -neighborhood¹¹ $N(S, r)$ of S :

⁷ The corresponding continuous filtration as defined by 1. may be significantly different from $\{K_r\}_{r \geq 0}$. The information about the sequence of changes between each pair of integer scales is lost.

⁸ An interpretation prevalent in the context of continuous filtrations.

⁹ A scale r of a continuous filtration is **critical**, if at least one simplex appears at r .

¹⁰ $\{r_1, r_2, \dots, r_k\} = \text{Im } f$.

☞ From this point on, whenever we mention an unspecified filtration, or consider a transition from a finite to continuous filtration or vice versa, the underlying interplays we have in mind are 1. and 2. (b). For an example see Figure 10.4.

Figure 10.4: The Čech filtration on three vertices forming an equilateral triangle of side length 1 nominally consists of infinitely many simplicial complexes. However, only at scales $0, 1/2$ and $1/\sqrt{3}$ do the changes occur and hence the corresponding discrete filtration (according to 2 (b) above) consists of simplices at those scales, depicted by the first three complexes in the figure. The annotation function is provided on the right, its image consists of the mentioned scales.

¹¹ Also called the r -offset of S .

$$\text{Cech}(S, r) = \mathcal{N}(\{B(s, r)\}_{s \in S}) \simeq \bigcup_{s \in S} B(s, r) = N(S, r).$$

It turns out that the conclusion of the nerve theorem behaves consistently¹² with the maps, which results in the following fact: if for $r_1 < r_2$ the inclusion $N(S, r_1) \hookrightarrow N(S, r_2)$ is a homotopy equivalence, then¹³ so is the inclusion $\text{Cech}(S, r_1) \hookrightarrow \text{Cech}(S, r_2)$.

The Cech filtration thus models the homotopy type of growing off-sets: if on some interval the growth of r results in homotopy equivalent growth of offsets, then it also results in a homotopy equivalent growth of the Cech complexes.

Interleaving distance for filtrations

We conclude this section by recalling the interleaving distance between filtrations. The concepts has already been defined in Chapter 5 for Rips and Cech filtrations. With the established general notions we can use the same definition for filtrations in general.

Definition 10.1.2. Choose $\varepsilon > 0$. Continuous filtrations $\{K_r\}_{r \geq 0}$ and $\{L_r\}_{r \geq 0}$ are ε -interleaved if there exist simplicial maps $\varphi_r: K_r \rightarrow L_{r+\varepsilon}$ and $\psi_r: L_r \rightarrow K_{r+\varepsilon}$ such that $\varphi_{r+\varepsilon} \circ \psi_r: L_r \rightarrow L_{r+2\varepsilon}$ and $\psi_{r+\varepsilon} \circ \varphi_r: K_r \rightarrow K_{r+2\varepsilon}$ are equal to the corresponding inclusions.

$$\begin{array}{ccccccc} \dots & \longrightarrow & K_r & \longrightarrow & K_{r+\varepsilon} & \longrightarrow & K_{r+2\varepsilon} \longrightarrow \dots \\ & & \searrow \varphi_r & & \nearrow & & \\ & & L_r & \longrightarrow & L_{r+\varepsilon} & \longrightarrow & L_{r+2\varepsilon} \longrightarrow \dots \\ & & \swarrow \psi_r & & \nearrow & & \end{array}$$

Given two filtrations their **interleaving distance** is defined as the minimum¹⁴ of all values $\varepsilon > 0$, for which the filtrations are ε -interleaved. It turns out that the interleaving distance is a metric¹⁵.

In Chapter 5 we proved that Rips and Cech filtrations equipped with the interleaving distance are continuous (stable) with respect to perturbations of the underlying points. Generalizing this result we now prove the sublevel filtrations are continuous with respect to perturbations of the filtration function in the max metric¹⁶.

Proposition 10.1.3. Let K be a simplicial complex. Assume $f, g: K \rightarrow [0, \infty)$ are filtration functions. Then the sublevel filtrations of K corresponding to f and g are $\|f - g\|_\infty$ interleaved.

Proof. In order to align our notation with the diagram above for $\varepsilon = \|f - g\|_\infty$ define $K_r = \{\sigma \in K \mid f(\sigma) \leq r\} \leq K$ and $L_r = \{\sigma \in K \mid g(\sigma) \leq r\} \leq K$. The interleaving maps φ, ψ are defined to be identities on vertices. The maps are well defined by the following argument:

¹² The formal term corresponding to this consistency is “functoriality” and the relevant extension of the nerve theorem is referred to as “persistent nerve theorem” or “functorial nerve theorem”.

¹³ At this point we crucially use the fact the Euclidean balls always form a good cover as required by the nerve theorem.

☞ Suppose filtrations $\{K_r\}_{r \geq 0}$ and $\{L_r\}_{r \geq 0}$ consist of subcomplexes of a simplicial complex K and let $\varepsilon \geq 0$. It is easy to see that if for each r we have $K_r \leq L_{r+\varepsilon}$ and $L_r \leq K_{r+\varepsilon}$, then the filtrations are ε -interleaved. An argument of this sort is used in the proof of Proposition 10.1.3.

¹⁴ It is not hard to prove that the minimum exists due to the additional requirement imposed on our filtrations.

¹⁵ In order to maintain this view we declare two filtrations to be **isomorphic** if they are 0-interleaved. The interleaving distance is a metric on the isomorphy classes of filtrations.

¹⁶ Given two functions $f, g: K \rightarrow \mathbb{R}$ defined on all simplices of a finite simplicial complex K , the **max distance** between them is

$$\|f - g\|_\infty = \max_{\sigma \in K} |f(\sigma) - g(\sigma)|.$$

- For each vertex $v \in K$: if $v \in K_r$ then $v \in L_{r+||f-g||_\infty}$ by the definition of the max distance. In a similar fashion, if $v \in L_r$ then $v \in K_{r+||f-g||_\infty}$. Hence maps φ, ψ are well defined on vertices.
- The same argument for simplices¹⁷ implies maps φ, ψ are simplicial.

□

¹⁷ For example, if a simplex $\sigma \in K$ is contained in K_r , it is also contained in $L_{r+||f-g||_\infty}$.

10.2 Persistence modules

Persistent homology is obtained by applying homology to a filtration. In this section we present the properties of the resulting algebraic objects (persistence modules) which model persistent homology. Just as filtrations model the growth of simplicial complexes, persistence modules model the evolution¹⁸ of vector spaces¹⁹.

For the rest of this chapter we fix a field \mathbb{F} , which will provide coefficients to all mentioned vector spaces, including homology groups.

Persistence modules

Definition 10.2.1. A **persistence module** is a collection of (finite dimensional) vector spaces $\{V_r\}_{r \geq 0}$ along with linear maps

$$h_{r,q}: V_r \rightarrow V_q, \quad \forall r \leq q$$

satisfying $h_{r,q} = h_{r,s} \circ h_{s,q}$ and $h_{r,r} = id_{V_r}$ for all $r \leq q \leq s$.

Scale $r \geq 0$ is said to be **regular** if there exists $\varepsilon > 0$ such that maps $h_{p,q}$ are isomorphisms for all $p, q \in (r - \varepsilon, r + \varepsilon)$ or (in the case $r = 0$) for all $p, q \in [0, \varepsilon]$, i.e., the maps h are isomorphisms close to r . Scale r is **critical** if it is not regular.

Our interest in persistence modules stems from the fact that they are the underlying algebraic structure of persistent homology of continuous filtrations. In order to simplify our treatment we thus restrict to persistence modules that appear as persistent homology of continuous filtrations as defined above. In particular, each persistence module treated here will be **assumed to have the following properties**:

1. There exists $R > 0$ such that for each $R \leq r < q$ maps $h_{r,q}$ are isomorphisms²⁰, i.e., eventually all maps h are isomorphisms.
2. For each $r > 0$ there exists $r' > r$ such that for all $q \in [r, r')$ the maps $h_{r,q}$ are isomorphisms²¹.
3. There exist finitely²² many critical scales.

☞ In general literature persistence modules may consist of infinite dimensional vector spaces.

☞ Critical scales of a continuous filtration are a superset of critical scales of its persistent homology as any change in homology requires a change of the underlying complex, but not vice versa.

☞ Properties 2. and 3. imply $[0, \infty)$ can be decomposed into finitely many intervals of the form $[*_1, *_2)$ on which all maps h are isomorphisms.

²⁰ An analogous property holds for continuous filtrations as they filter a finite simplicial complex, i.e., given a filtration function f , all sublevel complexes K_r for $r > \max |f|$ coincide.

²¹ This corresponds to the analogous property assumed for our continuous filtrations.

²² This property corresponds to the fact that continuous filtrations filter a finite simplicial complex.

Definition 10.2.2. Persistence modules $\{V_r\}_{r \geq 0}$ and $\{W_r\}_{r \geq 0}$ are **isomorphic** if for each $r \geq 0$ there exist isomorphisms $V_r \rightarrow W_r$ such that for each $0 \leq r_1 < r_2 < \dots$ the following diagram commutes

$$\begin{array}{ccccccc} \cdots & \longrightarrow & V_{r_j} & \longrightarrow & V_{r_{j+1}} & \longrightarrow & V_{r_{j+2}} \longrightarrow \cdots \\ & & \downarrow \cong & & \downarrow \cong & & \downarrow \cong \\ \cdots & \longrightarrow & W_{r_j} & \longrightarrow & W_{r_{j+1}} & \longrightarrow & W_{r_{j+2}} \longrightarrow \cdots \end{array}$$

Decomposition

It is often advantageous to decompose²³ mathematical objects into simple pieces and thus obtain a canonical form. In the previous chapter we decomposed persistent homology into pieces represented by bars. In this subsection we will formalize such a decomposition for persistence modules.

We first explain what we mean by a “decomposition”.

Definition 10.2.3. The **direct sum** of persistence modules $\{V_r\}_{r \geq 0}$ and $\{V'_r\}_{r \geq 0}$ along with respective linear maps $h_{r,q}$ and $h'_{r,q}$, is a persistence module consisting of:

- spaces $W_r = V_r \oplus V'_r$ and
- maps $\tilde{h}_{r,q} = (h_{r,q}, h'_{r,q})$

for all $0 \leq r < q$.

We next present interval modules, which are the pieces represented by bars.

Definition 10.2.4. Let $0 \leq p < q$. An **interval module** $\mathbb{F}_{p,q}$ corresponding to the pair (p, q) is a persistence module $\{V_r\}_{r \geq 0}$ defined as follows:

- $V_r = \mathbb{F}$ for $r \in [p, q)$ and $V_r = 0$ else.
- Maps $h_{s,s'}$ are isomorphisms whenever possible.

Theorem 10.2.5. [Structure theorem for persistent homology] Each persistence module is isomorphic to a direct sum of interval modules. The decomposition is unique up to the permutation of the intervals.

²³ Functions are decomposed into monomials (Taylor series) or trigonometric functions (Fourier series). Closed connected surfaces other than the sphere can be decomposed as a direct sum of tori or projective planes. Every n -dimensional vector space is of a form \mathbb{F}^n and in one of the previous appendices we mentioned how finitely generated Abelian groups can be decomposed into smallest indecomposable groups: groups of the form \mathbb{Z}_p and \mathbb{Z} .

☞ Rewriting the condition on maps in Definition 10.2.4:

- for $p \leq s \leq s' < q$ map $h_{s,s'}$ is the identity on \mathbb{F} .
- else $h_{s,s'}$ is the zero map.

☞ It is easy to verify that interval modules $\mathbb{F}_{p,q}$ and $\mathbb{F}_{p',q'}$ are isomorphic iff $p = p'$ and $q = q'$.

Barcodes and bars introduced in the previous chapter correspond to this decomposition and interval modules. Theorem 10.2.5 is an algebraic expression of the existence of barcodes. It states that the persistence module can be decomposed into the intervals and is completely determined²⁴ by the interval modules of its decomposition.

Interleaving distance for persistence modules

The interleaving distance has already been defined for filtrations. Conceptually the same definition applies to persistence modules.

Definition 10.2.6. Choose $\varepsilon > 0$. Persistence modules $\{V_r\}_{r \geq 0}$ and $\{W_r\}_{r \geq 0}$ along with their respective linear maps $h_{r,q}$ and $h'_{r,q}$ are ε -interleaved if there exist linear maps $\varphi_r: V_r \rightarrow W_{r+\varepsilon}$ and $\psi_r: W_r \rightarrow V_{r+\varepsilon}$ such that $\varphi_{r+\varepsilon} \circ \psi_r: W_r \rightarrow W_{r+2\varepsilon}$ and $\psi_{r+\varepsilon} \circ \varphi_r: V_r \rightarrow V_{r+2\varepsilon}$ are equal $h'_{r,r+\varepsilon}$ and $h_{r,r+\varepsilon}$ correspondingly.

Given two persistence modules their **interleaving distance** d_I is defined as the minimum of all values $\varepsilon > 0$, for which the filtrations are ε -interleaved.

$$\begin{array}{ccccccc} \cdots & \longrightarrow & V_r & \longrightarrow & V_{r+\varepsilon} & \longrightarrow & V_{r+2\varepsilon} \longrightarrow \cdots \\ & & \searrow \varphi_r & & \nearrow & & \nearrow \\ & & W_r & \longrightarrow & W_{r+\varepsilon} & \longrightarrow & W_{r+2\varepsilon} \longrightarrow \cdots \end{array}$$

It is not hard to prove that the minimum in the definition of the interleaving distance exists due to the additional requirement imposed on persistence modules. It is easy to verify that the interleaving distance is a metric on the isometry classes of persistence modules. As such the interleaving distance is the metric²⁵ of choice on persistent homologies.

The functoriality of homology implies that ε -interleaved filtrations²⁶ induce ε -interleaved persistence modules. Another setting in which ε -interleaved persistence modules (but not necessarily ε -interleaved filtrations) are obtained is that of spaces, which are “close” to each other. Let us first define closeness.

