# Computational Topology

Rote, Günter; Vegter, Gert

*Published in:*
Effective Computational Geometry for Curves and Surfaces



*Publication date:*
2006

# 7

# Computational Topology: An Introduction

Günter Rote and Gert Vegter*

## 7.1 Introduction

Topology studies point sets and their invariants under continuous deformations, invariants such as the number of connected components, holes, tunnels, or cavities. Metric properties such as the position of a point, the distance between points, or the curvature of a surface, are irrelevant to topology. Computational topology deals with the complexity of topological problems, and with the design of efficient algorithms for their solution, in case these problems are tractable. These algorithms can deal only with spaces and maps that have a finite representation. To this end we restrict ourselves to simplicial complexes and maps. In particular we study algebraic invariants of topological spaces like Euler characteristics and Betti numbers, which are in general easier to compute than topological invariants.

Many computational problems in topology are algorithmically undecidable. The mathematical literature of the 20th century contains many (beautiful) topological algorithms, usually reducing to decision procedures, in many cases with exponential-time complexity. The quest for efficient algorithms for topological problems has started rather recently. The overviews by Dey, Edelsbrunner and Guha [116], Edelsbrunner [129], Vegter [325], and the book by Zomorodian [346] provide further background on this fascinating area.

This chapter provides a tutorial introduction to computational aspects of algebraic topology. It introduces the language of combinatorial topology, relevant for a rigorous mathematical description of geometric objects like meshes, arrangements and subdivisions appearing in other chapters of this book, and in the computational geometry literature in general.

Computational methods are emphasized, so the main topological objects are simplicial complexes, combinatorial surfaces and submanifolds of some Euclidean space. These objects are introduced in Sect. 7.2. Here we also introduce the notions of homotopy and isotopy, which also feature in other

* Chapter coordinator

parts of this book, like Chapter 5. Most of the computational techniques are introduced in Sect. 7.3. Topological invariants, like Betti numbers and Euler characteristic, are introduced and methods for computing such invariants are presented. Morse theory plays an important role in many recent advances in computational geometry and topology. See, e.g., Sect. 5.5.2. This theory is introduced in Sect. 7.4.

Given our focus on computational aspects, topological invariants like Betti numbers are defined using simplicial homology, even though a more advanced study of deeper mathematical aspects of algebraic topology could better be based on singular homology, introduced in most modern textbooks on algebraic topology. Other topological invariants, like homotopy groups, are harder to compute in general; These are not discussed in this chapter.

The chapter is far from a complete overview of computational algebraic topology, and it does not discuss recent advances in this field. However, reading this chapter paves the way for studying recent books and papers on computational topology. Topological algorithms are currently being used in applied fields, like image processing and scattered data interpolation. Most of these applications use some of the tools presented in this chapter.

## 7.2 Simplicial complexes

*Topological spaces.*

In this chapter a *topological space* $X$ (or space, for short) is a subset of some Euclidean space $\mathbb{R}^d$, endowed with the induced topology of $\mathbb{R}^d$. In particular, an $\varepsilon$-neighborhood ($\varepsilon > 0$) of a point $x$ in $X$ is the set of all points in $X$ within Euclidean distance $\varepsilon$ from $x$. A subset $O$ of $X$ is *open* if every point of $O$ contains an $\varepsilon$-neighborhood contained in $O$, for some $\varepsilon > 0$. A subset of $X$ is *closed* if its complement in $X$ is open. The *interior* of a set $X$ is the set of all points having an $\varepsilon$-neighborhood contained in $X$, for some $\varepsilon > 0$. The *closure* of a subset $X$ of $\mathbb{R}^d$ is the set of points $x$ in $\mathbb{R}^d$ every $\varepsilon$-neigborhood of which has non-empty intersection with $X$. The *boundary* of a subset $X$ is the set of points in the closure of $X$ that are not interior points of $X$. In particular, every $\varepsilon$-neighborhood of a point in the boundary of $X$ has non-empty intersection with both $X$ and the complement of $X$. See [26, Sect. 2.1] for a more complete introduction of the basic concepts and properties of point set topology.

The space $\mathbb{R}^d$ is called the *ambient space* of $X$. Examples of topological spaces are:

1. The interval $[0, 1]$ in $\mathbb{R}$;
2. The open unit $d$-ball: $\mathbb{B}^d = \{(x_1, \ldots, x_d) \in \mathbb{R}^d \mid x_1^2 + \cdots + x_d^2 < 1\}$;
3. The closed unit $d$-ball: $\overline{\mathbb{B}}^d = \{(x_1, \ldots, x_d) \in \mathbb{R}^d \mid x_1^2 + \cdots + x_d^2 \leq 1\}$ (the closure of $\mathbb{B}^d$);

4. The unit $d$-sphere $\mathbb{S}^d = \{(x_1, \ldots, x_{d+1}) \in \mathbb{R}^{d+1} \mid x_1^2 + \cdots + x_{d+1}^2 = 1\}$ (the boundary of the $(d+1)$-ball);
5. A *$d$-simplex*, i.e., the convex hull of $d+1$ affinely independent points in some Euclidean space (obviously, the dimension of the Euclidean space cannot be smaller than $d$). The number $d$ is called the dimension of the simplex. Fig. 7.1 shows simplices of dimensions up to and including three.

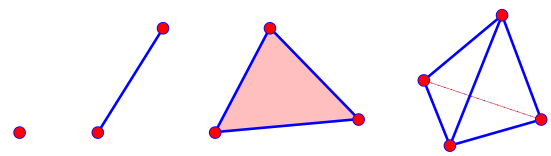


**Fig. 7.1.** Simplices of dimension zero, one, two and three.

*Homeomorphisms.*

A *homeomorphism* is a 1–1 map $h \colon X \to Y$ from a space $X$ to a space $Y$ with a continuous inverse. (In this chapter a *map* is always continuous by definition.) In this case we say that $X$ is *homeomorphic to* $Y$, or, simply, that $X$ and $Y$ are *homeomorphic*.

1. The unit $d$-sphere is homeomorphic to the subset $\Sigma$ of $\mathbb{R}^m$ defined by $\Sigma = \{(x_1, \ldots, x_{d+1}, 0, \ldots, 0) \in \mathbb{R}^m \mid x_1^2 + \cdots + x_{d+1}^2 = 1\}$ $(m > d)$. Indeed, the map $h \colon \mathbb{S}^d \to \Sigma$, defined by $h(x_1, \ldots, x_{d+1}) = (x_1, \ldots, x_{d+1}, 0, \ldots, 0)$, is a homeomorphism. Loosely speaking, the ambient space does not matter from a topological point of view.
2. The map $h \colon \mathbb{R}^k \to \mathbb{R}^m$, $m > k$, defined by
$$h(x_1, \ldots, x_k) = (x_1, \ldots, x_k, 0, \ldots, 0),$$
is *not* a homeomorphism.
3. Any invertible affine map between two Euclidean spaces (of necessarily equal dimension) is a homeomorphism.
4. Any two $d$-simplices are homeomorphic. (If the simplices lie in the same ambient space of dimension $d-1$, there is a unique invertible affine map sending the vertices of the first simplex to the vertices of the second simplex. For other, possibly unequal dimensions of the ambient space one can construct an invertible affine map between the affine hulls of the simplices.)
5. The boundary of a $d$-simplex is homeomorphic to the unit $d$-sphere. (Consider a $d$-simplex in $\mathbb{R}^{d+1}$. The projection of its boundary from a fixed point in its interior onto its circumscribed $d$-sphere is a homeomorphism. See Fig. 7.2. The circumscribed $d$-sphere is homeomorphic to the unit $d$-sphere.)

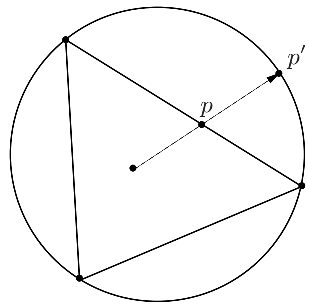



**Fig. 7.2.** The point $p$ on the boundary of a 3-simplex is mapped onto the point $p'$ on the 2-sphere. This mapping defines a homeomorphism between the 2-simplex and the 2-sphere.

*Simplices.*

Consider a $k$-simplex $\sigma$, which is the convex hull of a set $A$ of $k+1$ independent points $a_0, \ldots, a_k$ in some Euclidean space $\mathbb{R}^d$ (so $d \geq k$). $A$ is said to *span* the simplex $\sigma$. A simplex spanned by a subset $A'$ of $A$ is called a *face* of $\sigma$. If $\tau$ is a face of $\sigma$ we write $\tau \preceq \sigma$. The face is *proper* if $\emptyset \neq A' \neq A$. The *dimension* of the face is $|A'|-1$. A 0-dimensional face is called a *vertex*, a 1-dimensional face is called an *edge*. An *orientation* of $\sigma$ is induced by an ordering of its vertices, denoted by $\langle a_0 \cdots a_k \rangle$, as follows: For any permutation $\pi$ of $0, \ldots, k$, the orientation $\langle a_{\pi(0)} \cdots a_{\pi(k)} \rangle$ is equal to $(-1)^{\text{sign}(\pi)} \langle a_0 \cdots a_k \rangle$, where $\text{sign}(\pi)$ is the number of transpositions of $\pi$ (so each simplex has two distinct orientations). A simplex together with a specific choice of orientation is called an *oriented simplex*. If $\tau$ is a $(k-1)$-dimensional face of $\sigma$, obtained by omitting the vertex $a_i$, then the *induced orientation* on $\tau$ is $(-1)^i \langle a_0 \cdots \hat{a}_i \cdots a_k \rangle$, where the hat indicates omission of $a_i$.

*Simplicial complexes.*

A *simplicial complex* $K$ is a finite set of simplices in some Euclidean space $\mathbb{R}^m$, such that (i) if $\sigma$ is a simplex of $K$ and $\tau$ is a face of $\sigma$, then $\tau$ is a simplex of $K$, and (ii) if $\sigma$ and $\tau$ are simplices of $K$, then $\sigma \cap \tau$ is either empty or a common face of $\sigma$ and $\tau$. The *dimension* of $K$ is the maximum of the dimensions of its simplices. The *underlying space* of $K$, denoted by $|K|$, is the union of all simplices of $K$, endowed with the subspace topology of $\mathbb{R}^m$. The *i-skeleton* of $K$, denoted by $K^i$, is the union of all simplices of $K$ of dimension at most $i$. A *subcomplex* $L$ of $K$ is a subset of $K$ that is a simplicial complex. A *triangulation* of a topological space $X$ is a pair $(K, h)$, where $K$ is a simplicial complex and $h$ is a homeomorphism from the underlying space $|K|$ to $X$. The *Euler characteristic* of a simplicial $d$-complex $K$, denoted by

$\chi(K)$, is the number $\sum_{i=0}^{d}(-1)^i\alpha_i$, where $\alpha_i$ is the number of $i$-simplices of $K$. Examples of simplicial complexes are:

1. A *graph* is a 1-dimensional simplicial complex (think of a graph as being embedded in $\mathbb{R}^3$). The complete graph with $n$ vertices is the 1-skeleton of an $(n-1)$-simplex.
2. The Delaunay triangulation of a set of points in general position in $\mathbb{R}^d$ is a simplicial complex.

*Combinatorial surfaces.*

A *Combinatorial closed surface* is a finite two-dimensional simplicial complex in which each edge (1-simplex) is incident with two triangles (2-simplices), and the set of triangles incident to a vertex can be cyclically ordered $t_0, t_1, \ldots, t_{k-1}$ so that $t_i$ has exactly one edge in common with $t_{i+1 \bmod k}$, and these are the only common edges. Stillwell [319, page 69 ff] contains historical background and the basic theorem on the classification combinatorial surfaces.

*Homotopy and Isotopy: Continuous Deformations.*

Homotopy is a fundamental topological concept that describes equivalence between curves, surfaces, or more general topological subspaces within a given topological space, up to "continuous deformations".

Technically, homotopy is defined between two maps $g, h\colon X \to Y$ from a space $X$ into a space $Y$. The maps $g$ and $h$ are *homotopic* if there is a continuous map

$$f\colon X \times [0,1] \to Y$$

such that $f(x,0) = g(x)$ and $f(x,1) = h(x)$ for all $x \in X$. The map $f$ is then called a homotopy between $g$ and $h$. It is easy to see that homotopy is an equivalence relation, since a homotopy can be "inverted" and two homotopies can be "concatenated".