Definition 10.2.7. Let (X, d) be a metric space and assume $A, B \subset X$ are finite subsets. The **Hausdorff distance** $d_H(A, B)$ is defined as

$$d_H(A, B) = \max \left\{ \max_{b \in B} \min_{a \in A} d(a, b), \max_{a \in A} \min_{b \in B} d(a, b) \right\}.$$

The Hausdorff distance is a metric on all finite subspaces of a metric space X . It has a natural geometric meaning. Given the setting of Definition 10.2.7 find:

²⁴ And as a result, barcodes and persistence diagrams are complete descriptions of persistence modules.

²⁵ At this point it should be clear that continuity and small perturbations of persistent homology depend on the ability to perform continuous and small steps in the index set.

An interleaving distance defined on persistent homology of discrete filtrations or a single complex would have been, in the best of cases, restricted to the integer values, that do not accommodate the idea of continuity.

²⁶ We have already discussed how these appear by perturbing points when using Rips or Čech complexes, and by perturbing the filtration function when using the sublevel filtration.

- The minimal r_A such that $N(A, r_A) \supset B$, i.e., the r_A -neighborhood of A contains B .
- The minimal r_B such that $N(B, r_B) \supset A$.

We conclude that $d_H(A, B) = \min\{r_A, r_B\}$. Note that for each $a \in A$ there exists $b \in B$ such that $d(a, b) \leq d_H(a, b)$, and vice versa. An example is given in Figure 10.5, where a black set A and a red set B are displayed on top, while their respective neighborhoods are displayed in the middle and on the bottom.

Hausdorff distance measures the distances between finite subspaces of a metric space and heavily depends on a way in which these subspaces are embedded. For example, different isometric subspaces will be at a positive Hausdorff distance.

Similar to Hausdorff distance is the Gromov-Hausdorff distance.

Definition 10.2.8. Suppose A, B are finite metric spaces. The **Gromov-Hausdorff distance** $d_{GH}(A, B)$ is defined as

$$d_{GH}(A, B) = \inf_{\mu, \nu} \{d_H(\mu(A), \nu(B))\},$$

where the infimum is over all isometric embeddings $\mu: A \rightarrow X$ and $\nu: B \rightarrow X$ into a metric space X .

It turns out that the infimum in Definition 10.2.8 is always attained and that d_{GH} is a metric on the isometry classes²⁷ of finite metric spaces.

Proposition 10.2.9. Let A, B be finite metric spaces with $\varepsilon = d_{GH}(A, B)$. Then for each $q \in \{0, 1, \dots\}$:

1. $\{H_q(\text{Rips}(A, r))\}_{r \geq 0}$ and $\{H_q(\text{Rips}(B, r))\}_{r \geq 0}$ are 2ε -interleaved.
2. $\{H_q(\text{Cech}(A, r))\}_{r \geq 0}$ and $\{H_q(\text{Cech}(B, r))\}_{r \geq 0}$ are ε -interleaved.

Proof. We will only sketch the proof for $q = 1$ and Rips filtrations. The proof of other cases follows the same idea but requires some technical diligence. Without loss of generality we may assume A and B are subspaces of a metric space X and $\varepsilon = d_H(A, B)$.

We aim to define maps φ and ψ that constitute a commutative diagram:

$$\begin{array}{ccccccc} \cdots & \longrightarrow & H_1(\text{Rips}(A, r)) & \longrightarrow & H_1(\text{Rips}(A, r + 2\varepsilon)) & \longrightarrow & H_1(\text{Rips}(A, r + 4\varepsilon)) \longrightarrow \cdots \\ & & \searrow \varphi_r & & \swarrow & & \swarrow \\ \cdots & \longrightarrow & H_1(\text{Rips}(B, r)) & \longrightarrow & H_1(\text{Rips}(B, r + 2\varepsilon)) & \longrightarrow & H_1(\text{Rips}(B, r + 4\varepsilon)) \longrightarrow \cdots \end{array}$$

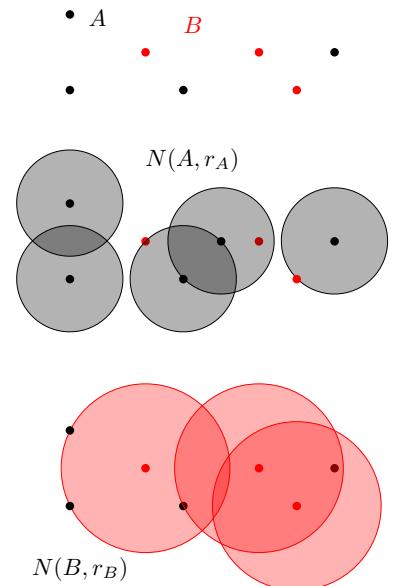


Figure 10.5: $d_H(A, B) = r_B > r_A$.

Observe that $d_{GH}(A, B) \leq d_H(A, B)$ for finite subspaces of a metric space X . As a result Proposition 10.2.9 also holds for d_H . However, Gromov-Hausdorff distance is typically harder to compute and thus it is occasionally more convenient to use d_H as the easily computable parameter of interleaving.

²⁷ In particular, $d_{GH}(A, B) = 0$ iff the spaces are isometric.

We first define maps on the vertices of the Rips complexes:

- For each $a \in A$ choose $b_a \in B$ such that $d(a, b_a) \leq \varepsilon$ and define $\varphi_r(a) = b_a, \forall r$.
- For each $b \in B$ choose $a_b \in A$ such that $d(b, a_b) \leq \varepsilon$ and define²⁸ $\psi_r(b) = a_b, \forall r$.

Given a 1-cycle $\alpha = \sum_i \langle a_i, a_{i+1} \rangle$ in $\text{Rips}(A, r)$ define $\varphi_r([\alpha]) = [\sum_i \langle b_{a_i}, b_{a_{i+1}} \rangle]$. This gives well defined maps φ_r (and also ψ_r) by the following arguments:

- $\sum_i \langle b_{a_i}, b_{a_{i+1}} \rangle$ is a cycle in $\text{Rips}(B, r + 2\varepsilon)$ as $d(a_i, a_{i+1}) \leq r$ implies $d(b_{a_i}, b_{a_{i+1}}) \leq r + 2\varepsilon$.
- If $[\sum_i \langle a_i, a_{i+1} \rangle] = [\sum_i \langle a'_i, a'_{i+1} \rangle]$ holds²⁹, then $\varphi([\sum_i \langle a_i, a_{i+1} \rangle]) = \varphi([\sum_i \langle a'_i, a'_{i+1} \rangle])$ as well³⁰.

At last we need to show that

$$\left[\sum_i \langle a_i, a_{i+1} \rangle \right] = \left[\sum_i \langle a''_i, a''_{i+1} \rangle \right]$$

in $H_1(\text{Rips}(A, r + 4\varepsilon))$ where $a''_i = a_{b_{a_i}}$. First note that $d(a_i, a''_i) \leq 2\varepsilon, \forall i$. The difference $\sum_i \langle a_i, a_{i+1} \rangle - \sum_i \langle a''_i, a''_{i+1} \rangle$ is a boundary as demonstrated by the blue 2-chain in Figure 10.6.

²⁸ As defined, maps φ_r and ψ_r do not define an interleaving between the Rips filtrations as in general $a_{b_a} \neq a$.

²⁹ This means $\sum_i \langle a_i, a_{i+1} \rangle - \sum_i \langle a'_i, a'_{i+1} \rangle = \partial \sum_j \langle x_j, y_j, z_j \rangle$.

³⁰ This holds as $\sum_i \langle b_{a_i}, b_{a_{i+1}} \rangle - \sum_i \langle b_{a'_i}, b_{a'_{i+1}} \rangle = \partial \sum_j \langle b_{x_j}, b_{y_j}, b_{z_j} \rangle$ and $\langle b_{x_j}, b_{y_j}, b_{z_j} \rangle$ are triangles in $\text{Rips}(B, r + 2\varepsilon)$.

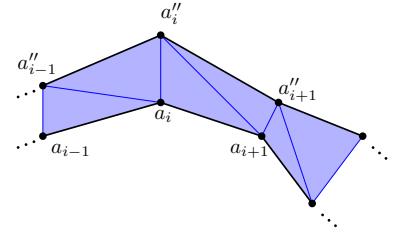


Figure 10.6: An excerpt from the proof of Proposition 10.2.9. Each edge connects points at distance at most $r + 2\varepsilon$.

³¹ With various versions discussing various forms of input.

10.3 Bottleneck distance and stability theorem

The many versions of the stability theorem for persistent homology state that persistent homology is continuous with respect to continuous change of the input parameters³¹. We have already seen examples of this sort: through Propositions 10.2.9 and 10.1.3 we can conclude that persistent homology behaves “continuously” in the interleaving distance. One of the main advantages of persistent homology is its visualization and so the final step towards a geometrically convenient form of the stability theorem is to interpret³² the interleaving distance in geometric terms as a distance on persistence diagrams³³. The resulting distance on persistence diagrams is called the bottleneck distance.

Bottleneck distance

We start by explaining notions and setting needed to define the bottleneck distance. Suppose $\mathcal{A} = (a_1, a_2, \dots, a_m)$ and $\mathcal{B} = (b_1, b_2, \dots, b_n)$ are persistence diagrams, i.e.:

³² A brief idea about a transition from the interleaving distance to the bottleneck distance is provided in appendix.

³³ For this setting, the visualization with persistence diagrams is much preferred to the visualization with the barcodes.

- each a_i and b_i is a point above the diagonal in the first quadrant in the plane, and
- each point may appear multiple times in any of the diagrams.

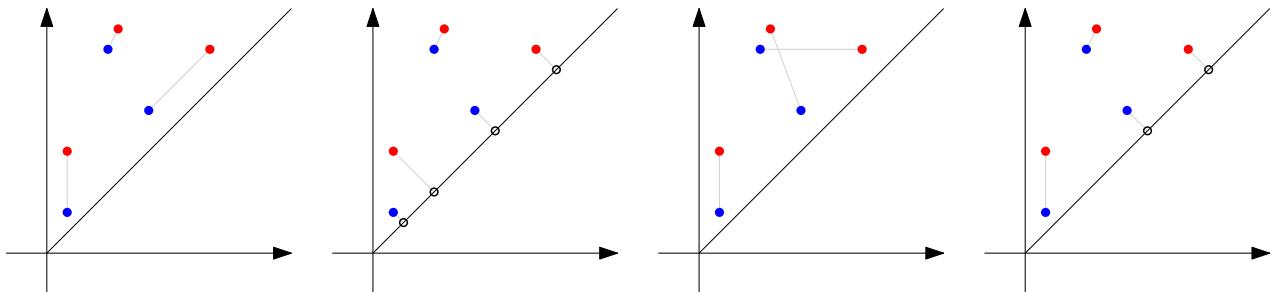
For a point $v = (x, y) \in \mathbb{R}^2$ let³⁴ $\bar{v} = ((x+y)/2, (x+y)/2) \in \mathbb{R}^2$. A **partial matching** between \mathcal{A} and \mathcal{B} is a bijective map $\varphi: \mathcal{A}' \rightarrow \mathcal{B}'$ where³⁵ $\mathcal{A}' \subseteq \mathcal{A}$ and $\mathcal{B}' \subseteq \mathcal{B}$. The **matching distance** of such a φ is defined as

$$d_M(\varphi) = \max \left\{ \max_{v \in \mathcal{A}'} \{d_\infty(v, \varphi(v))\}, \max_{v \in \mathcal{A} \setminus \mathcal{A}'} \{d_\infty(v, \bar{v})\}, \max_{v \in \mathcal{B} \setminus \mathcal{B}'} \{d_\infty(v, \bar{v})\} \right\}.$$

Let $\mu(\mathcal{A}, \mathcal{B})$ denote the collection of all partial matchings between \mathcal{A} and \mathcal{B} .

Definition 10.3.1. The **bottleneck distance** between persistence diagrams \mathcal{A} and \mathcal{B} is the minimal matching distance between them, i.e.,

$$d_B(\mathcal{A}, \mathcal{B}) = \min_{\varphi \in \mu(\mathcal{A}, \mathcal{B})} d_M(\varphi).$$



Examples of partial matchings are given in Figure 10.7. In order to demonstrate the additional pairs used in the definition of the bottleneck distance, the unmatched points are connected to the closest point on the diagonal. The matching with the smallest matching distance is the second from the left, a fact that can be verified in Figure 10.8, which illustrates the matching distances for matchings of three diagrams of Figure 10.7. The $d_\infty(a, b)$ distance between points a and b can be thought to represent one half of the side-length of the square centered at a which has b on its boundary. The maximal length of such sides is the smallest in the second case and the resulting quantity is the bottleneck distance d_B .

³⁴ \bar{v} represents the point on the diagonal $\Delta = \{(z, z) \mid z \in \mathbb{R}\}$ which is the closest to v in d_∞ (and also in d_2) metric.

³⁵ Again, a point can appear in \mathcal{A}' or \mathcal{B}' multiple times but not more times than in \mathcal{A} or \mathcal{B} respectively.

☞ Recall that the max distance $d_\infty((x_1, y_1), (x_2, y_2))$ between points in the plane is defined as

$$\max\{|x_1 - x_2|, |y_1 - y_2|\}.$$

Figure 10.7: Examples of partial matchings between the red and the blue persistence diagrams with points unmatched by φ being matched to the closest point on the diagonal.

☞ At this point it should become apparent why it is geometrically convenient to consider points on the diagonal represent the trivial persistence module. A side effect of this approach is that any two points on the diagonal represent the same trivial persistence module. In a way, the entire diagonal should thus be treated as a single point.