When $g$ and $h$ are two curves in $Y = \mathbb{R}^n$ defined over the same interval $X = [a,b]$, the homotopy $f$ defines, for each "time" $t$, $0 \le t \le 1$, a curve $f(\cdot,t)\colon [a,b] \to \mathbb{R}^n$ that interpolates smoothly between $f(\cdot,0) = g$ and $f(\cdot,1) = h$.[1]

To define homotopy for two surfaces or more general spaces $S$ and $T$, we start with the identity map on $S$ and deform it into a homeomorphism from $S$ to $T$. Two topological subspaces $S, T \subseteq X$ are called *homotopic* if there is a continuous mapping

$$\gamma\colon S \times [0,1] \to X$$

such that $\gamma(\cdot,0)$ is the identity map on $S$ and $\gamma(\cdot,1)$ is a homeomorphism from $S$ to $T$.

[1] In the case of curves with the same endpoints $g(a) = h(a)$ and $g(b) = h(b)$, one usually requires also that these endpoints remain fixed during the deformation: $f(a,t) = g(a)$ and $f(b,t) = g(b)$ for all $t$.

By the requirement that we have a homeomorphism at time $t = 1$, one can see that this definition is symmetric in $S$ and $T$. Note that we do not require $\gamma(\cdot, t)$ to be a homeomorphism at all times $t$. Thus, a clockwise cycle and a counterclockwise cycle in the plane are homotopic. In fact, all closed curves in the plane are homotopic: every cycle can be contracted into a point (which is a special case of a closed curve). A connected topological space with this property is called *simply connected*.

Examples of spaces which are not simply connected are a plane with a point removed, or a (solid or hollow) torus. For example, on the hollow torus in Fig. 7.3, the closed curve in the figure is not homotopic to its inverse.

If we require that $\gamma(\cdot, t)$ is a homeomorphism at all times during the deformation we arrive the stronger concept of *isotopy*. For example, the smooth closed curves without self-intersections in the plane fall into two isotopy classes, according to their orientation (clockwise or counterclockwise). Isotopy is usually what is meant when speaking about a "topologically correct" approximation of a given surface, as discussed in Sect. 5.1, where the stronger concept of *ambient isotopy* is also defined (Definition 1, p. 181).

A map $f\colon X \to Y$ is a *homotopy equivalence* if there is a map $g\colon Y \to X$ such that the composed maps $gf$ and $fg$ are homotopy equivalent to the identity map (on $X$ and $Y$, respectively). The map $g$ is a homotopy inverse of $f$. The spaces $X$ and $Y$ are called *homotopy equivalent*. A space is *contractible* if it is homotopy equivalent to a point.

1. The unit ball in a Euclidean space is contractible. Let $f\colon \{0\} \to \mathbb{B}^d$ be the inclusion map. The constant map $g\colon \mathbb{B}^d \to \{0\}$ is a homotopy inverse of $f$. To see this, observe that the map $fg$ is the identity, and $gf$ is homotopic to the identity map on $\mathbb{B}^d$, the homotopy being the map $F\colon \mathbb{B}^d \times [0,1] \to \mathbb{B}^d$ defined by $F(x,t) = tx$.
2. The solid torus is homotopy equivalent to the circle. More generally, the cartesian product of a topological space $X$ and a contractible space is homotopy equivalent to $X$.
3. A punctured $d$-dimensional Euclidean space $\mathbb{R}^d \setminus \{0\}$ is homotopy equivalent to a $(d-1)$-sphere.

Note that homotopy equivalent spaces need not be homeomorphic. However, such spaces share important topological properties, like having the same Betti numbers (to be introduced in the next section). Section 6.2.3 (p. 248) describes how this concept is applied in surface reconstruction.

## 7.3 Simplicial homology

*A calculus of closed loops.*

Intuitively, it is clear that the sphere and the torus have different shapes in the sense that these surfaces are not homeomorphic. A formal proof of this observation could be based on the Jordan curve theorem: take a simple closed curve on the torus that does not disconnect the torus. Such curves, the complement of which is connected, do exist, as can be seen from Fig. 7.3. If there exists a homeomorphism from the torus to the sphere, the image of the curve on the torus would be a simple closed curve on the sphere. By the Jordan curve theorem, the complement of this curve is disconnected. Since connectedness is preserved by homeomorphisms, the complements of the curves on the torus and the sphere are not homeomorphic. This contradiction proves that the torus and the sphere are not homeomorphic.

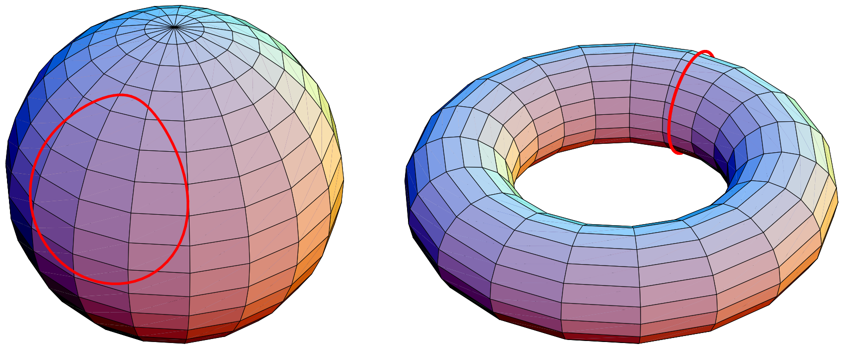


**Fig. 7.3.** Every simple closed curve on the sphere disconnects. Not every closed curve on the torus disconnects.

This proof seems rather ad hoc: it only proves that the sphere is not homeomorphic to a closed surface with holes, but it cannot be used to show that a surface with more than one hole is not homeomorphic to the torus. Homology theory provides a systematic way to generalize the argument above to more general spaces.

In this chapter we present basic concepts and properties of *simplicial homology theory*, closely related to simplicial complexes and suitable for computational purposes. An alternative, more abstract approach is followed in the context of *singular homology theory.* This theory is more powerful when proving general results like topological invariance of homology spaces. Since we focus on basic computational techniques we will not discuss this theory here, but refer the reader to standard textbooks on algebraic topology, like [200]. The equivalence of Simplicial and Singular Homology is proven in [200, Sect. 2.1].

*Chain spaces and simplicial homology.*

Let $K$ be a finite simplicial complex. In this chapter, an *simplicial $k$-chain* is a formal sum of the form $\sum_j a_j\sigma_j$ over the oriented $k$-simplices $\sigma_j$ in $K$, with coefficients $a_j$ in the field $\mathbb{Q}$ of rational numbers. In other words, it can be regarded as a rational vector whose entries are indexed by the oriented

$k$-simplices of $K$. Furthermore, by definition, $-\sigma = (-1)\sigma$ is the simplex obtained from $\sigma$ by reversing its orientation. With the obvious definition for addition and multiplication by scalars (i.e., rational numbers), the set of all simplicial $k$-chains forms a vector space $C_k(K, \mathbb{Q})$, called the *vector space of simplicial $k$-chains* of $K$. The dimension of this vector space is equal to the number of $k$-simplices of $K$. Therefore, the Euler characteristic of a $d$-dimensional simplicial complex $K$ can be expressed as an alternating sum of dimensions of the spaces of $k$-chains:

$$\chi(K) = \sum_{i=0}^{d} (-1)^i \dim C_k(K, \mathbb{Q}). \tag{7.1}$$

The *boundary operator* $\partial_k \colon C_k(K, \mathbb{Q}) \to C_{k-1}(K, \mathbb{Q})$ is defined as follows. For a single $k$-simplex $\sigma = \langle v_{i_0} \cdots v_{i_k} \rangle$, $k > 0$, let

$$\partial_k \sigma = \sum_{h=0}^{k} (-1)^h \langle v_{i_0} \cdots \hat{v}_{i_h} \cdots v_{i_k} \rangle,$$

and then let $\partial_k$ be extended linearly, viz., $\partial_k(\sum_j a_j \sigma_j) = \sum_j a_j \partial_k \sigma_j$. For consistency we define $C_{-1}(K, \mathbb{Q}) = 0$, and we let $\partial_0 : C_0(K, \mathbb{Q}) \to C_{-1}(K, \mathbb{Q})$ be the zero-map. The boundary operator is a linear map between vector spaces. It is easy to check that it verifies the relation $\partial_k \partial_{k+1} = 0$.

*Example: One-homologous chains.*

In the simplicial complex of Fig. 7.4 we consider the 2-chain $\gamma = \langle v_1 v_4 v_2 \rangle + \langle v_2 v_4 v_5 \rangle + \langle v_2 v_5 v_3 \rangle + \langle v_3 v_5 v_6 \rangle + \langle v_1 v_3 v_6 \rangle + \langle v_1 v_6 v_4 \rangle$. Then $\partial_2 \gamma = \alpha - \beta$, where $\alpha = \langle v_4 v_5 \rangle + \langle v_5 v_6 \rangle - \langle v_4 v_6 \rangle$ and $\beta = \langle v_1 v_2 \rangle + \langle v_2 v_3 \rangle - \langle v_1 v_3 \rangle$. Since $\partial_1 \alpha = 0$ and $\partial_1 \beta = 0$, it follows that $\partial_1 \partial_2 \gamma = 0$.

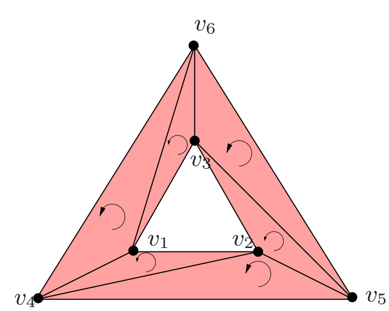



**Fig. 7.4.** One- and two-chains in an annulus.

The vector space $Z_k(K, \mathbb{Q}) = \ker \partial_k$ is called the vector space of *simplicial $k$-cycles*. The vector space $B_k(K, \mathbb{Q}) = \operatorname{im} \partial_{k+1}$ is called vector space of simplicial $k$-*boundaries*. Since the boundary of a boundary is 0, $B_k(K,\mathbb{Q})$ is a subspace of $Z_k(K,\mathbb{Q})$. The quotient vector space $H_k(K,\mathbb{Q}) = Z_k(K,\mathbb{Q})/B_k(K,\mathbb{Q})$ is the $k$-*th homology vector space* of $K$. In particular, two $k$-cycles $\alpha$ and $\beta$ are $k$-*homologous* if their difference is a $k$-boundary, i.e., if there is a $k+1$-chain $\gamma$ such that $\alpha - \beta = \partial_{k+1}\gamma$. The homology class of $\alpha \in Z_k(K,\mathbb{Q})$ is denoted by $[\alpha]$. The $k$-*th Betti number* of the simplicial complex $K$, denoted by $\beta_k(K,\mathbb{Q})$, is the dimension of $H_k(K,\mathbb{Q})$. In particular:

$$\beta_k(K,\mathbb{Q}) = \dim Z_k(K,\mathbb{Q}) - \dim B_k(K,\mathbb{Q}). \tag{7.2}$$

**Remark.** In this chapter, the coefficients of simplicial chains are rational numbers. One usually takes these coefficients in a ring, like the set of integers. In that case one obtains homology *groups* in stead of homology vector spaces. Then, the Betti numbers are the ranks of these groups.

*Example: Zero-homology of a connected simplicial complex.*

Consider the connected simplicial complex $K$ of Fig. 7.5. The 0-chains $\alpha = \langle v_6\rangle$ and $\beta = \langle v_2\rangle$ are 0-homologous since their difference is the boundary of the 1-chain $\gamma = -\langle v_1v_2\rangle + \langle v_1v_4\rangle + \langle v_4v_6\rangle$, since $\partial_1\gamma = -(\langle v_2\rangle - \langle v_1\rangle) + (\langle v_4\rangle - \langle v_1\rangle) + (\langle v_6\rangle - \langle v_4\rangle) = \alpha - \beta$. In the same way one shows that every 0-chain of

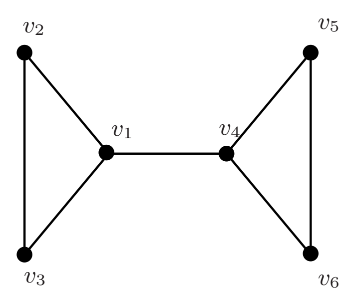



**Fig. 7.5.** Zero-homology of a graph.

the form $\langle v_i\rangle$, $1 \le i \le 6$, is homologous to $\alpha$. This implies that every 0-chain of $K$ is of the form $c\langle\alpha\rangle$, for some $c \in \mathbb{Q}$. Hence: $H_0(K,\mathbb{Q}) = \mathbb{Q}$. It is not hard to generalize this property to all *connected* simplicial complexes: if $K$ is a finite connected simplicial complex, then $H_0(K,\mathbb{Q}) = \mathbb{Q}$.

*Example: One-homologous chains.*

The boundary chains of the annulus in Fig. 7.4 are one-homologous. Indeed, the difference of the boundary chains $\alpha = \langle v_4v_5\rangle + \langle v_5v_6\rangle - \langle v_4v_6\rangle$ and $\beta = \langle v_1v_2\rangle + \langle v_2v_3\rangle - \langle v_1v_3\rangle$ is the boundary of the 2-chain $\gamma = \langle v_1v_4v_2\rangle + \langle v_2v_4v_5\rangle + \langle v_2v_5v_3\rangle + \langle v_3v_5v_6\rangle + \langle v_1v_3v_6\rangle + \langle v_1v_6v_4\rangle$.

*Betti numbers.*

We present a few examples, demonstrating the computation of Betti numbers directly from the definition.

*1. Connected simplicial complex.* If $K$ is a connected simplicial complex, then $\beta_0(K, \mathbb{Q}) = 1$. In fact, we already did the example on 0-homologous chains of a connected simplicial complex $K$, proving that $H_0(K, \mathbb{Q}) = \mathbb{Q}$.

*2. Betti numbers of a tree.* The tree of Fig. 7.6 is a simplicial complex $K$ with edges oriented according to the direction of the arrows, i.e., $e_1 = \langle v_1 v_2 \rangle$ and so on. Since it is connected, we have $\beta_0(K, \mathbb{Q}) = 1$. Furthermore, the matrix

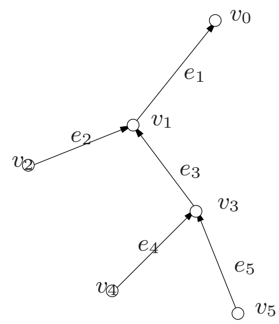



**Fig. 7.6.** A tree.

of the boundary operator $\partial_1 \colon C_1(K, \mathbb{Q}) \to C_0(K, \mathbb{Q})$ with respect to the basis $e_1, e_2, e_3, e_4, e_5$ of $C_1(K, \mathbb{Q})$ and $\langle v_0 \rangle, \langle v_1 \rangle, \langle v_2 \rangle, \langle v_3 \rangle, \langle v_4 \rangle, \langle v_5 \rangle$ of $C_0(K, \mathbb{Q})$ is

| $\partial_1$ | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_5$ |
|---|---|---|---|---|---|
| $\langle v_0 \rangle$ | 1 | 0 | 0 | 0 | 0 |
| $\langle v_1 \rangle$ | −1 | 1 | 1 | 0 | 0 |
| $\langle v_2 \rangle$ | 0 | −1 | 0 | 0 | 0 |
| $\langle v_3 \rangle$ | 0 | 0 | −1 | 1 | 1 |
| $\langle v_4 \rangle$ | 0 | 0 | 0 | −1 | 0 |
| $\langle v_5 \rangle$ | 0 | 0 | 0 | 0 | −1 |

(E.g., $\partial_1(e_1) = \langle v_0 \rangle - \langle v_1 \rangle = 1 \cdot \langle v_0 \rangle + (-1) \cdot \langle v_1 \rangle + 0 \cdot \langle v_2 \rangle + 0 \cdot \langle v_3 \rangle + 0 \cdot \langle v_4 \rangle + 0 \cdot \langle v_5 \rangle$.) Since the columns of this matrix are independent (why?), the image of $\partial_1$ has dimension 5. Therefore, $\beta_1(K, \mathbb{Q}) = \dim \ker \partial_1 = \dim C_1(K, \mathbb{Q}) - \dim \operatorname{im} \partial_1 = 0$.

*3. Betti numbers of the 2-sphere.* The simplicial complex $K$ of Fig. 7.7 is the boundary of a 3-simplex, consisting of four 2-simplices, six 1-simplices and four

0-simplices. For convenience it is shown flattened on the plane, after cutting the edges incident to 0-simplex $v_4$. The underlying space $|K|$ is homeomorphic to the 2-sphere. Vertices with the same label have to be identified, like edges between vertices with the same label. The matrix of the boundary operator

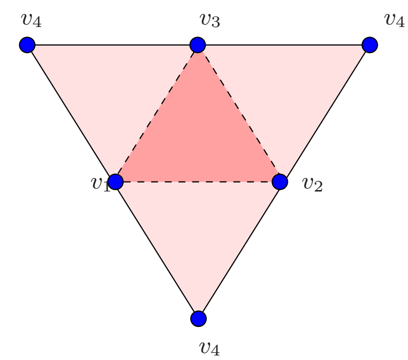



**Fig. 7.7.** A 2-sphere.

$\partial_1$ with respect to the canonical bases of $C_1(K, \mathbb{Q})$ and $C_0(K, \mathbb{Q})$ is

| $\partial_1$ | $\langle v_1 v_2 \rangle$ | $\langle v_1 v_3 \rangle$ | $\langle v_1 v_4 \rangle$ | $\langle v_2 v_3 \rangle$ | $\langle v_2 v_4 \rangle$ | $\langle v_3 v_4 \rangle$ |
|---|---|---|---|---|---|---|
| $\langle v_1 \rangle$ | $-1$ | $-1$ | $-1$ | 0 | 0 | 0 |
| $\langle v_2 \rangle$ | 1 | 0 | 0 | $-1$ | $-1$ | 0 |
| $\langle v_3 \rangle$ | 0 | 1 | 0 | 1 | 0 | $-1$ |
| $\langle v_4 \rangle$ | 0 | 0 | 1 | 0 | 1 | 1 |

It follows that $\dim C_0(K, \mathbb{Q}) = 4$, $\dim \operatorname{im} \partial_1 = 3$, and $\dim \ker \partial_1 = 3$. The matrix of the boundary operator $\partial_2$ with respect to the canonical bases of $C_2(K, \mathbb{Q})$ and $C_1(K, \mathbb{Q})$ is

| $\partial_2$ | $\langle v_1 v_2 v_3 \rangle$ | $\langle v_1 v_3 v_4 \rangle$ | $\langle v_1 v_4 v_2 \rangle$ | $\langle v_2 v_4 v_3 \rangle$ |
|---|---|---|---|---|
| $\langle v_1 v_2 \rangle$ | 1 | 0 | $-1$ | 0 |
| $\langle v_1 v_3 \rangle$ | $-1$ | 1 | 0 | 0 |
| $\langle v_1 v_4 \rangle$ | 0 | $-1$ | 1 | 0 |
| $\langle v_2 v_3 \rangle$ | 1 | 0 | 0 | $-1$ |
| $\langle v_2 v_4 \rangle$ | 0 | 0 | $-1$ | 1 |
| $\langle v_3 v_4 \rangle$ | 0 | 1 | 0 | $-1$ |

Therefore, $\dim \operatorname{im} \partial_2 = 3$ and $\dim \ker \partial_2 = 1$. Combining the previous results, we conclude that $\beta_0(K, \mathbb{Q}) = 1$, $\beta_1(K, \mathbb{Q}) = 0$ and $\beta_2(K, \mathbb{Q}) = 1$.

*4. Betti numbers of the torus.* Consider the simplicial complex of Fig. 7.8, which is a triangulation of the torus. It has 7 vertices, 21 oriented edges, and 14 oriented faces. The matrix of $\partial_2$ with respect to the canonical bases of

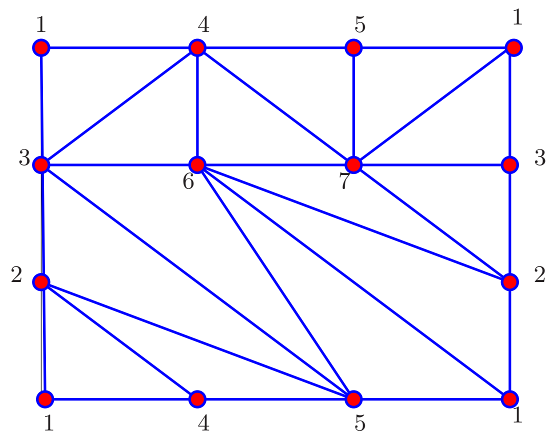



**Fig. 7.8.** A triangulation of the torus.

$C_1(K, \mathbb{Q})$ and $C_2(K, \mathbb{Q})$ is

| $\partial_2$ | $\overline{142}$ | $\overline{245}$ | $\overline{253}$ | $\overline{356}$ | $\overline{165}$ | $\overline{126}$ | $\overline{276}$ | $\overline{237}$ | $\overline{173}$ | $\overline{157}$ | $\overline{475}$ | $\overline{467}$ | $\overline{134}$ | $\overline{364}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\overline{12}$ | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\overline{13}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | −1 | 0 | 0 | 0 | 1 | 0 |
| $\overline{14}$ | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | −1 | 0 |
| $\overline{15}$ | 0 | 0 | 0 | 0 | −1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| $\overline{16}$ | 0 | 0 | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\overline{17}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 |
| $\overline{23}$ | 0 | 0 | −1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\overline{24}$ | −1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\overline{25}$ | 0 | −1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\overline{26}$ | 0 | 0 | 0 | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\overline{27}$ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\overline{34}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | −1 |
| $\overline{35}$ | 0 | 0 | −1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\overline{36}$ | 0 | 0 | 0 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| $\overline{37}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 | 0 |
| $\overline{45}$ | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | −1 | 0 | 0 | 0 |
| $\overline{46}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | −1 |
| $\overline{47}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | −1 | 0 | 0 |
| $\overline{56}$ | 0 | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\overline{57}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | −1 | 0 | 0 | 0 |
| $\overline{67}$ | 0 | 0 | 0 | 0 | 0 | 0 | −1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

The matrix of $\partial_1$ with respect to the canonical bases of $C_0(K, \mathbb{Q})$ and $C_1(K, \mathbb{Q})$ is obtained similarly (preferably using a computer algebra system). Computing the dimensions of the kernel and image of these operators we finally get

$$\beta_0(K, \mathbb{Q}) = 1, \quad \beta_1(K, \mathbb{Q}) = 2, \quad \beta_2(K, \mathbb{Q}) = 1$$

*Euler characteristic and Betti numbers.*

One of the fundamental results of simplicial homology theory states that Betti numbers of the underlying space of finite simplicial complex does not depend on the triangulation.

**Theorem 1.** *Betti numbers are* ***homotopy invariants****: if $K$ and $L$ are simplicial complexes with homotopy equivalent underlying spaces, then the $i$-th homology vector spaces of $K$ and $L$ are isomorphic. In particular,*

$$\beta_i(K, \mathbb{Q}) = \beta_i(L, \mathbb{Q}), \text{ for all } i.$$

The proof of this theorem is beyond the scope of these introductory notes. One usually introduces the more general *singular homology groups* for a topological space $X$, which are independent of any triangulation. Then one proves that these groups are isomorphic to the simplicial homology groups, obtained by taking simplicial chains with integer coefficients in stead of rational coefficients. In particular, the corresponding Betti numbers, being the ranks of these groups, are equal.

**Theorem 2.** *Let $K$ be a $d$-dimensional simplicial complex. Then*

$$\chi(K) = \sum_{i=0}^{d} (-1)^i \beta_i(K, \mathbb{Q}).$$

*Proof.* Recall from (7.1) that $\chi(K) = \sum_{i=0}^{d} (-1)^i \dim C_k(K, \mathbb{Q})$. Since

$$H_i(K, \mathbb{Q}) = \ker \partial_i \big/ \operatorname{im} \partial_{i+1}.$$

we see that

$$\begin{aligned} \beta_i(K, \mathbb{Q}) &= \dim H_i(K, \mathbb{Q}) \\ &= \dim \ker \partial_i - \dim \operatorname{im} \partial_{i+1} \\ &= \dim C_i(K, \mathbb{Q}) - \dim \operatorname{im} \partial_i - \dim \operatorname{im} \partial_{i+1}. \end{aligned}$$

Now:

$$\sum_{i=0}^{d} (-1)^i (\dim \operatorname{im} \partial_i + \dim \operatorname{im} \partial_{i+1}) = 0.$$

Hence:

$$\sum_{i=0}^{d} (-1)^i \beta_i(K, \mathbb{Q}) = \chi(K, \mathbb{Q}).$$

The claimed identities follow from the preceding derivation.

If $X$ is a topological space with a simplicial complex $K$ triangulating it, then we define $\chi(X) = \chi(K, \mathbb{Q})$. It follows from Theorem 1 and Theorem 2 that the Euler characteristic does not depend on the specific choice of the triangulation $K$.

*Incremental algorithm for computation of Betti numbers.*

As can be seen in the case of a simple space like the torus, the matrices of the boundary map become rather large, even for simple examples. Therefore alternative approaches have been developed for special cases. We start with an incremental approach, in which the simplical complex is constructed by adding simplices one at a time, making sure that during the process all partial constructs are indeed simplicial complexes. The key idea is to maintain the Betti numbers of the partial complexes. The following result indicates how to do this.