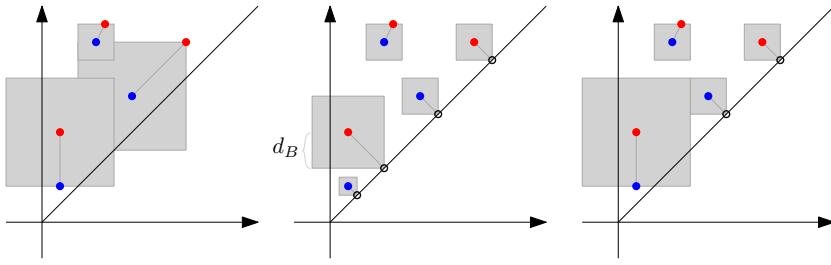


Figure 10.8: The distances between the matched pairs are demonstrated by the squares arising as the balls of the d_∞ metric. The diagram with the smallest maximal square amongst all matchings (even the ones not displayed here) is the middle one. Hence the resulting bottleneck distance d_B arises from the middle diagram.

Theorem 10.3.2 (Isometry theorem). *The interleaving distance between persistence modules equals the bottleneck distance between the corresponding persistence diagrams.*

Example 10.3.3. Let \mathcal{A} be the persistence diagram presented by the four blue points in Figure 10.9. If persistence diagram \mathcal{B} satisfies $d_B(\mathcal{A}, \mathcal{B}) \leq \varepsilon$, then \mathcal{B} consists of the following:

- For each blue point there exists one designated³⁶ red point within the grey square (i.e., the ε -ball in d_∞) around it.
- Arbitrarily many points within the grey ε -band³⁷ along the diagonal.

Stability theorem

Theorem 10.3.4. [Stability theorem] Assume persistence diagrams \mathcal{A} and \mathcal{B} represent persistent homologies of filtrations \mathcal{V} and \mathcal{W} obtained by one of the following procedures:

1. As the sublevel filtrations of filtration functions f and g satisfying condition $\|f - g\|_\infty \leq \varepsilon$, see Proposition 10.1.3.
2. As the Rips filtrations of metric spaces X and Y satisfying condition $d_{GH}(X, Y) \leq \varepsilon/2$, see 1. of Proposition 10.2.9.
3. As the Čech filtrations of metric spaces X and Y satisfying condition $d_{GH}(X, Y) \leq \varepsilon$, see 2. of Proposition 10.2.9.

Then $d_B(\mathcal{A}, \mathcal{B}) \leq \varepsilon$.

Figure 10.10 is a schematic representation of the discussion leading to the stability theorem as presented here.

The moral of the theorem is that small perturbations of the input lead to small changes in persistence diagrams. On the other hand, critical simplices and homology representatives may be unstable.

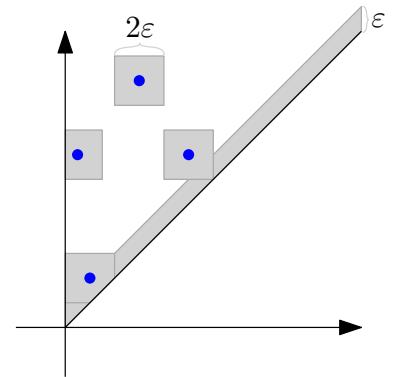


Figure 10.9: A schematic representation of the ε -neighborhood of the diagram consisting of blue points as discussed in Example 10.3.3.

³⁶ If some of the squares had non-empty intersection, then within that intersection there might be more points of \mathcal{B} , so a single square might contain more red points. However, one of them, potentially a diagonal point, has to be the designated one, i.e., the point to which the blue point in question is matched in an optimal matching.

³⁷ This band is actually the ε -neighborhood of the trivial (empty) persistence diagram. Again, this implies that the squares intersecting the band may contain more points.

☞ The stated version combines several separate versions of the stability theorem found throughout the literature by stating several different initial assumptions.

☞ While the presented results explain stability in terms of the bottleneck distance, there is another family of distances on persistence diagrams called the **Wasserstein distances**. For example, the 1-Wasserstein distance is obtained by defining the matching distance as the sum (rather than max) of individual terms. Under appropriate assumptions the persistence diagrams are also stable when using Wasserstein distances.

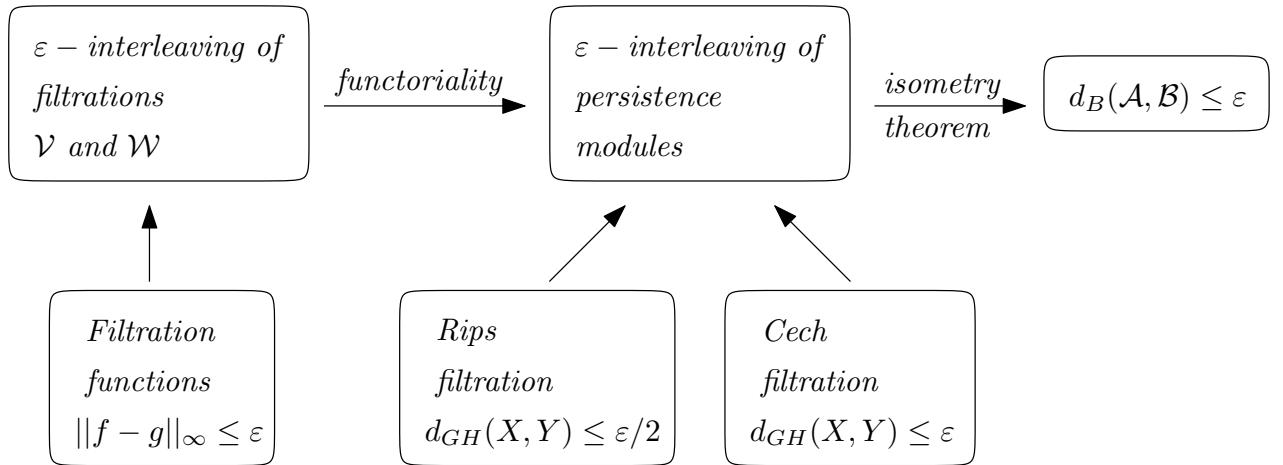


Figure 10.10: The diagram summarizing the stability theorem and strategy of its proof that have been discussed.

10.4 Interpretations and examples

With the stability theorem, persistent homology may be thought of as a stable description of a geometric shape. Stability itself justifies the following observations:

- Given a geometric shape, ever better approximating point-clouds induce persistent diagrams ever closer³⁸ (converging) to the persistence diagram of the shape.
- The points of higher persistence (i.e., the longer bars in the barcode) represent more stable³⁹ features and are thus typically deemed to be of higher importance, leading to simplification schemes on data, such as denoising.

In this section we will present several examples⁴⁰ of persistent homology arising via Rips complexes and comment on their structure.

1-dimensional persistence of geodesic spaces

Let X be a closed⁴¹ geodesic manifold or, more generally, the body of a finite simplicial complex equipped with a geodesic metric. Assume S_n is a sequence⁴² of finite metric spaces converging towards X in the Gromov-Hausdorff metric. Let \mathcal{A}_n denote the 1-dimensional persistence diagram obtained from S_n via Rips filtration and coefficients in \mathbb{F} . It turns out that the limiting diagram⁴³ $\mathcal{A} = \lim_{n \rightarrow \infty} \mathcal{A}_n$ encodes a shortest base of $H_1(X; \mathbb{F})$: for each member⁴⁴ α of a shortest homology base of X we obtain a bar $[0, |\alpha|/3]$, where $|\alpha|$ is the length of α , see Figure 10.11. Without going through all the details let us demonstrate the situation through examples.

³⁸ Given a closed manifold X , a sufficiently small scale r , and a point-cloud S sufficiently close to X in d_{GH} , the corresponding Rips and Čech complexes are homotopy equivalent to X . As a result, the persistent homology around scale r of a good approximation reveals homology of X .

³⁹ I.e., they remain non-trivial under larger perturbations.

⁴⁰ Generated by Ripserer.jl, with coefficients in \mathbb{Z}_2 .

⁴¹ This implies it admits a finite triangulation.

⁴² Such a sequence may be, roughly speaking, obtained by constructing ever finer finite approximations of X and inducing an approximation of a geodesic metric on them.

⁴³ \mathcal{A} can be obtained as persistence diagram of the Rips filtration of X , a construction which involves infinite simplicial complexes and is formally beyond the scope of this book.

⁴⁴ Members are formally cycles whose length in this case is the length of the corresponding loop in X . One can choose a triangulation for which these simplicial loops are the shortest possible.

Figures 10.12 and 10.13 represent three surfaces in \mathbb{R}^3 approximated by a finite collection of points. An approximation of a geodesic metric is induced on the points and used to compute 1-dimensional persistent homology. The right part of the figures represents the longest one or two bars obtained from each of the computations. Starting with a discrete set of points, a multitude of short bars is also generated, but are the artefact of a finite approximation rather than topologically significant features. By the stability theorem their lifespans decrease (although their numbers increase) as we improve the approximation density.

Let us interpret the results:

1. Our chosen samples are dense enough for the longest bars to detect the shortest 1-dimensional homology bases, which in this case coincide for all coefficients.
2. The bars would ideally be born at 0 and run until⁴⁵ one-third of the lengths of the corresponding homology generators. With increased density the resulting barcode would approach this scenario.
3. The visualizations of approximating points also contain a loop or two: these are obtained by connecting the vertices of the critical triangles by the shortest paths through our points. For bars corresponding to the basis, this gives an approximation of the shortest homology basis. In the spirit of the stability theorem, the finer the approximation by points, the closer approximation of the loops we obtain.

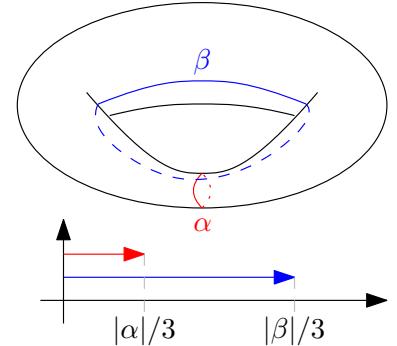


Figure 10.11: The 1-dimensional persistent homology of a torus detects its shortest homology basis.

⁴⁵ Žiga Virk. 1-dimensional intrinsic persistence of geodesic spaces. *Journal of Topology and Analysis*, 12(01):169–207, 2020. doi: [10.1142/S1793525319500444](https://doi.org/10.1142/S1793525319500444)

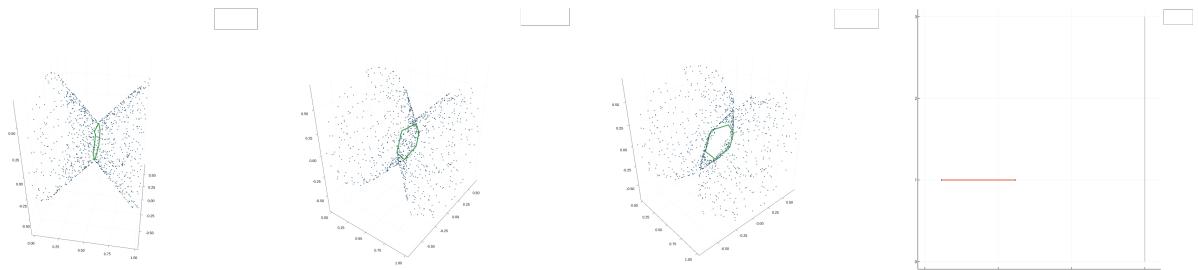


Figure 10.12: The longest bar of 1-dimensional persistent homology.

4. Going beyond the basis, we see that the next bar in Figure 10.13 detected a hole in our approximating points. The lifespan of this bar would decrease towards zero with ever better approximations.

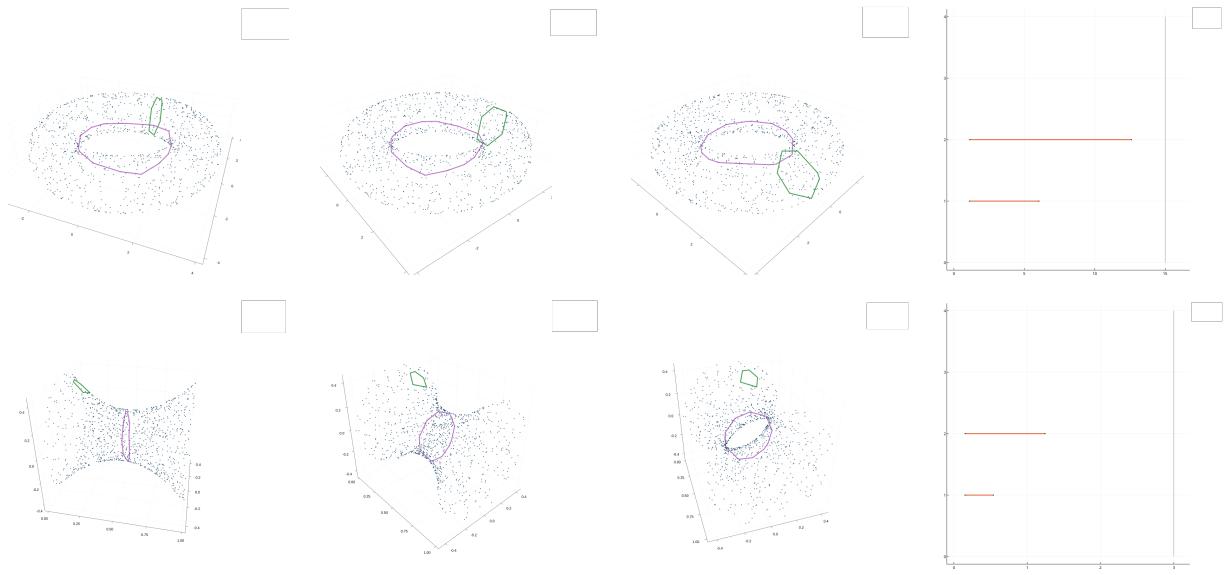


Figure 10.13: The longest two bars of 1-dimensional persistent homology.

Stability demonstrated

Figure 10.14 contains four approximations of a circle by discrete sets and the corresponding persistence diagrams arising from the Rips filtration. The stability theorem states that the induced persistence diagrams should be close to each other, and the figure demonstrates this is indeed the case.