**Proposition 1.** *Let $K$ be a simplicial complex, and let $K'$ be a simplicial complex such that $K' = K \cup \sigma$ for some $k$-simplex $\sigma$. Let $\partial_i$ and $\partial'_i$ be the boundary operators of the chain complexes associated with $K$ and $K'$, respectively. Furthermore, let $\gamma = \partial'_k \sigma$. If $\gamma$ also bounds in $K$, i.e., $\partial'_k \sigma \in \operatorname{im} \partial_k$, then*

$$\beta_i(K', \mathbb{Q}) = \begin{cases} \beta_i(K, \mathbb{Q}) & \text{if } i \neq k \\ \beta_k(K, \mathbb{Q}) + 1 & \text{if } i = k \end{cases}$$

*If $\gamma$ does not bound in $K$, i.e., $\partial'_k \sigma \notin \operatorname{im} \partial_k$, then*

$$\beta_i(K', \mathbb{Q}) = \begin{cases} \beta_i(K, \mathbb{Q}) & \text{if } i \neq k-1 \\ \beta_{k-1}(K, \mathbb{Q}) - 1 & \text{if } i = k-1 \end{cases}$$

*Proof.*

$$\begin{array}{ccccccc} \cdots & \xrightarrow{\partial'_{k+1}} & C_k(K', \mathbb{Q}) & \xrightarrow{\partial'_k} & C_{k-1}(K', \mathbb{Q}) & \xrightarrow{\partial'_{k-1}} & \cdots \\ & & \| & & \| & & \\ & & C_k(K, \mathbb{Q}) \oplus \mathbb{Q}[\sigma] & & C_{k-1}(K, \mathbb{Q}) & & \end{array}$$

**Case 1:** $\partial'_k \sigma \in \operatorname{im} \partial_k$. Then $\operatorname{im} \partial'_i = \operatorname{im} \partial_i$, for all $i$, so $\dim \operatorname{im} \partial'_i = \dim \operatorname{im} \partial_i$, for all $i$. Therefore:

$$\begin{aligned} \dim \ker \partial'_k &= \dim C_k(K', \mathbb{Q}) - \dim \operatorname{im} \partial'_k \\ &= 1 + \dim C_k(K, \mathbb{Q}) - \dim \operatorname{im} \partial_k \\ &= 1 + \dim \ker \partial_k \end{aligned}$$

Furthermore, for $i \neq k$ we have $\dim \ker \partial'_i = \dim \ker \partial_i$. Hence (recall $\dim H_i(K', \mathbb{Q}) = \dim \ker \partial'_i - \dim \operatorname{im} \partial'_{i+1}$):

$$\beta_i(K', \mathbb{Q}) = \dim H_i(K', \mathbb{Q}) = \begin{cases} \dim H_i(K, \mathbb{Q}) & \text{if } i \neq k \\ 1 + \dim H_k(K, \mathbb{Q}) & \text{if } i = k \end{cases}$$

**Case 2:** $\partial'_k \sigma \notin \operatorname{im} \partial_k$. Then

$$\dim \operatorname{im} \partial_i' = \begin{cases} \dim \operatorname{im} \partial_i & \text{if } i \neq k \\ \dim \operatorname{im} \partial_k + 1 & \text{if } i = k. \end{cases}$$

Hence:

$$\begin{aligned} \dim \ker \partial_i' &= \dim C_i(K', \mathbb{Q}) - \dim \operatorname{im} \partial_i' \\ &= \begin{cases} \dim C_i(K, \mathbb{Q}) - \dim \operatorname{im} \partial_i & \text{if } i \neq k \\ 1 + \dim C_k(K, \mathbb{Q}) - (1 + \dim \operatorname{im} \partial_k) & \text{if } i = k \end{cases} \\ &= \dim \ker \partial_i \end{aligned}$$

This result yields an incremental algorithm for the computation of Betti numbers. Whether this algorithm is efficient depends on the implementation of the test '$\partial_k' \sigma \notin \operatorname{im} \partial_k$'. The paper [110] presents an efficient implementation of this algorithm for subcomplexes of the three-sphere. This incremental method can be used to compute the Betti numbers of some familiar spaces. Before showing how to do this, we introduce some additional tools that are helpful in the computation of Betti numbers.

*Chain maps and chain homotopy.*

Just like maps between spaces provide information about the topology of these spaces, maps between homology spaces provide information about the homology of these spaces. The key stepping stone towards these maps are chain maps.

Let $K$ and $L$ be finite simplicial complexes. A chain map from $K$ to $L$ is a sequence of linear maps $f_k \colon C_k(K, \mathbb{Q}) \to C_k(L, \mathbb{Q})$ such that $\partial_{k+1} \circ f_{k+1} = f_k \circ \partial_{k+1}$. In other words, the sequence $\{f_k\}$ is a chain map if the following diagram is commutative:

$$\begin{array}{ccccccc} \ldots & \xrightarrow{\partial_{k+2}} & C_{k+1}(K, \mathbb{Q}) & \xrightarrow{\partial_{k+1}} & C_k(K, \mathbb{Q}) & \xrightarrow{\partial_k} & \ldots \\ & & \Big\downarrow f_{k+1} & & \Big\downarrow f_k & & \\ \ldots & \xrightarrow{\partial_{k+2}} & C_{k+1}(L, \mathbb{Q}) & \xrightarrow{\partial_{k+1}} & C_k(L, \mathbb{Q}) & \xrightarrow{\partial_k} & \ldots \end{array}$$

This chain map is denoted by $f \colon C(K, \mathbb{Q}) \to C(L, \mathbb{Q})$. In fact, a chain map is a family of maps, containing one linear map for each dimension.

**Proposition 2.** *Let $K$, $L$ and $M$ be finite simplicial complexes.*

*1. The sequence of identity maps $\mathrm{id}_k : C_k(K, \mathbb{Q}) \to C_k(K, \mathbb{Q})$ is a chain map.*

*2. The composition of a chain map from $K$ to $L$ and a chain map from $L$ to $M$ is a chain map from $K$ to $M$.*

The proof of this result is straightforward and left as an exercise (Exercise 4). Let $f \colon C(K, \mathbb{Q}) \to C(L, \mathbb{Q})$ be a chain map. The linear map $f_* \colon H(K, \mathbb{Q}) \to H(L, \mathbb{Q})$ is defined by

$$f_{*k}([\alpha]) = [f_k(\alpha)],$$

for $\alpha \in Z_k(K, \mathbb{Q})$. We say that $f_*$ is the map induced by $f$ at the level of homology. Using commutativity of the diagram above, it is easy to see that this map is well-defined, i.e., that $[f_k(\alpha)]$ is independent of the choice of the representative $\alpha$ of the homology class $[\alpha]$. This map has some natural properties, following in a straightforward way from the definition.

**Proposition 3.** *Let $K$, $L$ and $M$ be finite simplicial complexes.*

*1. The identity chain map generates the identity map at the level of homology.*

*2. The map induced by a composition of chain maps is the composition of the maps induced by each chain map. In other words, for chain maps $f\colon C(K, \mathbb{Q}) \to C(L, \mathbb{Q})$ and $g\colon C(L, \mathbb{Q}) \to C(M, \mathbb{Q})$:*

$$(g \circ f)_* = g_* \circ f_*.$$

A *chain homotopy* between two chain maps $f, g\colon C(K, \mathbb{Q}) \to C(L, \mathbb{Q})$ is a sequence $\{T_k\}$ of linear maps $T_k\colon C_k(K, \mathbb{Q}) \to C_{k+1}(L, \mathbb{Q})$ such that

$$T_{k-1} \circ \partial_k + \partial_{k+1} \circ T_k = f_k - g_k.$$

If such a chain homotopy exists, then $f$ and $g$ are called *chain-homotopic*. We shall frequently use the following result, the proof of which is a simple exercise in Linear Algebra (see Exercise 4).

**Proposition 4.** *Chain homotopic chain maps induce the same linear map at the level of homology.*

*Simplical collapse.*

We now consider *simplicial collapse*, a very simple transformation of simplicial complexes which does not alter homology in positive dimensions. This operation allows us to compute the Betti numbers of a simplicial complex $K$ by simplifying $K$ until we obtain another simplical complex $L$ for which the Betti numbers are known or easy to compute.

Let $K$ be a finite simplicial complex, and let $\alpha$ and $\beta$ be two simplices of $K$ such that $\alpha$ is a face of $\beta$, and $\alpha$ is not a face of any other simplex of $K$. Let $L$ be the subcomplex of $K$ obtained by deleting the simplices $\alpha$ and $\beta$. The transformation from $K$ to $L$ is called an *elementary collapse*. See Fig. 7.9.

More generally, we say that $K$ *collapses onto* a subcomplex $L$, denoted by $K \searrow L$, if there is a finite sequence of elementary collapses transforming $K$ into $L$.

**Proposition 5.** *Let $K$ and $L$ be finite simplicial complexes such that $K$ collapses onto $L$. Then $H_k(K, \mathbb{Q})$ and $H_k(L, \mathbb{Q})$ are isomorphic.*

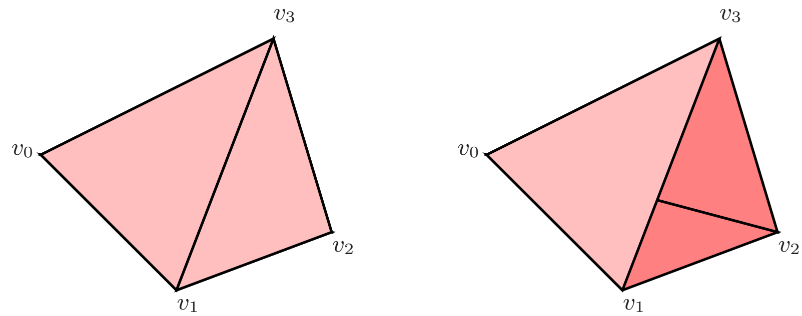



**Fig. 7.9.** An elementary collapse removes the simplices $v_0v_1v_2v_3$ and $v_1v_2v_3$ from the leftmost simplex.

*Proof.* We give the proof for positive $k$, the case $k = 0$ being trivial. Our strategy consists of finding a chain homotopy inverse to the inclusion chain map $\iota\colon C(L,\mathbb{Q}) \to C(K,\mathbb{Q})$. To this end let $\alpha$ be a $k$-simplex, positively oriented in the boundary $\partial\beta$ of the $k+1$-simplex $\beta$. Introduce the map $f\colon C(K,\mathbb{Q}) \to C(L,\mathbb{Q})$ by putting $f_k(\alpha) = \alpha - \partial\beta$, $f_{k+1}(\beta) = 0$, $f_i(\sigma) = \sigma$ for every $i$-simplex different from $\alpha$ and $\beta$, and extending linearly. It is not hard to prove that $f$ is a chain map. Furthermore, $f \circ \iota$ is the identity chain map on $C(L,\mathbb{Q})$.

Let the sequence of linear maps $P_i\colon C_i(K,\mathbb{Q}) \to C_{i+1}(K,\mathbb{Q})$ be defined by $P_k(\alpha) = \beta$, and $P_i(\sigma) = 0$ for each $i$-simplex $\sigma$ different from $\alpha$. A straightforward computation shows that the sequence $\{P_i\}$ is a chain homotopy between the identity map on $C(K,\mathbb{Q})$ and the chain map $\iota \circ f$. From this we conclude that $\iota_i\colon H_i(L,\mathbb{Q}) \to H_i(K,\mathbb{Q})$ is an isomorphism, for $i > 0$. In particular, $K$ and $L$ have the same Betti numbers in positive dimension.

*Example: Betti numbers of the projective plane.*

The incremental algorithm, combined with the method of simplicial collapse, allows for rather painless computation of Betti numbers of familiar spaces. In this example we compute the Betti numbers of the projective plane $\mathbb{R}P_2$. The simplicial complex $K$ of Fig. 7.10 is the unique triangulation of the projective plane with a minimal number of vertices. The vertices and edges on the boundary of the six-gon are identified in pairs, as indicated by the double occurrence of the vertex-labels $v_1$, $v_2$ and $v_3$. The arrows indicate the orientation of the simplices forming the basis of the chain space $C_2(K)$. We orient the edges of the simplex from the vertex with lower index to the vertex with higher index.