A few comments on the diagrams:

1. There is a point $(0, 3)$ representing $(0, \infty)$ indicating the one persistent component.
2. The main dominating feature is the very persistent point terminating at around 1.5. It represents the 1-dimensional hole, i.e., the homology of S^1 . Its precise coordinates tell us more about the geometry of the sample.
 - The birth is between 0 and .5. The precise birth depends on the edges of the Rips complex going “around the circle”⁴⁶. The maximal gap needed for such circumcision is the birth time. We can see that such a gap is smallest in the upper-right case resulting in early birth. On the other hand, the gap is largest in the upper-left case⁴⁷ and results in a later birth.
 - The terminal scale of this feature is the minimal diameter of an “almost equilateral triangle” reaching “around the circle”.
3. The other points on persistence diagrams are of low persistence and appear⁴⁸ as an artefact of discretization.

⁴⁶ ...and thus generating the 1-cycle.

⁴⁷ The gap of this sample appears in the upper-right part

⁴⁸ For example, as the Rips complex on n points is a discrete collection of n points at small scales, each such diagram will have n many points indicating persistent 0-dimensional homology.

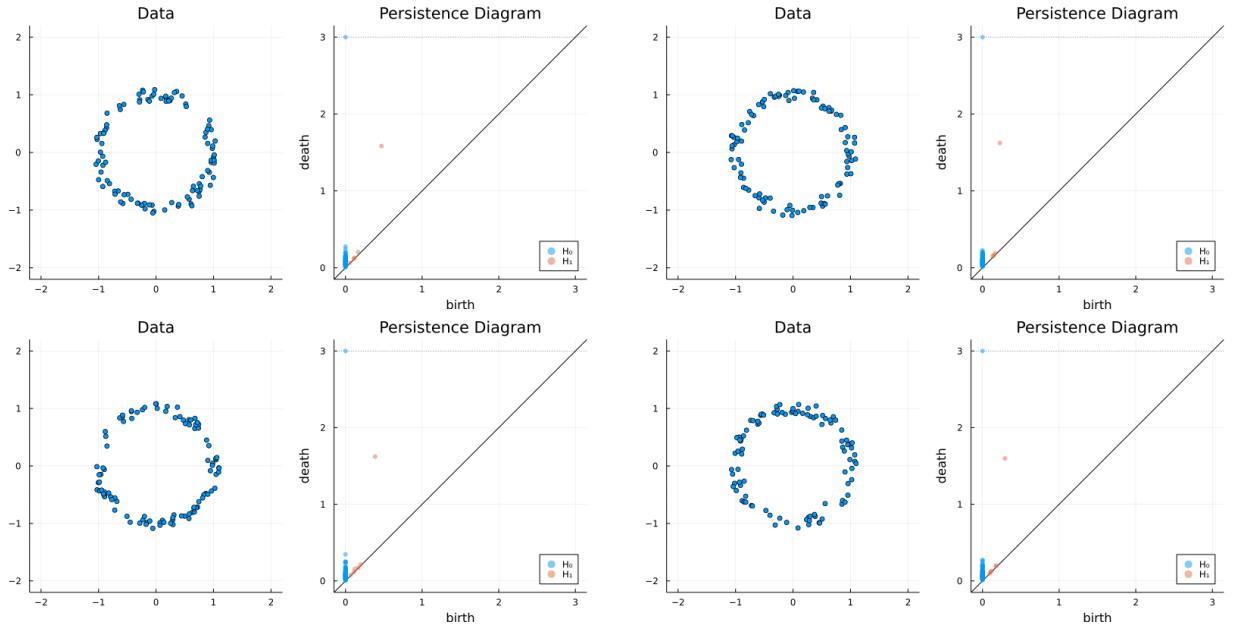


Figure 10.14: Four approximations of a circle and the corresponding persistence diagrams.

Spheres

We next present examples approximating spheres. In Figure 10.15 we present persistence diagrams via Rips complexes of a sample of 100 points on unit spheres: S^2 (on left) and S^3 (on right); using Euclidean (top) or geodesic distance (bottom).

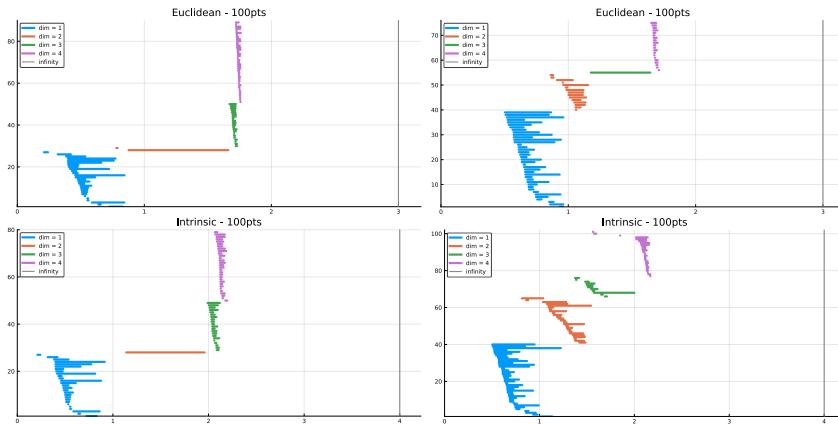


Figure 10.15: Persistence diagrams via Rips complexes of samples of one hundred points sampled from unit S^2 (on left) and S^3 (on right) using Euclidean or intrinsic (geodesic) distance.

As the only non-trivial homology (except for dimension 0) of S^2 is $H_2(S^2; \mathbb{F}) \cong \mathbb{F}$, we expect a long persistent line in that dimension, which is indeed the case. In fact, the long 2-dimensional bar clearly indicates that in both cases the most prominent homology is of rank 1 in dimension 2. The same holds for S^3 although a denser sample would

In persistence diagrams in Figure 10.15 there are a lot of short bars. Some of these are artefacts of discretization, other indicate a more complex structure of persistent homology reaching beyond the interpretation of the size of homology representatives of the underlying space. Interpreting such bars is a very active research topic.

have made the same observation easier in the geodesic case.

To demonstrate the improvement induced by larger density we present in Figure 10.17 a sequence of diagrams with increasing density. The underlying space is a cut-off unit sphere, i.e., a 2-dimensional sphere with a cap above the parallel of circumference approximately 1.5 removed, see Figure 10.16. We take a sample of 100, 200, and 400 points, generate a geodesic distance, and generate persistence diagrams via Rips filtrations.

Here is what we would expect from resulting persistence diagrams:

1. The cut-off sphere is contractible for small scales and thus initial clutter of 1-dimensional bars should be decreasing in size as we increase density.
2. At a certain scale an offset of the cut-off sphere will fill in the top and create a void and hence⁴⁹ a 2-dimensional homology in the Čech complex. As Rips complexes are interleaved with Čech complexes⁵⁰, we might hope that the same 2-dimensional bar might appear in our case. That is indeed the case and the mentioned long bar grows with increasing density of the sample.
3. As is frequently the case, there appear multitudes of short bars we choose to ignore at this point.

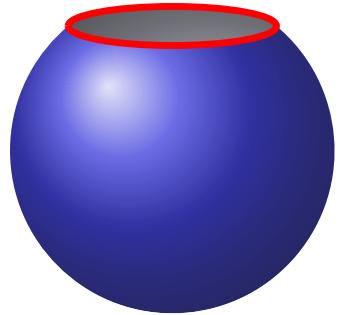


Figure 10.16: A cut-off sphere as an underlying space leading to diagrams in Figure 10.17.

⁴⁹ See Example 10.1.1.

⁵⁰ ...and hence the persistence diagrams are not too different.

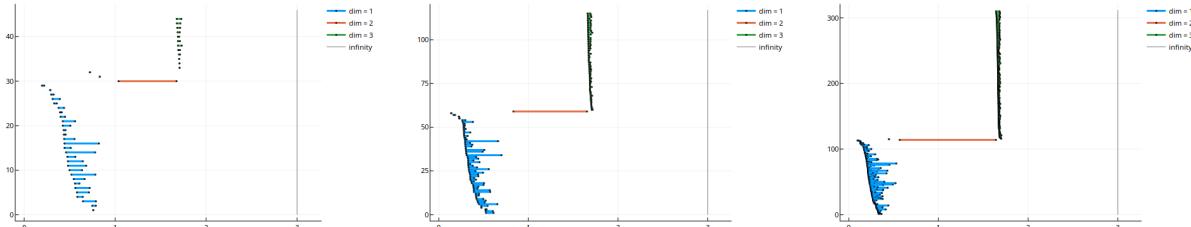


Figure 10.17: Four persistence diagrams via Rips complexes of a cut-off sphere, based on samples of 100, 200, and 400 points.

De-noising a function

Suppose we are given an approximation of a function f in the form of a discrete set of equally spaced measurements. We can connect the resulting points by edges and obtain a graph G representing our measurements, see the left side of Figure 10.18. Suppose we want to extract the global behaviour of f as in the center of Figure 10.18 by removing the local oscillations we consider to be noise. A way to achieve it would be to construct the sublevel⁵¹ filtration of the simplicial complex G and choose the threshold ε for the noise level. We would then draw the corresponding 0-dimensional persistence diagram and ignore the points in the shaded ε -neighborhood⁵² of the diagonal, see the right side of Figure 10.18. As a result we obtain

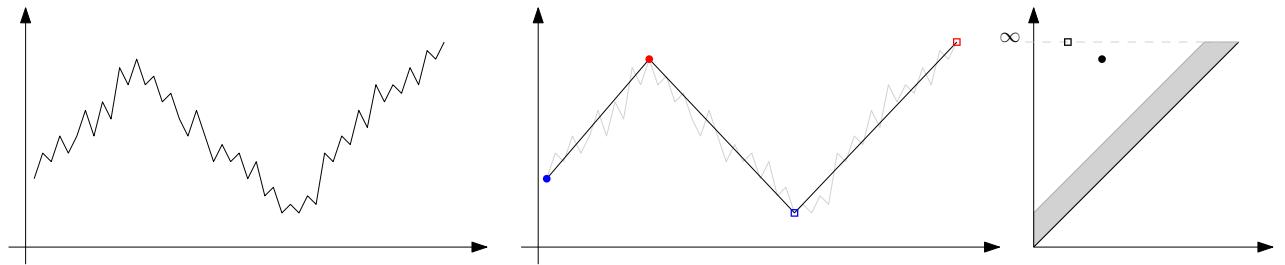
⁵¹ A vertex of G appears at the function value it represents. An edge of G appears as soon as both of its vertices appear.

⁵² Each local minimum in our approximation except for \bullet and \square generates a point in this neighborhood.

two prominent points \bullet, \square in the persistence diagram. The de-noised function presented in the center of Figure 10.18 can now be obtained by connecting the critical simplices corresponding to these points:

1. Blue birth simplices get connected to the higher endpoint of the red terminal simplex.
2. The only exception is \square , which is not a terminal simplex, but gets added as the highest point in the graph in order to finalize⁵³ our approximation by connecting it to \square .

 By adjusting the threshold ε we can adjust the level of details we want to preserve.



⁵³ The component represented by \square can not get terminated as a homology element.

10.5 Concluding remarks

Recap (highlights) of this chapter

- continuous filtrations;
- persistence modules;
- interleaving;
- Stability theorem;

Background and applications

There are three different proofs of the stability theorem in the literature. The initial proof was combinatorial⁵⁴ and did not include the isometry theorem or interleavings but rather just the continuity of persistence diagrams. It motivated a more algebraically oriented proof using interleavings⁵⁵. The third and most direct proof uses explicit matching⁵⁶.

The existence of a decomposition of a persistence module into indecomposable parts is a particular case of a standard approach referred to as the Krull Remak Schmidt principle. The fact that the indecomposable are precisely the interval modules is a special case of the Gabriel's theorem. The fact that the indecomposable parts of

Figure 10.18: A noisy function, its reconstruction and the corresponding persistence diagram. The shaded region contains a multitude of points we choose to ignore in our reconstruction.

⁵⁴ David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007. doi: [10.1007/s00454-006-1276-5](https://doi.org/10.1007/s00454-006-1276-5)

⁵⁵ Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the Twenty-Fifth Annual Symposium on Computational Geometry*, SCG '09, pages 237–246, New York, NY, USA, 2009. ACM. doi: [10.1145/1542362.1542407](https://doi.org/10.1145/1542362.1542407)

⁵⁶ Ulrich Bauer and Michael Lesnick. Induced matchings of barcodes and the algebraic stability of persistence. In *Proceedings of the Thirtieth Annual Symposium on Computational Geometry*, SOCG'14, pages 355–364, New York, NY, USA, 2014. ACM. doi: [10.1145/2582112.2582168](https://doi.org/10.1145/2582112.2582168)

multi-parameter persistent homology are not as simple as the interval modules is the major obstacle to exhaustive applications of multi-parameter persistence.

For a longer treatment of stability see a book⁵⁷. There is also a recent treatment simplifying some of these ideas⁵⁸.

The material presented up to this point represents the core ideas and properties of persistent homology. From this point on the topics diverge, with some of the major motivations being:

- Theoretical treatment: persistent homology represents a parameterized version of homology and as such there are many ways to explore the structure further, either by generalizing the framework⁵⁹, determining what geometrical properties it encodes⁶⁰ and expanding the ideas into other theoretical contexts.
- Practical treatment, mostly associated with data analysis: in this context persistent homology is often viewed as a stable shape descriptor. As a result considerable effort is being invested to incorporate persistent homology into the flow of data analysis, either by adjusting it to specific data types⁶¹, establishing meaningful probabilistic⁶² and statistical⁶³ analysis, and to extract relevant features.
- Computational treatment: the aim of this context is to optimize the computational resources required to obtain persistence (or at least a part of it) by developing faster algorithms often incorporating various shortcuts (for example, the twist⁶⁴) or additional structure⁶⁵. Currently available software for computing persistent homology includes (but is not restricted to) Ripser and related Ripserer.jl, Ripser.py, and Cubical Ripser, Dionysus, PHAT, GUDHI, javaPlex, Perseus, Eirene, etc.

This list of topics and software is by no means exhaustive.

Appendix: From the interleaving distance to the bottleneck distance

In this appendix we will provide an explanation that leads to the bottleneck distance by determining the interleaving distance between pairs of interval modules.

Case 1: distance between an interval module and the zero persistence module. The situation is presented in Figures 10.19 and 10.20. For $0 \leq p < q$ let us discuss the interleaving of the interval module $\mathbb{F}_{p,q}$ (the bold portion of the figures) and the trivial persistence module (the grey portion below).

⁵⁷ Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Y. Oudot. The Structure and Stability of Persistence Modules. *Springer Briefs in Mathematics*. Springer, 2016. doi: [10.1007/978-3-319-42545-0](https://doi.org/10.1007/978-3-319-42545-0)

⁵⁸ Primož Škraba and Katharine Turner. Notes on an elementary proof for the stability of persistence diagrams, arXiv: 2103.10723, 2021

⁵⁹ For example, with zig-zag persistence, multi-parameter persistence, introduction of new constructions of complexes (Witness, selective Rips complex), etc.

⁶⁰ For example it encodes, at least to some degree, shortest homology bases, intrinsic volumes, geometric shapes, curvatures, dimension, etc.

⁶¹ Besides point-clouds, these include time series, high-dimensional data, dynamical systems, sensor networks, etc.

⁶² It turns out there are significant phase transitions in persistent homology of random processes, etc.

⁶³ This is typically done by mapping persistence diagrams into Hilbert space via any of the multitude maps available, for example persistence landscapes, persistence silhouette, persistence silhouette, persistence images, etc.

⁶⁴ Chao Chen and Michael Kerber. Persistent homology computation with a twist. In *Proceedings 27th European Workshop on Computational Geometry*, volume 11, pages 197–200, 2011

⁶⁵ For example, when computing one-dimensional persistence of geodesic spaces.

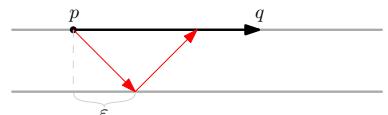


Figure 10.19: interval module $\mathbb{F}_{p,q}$ is not ϵ -interleaved with the trivial interval if $\epsilon < (q - p)/2$.

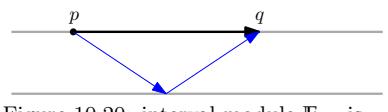


Figure 10.20: interval module $\mathbb{F}_{p,q}$ is ϵ -interleaved with the trivial interval if $\epsilon \geq (q - p)/2$.

- If the interleaving parameter was $\varepsilon < (q - p)/2$ as in Figure 10.19, the composition of the red diagonal maps:
 - is the trivial map as it factors through the trivial vector space below;
 - should have been identity on \mathbb{F} by the interleaving condition, as its domain and target are in $[p, q]$.

These two observations contradict each other hence the interleaving parameter is at least $(q - p)/2$.

- For $\varepsilon = (p - q)/2$ though, the composition of the diagonal maps increases the scale parameter by $p - q$, and any such structure map of $\mathbb{F}_{p,q}$ is trivial, hence the ε -interleaving consisting of trivial maps exists, see Figure 10.20.

We conclude that the interleaving distance is $\varepsilon = (p - q)/2$.

Case 2: general case. From Case 1 we conclude that for $0 \leq p' < q'$ the following holds: If $\mathbb{F}_{p,q}$ is ε -interleaved with $\mathbb{F}_{p',q'}$ for $\varepsilon < (p - q)/2$, then $[p', q') \supset [p + \varepsilon, q - \varepsilon]$ see Figure 10.21. By symmetry the opposite also holds: $[p, q) \supset [p' + \varepsilon, q' - \varepsilon]$. It is easy to see these conditions are also sufficient. For minimal ε for which these two conditions hold we obtain the ε -interleaving by mapping the designated generator of $\mathbb{F}_{p,q}$ to the designated generator of $\mathbb{F}_{p',q'}$ whenever possible, with other maps being trivial, see Figure 10.23. It should be apparent from Figure 10.22 that the ε in question is $\max\{|p - q|, |p' - q'|\}$.

We conclude that $\mathbb{F}_{p,q}$ and $\mathbb{F}_{p',q'}$ are $\max\{|p - q|, |p' - q'|\}$ interleaved. However, Case 1 demonstrates that $\mathbb{F}_{p,q}$ and $\mathbb{F}_{p',q'}$ are also $\max\{(p - q)/2, (p' - q')/2\}$ -interleaved by the trivial maps, the interleaving distance between $\mathbb{F}_{p,q}$ and $\mathbb{F}_{p',q'}$ is

$$\min\{\max\{|p - q|, |p' - q'|\}, \max\{(p - q)/2, (p' - q')/2\}\}.$$

We now interpret the obtained distance in the context of persistence diagrams representing interval modules. First note that

$$(p - q)/2 = d_\infty \left((p, q), \left(\frac{p+q}{2}, \frac{p+q}{2} \right) \right)$$

is the d_∞ distance between (p, q) and the diagonal Δ .

Case 1. The interleaving distance between $\mathbb{F}_{p,q}$ and the zero persistence module is realized by matching point (p, q) to the closest point on the diagonal and computing the resulting d_∞ distance, see Figure 10.24.

Case 2. The distance between $\mathbb{F}_{p,q}$ and $\mathbb{F}_{p',q'}$ is the smaller of the following two:

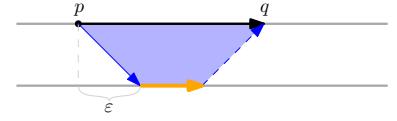


Figure 10.21: ε -interleaving implies the orange part is non-trivial as the interleaving maps in the blue region have to be non-trivial.

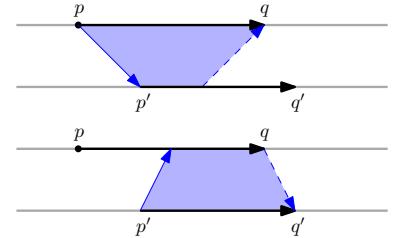


Figure 10.22: Condition of Figure 10.21 induces two shapes. The interleaving distance is the larger ε parameter these shapes induce.

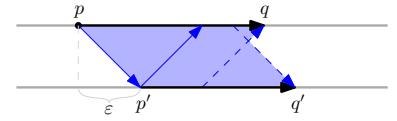


Figure 10.23: The interleaving for $\varepsilon = \max\{|p - q|, |p' - q'|\}$. The non-trivial maps are in the shaded region.

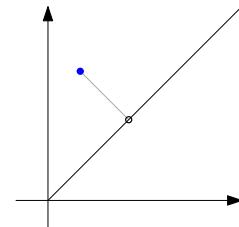


Figure 10.24: Matching a point to Δ .

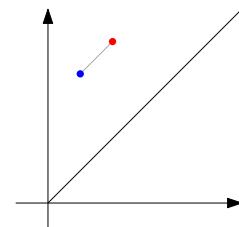


Figure 10.25: Matching two points.

1. Either $\max\{|p - q|, |p' - q'|\} = d_\infty((p, q), (p', q'))$, which is the d_∞ distance between the points, see Figure 10.25.
2. Or $\max\{(p - q)/2, (p' - q')/2\}$ which can be interpreted as follows:
match each of the points to the closest point on the diagonal Δ and
take the maximal d_∞ distance, see Figure 10.26.

We have thus interpreted the interleaving distance between interval modules in the context of persistence modules and obtained the bottleneck distance for diagrams containing at most one point. The crucial ingredients of the interpretation are the matching and d_∞ . Theorem 10.3.2 essentially states that an optimal interleaving between any pair of persistence modules essentially consists of such matchings: match some pairs of interval modules from both persistence modules, and then match the remaining interval modules to Δ . The interleaving distance (and thus the bottleneck distance) corresponds to the matching whose d_∞ -distance of its maximal matching is minimal.

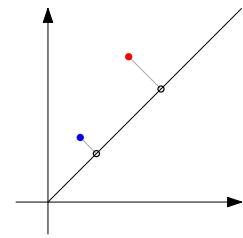


Figure 10.26: Matching each of the two points to Δ .

11

Discrete Morse theory

HOMOLOGY AND PERSISTENT HOMOLOGY detect holes in spaces through the use of algebraic constructions: a simplicial complex generates a chain complex and the resulting homology construction detects holes. However, functions and vector fields also contain information about the topology of the domain. In the smooth setting this information is contained in critical points of functions and zeros of vector fields, a situation which is beautifully described by Morse theory.

In this chapter we will describe discrete Morse theory. As the name suggests we will delve into the world of discrete functions and discrete vector fields defined on simplicial complexes. Our main goal will be to describe how these encode homology, often leading to simplified representations and faster computations than the standard methods.

11.1 Motivation

We first recall the definition of elementary collapses.

Definition 11.1.1. A simplex in a simplicial complex is a **free face** if it is a face of precisely one simplex. This implies that the coface in question is a maximal simplex

Let K be a simplicial complex, $\sigma^{k-1} \subset \tau^k \in K$, and assume σ is a free face in K . A removal $K \rightarrow K \setminus \{\sigma, \tau\}$ is called an **elementary collapse**.

Complex K is **collapsible to a subcomplex** $L \leq K$ if there is a **collapse** (i.e., a sequence of elementary collapses) resulting in the subcomplex L . Complex K is **collapsible** if it is collapsible to a point.

☞ If σ is a free face in a simplicial complex K then its only coface is a maximal simplex in K .

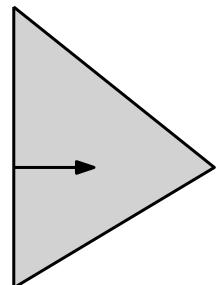


Figure 11.1: An elementary collapse indicated by an arrow from σ into τ .

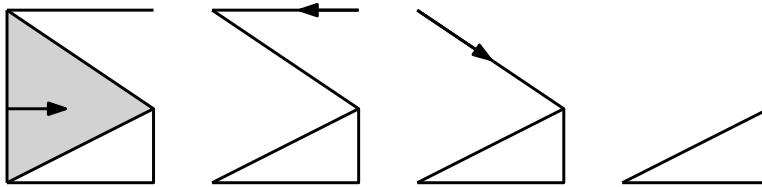
Remark 11.1.2. We have already proved in Lemma 3.4.6 that an elementary collapse results in a homotopically equivalent space. As a result, if a simplicial complex K is collapsible to a subcomplex $L \leq K$

then $L \simeq K$. In particular, each collapsible simplicial complex is contractible. The converse does not hold as there exist contractible simplicial complexes without a free face, for example *Dunce hat* (Figure 11.2) and *Bing's house*.

Given a simplicial complex K it would be of interest to simplify (i.e. collapse) it as much as possible. This would, for example, simplify the computation of homology groups. One would go about such simplification by repeating the following sequence as long as possible: find a free face and perform the corresponding collapse. An example is given in Figure 11.3, where the collapses of the first three steps are indicated by the arrows. One can encode such a collapse by:

- drawing all the arrows¹ indicating collapses, or
- annotate simplices by numbers² so that the countdown-sequence encodes³ the collapsing sequence.

Both of these encodings are demonstrated on the right side of Figure 11.3.



Eventually a collapsing sequence ends when there are no more free faces. At this point we can resort to another trick that will on one hand change the structure of a complex, yet still simplify its description in a way. Choose any simplex, declare it to be a critical simplex, remove it from the complex, and continue with collapsing. In the end we will form a “complex” consisting of critical simplices. The details of the construction will be described throughout this section. At this point we only illustrate a geometric interpretation of this idea in terms of “stretching” simplices.

For our motivational purposes let us continue in Figure 11.4 with the example from Figure 11.3. We are left with a triangle. We choose one of its edges to be a critical edge and continue with collapsing. We can imagine that each collapse stretches the critical edge until, at the end, we are left with two critical simplices: an edge and a point jointly forming a circle. The resulting space is homotopy equivalent to our original simplicial complex of Figure 11.3, has a simple representation, but is not a simplicial complex. However, its homology can be

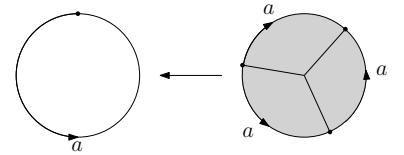


Figure 11.2: *Dunce hat* is obtained by glueing the boundary of a disc along a circle: twice alone one direction and once along the other direction. The obtained space can be triangulated but contains no free face meaning it is not collapsible. However, it turns out to be contractible.

¹ An example of a discrete vector field.

² A rough example of a discrete Morse function.

³ 10 collapses to 9; 8 collapses to 7, etc.

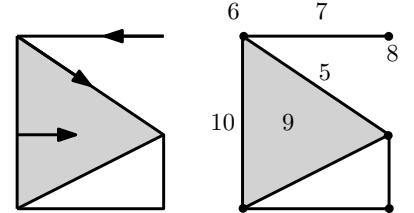


Figure 11.3: A simplification of a simplicial complex using elementary collapses and the encoding of the resulting collapse by arrows (discrete vector field) and annotations (discrete Morse function).

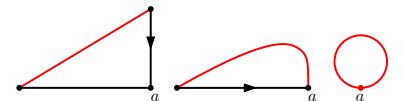


Figure 11.4: Declaring the red edge to be critical, we can collapse the other two edges and obtain a representation of a circle using only two critical “simplices”.

computed in the same way as simplicial homology so in effect, we have transformed the 1-dimensional boundary matrix from 6 columns to 1 column. For another example see Figure 11.5.