Let $L$ be the simplicial complex obtained from $K$ by deleting the oriented simplex $\tau = \langle v_4v_5v_6\rangle$. The Betti numbers of $L$ are easy to compute, since a sequence of simplicial collapses transforms $L$ into the subcomplex $L_0$ with

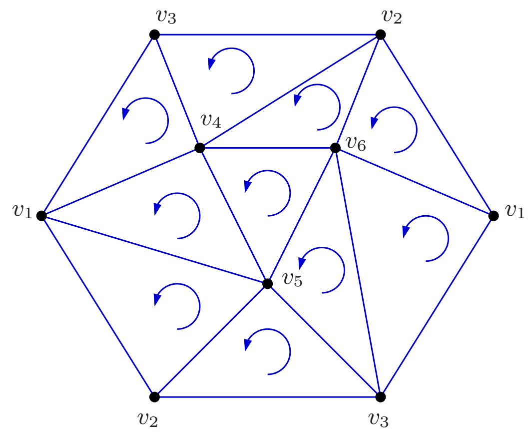



**Fig. 7.10.** A triangulation of the projective plane.

vertices $v_1$, $v_2$ and $v_3$, and oriented edges $\langle v_1 v_2 \rangle$, $\langle v_2 v_3 \rangle$ and $\langle v_1 v_3 \rangle$. The simplicial complex $L_0$ is a 1-sphere, so $\beta_0(L) = \beta_0(L_0) = 1$, $\beta_1(L) = \beta_1(L_0) = 1$, and $\beta_i(L) = \beta_i(L_0) = 0$ for $i > 1$.

To relate the Betti numbers of $K$ with those of $L$, we have to determine whether $\tau' = \partial_2 \tau$ is a boundary in $L$. Consider the special 2-chain $\alpha$, which is the formal sum of all oriented 2-simplices in $L$. Taking the boundary of $\alpha$, we see that all oriented 1-simplices not in $\partial_2 \tau$ occur twice, those in the interior of the six-gon in Fig. 7.10 with opposite coefficients and those in the boundary with the same coefficient. In other words, $\partial_2 \alpha = 2\gamma - \partial_2 \tau$, where $\gamma$ is the 1-cycle $\langle v_1 v_2 \rangle + \langle v_2 v_3 \rangle - \langle v_1 v_3 \rangle$ of $L$. Therefore, $[\tau'] = 2[\gamma]$ in $H_1(L)$. Since $[\gamma]$ forms a basis for $H_1(L)$, we conclude that $[\tau'] \neq 0$ in $H_1(L)$. Hence $\tau'$ is *not* a boundary in $L$. Applying the incremental algorithm we see that $\beta_0(K) = \beta_0(L) = 1$, $\beta_1(K) = \beta_1(L) - 1 = 0$, and $\beta_2(K) = \beta_2(L) = 0$.

*Example: Betti numbers depend on field of scalars.*

Homology theory can be set up with coefficients in a general field. A priory, this leads to different Betti numbers. This is illustrated by revisiting the simplicial complex $K$ of Fig. 7.10, and applying the same procedure to compute the Betti numbers over $\mathbb{Z}_2$. Using the same notation as in the preceding example, we see that $[\tau'] = 2[\gamma] = 0$ in $H_1(L, \mathbb{Z}_2)$, so $\tau'$ is a boundary in $C_2(L, \mathbb{Z}_2)$. Applying the incremental algorithm again we conclude that $\beta_i(K, \mathbb{Z}) = \beta_i(L, \mathbb{Z}) = 1$, for $i = 0, 1$, and $\beta_2(K, \mathbb{Z}) = \beta_2(L, \mathbb{Z}) + 1 = 1$. Note that the Euler characteristic is independent of the coefficient field.

## 7.4 Morse Theory

Finite dimensional Morse theory deals with the relation between the topology of a smooth manifold and the critical points of smooth real-valued functions on the manifold. It is the basic tool for the solution of fundamental problems in differential topology. Recently, basic notions from Morse theory have been used in the study of the geometry and topology of large molecules. We review some basic concepts from Morse theory, like in [325]. More elaborate treatments are [250] and [245].

### 7.4.1 Smooth functions and manifolds

*Differential of a smooth map.*

A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is called *smooth* if all derivatives of any order exist. A map $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ is called smooth if its component functions are smooth. The *differential* of $\varphi$ at a point $q \in \mathbb{R}^n$ is the *linear map* $d\varphi_q\colon \mathbb{R}^n \to \mathbb{R}^m$ defined as follows. For $v \in \mathbb{R}^n$, let $\alpha\colon I \to \mathbb{R}^n$, with $I = (-\varepsilon, \varepsilon)$ for some positive $\varepsilon$, be defined by $\alpha(t) = \varphi(q + tv)$, then $d\varphi_q(v) = \alpha'(0)$. Let $\varphi(x_1, \dots, x_n) = (\varphi_1(x_1, \dots, x_n), \dots, \varphi_m(x_1, \dots, x_n))$. The differential $d\varphi_q$ is represented by the Jacobian matrix

$$\begin{pmatrix} \frac{\partial \varphi_1}{\partial x_1}(q) & \dots & \frac{\partial \varphi_1}{\partial x_n}(q) \\ \vdots & & \vdots \\ \frac{\partial \varphi_m}{\partial x_1}(q) & \dots & \frac{\partial \varphi_m}{\partial x_n}(q) \end{pmatrix}.$$

*Regular surfaces in* $\mathbb{R}^3$.

A subset $S$ in $\mathbb{R}^3$ is a *smooth surface* if we can cover the surface with open coordinate neighborhoods. More precisely, a coordinate neighborhood of a point $p$ on the surface is a subset of the form $V \cap S$, where $V$ is an open subset of $\mathbb{R}^3$, for which there exists a smooth map $\varphi\colon U \to \mathbb{R}^3$ defined on an open subset $U$ of $\mathbb{R}^2$, such that where $V$ is an open subset of $\mathbb{R}^3$ containing $p$, for which there exists a smooth map $\varphi\colon U \to \mathbb{R}^3$ defined on an open subset $U$ of $\mathbb{R}^2$, such that

(i) The map $\varphi$ is a homeomorphism from $U$ onto $V \cap S$;
(ii) If $\varphi(u, v) = (x(u, v), y(u, v), z(u, v))$, then the two tangent vectors

$$\begin{pmatrix} \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial u} \\ \frac{\partial z}{\partial u} \end{pmatrix}, \quad \begin{pmatrix} \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial v} \end{pmatrix}$$

are non-zero and not parallel.

The map $\varphi$ is called a parametrization or a system of local coordinates in $p$. The set $S$ is a smooth surface if each point of $S$ has a coordinate neighborhood. Note that condition (ii) is equivalent to the fact that the differential of $\varphi$ at $(u, v)$ is an injective map.

*Example: spherical coordinates.* Let $S$ be a 2-sphere in $\mathbb{R}^3$ with radius $R$ and center $(0, 0, 0) \in \mathbb{R}^3$. Consider the set $U = \{ (u, v) \mid 0 < u < 2\pi, -\pi/2 < v < \pi/2 \}$. The map $\varphi \colon U \to S$, given by

$$\varphi(u, v) = (R \cos u \cos v, R \sin u \cos v, R \sin v).$$

corresponds to the well-known spherical coordinates. Note that $\varphi(U)$ is the 2-sphere minus a meridian. Each point of $\varphi(U)$ has a system of local coordinates given by $\varphi$.

*Example: coordinates on the upper and lower hemisphere.* Again, let $S$ be the sphere with radius $R$ and center at the origin of $\mathbb{R}^3$, and let $U = \{ (x, y) \mid x^2 + y^2 < R^2 \}$. The (open) upper and lower hemispheres of the torus are the graph of a smooth function. More precisely, each point of the upper hemisphere has local coordinates given by the map

$$\varphi(x, y) = (x, y, \sqrt{R^2 - x^2 - y^2}).$$

A similar expression defines local coordinates at each point of the lower hemisphere. Covering the sphere by six hemispheres yields a system (at least one) of local coordinate system for each point of the sphere. Therefore, the sphere is a regular surface.

*Example: coordinates on the torus of revolution.* Let $S$ be the torus obtained by rotating the circle in $x, y$-plane with center $(0, R, 0)$ and radius $r$ around the $x$-axis, where $R > r$. We show that $S$ is a smooth surface by introducing a system of local coordinates for all points of the torus. To this end, let $U = \{(u, v) \mid 0 < u, v < 2\pi\}$ and let $\varphi \colon U \to \mathbb{R}^3$ be the map defined by

$$\varphi(u, v) = (r \sin u, (R - r \cos u) \sin v, (R - r \cos u) \cos v).$$

It is not hard to check that $\varphi(U) \subset S$. In fact, the map $\varphi$ covers the torus except for one meridian and one parallel circle. It is easy to find local coordinates in points of these two circles by translating the parameter domain $U$ a little bit. Therefore, the torus is a regular surface.

*Example: Local form of torus of revolution near* $(0, 0, \pm(R - r))$. As in the example of hemispheres, parts of the torus are graphs of a smooth function. In particular, the points $(0, 0, \pm(R - r))$ have local coordinates of the form $\varphi(x, y) = (x, y, f_\pm(x, y))$, where

$$f_\pm(x, y) = \pm\sqrt{R^2 + r^2 - x^2 - y^2 - 2R\sqrt{r^2 - x^2}}.$$

*Submanifolds of* $\mathbb{R}^n$*.*

More generally, a subset $M$ of $\mathbb{R}^n$ is an $m$-dimensional smooth *submanifold* of $\mathbb{R}^n$, $m \leq n$, if for each $p \in M$, there is an open set $V$ in $\mathbb{R}^n$, containing $p$, and a map $\varphi\colon U \to M \cap V$ from an open subset $U$ in $\mathbb{R}^m$ onto $V \cap M$ such that (i) $\varphi$ is a smooth homeomorphism, (ii) the differential $d\varphi_q\colon \mathbb{R}^m \to \mathbb{R}^n$ is injective for each $q \in U$. Again, the map $\varphi$ is called a parametrization or a system of local coordinates on $M$ in $p$. In particular, the space $\mathbb{R}^n$ is a submanifold of $\mathbb{R}^n$. A subset $N$ of a submanifold $M$ of $\mathbb{R}^n$ is a submanifold of $M$ if it is a submanifold of $\mathbb{R}^n$. The difference of the dimensions of $M$ and $N$ is called the *codimension* of $N$ (in $M$).

*Example: linear subspaces are submanifolds.* The Euclidean space $\mathbb{R}^m$ is a smooth submanifold of $\mathbb{R}^n$, for $m \leq n$. For $m < n$, we identify $\mathbb{R}^m$ with the subset $\{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_{m+1} = \cdots = x_n = 0\}$ of $\mathbb{R}^n$.

*Example:* $\mathbb{S}^{n-1}$ *is a smooth submanifold of* $\mathbb{R}^n$*.* A smooth parametrization of $\mathbb{S}^{n-1}$ at $(0, \ldots, 0, 1) \in \mathbb{S}^{n-1}$ is given by $\varphi\colon U \to \mathbb{R}^n$, with

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

and

$$\varphi(x_1, \ldots, x_{n-1}) = (x_1, \ldots, x_{n-1}, \sqrt{1 - x_1^2 - \cdots - x_{n-1}^2}).$$

In fact, $\varphi$ is a parametrization in every point of the upper hemisphere, i.e., the intersection of $\mathbb{S}^{n-1}$ and the upper half space $\{(y_1, \ldots, y_n) \mid y_n > 0\}$.

*Example: codimension one submanifolds.* The equator $\mathbb{S}^1 = \{(x_1, x_2, 0) \mid x_1^2 + x_2^2 = 1\}$ is a codimension one submanifold of $\mathbb{S}^2 = \{(x_1, x_2, x_3) \mid x_1^2 + x_2^2 + x_3^2 = 1\}$. More generally, every intersection of the 2-sphere with a plane at distance less than one from the origin is a codimension one submanifold.

*Tangent space of a manifold.*

The tangent vectors at a point $p$ of a manifold form a vector space, called the tangent space of the manifold at $p$. More formally, a *tangent vector* of $M$ at $p$ is the tangent vector $\alpha'(0)$ of some smooth curve $\alpha\colon I \to M$ through $p$. Here a smooth curve through a point $p$ on a smooth submanifold $M$ of $\mathbb{R}^n$ is a smooth map $\alpha\colon I \to \mathbb{R}^n$, with $I = (-\varepsilon, \varepsilon)$ for some positive $\varepsilon$, satisfying $\alpha(t) \in M$, for $t \in I$, and $\alpha(0) = p$. The set $T_pM$ of all tangent vectors of $M$ at $p$ is the *tangent space* of $M$ at $p$.

If $\varphi\colon U \to M$ is a smooth parametrization of $M$ at $p$, with $0 \in U$ and $\varphi(0) = p$, then $T_pM$ is the $m$-dimensional subspace $d\varphi_0(\mathbb{R}^m)$ of $\mathbb{R}^n$, which passes through $\varphi(0) = p$. Let $\{e_1, \ldots, e_m\}$ be the standard basis of $\mathbb{R}^m$; define the tangent vector $\overline{e}_i \in T_pM$ by $\overline{e}_i = d\varphi_0(e_i)$. Then $\{\overline{e}_1, \ldots, \overline{e}_m\}$ is a basis of $T_pM$.