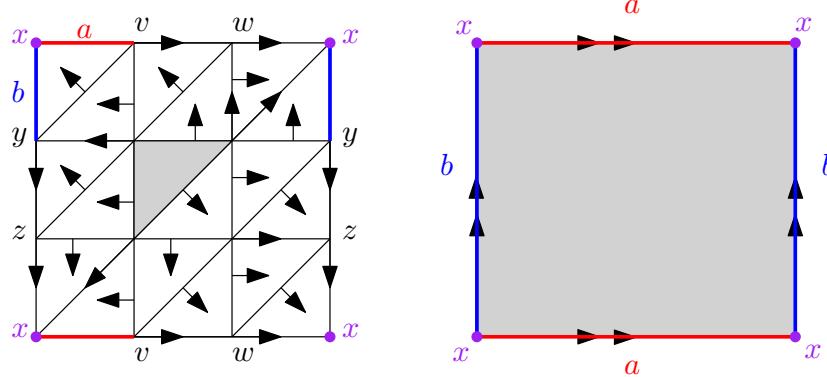


Figure 11.5: Stretching critical simplices of a standard triangulation of the torus (on the left) along the indicated collapses results in a standard representation of a torus as a square with identified sides (on the right).

Critical simplices can be thought of as zeros of the resulting discrete vector field. We have already seen in the hairy ball theorem that there is a connection between zeros of smooth vector fields and topology of the domain.

11.2 Discrete Morse functions and discrete vector fields

We start by defining functions that encode the collapsing sequences and deformations.

Definition 11.2.1. Let K be an abstract simplicial complex. A function $f: K \rightarrow \mathbb{R}$ is a **discrete Morse function** [DMF] if $\forall \sigma^k \in K$:

1. $e_1 = |\{\tau^{k-1} \in K \mid f(\tau) \geq f(\sigma)\}| \leq 1$ and
2. $e_2 = |\{\tau^{k+1} \in K \mid f(\tau) \leq f(\sigma)\}| \leq 1$.

An abstract simplicial complex is a collection of simplices hence a real function defined on it maps each simplex into a real number.

A function $g: K \rightarrow \mathbb{R}$ respects dimension⁴ if for each $\sigma^{k-1} \subset \tau^k \in K$ we have $g(\sigma) < g(\tau)$. Such a function is a DMF. On the other hand, each DMF almost respects dimension in the sense⁵ that for each simplex τ^k at most one exceptional facet and at most one exceptional coface of dimension $k+1$ are allowed. The following proposition demonstrates that the two exceptions cannot occur simultaneously.

Proposition 11.2.2. Given the notation of Definition 11.2.1, either $e_1 = 0$ or $e_2 = 0$.

Proof. Aiming for the contradiction, assume that for $\sigma \in K$ and for vertices $v_1, v_2 \in K^{(0)}$ we have

$$f(\sigma) \geq f(\sigma \cup \{v_1\}) \geq f(\sigma \cup \{v_1, v_2\}). \quad (11.1)$$

But then $\sigma \cup \{v_2\} \in K$ and we have:

⁴ As an example think of $g(\sigma) = \dim(\sigma)$.

⁵ Putting it differently, for each simplex the values of the function strictly decrease by passing to its faces with at most one exception, and the values of the function strictly increases by passing to its cofaces with at most one exception.

1. $f(\sigma \cup \{v_2\}) > f(\sigma)$ as the exceptional coface of σ is $\sigma \cup \{v_1\}$.
2. $f(\sigma \cup \{v_2\}) < f(\sigma \cup \{v_1, v_2\})$ as the exceptional face of $\sigma \cup \{v_1, v_2\}$ is $\sigma \cup \{v_1\}$.

These two conclusions combine into $f(\sigma) < f(\sigma \cup \{v_1, v_2\})$ which contradicts equation 11.1. \square

Proposition 11.2.2 implies that simplices with exceptions form disjoint pairs. We will refer to such pairs as **regular pairs**. A regular pair consists of a simplex τ and its facet σ . It encodes an “arrow” $\sigma \rightarrow \tau$ in the sense of the motivational section and is thus presented as such, see Figure 11.6 for an example. A simplex without any exception⁶ is called⁷ **critical simplex**. Given a DMF on a simplicial complex, each simplex is either critical or contained in a unique regular pair.

Definition 11.2.3. Let K be an abstract simplicial complex. A **discrete vector field** is a disjoint collection of pairs (σ_i, τ_i) of simplices from K such that for each i simplex σ_i is a facet of τ_i . Each pair of a discrete vector field is referred to as an arrow.

The disjointness condition means that each simplex can be the member of at most one pair of a discrete vector field. The collection of regular pairs of a DMF forms⁸ a discrete vector field, see Figure 11.6. A discrete vector field is called a **gradient vector field**⁹ if it is induced by some DMF in this manner. The arrows constituting a discrete vector field will be sometimes referred¹⁰ to as regular pairs.

Gradient vector fields

Definition 11.2.5. Let K be a simplicial complex and $p \in \mathbb{N}$. Given a discrete vector field on K consisting of pairs $\{(\sigma_i, \tau_i)\}_{i \in J}$, a **p -path** is a sequence

$$\sigma_{i_1}^{p-1} \rightarrow \tau_{i_1}^p \geq \sigma_{i_2}^{p-1} \rightarrow \tau_{i_2}^p \geq \sigma_{i_3}^{p-1} \rightarrow \cdots \rightarrow \tau_{i_k}^p \geq \sigma_{i_{k+1}}^{p-1}$$

such that for each j :

- $(\sigma_{i_j}^{p-1}, \tau_{i_j}^p)$ is an arrow in the discrete vector field, and
- $\sigma_{i_j}^{p-1}$ is a facet of $\tau_{i_{j-1}}^p$.

Such a p -path is a **cycle** if $\sigma_1^{p-1} = \sigma_{k+1}^{p-1}$ and $k \geq 1$.
A discrete vector field is **acyclic** if it admits no cycle.

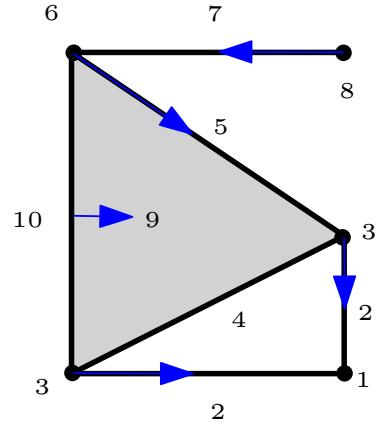


Figure 11.6: An example of a DMF and the resulting discrete vector field in blue.

⁶I.e., for which $e_1 = e_2 = 0$.

⁷A critical simplex is not contained in any regular pair.

Proposition 11.2.4. Let f be a DMF on a simplicial complex K . For each i let n_i denote the number of critical simplices of dimension i . Then $\chi = n_0 - n_1 + n_2 - \dots$

Proof. Removing a regular pair of simplices does not change χ because simplices are of adjacent dimensions. \square

⁸The converse is not true in general as we will explain the the next subsection.

⁹We will be omitting adjective “discrete” when mentioning gradient vector fields.

¹⁰The reason is twofold: to emphasize that the pair is a part of the structure of a discrete vector field, and to stress the analogy with regular pairs of a DMF.

A few observations concerning Definition 11.2.5:

1. A critical simplex can only appear as the last simplex of a p -path in a discrete vector field.
2. Given a DMF f , function values decrease along any p -path in the induced discrete vector field, i.e.:

$$f(\sigma_{i_j}) \geq f(\tau_{i_j}) > f(\sigma_{i_{j+1}}), \quad \forall j.$$

In particular, $f(\sigma_{i_1}) > f(\sigma_{i_m})$ for all $m > 1$.

3. Observation 2. implies that each gradient vector field is acyclic.
The following theorem proves the converse.

Theorem 11.2.6. *Each acyclic discrete vector field on a simplicial complex K is a gradient vector field, i.e. it is induced by some DMF.*

A proof is given in the appendix. As a result we obtain the following theorem.

Theorem 11.2.7. *A discrete vector field is gradient vector field iff it is acyclic.*

We conclude this section by demonstrating how acyclic discrete vector fields encode collapses.

Proposition 11.2.8. *Suppose the critical simplices of an acyclic discrete vector field on K form a subcomplex $L \leq K$. Then there exists a collapse $K \rightarrow L$ and thus $K \simeq L$.*

Proof. We claim there exists a regular pair (σ, τ) such that σ is a free face. Assuming for a moment this claim is true, we can remove pair (σ, τ) by performing an elementary collapse and proceed by using the claim on the resulting complex. Thus the inductive argument and the claim suffice to prove the proposition.

We now turn our attention to proving the claim. Let n denote the maximal dimension of a simplex in $K \setminus L$. There exists an n -path in the discrete vector field. Take a maximal¹¹ such path and let $\sigma \rightarrow \tau$ denote the first regular pair in it. Simplex σ is a free face by the following argument:

- $\sigma \leq \tau$ as the simplices form a regular pair.
- If σ was a facet of another simplex τ' in $K \setminus L$, then the n -simplex τ' would be contained in another regular pair¹² (σ', τ') , which could be used to prolong our n -path. This contradicts the maximality of the chosen n -path.

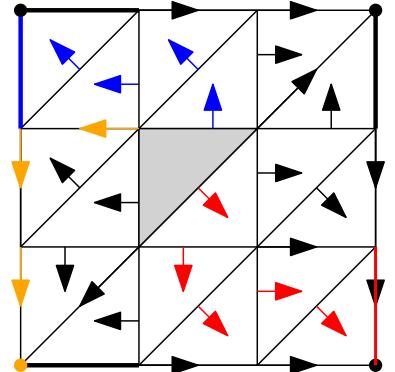


Figure 11.7: A 2-path in blue ending in a critical edge, a 2-path in red ending in a non-critical edge, and a 1-path in orange ending in a critical vertex.

Different DMFs on a simplicial complex K may induce the same discrete vector field. For example, if f is a DMF, then so are ef and $3f - 5$, and all of them induce the same discrete vector field. Our primary interest in discrete vector fields lies in their encodings of collapses and deformations-simplifications of a simplicial complex. A DMF represents a convenient but not unique way of encoding a discrete vector field.

Corollary 11.2.9. *If an acyclic discrete vector field on K has a single critical simplex, then that simplex is a vertex and K is collapsible.*

Proof. The statement follows directly from Proposition 11.2.8. \square

¹¹ Such a path contains a first regular pair because the discrete vector field is acyclic.

¹² There are no $(n+1)$ -simplices in $K \setminus L$.

- If σ was a facet of another simplex τ' in L , then $\sigma \in L$ as L is a subcomplex, a contradiction.

□

11.3 Morse homology

In this section we will explain the procedure that leads to the computation of homology from a gradient vector field. The geometric idea behind the theory was presented at the beginning of this chapter: collapsing regular pairs to stretch critical simplices, with the resulting space having the same homotopy type as the original simplicial complex but fewer “simplices”. In our treatment we will refrain¹³ from formally defining the resulting space and instead construct the resulting chain complex directly. However, it might still be helpful to keep the geometric idea in mind to help navigate the algebraic construction.

Morse chain complex

For the rest of this section we fix a simplicial complex K , a gradient vector field on K , and an (algebraic) field \mathbb{F} to provide efficient algebraic constructions. For each i let n_i denote the number of critical i -simplices.

Definition 11.3.1. Let $p \in \{0, 1, \dots\}$. A **Morse p -chain** is a formal sum $\sum_{i=1}^{n_p} \lambda_i \sigma_i^p$ with $\lambda_i \in \mathbb{F}$ and σ_i^p being an oriented critical simplex of dimension p in K for each i .

The p -dimensional **Morse chain group** \mathfrak{C}_p is the vector space consisting of all Morse p -chains.

Observe that $\mathfrak{C}_p \cong \mathbb{F}^{n_p}$. In order to obtain a chain complex we also need to define boundary maps. These are based on oriented¹⁴ paths in discrete vector fields.

Definition 11.3.2. Let $p \in \{0, 1, \dots\}$. An **oriented p -path** from an oriented simplex σ_1^{p-1} to an oriented simplex σ_{k+1}^{p-1} is a p -path

$$\sigma_1^{p-1} \rightarrow \tau_1^p \geq \sigma_2^{p-1} \rightarrow \tau_2^p \geq \sigma_3^{p-1} \rightarrow \cdots \rightarrow \tau_k^p \geq \sigma_{k+1}^{p-1}$$

consisting of oriented simplices, such that for each j the orientation induced by τ_j on its facets:

1. matches σ_j , and
2. does not match σ_{j+1} .

¹³ A formal definition of resulting spaces would require a significant amount of additional material from algebraic topology. This would include a formal treatment of CW complexes, i.e., spaces obtained by inductively glueing discs. We have actually mentioned several presentations of such constructions when presenting torus, Klein bottle and projective plane by drawing a square with identifications along the edges, when defining the dunce hat, and in one of the previous appendices in the context of relative homology.

☞ As with the usual homology, multiplying an oriented simplex by -1 changes its orientation.

¹⁴ Paths in a discrete vector field are directed by definition. The adjective “orientable” refers to the fact that the simplices forming the path are oriented in a certain way.

☞ One could say that the simplices τ_j in an oriented p -path are oriented consistently along the path.

Given an oriented critical p -simplex τ , let $\delta(\tau)$ denote the collection of all of its facets with the induced orientation arising from τ . For each¹⁵ oriented critical $(p-1)$ -simplex σ define

$$\alpha_{\tau,\sigma} = \sum_{\sigma' \in \delta(\tau)} |\{ \text{oriented paths from } \sigma' \text{ to } \sigma \}|$$

as the number of oriented p -paths from elements of $\delta(\tau)$ to σ .

Definition 11.3.3. *The boundary map \mathfrak{d} of the Morse chain complex is defined as follows: for each oriented critical p -simplex τ define*

$$\mathfrak{d}_p \tau = \sum_{i=1}^{n_{p-1}} (\alpha_{\tau,\sigma_i} - \alpha_{\tau,-\sigma_i}) \sigma_i,$$

where $\sigma_1, \dots, \sigma_{n_{p-1}}$ are critical $(p-1)$ -simplices with a fixed orientation.

¹⁵ Given an oriented critical $(p-1)$ -simplex σ observe that $\alpha_{\tau,\sigma}$ counts different paths than $\alpha_{\tau,-\sigma}$.