*Example: tangent space of the sphere.* The tangent space of the unit sphere $\mathbb{S}^{n-1} = \{(x_1, \ldots, x_n) \mid x_1^2 + \cdots + x_n^2 = 1\}$ at a point $p$ is the hyperplane through $p$, perpendicular to the normal vector of the sphere at $p$.

*Smooth function on a submanifold.*

A function $f\colon M \to \mathbb{R}$ on an $m$-dimensional smooth submanifold $M$ of $\mathbb{R}^n$ is smooth at $p \in M$ if there is a smooth parametrization $\varphi\colon U \to M \cap V$, with $U$ an open set in $\mathbb{R}^m$ and $V$ an open set in $\mathbb{R}^n$ containing $p$, such that the function $f \circ \varphi\colon U \to \mathbb{R}$ is smooth. A function on a manifold is called smooth if it is smooth at every point of the manifold.

*Example: height function on a surface.* The height function $h\colon S \to \mathbb{R}$ on a surface $S$ in $\mathbb{R}^3$ is defined by $h(x, y, z) = z$, for $(x, y, z) \in \mathbb{S}$. Let $\varphi(u, v) = (x(u, v), y(u, v), z(u, v))$ be a system of local coordinates in a point of the surface, then $h \circ \varphi(u, v) = z(u, v)$ is smooth. Therefore, the height function is a smooth function on $S$.

*Regular and critical points.*

A point $p \in M$ is a *critical point* of a smooth function $f\colon M \to \mathbb{R}$ if there is a local parametrization $\varphi\colon U \to \mathbb{R}^n$ of $M$ at $p$, with $\varphi(0) = p$, such that 0 is a critical point of $f \circ \varphi\colon U \to \mathbb{R}$ (i.e., the differential of $f \circ \varphi$ at $q$ is the zero function on $\mathbb{R}^n$). This condition does not depend on the particular parametrization.
A real number $c \in \mathbb{R}$ is a *regular value* of $f$ if $f(p) \neq c$ for all critical points $p$ of $f$, and a *critical value* otherwise.

*Example: critical points of height function on the sphere.* Consider the height function on the unit sphere in $\mathbb{R}^3$. Spherical coordinates define a parametrization $\varphi(u, v)$ in every point, except for the poles $(0, 0, \pm 1)$. With respect to this parametrization the height function $h$ has the expression $\tilde{h}(u, v) = h(\varphi(u, v)) = \sin v$, so none of these points is singular (since $-\pi/2 < v < \pi/2$ away from the poles). Near the poles $(0, 0, \pm 1)$ we consider the sphere as the graph of a function, corresponding to the parametrization $\psi(x, y) = (x, y, \sqrt{1 - x^2 - y^2})$. The height function is expressed in these local coordinates as $\tilde{h}(x, y) = h(\psi(x, y)) = \pm\sqrt{1 - x^2 - y^2}$, so the singular points of $h$ are $(0, 0, -1)$ (minimum), and $(0, 0, 1)$ (maximum).

*Example: critical points of height function on the torus.* The torus $M$ in $\mathbb{R}^3$, obtained by rotating a circle in the $x, y$-plane with center $(0, R, 0)$ and radius $r$ around the $x$-axis, is a smooth 2-manifold. Let $U = \{(u, v) \mid -\pi/2 < u, v < 3\pi/2\} \subset \mathbb{R}^2$, and let the map $\varphi\colon U \to \mathbb{R}^3$ be defined by

$$\varphi(u, v) = (r \sin u, (R - r \cos u) \sin v, (R - r \cos u) \cos v).$$

Then $\varphi$ is a parametrization at all points of $M$, except for points on one latitudinal and one longitudinal circle. The height function on $M$ is the function $h\colon M \to \mathbb{R}$ defined by $\tilde{h}(u, v) = h(\varphi(u, v)) = (R - r \cos u) \cos v$, so the singular points of $h$ are:

| $(u,v)$ | $\varphi(u,v)$ | type of singularity |
|---|---|---|
| $(0,0)$ | $(0,0,R-r)$ | saddle point |
| $(0,\pi)$ | $(0,0,-R+r)$ | saddle point |
| $(\pi,0)$ | $(0,0,R+r)$ | maximum |
| $(\pi,\pi)$ | $(0,0,-R-r)$ | minimum |

The type of a singular point will be introduced in Sect. 7.4.2.

*Implicit surfaces and manifolds.*

In many cases a set is given as the zero set of a smooth function (or a system of functions). If this zero set contains no singular point of the function, then it is a manifold:

**Proposition 6.** (Implicit Function Theorem). *Let $f\colon M \to \mathbb{R}$ be a smooth function on the smooth submanifold $M$ of $\mathbb{R}^n$. If $c$ is a regular value of $f$, then the level set $f^{-1}(c)$ is a smooth submanifold of $M$ of codimension one.*

A proof can be found in any book on analysis on manifolds, like [317].

*Example: implicit surfaces in three-space.* The unit sphere in three space is a regular surface, since 0 is a regular value of the function $f(x,y,z) = x^2+y^2+z^2-1$. The torus of revolution is a regular surface, since 0 is a regular value of the function $g(x,y,z) = (x^2+y^2+z^2-R^2-r^2)^2 - 4R^2(r^2-x^2)$.

*Hessian at a critical point.*

Let $M$ be a smooth submanifold of $\mathbb{R}^n$, and let $f\colon M \to \mathbb{R}$ be a smooth function. The *Hessian* of $f$ at a critical point $p$ is the quadratic form $H_pf$ on $T_pM$ defined as follows. For $v \in T_pM$, let $\alpha\colon (-\varepsilon,\varepsilon) \to M$ be a curve with $\alpha(0) = p$, and $\alpha'(0) = v$. Then

$$H_pf(v) = \left.\frac{d^2}{dt^2}\right|_{t=0} f(\alpha(t)).$$

The right hand side does not depend on the choice of $\alpha$. To see this, let $\varphi\colon U \to M$ be a smooth parametrization of $M$ at $p$, with $0 \in U$ and $\varphi(0) = p$, and let $v = v_1\overline{e}_1 + \cdots + v_m\overline{e}_m \in T_pM$, where $\overline{e}_i = d\varphi_0(e_i)$. Then

$$H_pf(v) = \sum_{i,j=1}^{m} \frac{\partial^2(f\circ\varphi)}{\partial x_i \partial x_j}(0)v_iv_j.$$

In particular, the matrix of $H_f(p)$ with respect to this basis is

$$\begin{pmatrix} \dfrac{\partial^2(f\circ\varphi)}{\partial x_1^2}(0) & \dots & \dfrac{\partial^2(f\circ\varphi)}{\partial x_1 \partial x_m}(0) \\ \vdots & & \vdots \\ \dfrac{\partial^2(f\circ\varphi)}{\partial x_1 \partial x_m}(0) & \dots & \dfrac{\partial^2(f\circ\varphi)}{\partial x_m^2}(0) \end{pmatrix}. \tag{7.3}$$

It is not hard to check that the numbers of positive and negative eigenvalues of the Hessian do not depend on the choice of $\varphi$, since $p$ is a critical point of $f$.

*Non-degenerate critical point.*

The critical point $p$ of $f\colon M \to \mathbb{R}$ is *non-degenerate* if the Hessian $H_p f$ is non-degenerate. The *index* of the non-degenerate critical point $p$ is the number of negative eigenvalues of the Hessian at $p$. If $M$ is 2-dimensional, then a critical point of index 0, 1, or 2, is called a *minimum*, *saddle point*, or *maximum*, respectively.

### 7.4.2 Basic Results from Morse Theory

*Morse function.*

A smooth function on a manifold is a *Morse function* if all critical points are non-degerate. The *k-th Morse number* of a Morse function $f$, denoted by $\mu_k(f)$, is the number of critical points of $f$ of index $k$.

*Example: quadratic function on* $\mathbb{R}^m$. The function $f\colon \mathbb{R}^m \to \mathbb{R}$, defined by $f(x_1, \ldots, x_m) = -x_1^2 - \ldots - x_k^2 + x_{k+1}^2 + \ldots + x_m^2$, is a Morse function, with a single critical point $(0, \ldots, 0)$. This point is a non-degenerate critical point, since the Hessian matrix at this point is $\mathrm{diag}(-2, \ldots, -2, 2, \ldots, 2)$, with $k$ entries on the diagonal equal to $-2$. In particular, the index of the critical point is $k$.

*Example: singularities of the height function on* $S^{m-1}$. The height function on the standard unit sphere $\mathbb{S}^{m-1}$ in $\mathbb{R}^m$ is a Morse function. This function is defined by $h(x_1, \ldots, x_m) = x_m$ for $(x_1, \ldots, x_m) \in \mathbb{S}^{m-1}$, With respect to the parametrization $\varphi(x_1, \ldots, x_{m-1}) = (x_1, \ldots, x_{m-1}, \sqrt{1 - x_1^2 - \cdots - x_{m-1}^2})$, the expression of the height function is

$$h \circ \varphi(x_1, \ldots, x_{m-1}) = \sqrt{1 - x_1^2 - \cdots - x_{m-1}^2}.$$

Therefore, the only critical point of $h$ on the upper hemisphere is $(0, \ldots, 0, 1)$. The Hessian matrix (7.3) is the diagonal matrix $\mathrm{diag}(-1, -1, \ldots, -1)$, so this critical point has index $m-1$. Similarly, $(0, \ldots, 0, -1)$ is the only critical point on the lower hemisphere. It is a critical point of index 0.

*Example: singularities of the height function on the torus.* The singular points of the height function on the torus of revolution with radii $R$ and $r$ are $(0, 0, -R-r)$, $(0, 0, -R+r)$, $(0, 0, R-r)$, and $(0, 0, R+r)$. See also Sect. 7.4.1. A parametrization of this torus near the singular points $\pm(R-r)$ is $\varphi(x, y) = (x, y, f_\pm(x, y))$, where $f_\pm(x, y) = \pm\sqrt{R^2 + r^2 - x^2 - y^2 - 2R\sqrt{r^2 - x^2}}$. The expression $h(x, y) = f_\pm(x, y)$ of the height function with respect to these local coordinates at $(x, y) = (0, 0)$ is

$$h(x,y) = \pm\big(R - r - \frac{1}{2r}x^2 + \frac{1}{2(R-r)}y^2\big) + \text{Higher Order Terms.}$$

Hence the singular points corresponding to $(x,y) = (0,0)$, i.e., $(0,0,\pm(R-r))$, are saddle points, i.e., singular points of index one. Similarly, the singular point $(0,0,R+r)$ is a maximum (index two), and the singular point $(0,0,-R-r)$ is a minimum (index zero), and the

*Regular level sets.*

Let $M$ be an $m$-dimensional submanifold of $\mathbb{R}^n$, and let $f\colon M \to \mathbb{R}$ be a smooth function. The set $f^{-1}(h) := \{q \in M | f(q) = h\}$ of points where $f$ has a fixed value $h$ is called a *level set* (at level $h$). If $h \in \mathbb{R}$ is a regular value of $f$, then $f^{-1}(h)$ is a smooth $(m-1)$-dimensional submanifold of $\mathbb{R}^n$.
Similarly, we define the *lower level set* (also called *excursion set*) at some level $h \in \mathbb{R}$ as $M_h = \{\, q \in M \mid f(q) \leq h \,\}$. If $f$ has no critical values in $[a,b]$, for $a < b$, then the subsets $M_a$ and $M_b$ of $M$ are homeomorphic (and even isotopic).

*The Morse Lemma.*

Let $f\colon M \to \mathbb{R}$ be a smooth function on a smooth $m$-dimensional submanifold $M$ of $\mathbb{R}^n$, and let $p$ be a non-degenerate critical point of index $k$. Then there is a smooth parametrization $\varphi\colon U \to M$ of $M$ at $p$, with $U$ an open neighborhood of $0 \in \mathbb{R}^m$ and $\varphi(0) = p$, such that

$$f \circ \varphi(x_1, \ldots, x_m) = f(p) - x_1^2 - \cdots - x_k^2 + x_{k+1}^2 + \cdots + x_m^2.$$

In particular, a critical point of index 0 is a local minimum of $f$, whereas a critical point of index $m$ is a local maximum of $f$. See Fig. 7.11.

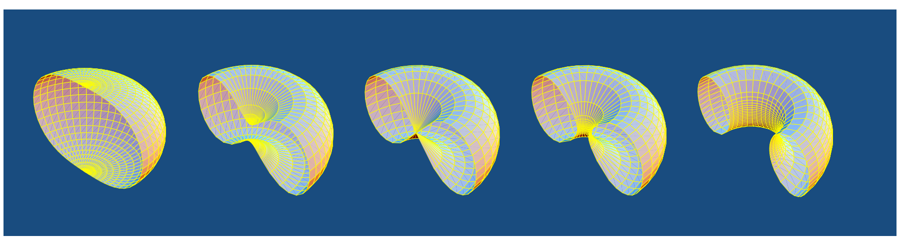


**Fig. 7.11.** Passing a critical level set of a Morse function in three-space. The critical point has index 1. A local model of the function near the critical point is $f(x_1, x_2, x_3) = -x_1^2 + x_2^2 + x_3^2$, with the $x_1$-axis running vertically.

*Abundance of Morse functions.*

(i) *Morse functions are generic.* Every smooth compact submanifold of $\mathbb{R}^n$ has a Morse function. (In fact, if we endow the set $C^\infty(M)$ of smooth functions on $M$ with the so-called Whitney topology, then the set of Morse functions on $M$ is an open and dense subset of $C^\infty(M)$. In particular, there are Morse functions arbitrarily close to any smooth function on $M$.)
(ii) *Generic height functions are Morse functions.* Let $M$ be an $m$-dimensional submanifold of $\mathbb{R}^{m+1}$ (e.g., a smooth surface in $\mathbb{R}^3$). For $v \in \mathbb{S}^m$, the height-function $h_v\colon M \to \mathbb{R}$ with respect to the direction $v$ is defined by $h_v(p) = \langle v, p \rangle$. The set of $v$ for which $h_v$ is not a Morse function has measure zero in $\mathbb{S}^m$.

*Passing critical levels.*

One can build complicated spaces from simple ones by attaching a number of cells. Let $X$ and $Y$ be topological spaces, such that $X \subset Y$. We say that $Y$ is obtained by *attaching a $k$-cell to $X$* if $Y \setminus X$ is homeomorphic to an open $k$-ball. More precisely, there is a map $f\colon \overline{\mathbb{B}}^k \to \overline{Y \setminus X}$, such that $f(\mathbb{S}^{k-1}) \subset X$ and the restriction $f \mid \mathbb{B}^k$ is a homeomorphism $\mathbb{B}^k \to Y \setminus X$. Let $f\colon M \to \mathbb{R}$ be a smooth Morse function with exactly one critical level in $(a, b)$, and $a$ and $b$ are regular values of $f$. Then $M_b$ is homotopy equivalent to $M_a$ with a cell of dimension $k$ attached, where $k$ is the index of the critical point in $f^{-1}([a, b])$. See Fig. 7.12.

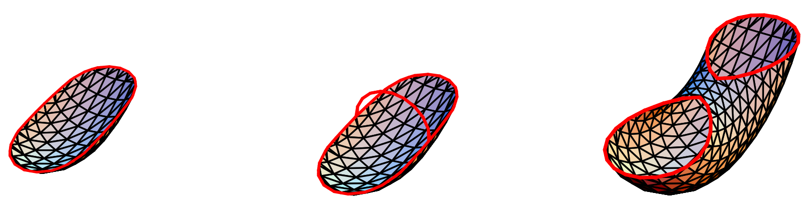


**Fig. 7.12.** Passing a critical level of index 1 corresponds to attaching a 1-cell. Here $M$ is the 2-torus embedded in $\mathbb{R}^3$, in standard vertical position, and $f$ is the height function with respect to the vertical direction. Left: $M_a$, for $a$ below the critical level of the lower saddle point of $f$. Middle: $M_a$ with a 1-cell attached to it. Right: $M_b$, for $b$ above the critical level of the lower saddle point of $f$. This set is homotopy equivalent to the set in the middle part of the figure.

*Morse inequalities.*

Let $f$ be a Morse function on a compact $m$-dimensional smooth submanifold of $\mathbb{R}^n$. For each $k$, $0 \leq k \leq m$, the $k$-th Morse number of $f$ dominates the

$k$-th Betti number of $M$:

$$\mu_k(f) \geq \beta_k(M, \mathbb{Q}).$$

An intuitive explanation is based on the observation that passing a critical level of a critical point of index $k$ is equivalent corresponds to the attachment of a $k$-cell at the level of homotopy equivalence. Therefore, either the $k$-th Betti number increases by one, or the $k-1$-st Betti number decreases by one, cf the incremental algorithm for computing Betti numbers in Sect. 7.3, while none of the other Betti numbers changes. Since only the $k$-th Morse number changes, more precisely, increases by one, the Morse inequalities are invariant upon passage of a critical level.

In the same spirit one can show that the Morse numbers of $f$ are related to the Betti numbers and the Euler characteristic of $M$ by the following identity:

$$\sum_{k=1}^{m} (-1)^k \mu_k(f) = \sum_{k=1}^{m} (-1)^k \beta_k(M, \mathbb{Q}) = \chi(M).$$

*Gradient vector fields.*

Consider a smooth function $f\colon M \to \mathbb{R}$, where $M$ is a smooth $m$-dimensional submanifold of $\mathbb{R}^n$. The *gradient of* $f$ is a smooth map $\operatorname{grad} f\colon M \to \mathbb{R}^n$, which assigns to each point $p \in M$ a vector $\operatorname{grad} f(p) \in T_pM \subset \mathbb{R}^n$, such that

$$\langle \operatorname{grad} f(p), v\rangle = df_p(v), \quad \text{for all } v \in T_pM.$$

Since $df_p(v)$ is a linear form in $v$, the vector $\operatorname{grad} f(p)$ is well defined by the preceding identity. This definition has a few straightforward implications. The gradient of $f$ vanishes at a point $p$ if and only if $p$ is a singular point of $f$. If $p$ is not a singular point of $f$, then $df_p(v)$ is maximal for a unit vector $v \in T_pM$ iff $v = \operatorname{grad} f(p)/\|\operatorname{grad} f(p)\|$. In other words, $\operatorname{grad} f(p)$ is the direction of steepest ascent of $f$ at $p$. Furthermore, if $c \in \mathbb{R}$ is a regular value of the function $f$, then $\operatorname{grad} f$ is perpendicular to the level set $f^{-1}(c)$ at every point.

To express $\operatorname{grad} f$ in local coordinates, let $\varphi\colon U \to M$ be a system of local coordinates at $p \in M$. Let $\overline{e_1}, \ldots, \overline{e_m}$ be the basis of $T_pM$ corresponding to the standard basis $e_1, \ldots, e_m$ of $\mathbb{R}^m$. In other words: $\overline{e_i} = d\varphi_q(e_i)$, where $q \in U$ is the pre-image of $p$ under $\varphi$. We denote the standard coordinates on $\mathbb{R}^m$ by $x_1, \ldots, x_m$. Then

$$\operatorname{grad} f(p) = \sum_{i=1}^{m} a_i(q)\overline{e_i},$$

where $a_i\colon U \to \mathbb{R}$ is the smooth function defined by the set of linear equations

$$\sum_{j=1}^{m} g_{ij}(q) a_j(q) = \frac{\partial (f \circ \varphi)}{\partial x_i}(q), \qquad (1 \leq i \leq m),$$

with $g_{ij}(q) = \langle \overline{e_i}, \overline{e_j} \rangle$. Since the coefficients are the entries $\langle \overline{e_i}, \overline{e_j} \rangle$ of a Gram matrix, the system is non-singular. Note that $a_i = \dfrac{\partial (f \circ \varphi)}{\partial x_i}$ if the system of coordinates is orthonormal at $p$, that is $g_{ij}(q) = 1$, if $i = j$ and $g_{ij}(q) = 0$, if $i \neq j$. This holds in particular if $U = M = \mathbb{R}^n$ and $\varphi$ is the identity map on $U$, so the definition agrees with the usual definition in a Euclidean space.

*Integral lines, and their local structure near singular points.*

In the sequel $M$ is a *compact* submanifold of $\mathbb{R}^n$. The gradient of a smooth function $f$ on $M$ is a smooth vector field on $M$. For every point $p$ of $M$, there is a unique curve $x \colon \mathbb{R} \to M$, such that $x(0) = p$ and $x'(t) = \operatorname{grad} f(x(t))$, for all $t \in \mathbb{R}$. The image $x(\mathbb{R})$ is called the *integral curve* of the gradient vector field through $p$.

**Lemma 1.** *Let $f \colon M \to \mathbb{R}$ be a smooth function on a submanifold $M$ of $\mathbb{R}^n$.*

1. *The integral curves of a gradient vector field of $f$ form a partition of $M$.*
2. *The integral curve $x(t)$ through a singular point $p$ of $f$ is the constant curve $x(t) = p$.*
3. *The integral curve $x(t)$ through a regular point $p$ of $f$ is injective, and both $\lim_{t\to\infty} x(t)$ and $\lim_{t\to-\infty} x(t)$ exist. These limits are singular points of $f$.*
4. *The function $f$ is strictly increasing along the integral curve of a regular point of $f$.*
5. *Integral curves are perpendicular to regular level sets of $f$.*

The proof is a bit technical, so we skip it. See [191] for details. The first property implies that the integral curves through two points of $M$ are disjoint or coincide. The third property implies that a gradient vector field does not have closed integral curves. The limit $\lim_{t\to\infty} x(t)$ is called the $\omega$-limit of $p$, and is denoted by $\omega(p)$. Similarly, $\lim_{t\to-\infty} x(t)$ is the $\alpha$-limit of $p$, denoted by $\alpha(p)$. Note that all points on an integral curve have the same $\alpha$-limit and the same $\omega$-limit. Therefore, it makes sense to refer to these points as the $\alpha$-limit and $\omega$-limit of the integral curve. It follows from Lemma 1.2 that $\omega(p) = p$ and $\alpha(p) = p$ for a singular point $p$.

*Stable and unstable manifolds.*

The structure of integral lines of a gradient vector field $\operatorname{grad} f$ near a singular point can be quite complicated. However, for Morse functions, the situation is simple. To gain some intuition, let us consider the simple example of the function $f(x_1, x_2) = x_1^2 - x_2^2$ on a neighborhood of the non-degenerate singular point $0 \in \mathbb{R}^2$. The gradient vector field is $2x_1e_1 - 2x_2e_2$, where $e_1, e_2$ is the standard basis of $\mathbb{R}^2$. The integral line $(x_1(t), x_2(t))$ through a point $p = (p_1, p_2)$ is determined by $x_1(0) = p_1$, $x_2(0) = p_2$, and

$$\begin{cases} x_1'(t) = 2x_1(t) \\ x_2'(t) = -2x_2(t) \end{cases}$$

Therefore, the integral curve through $p$ is $(x_1(t), x_2(t)) = (p_1e^{2t}, p_2e^{-2t})$, which is of the form $x_1x_2 = c$. See Fig. 7.13 (Left). The singular point

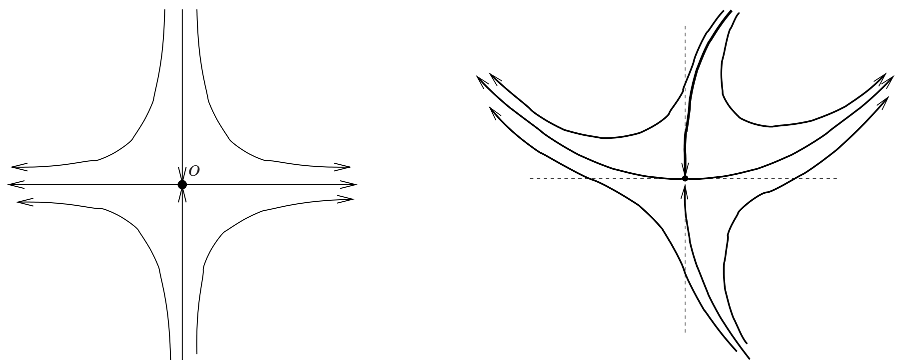



**Fig. 7.13.** Left: Integral curves of the gradient of $f(x_1, x_2) = x_1^2 - x_2^2$ on a neighborhood of the singular point $(0, 0) \in \mathbb{R}^2$. Right: Integral curves of the gradient vector field near a general saddle point of a function on $\mathbb{R}^2$.

$o = (0, 0)$ is the $\alpha$-limit of all points on the horizontal axis, and the $\omega$-limit of all points on the vertical axis. The general structure of integral curves near a saddle point is similar, as indicated by Fig. 7.13 (Right). The stable curve of $p$ consists of all points with $\omega$-limit equal to $p$. The unstable curve is defined similarly. These curves intersect each other at $p$, and are perpendicular there.

More generally, the stable manifold of a singular point $p$ is the set $W^s(p) = \{q \in M \mid \omega(q) = p\}$. Similarly, the unstable manifold of $p$ is the set $W^u(p) = \{q \in M \mid \alpha(q) = p\}$. Note that both $W^s(p)$ and $W^u(p)$ contain the singular point $p$ itself. Furthermore, the intersection of the stable and unstable manifolds of a singular point consists just of the singular point: $W^s(p) \cap W^u(p) = \{p\}$. Stable and unstable manifolds of gradient systems are submanifolds [211, Chapter 6]. The dimension of $W^s(p)$ is equal to the number of negative eigenvalues of the Hessian of $f$ at $p$, whereas the dimension of $W^u(p)$ is equal to the number of positive eigenvalues of this Hessian. Stable and unstable manifolds of gradient systems are submanifolds [211, Chapter 6].