 The oriented paths constituting the boundary map model how arrows stretch the boundary of a p -critical simplex towards critical $(p-1)$ -simplices.

Examples will be provided below when demonstrating the computation of Morse homology, see also Figures 11.8 and 11.9. It turns out that $\mathfrak{d}^2 = 0$.

Definition 11.3.4. *The Morse chain complex is the chain complex defined as*

$$\dots \xrightarrow{\mathfrak{d}} \mathfrak{C}_n \xrightarrow{\mathfrak{d}} \mathfrak{C}_{n-1} \xrightarrow{\mathfrak{d}} \dots \xrightarrow{\mathfrak{d}} \mathfrak{C}_1 \xrightarrow{\mathfrak{d}} \mathfrak{C}_0 \xrightarrow{\mathfrak{d}} 0.$$

Morse homology

We may now define Morse homology as the homology arising from the Morse chain complex.

Definition 11.3.5. *Let $p \in \{0, 1, \dots\}$. The Morse homology of a gradient vector field on K is defined as*

$$\mathfrak{H}_p(K; \mathbb{F}) = \ker \mathfrak{d}_p / \operatorname{Im} \mathfrak{d}_{p+1}.$$

Theorem 11.3.6. *Morse homology is isomorphic to the standard (simplicial) homology:*

$$\mathfrak{H}_p(K; \mathbb{F}) \cong H_p(K; \mathbb{F}).$$

Corollary 11.3.7 (Weak Morse inequalities). *For each p the number of critical p simplices is greater or equal to the corresponding Betti number: $n_p \geq b_p$.*

Example 11.3.8. Given the situation of Figure 11.8 there is one critical edge $\langle b, e \rangle$ and one critical vertex $\langle a \rangle$. Thus $\mathfrak{C}_1 \cong \mathfrak{C}_0 \cong \mathbb{F}$ with the other Morse chain groups being trivial.

Let us determine $\mathfrak{d}\langle b, e \rangle$:

1. $\delta(\langle b, e \rangle) = \{\langle e \rangle, -\langle b \rangle\}.$
2. There is one oriented 1-path from $\langle e \rangle$ to $\langle a \rangle$:

$$\langle e \rangle \rightarrow \langle e, a \rangle \geq \langle a \rangle.$$

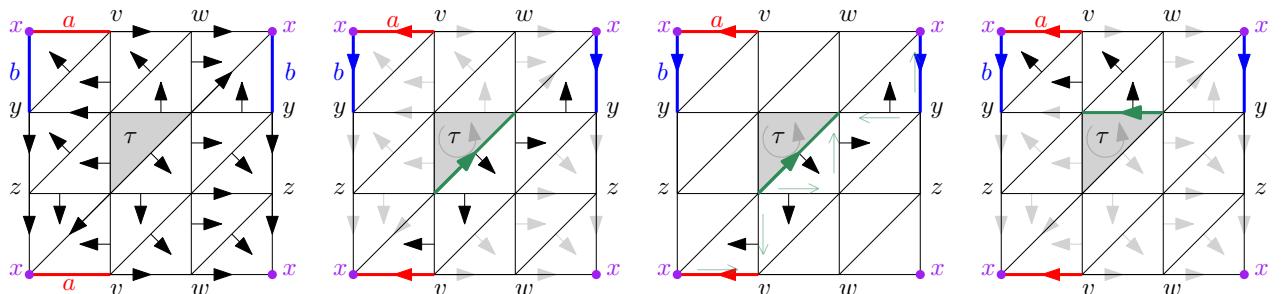
3. In a similar fashion there is one oriented 1-path from $-\langle b \rangle$ to $-\langle a \rangle$.
4. Observations 2. and 3. imply $\alpha_{\langle b, e \rangle, \langle a \rangle} = 1$ and $\alpha_{\langle b, e \rangle, -\langle a \rangle} = 1$.
5. $\mathfrak{d}\langle b, e \rangle = (\alpha_{\langle b, e \rangle, \langle a \rangle} - \alpha_{\langle b, e \rangle, -\langle a \rangle}) \cdot \langle a \rangle = 0$

The resulting Morse chain complex is of the form

$$\cdots \rightarrow 0 \rightarrow \mathbb{F} \xrightarrow{0} \mathbb{F} \rightarrow 0.$$

The resulting homology is trivial in dimensions two and above, and nontrivial below: $\mathfrak{H}_0(K) \cong \mathfrak{H}_1(K) \cong \mathbb{F}$.

Example 11.3.9. Let us compute the Morse homology of a torus.



Given the triangulation and the gradient vector field on a torus presented on the leftmost part of Figure 11.9 we determine the following critical simplices:

- critical vertex x in purple;
- critical edges a (red) and b (blue);
- critical triangle τ .

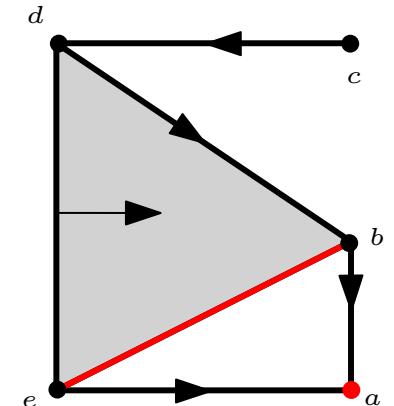


Figure 11.8: A gradient vector field on a simplicial complex. Critical edges are colored in red. There is one oriented 1-path from $\langle e \rangle$ to $\langle a \rangle$ and one oriented 1-path from $\langle b \rangle$ to $\langle a \rangle$.

Figure 11.9: A triangulation of a torus, a gradient vector field, and paths generating the Morse boundary from Example 11.3.8.

We orient the critical simplices according to visualizations in the other parts in Figure 11.9. The resulting Morse chain complex is of the form

$$\cdots \rightarrow 0 \rightarrow \mathbb{F} \rightarrow \mathbb{F}^2 \rightarrow \mathbb{F} \rightarrow 0.$$

We next determine the Morse boundary of τ . Only two simplices of $\delta(\tau)$ are the starting simplices of oriented 2-chains ending in a critical edge:

- From the diagonal edge of $\delta(\tau)$ there are two oriented 2-paths¹⁶ to critical edges¹⁷ $-\langle a \rangle$ and $-\langle b \rangle$.
- From the top edge of $\delta(\tau)$ there are two oriented 2-paths¹⁸ to critical edges¹⁹ $\langle a \rangle$ and $\langle b \rangle$.
- There are oriented 2-paths starting in the vertical edge of $\delta(\tau)$ but none of them ends in a critical edge.

Combining these three cases we conclude

$$\mathfrak{d}(\tau) = -\langle a \rangle - \langle b \rangle + \langle a \rangle + \langle b \rangle = 0.$$

In a similar way we conclude that $\mathfrak{d}a = \mathfrak{d}b = 0$. The resulting Morse chain complex is of the form

$$\cdots \rightarrow 0 \rightarrow \mathbb{F} \xrightarrow{0} \mathbb{F}^2 \xrightarrow{0} \mathbb{F} \rightarrow 0.$$

The resulting homology is trivial in dimensions three and above, and nontrivial below: $\mathfrak{H}_0(K) \cong \mathfrak{H}_2(K) \cong \mathbb{F}$ and $\mathfrak{H}_1(K) \cong \mathbb{F}^2$.

Generating DMFs and gradient vector fields

Using discrete Morse theory depends on the ability to generate DMFs and gradient vector fields with as few critical simplices as possible. The weak Morse inequalities show that the lower bounds for the numbers of critical simplices are Betti numbers. A DMF on a simplicial complex is **perfect** if the number of critical p -simplices coincides with the p^{th} Betti number. In terms of the numbers of critical simplices, perfect DMFs are optimal DMFs. Not every simplicial complex admits a perfect DMF: an example is the Dunce hat.

There is a simple algorithm to generate a perfect DMF on a graph. For each component generate a gradient vector field as follows:

- Find a spanning tree.
- Choose a critical vertex.
- Define a gradient vector field pointing towards the critical vertex along the spanning tree, see Figure 11.11.

¹⁶ Drawn in black on the center-left part of Figure 11.9.

¹⁷ The center-right part of Figure 11.9 contains opaque green arrows indicating the orientations of the edges contained in the oriented 2-paths. The terminal edges of the oriented 2-paths are $-\langle a \rangle$ and $-\langle b \rangle$.

¹⁸ Drawn in black on the rightmost part of Figure 11.9.

¹⁹ The two oriented 2-paths differ only in the last simplex.

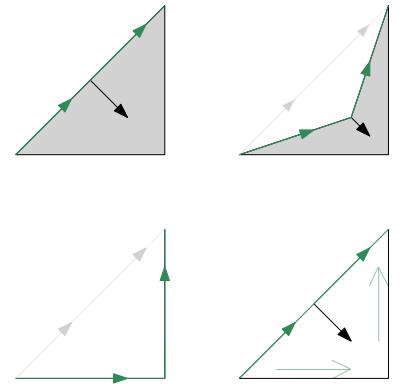


Figure 11.10: In this figure we provide a geometric justification for the way the orientation carries forward through oriented 2-paths. The bottom-right part is a snapshot from the center-right part Figure 11.9 indicating how the orientation of the diagonal edge carries on through the arrow to the other two edges of the triangle. The first three parts of this figure indicate how such an orientation on the two edges is obtained by deforming the oriented diagonal edge along the arrow of a discrete vector field.

The mentioned construction can be generalized to higher dimensional simplicial complexes: keep adding arrows while making sure that the acyclicity condition is preserved. However, better results are typically obtained through more elaborate designs.

11.4 Concluding remarks

Recap (highlights) of this chapter

- Discrete Morse functions
- Gradient vector fields
- Morse homology

Background and applications

Smooth Morse theory was developed²⁰ by Marston Morse in the first part of the twentieth century. Amongst its results it relates critical points and gradient flows of a generic function on a manifold to the homology of a manifold. Its discrete version²¹ has been introduced at the turn of the millennium. The past two decades saw a considerable development of discrete Morse theory from various directions, including computational aspects, developing analogies between discrete and smooth results, etc. For a textbook presenting smooth and discrete point of view see a book²², for recent applications see a book²³. An algorithm for generating a discrete Morse function is given in a paper²⁴.

An echo of the smooth Morse theory is the hairy ball theorem: the topology of a domain is connected to zeros of vector fields and thus to extrema of functions. In a similar way, an echo of the discrete Morse theory is our proof of the Euler-Poincaré formula, where we essentially only counted the maxima of the x -coordinate function. Theorem 11.2.6 is a discrete variant of the assigning of a potential function to a vector field.

Generalized discrete Morse theories can be used to prove²⁵ that Čech complexes collapse onto alpha complexes in Euclidean spaces. Several computer programs use discrete Morse theory to a different degree to assist²⁶ with computations of homology. The theory can also be used as a preprocessing tool or a framework within which to analyze discrete functions.

A proof of Theorem 11.2.6

We first introduce some preliminary notions. Given a simplicial complex K the **Hasse diagram** of K is a directed graph defined as

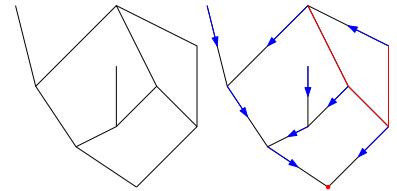


Figure 11.11: A graph (left) and a blue spanning tree (right) with a gradient vector field pointing to the chosen critical vertex (red). The edges not contained in the tree (red) are the critical edges.

²⁰ Marston Morse. The Calculus of Variations in the Large. *American Mathematical Society Colloquium Publication*. Vol. 18. New York, 1934

²¹ Robin Forman. A user's guide to discrete morse theory. *Sém. Lothar. Combin.*, 48, 12 2001

²² Kevin P. Knudson. Morse Theory: Smooth and Discrete. *World scientific*, 2015. doi: [10.1142/9360](https://doi.org/10.1142/9360)

²³ Tamal K. Dey and Yusu Wang. Computational Topology for Data Analysis. *Cambridge: Cambridge University Press*, 2022. doi: [10.1017/9781009099950](https://doi.org/10.1017/9781009099950)

²⁴ Henry King, Kevin Knudson, and Neža Mramor. Generating Discrete Morse Functions from Point Data. *Experiment. Math.* 14 (4) 435 – 444, 2005

²⁵ Ulrich Bauer and Herbert Edelsbrunner. The morse theory of Čech and Delaunay complexes. *Transactions of the American Mathematical Society*, 369:1, 06 2016. doi: [10.1090/tran/6991](https://doi.org/10.1090/tran/6991)

²⁶ For example, using simplification using emergent pairs in Ripser. On the other hand Perseus is actually based on a discrete Morse theory.

follows:

1. The nodes are the simplices of K ;
2. Directed edges correspond to pairs (simplex, a facet). In particular, each n -simplex is the source of $n + 1$ directed edges.

An example is provided in Figure 11.12. The directed edges in the graph represent the containment of a facet.

Given an empty discrete vector field on a simplicial complex K , the directed edges also represent the direction of descent of the corresponding DMF. In this trivial case, all the simplices are critical, there are no exceptions and the DMF respects dimension. An obvious choice of a DMF in this case is the dimension function of a simplex. For illustrative purposes let us discuss how we could obtain such a function from the Hasse diagram in an inductive manner:

- Assign the smallest value, say 0, to all minimal nodes of a directed graph;
- Assign the second smallest value, say 1, to all nodes whose lower set²⁷ has already been enumerated;
- Proceed by induction: In step number n assign the n^{th} smallest value, say $n - 1$, to all nodes whose lower set has already been enumerated;

This inductive construction of function works for any acyclic directed graph and will be used in our eventual proof.

Given a non-trivial discrete vector field on a simplicial complex K we define a modified Hasse diagram of K by reverting the direction of the directed edges corresponding to the regular pairs, see Figure 11.12 for an example. A modified Hasse diagram encodes the sufficient conditions on a DMF to generate the initial discrete vector field. The above inductive procedure on such a diagram will produce a required DMF iff the diagram itself is acyclic as a directed graph.

Lemma 11.4.1. *The modified Hasse diagram of an acyclic discrete vector field is acyclic.*

Proof. Since each simplex can be a member of at most one regular pair in a discrete vector field, a cycle in the modified Hasse diagram H cannot contain consecutive directed edges corresponding to regular pairs. As directed edges of H either end in a simplex of dimension 1 higher (in the case of regular pairs) or lower (in the case of edges encoding the facet relation) than the dimension of the initial simplex, any cycle in H has to be an alternating concatenation of these two types. As a result, a cycle in H corresponds to a p -cycle in the initial

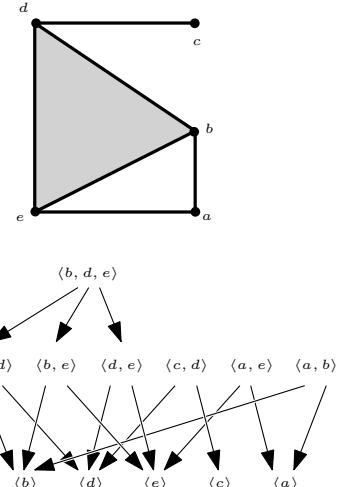


Figure 11.12: A simplicial complex and its Hasse diagram. Hasse diagrams are typically drawn in levels corresponding to the dimensions of simplices.

²⁷ The lower set of a node σ is the collection of all nodes which appear as the target of a directed edge starting at σ .

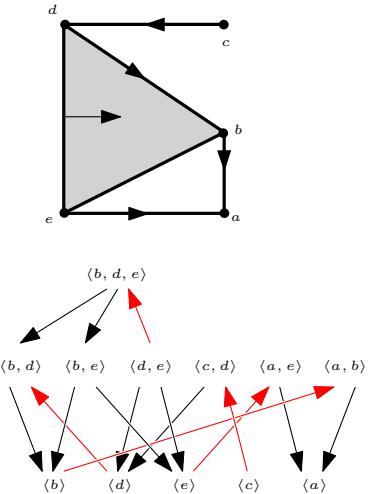


Figure 11.13: The modified Hasse diagram, the reverted edges are red.

discrete vector field, which is non-existent by the main assumption. \square

A proof of Theorem 11.2.6. Given an acyclic discrete vector field, the corresponding modified Hasse diagram is acyclic by Lemma 11.4.1. Thus the inductive procedure above results in a suitable DMF, see Figure 11.14. \square

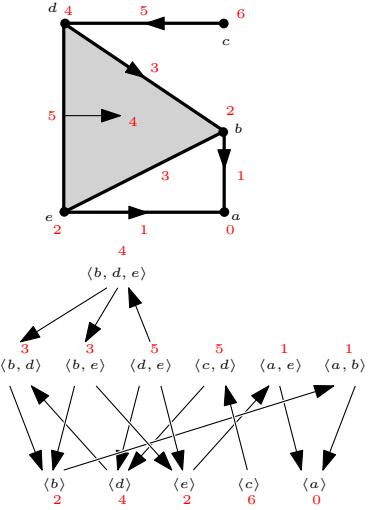


Figure 11.14: The modified Hasse diagram and the DMF (in red) constructed by the inductive procedure.

Index

- Abelian group, 81
- abstract simplex, 37
- abstract simplicial complex, 37
- affine combination, 33
- affine hull, 33
- affine independence, 34
- Alexander duality, 102
- alpha complex, 65
- annotation, 138
- ball, 12
- barcode, 121, 125
- barycentric coordinates, 21, 34
- barycentric subdivision, 20
- basis, 79
- Betti number, 89
- birth simplex, 129
- body, 36
- bottleneck distance, 146
- boundary, 89
- boundary map, 87
- boundary of a manifold, 48
- boundary point, 48
- Brouwer fixed point, 111
- Brouwer fixed point theorem, 111
- chain, 86
- chain complex, 88
- chain group, 86
- classification algorithm for surfaces, 55
- close manifold, 48
- closed surface, 48
- coface, 35
- collapse, 157
- column echelon form, 93
- combinatorial manifold, 49
- component, 17
- connected space, 17
- connected sum, 52
- consistent orientation, 51
- continuous filtration, 138
- continuous map, 13
- contractible space, 16
- convex combination, 21
- convex hull, 21
- convex set, 21
- critical scale, 139, 141
- critical simplex, 160
- cubical complex, 106
- cycle, 89
- Čech complex, 61
- Čech filtration, 62
- Dante, 57
- Delaunay triangulation, 25
- diameter, 59
- dimension of a simplex, 35
- dimension of a simplicial complex, 36
- direct sum, 142
- disc, 14
- discrete Morse function, 159
- discrete vector field, 160
- distance, 11

- Dowker duality, 73
- elder rule, 122
- elementary collapse, 44, 96, 157
- elementary divisors, 104
- Euler characteristic, 22, 41, 97
- Euler-Poincaré formula, 22
- exact sequence, 114
- face, 35
- facet, 35
- field, 75
- fields of remainders, 76
- filtration, 123
- filtration function, 138
- free face, 157
- full simplex, 98
- functoriality, 111
- fundamental class, 99
- geodesic metric, 12
- geometric realization, 38
- geometric simplex, 35
- geometric simplicial complex, 35
- gradient vector field, 160
- Gromov-Hausdorff distance, 144
- group, 81
- hairy ball theorem, 112
- Hausdorff distance, 143
- homeomorphism, 14
- homology group, 89
- homology representative, 94
- homomorphism, 82
- homotopic maps, 15
- homotopy, 15
- homotopy equivalence, 16
- homotopy equivalent spaces, 16
- image, 79
- incremental expansion, 96
- induced maps, 110
- induced orientation, 51
- Inscribed angle theorem, 28
- interior of a manifold, 48
- interior point, 48
- interleaving, 68
- interleaving distance, 140, 143
- interval module, 142
- isometry, 13
- isomorphism, 79, 82, 142
- Jaccard distance, 12
- Jung's theorem, 62
- kernel, 79
- Klein bottle, 48, 100
- line sweep, 23
- link, 37
- locally Delaunay, 26
- manifold, 48
- mapper, 66
- matching distance, 146
- MaxMin edge, 27
- Mayer-Vietoris exact sequence, 115
- metric, 11
- metric space, 11
- MiniBall algorithm, 70
- Moebius band, 16
- Morse homology, 163
- multiplicity, 126
- nerve, 62
- nerve theorem, 64
- offsets, 139
- orientable surface, 51
- oriented simplex, 50
- oriented triangulation, 51
- p-path, 160
- partial matching, 146
- path, 13

- path component, 17
- path connected space, 17
- perfect DMF, 165
- persistence diagram, 126
- persistence module, 141
- persistent Betti numbers, 123
- persistent homology, 123
- pivot, 128
- projective plane, 48
- quotient, 76
- rank, 83
- regular pairs, 160
- relative homology, 119
- representing cycles, 94
- retraction, 111
- Rips complex, 59
- Rips filtration, 60
- row echelon form, 92
- simplicial approximation, 43
- simplicial map, 42
- skeleton, 36
- Smith normal form, 94, 104
- sphere, 14
- star, 37
- subcomplex, 36
- subdivision, 36
- sublevel filtration, 138
- surface, 48
- tangent vector field, 112
- terminal simplex, 129
- torus, 39
- triangulation, 19, 36
- vector space, 78
- Vietoris complex, 73
- Voronoi diagram, 25
- Voronoi region, 24
- Wasserstein distances, 147
- zig-zag lemma, 117

Bibliography

Yuliy Baryshnikov and Robert Ghrist. Target enumeration via Euler characteristic integrals. *SIAM Journal on Applied Mathematics*, 70(3):825–844, 2009. doi: [10.1137/070687293](https://doi.org/10.1137/070687293).

Ulrich Bauer and Herbert Edelsbrunner. The morse theory of Čech and Delaunay complexes. *Transactions of the American Mathematical Society*, 369:1, 06 2016. doi: [10.1090/tran/6991](https://doi.org/10.1090/tran/6991).

Ulrich Bauer and Michael Lesnick. Induced matchings of barcodes and the algebraic stability of persistence. In *Proceedings of the Thirtieth Annual Symposium on Computational Geometry*, SOCG’14, pages 355–364, New York, NY, USA, 2014. ACM. doi: [10.1145/2582112.2582168](https://doi.org/10.1145/2582112.2582168).

Raoul Bott and Loring W. Tu. Differential Forms in Algebraic Topology. *Springer New York*, New York, NY, 1982. doi: [10.1007/978-1-4757-3951-0](https://doi.org/10.1007/978-1-4757-3951-0).

Constantin Caratheodory. Über den Variabilitätsbereich der Koef- fizienten von Potenzreihen, die gegebene Werte nicht annehmen. *Math. Ann.* 64, no. 1, 95–115, 1907. doi: [10.1007/BF01449883](https://doi.org/10.1007/BF01449883).

Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the Twenty-Fifth Annual Symposium on Computational Geometry*, SCG ’09, pages 237–246, New York, NY, USA, 2009. ACM. doi: [10.1145/1542362.1542407](https://doi.org/10.1145/1542362.1542407).

Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Y. Oudot. The Structure and Stability of Persistence Modules. *Springer Briefs in Mathematics*. Springer, 2016. doi: [10.1007/978-3-319-42545-0](https://doi.org/10.1007/978-3-319-42545-0).

Chao Chen and Michael Kerber. Persistent homology computation with a twist. In *Proceedings 27th European Workshop on Computational Geometry*, volume 11, pages 197–200, 2011.

David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007. doi: [10.1007/s00454-006-1276-5](https://doi.org/10.1007/s00454-006-1276-5).

Mark de Berg, Marc van Kreveld, Mark Overmars, and Otfried Schwarzkopf. Computational Geometry: Algorithms and Applications. *Springer-Verlag*, second edition, 2000. doi: [10.1007/978-3-540-77974-2](https://doi.org/10.1007/978-3-540-77974-2).

Tamal K. Dey and Yusu Wang. Computational Topology for Data Analysis. *Cambridge: Cambridge University Press*, 2022. doi: [10.1017/9781009099950](https://doi.org/10.1017/9781009099950).

Clifford H. Dowker. Homology groups of relations. *Annals of Mathematics*, 56(1):84–95, 1952. doi: [10.2307/1969768](https://doi.org/10.2307/1969768).

David S. Dummit and Richard M. Foote. Abstract algebra. *Wiley*, 3rd edition, 2004.

Herbert Edelsbrunner and John Harer. Computational Topology: An Introduction. *Applied Mathematics. American Mathematical Society*, 2010. doi: [10.1090/mbk/069](https://doi.org/10.1090/mbk/069).

Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28(4):511–533, 2002. doi: [10.1007/s00454-002-2885-2](https://doi.org/10.1007/s00454-002-2885-2).

Robin Forman. A user’s guide to discrete morse theory. *Sém. Lothar. Combin.*, 48, 12 2001.

Jean Gallier and Dianna Xu. A Guide to the Classification Theorem for Compact Surfaces. *Springer Berlin Heidelberg*, 2013. doi: [10.1007/978-3-642-34364-3](https://doi.org/10.1007/978-3-642-34364-3).

Allen Hatcher. Algebraic topology. *Cambridge Univ. Press, Cambridge*, 2000.

Henry King, Kevin Knudson, and Neža Mramor. Generating Discrete Morse Functions from Point Data. *Experiment. Math.* 14 (4) 435 – 444, 2005.

L. Christine Kinsey. Topology of Surfaces. *Springer New York*, 1993. doi: [10.1007/978-1-4612-0899-0](https://doi.org/10.1007/978-1-4612-0899-0).

Kevin P. Knudson. Morse Theory: Smooth and Discrete. *World scientific*, 2015. doi: [10.1142/9360](https://doi.org/10.1142/9360).

Marston Morse. The Calculus of Variations in the Large. *American Mathematical Society Colloquium Publication*. Vol. 18. New York, 1934.

James Munkres. Elements of Algebraic Topology. *Perseus Books*, 1984. doi: [10.1201/9780429493911](https://doi.org/10.1201/9780429493911).

James R. Munkres. Topology. *Prentice Hall, Inc*, 2nd ed edition, 2000.

Steve Y. Oudot. Persistence Theory: From Quiver Representations to Data Analysis. *Number 209 in Mathematical Surveys and Monographs. American Mathematical Society*, 2015. doi: [10.1090/surv/209](https://doi.org/10.1090/surv/209).

Matt Parker. Humble pi: a comedy of maths errors. *Allen Lane*, 2019.

A. Roy, R. A. I. Haque, A. J. Mitra, M. Dutta Choudhury, S. Tarafdar, and T. Dutta. Understanding flow features in drying droplets via Euler characteristic surfaces—a topological tool. *Physics of Fluids*, 32(12):123310, 2020. doi: [10.1063/5.0026807](https://doi.org/10.1063/5.0026807).

Gurjeet Singh, Facundo Memoli, and Gunnar Carlsson. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007. doi: [10.2312/SPBG/SPBG07/091-100](https://doi.org/10.2312/SPBG/SPBG07/091-100).

Žiga Virk. 1-dimensional intrinsic persistence of geodesic spaces. *Journal of Topology and Analysis*, 12(01):169–207, 2020. doi: [10.1142/S1793525319500444](https://doi.org/10.1142/S1793525319500444).

Žiga Virk. Rips complexes as nerves and a functorial Dowker-nerve diagram. *Mediterranean Journal of Mathematics*, 18(2):58, 2021. doi: [10.1007/s00009-021-01699-4](https://doi.org/10.1007/s00009-021-01699-4).

Emo Welzl. Smallest enclosing disks (balls and ellipsoids). In Hermann Maurer, editor, *New Results and New Trends in Computer Science*, pages 359–370, Berlin, Heidelberg, 1991. Springer Berlin Heidelberg.

Primož Škraba and Katharine Turner. Notes on an elementary proof for the stability of persistence diagrams, arXiv: 2103.10723, 2021.