*The Morse-Smale complex.*

A Morse function on $M$ is called a *Morse-Smale function* if its stable and unstable manifolds intersect transversally, i.e., at a point of intersection the tangent spaces of the stable and unstable manifolds together span the tangent space of $M$. If $p$ and $q$ are distinct singular points, the intersection $W^s(p) \cap$

$W^u(q)$ consists of all regular integral curves with $\omega$-limit equal to $p$ and $\alpha$-limit equal to $q$. In particular, a Morse-Smale function on a two-dimensional manifold has no integral curves connecting two saddle points, since the stable manifold of one of the saddle points and the unstable manifold of the second saddle point would intersect non-transversally along this connecting integral curve.

Morse-Smale functions form an open and dense subset of the space of functions on a compact manifold [314].

The *Morse-Smale complex* associated with a Morse-Smale function $f$ on $M$ is the subdivision of $M$ formed by the connected components of the intersections $W^s(p) \cap W^u(q)$, where $p$ and $q$ range over all singular points of $f$, see Fig. 7.14. The Morse-Smale complex is a CW-complex. In geographical literature, the Morse-Smale complex is known as the *surface network*.

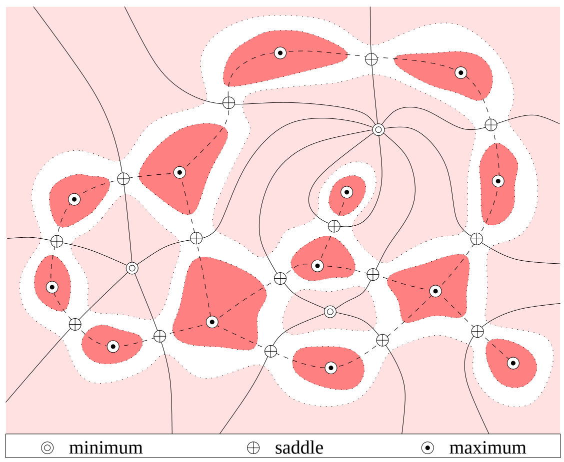



**Fig. 7.14.** The Morse-Smale complex of a function on the plane. The stable one-manifolds are solid, the unstable one-manifolds are dashed. (Courtesy Herbert Edelsbrunner.)

The Morse-Smale complex on a two-manifold consists of cells of dimension 0, 1 and 2, called vertices, edges and regions. According to the *Quadrangle Lemma* [131], each region of the Morse-Smale complex is a quadrangle with vertices of index 0, 1, 2, 1, in this order around the region. Hence the complex is not necessarily a regular CW-complex, since the boundary of a cell is possibly glued to itself along vertices and arcs.

Using a paradigm called *simulation of differentiability*, in [131] the concept of Morse-Smale complex is also defined for piecewise linear functions, and an algorithm for its construction is applied to geographic terrain data. In [130] this work is extend to piecewise linear 3-manifolds.

*Reeb graphs and contour trees.*

The level sets $f^{-1}(h)$ of a Morse function $f$ on a two-dimensional domain change as $h$ varies. At certain values of $h$, components of the level set may disappear, new components may appear, or a component may split into two components, or two components may merge. A component of a level set is called a *contour*. The Reeb graph (after the American journalist John Reeb, 1887–1920 [288]) encodes the changes of contours. It is obtained by contracting every contour to a single point. When $f$ is defined on a simply connected domain (for example, a box), the Reeb graph is a tree, and it is also referred to as the *contour tree*. Fig. 7.15b shows an example of a contour tree of a bivariate function $h = f(x, y)$ defined on a square domain. The vertical axis

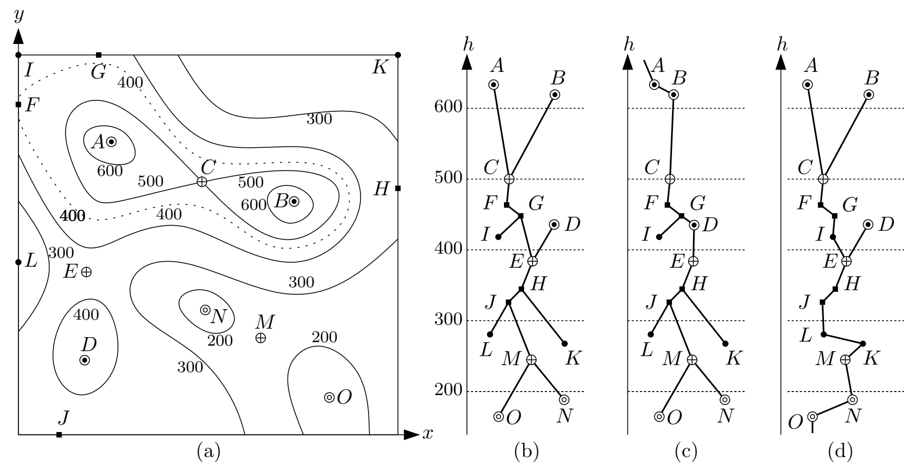



**Fig. 7.15.** (a) a contour map of level sets (isolines), (b) the corresponding contour tree, (c) the join tree, and (d) the split tree. As in Fig. 7.14, minima and maxima are indicated by empty and full circles, and crosses denote saddle points. The points where a contour touches the boundary play also a role in the contour tree (for example, they may be local minima or maxima) but they are not critical points in the sense of having derivative 0. The level sets in (a) are labeled with the height values, and these values are indicated in the trees of (b), (c), and (d). The critical point $F$ changes only the topology of a contour and not the number of contours; when the contour tree is viewed as a discrete structure, $F$ is not a vertex of the tree.

of the contour tree represents the value $h$ of the function. The intersection of a horizontal line at a given value $h$ with the contour tree yields all contours at that level, and the merging or splitting, appearance or disappearance of contours is reflected in vertices of degree 3 and 1 in the contour tree, respectively. Saddle points become vertices of degree 3, and minima and maxima become vertices of degree 1. A contour tree is therefore a good tool to visualize the behavior of a function on a global scale, in particular when it is a function of more than two variables, see [221, 220]. In these applications, $f$ is usually a continuous piecewise linear function interpolating data at given sample points. These functions are not smooth and therefore not Morse functions, but the notion of level sets and Reeb graphs extends without difficulty to this class of functions. It is not uncommon to have *multiple saddle points*, where more than two contours meet at the same time. The Reeb graph has then vertices of degree higher than three. More examples of contour trees are shown in Fig. 5.23 of Sect. 5.5.2.

Note that the Reeb graph only regards the number of components (the 0-homology) of the level sets, it does not reflect every change of topology. For example, in three dimensions, a contour might start as a ball, and as $h$ increases, it might extrude two arms that meet each other, forming a torus, without changing the connectivity between contours. (At this point, we have a saddle of index 1.) In two dimensions, this phenomenon happens only for points on the boundary of the domain, such as the point $F$ in Fig. 7.15.

Figure 7.15c displays the *join tree*, which is defined analogously to the contour tree, except that it describes the evolution of the lower level sets $M_h = f^{-1}([-\infty, h])$ instead of the "ordinary" level sets $f^{-1}(h)$. For example, at $h = 300$ we have three components in the lower level set, as indicated in Fig. 7.16. Since the lower level sets can only get bigger as $h$ increases, they can

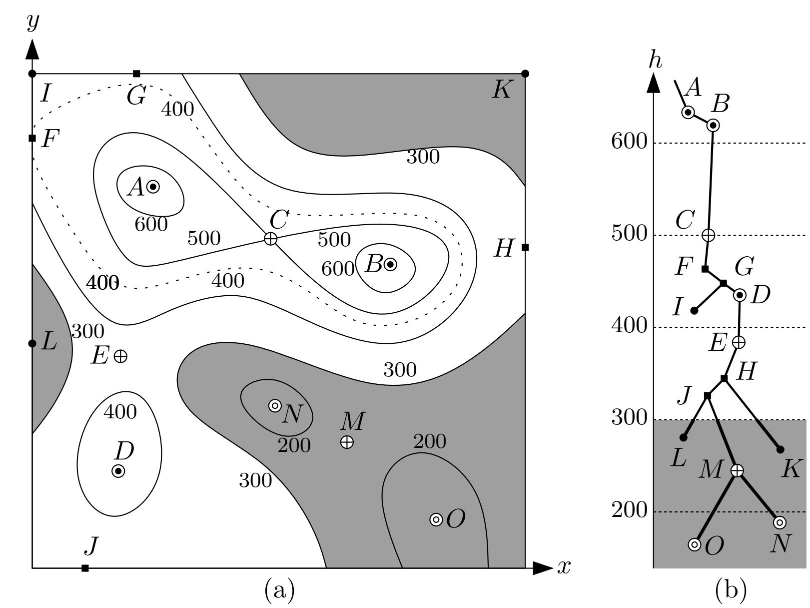
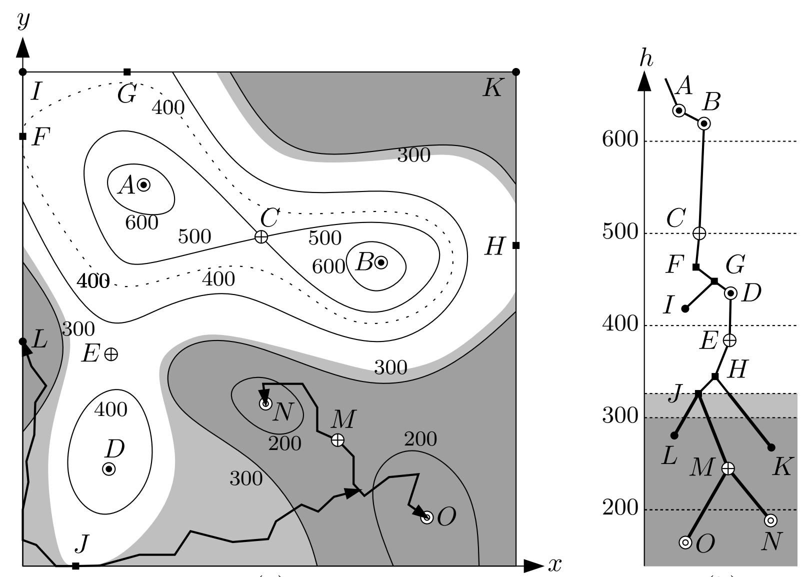



**Fig. 7.16.** (a) the lower level set at level 300 and (b) the corresponding part in the join tree.

only join and never split (hence the name join tree): the tree is a directed tree with the root at the highest vertex. The *split tree* (Fig. 7.15d) can be defined analogously for upper level sets. The join and the split tree are important because it is easier to construct these trees first instead of constructing the contour tree directly. As shown by Carr, Snoeyink, and Axen [74] the contour tree can then be built from the join tree and the split tree in linear time.

The simplest and fastest way to construct the join (and split) tree of a piecewise linear function is the method of *monotone paths*, as described in [88]. We sketch the main idea. This method requires an initial identification of all “critical” vertices: vertices where the topology of the level set changes locally as the level set passes through them. This condition can be checked by scanning the neighboring faces of each vertex independently. These vertices are candidates for becoming vertices of the join tree. They are sorted by function values and processed in increasing order. At each critical vertex $v$ which is not a minimum, we start a monotone decreasing path into each different “local component” of the lower level set in the neighborhood of $v$. For example, if we increase $h$ in Fig. 7.16, the next critical point that is processed is $J$, see Fig. 7.17. Into each of the two shaded regions, we start a descending




(a) (b)

**Fig. 7.17.** Identifying the components that are to be merged by growing descending paths.

path. Each path is continued until it reaches a local minimum (such as the point $L$) or a previously constructed path (such as the descending path from

$M$ that ends in $O$). If we have stored the appropriate information with each path, we can identify the components of the lower level sets that need to be merged (namely, the component $L$ and the component $MNO$; the component $K$ remains separate). Since each path can only descend, it is guaranteed that it cannot leave the lower level set into which it belongs, and therefore it identifies the correct component. It can happen that two descending paths reach the same component. In this case we only have a change of topology of the contour, without changing the number of contours.

This algorithm works in any dimension. If the piecewise linear function $f$ is defined on a triangulated mesh with $t$ cells and there are $n_c$ critical points, the algorithm $O(t + n_c \log n_c)$ time and $O(t)$ space.

Note that the descending paths do not have to follow the steepest direction; thus, unlike the integral curves of the gradient vector field, they can cross the boundaries of the Morse-Smale complex.

With few exeptions [95], the efficient computation of Reeb graphs has been studied mostly for functions on simply connected domains, and hence under the heading of contour *trees*.

## 7.5 Exercises

**Exercise 1 (Triangulations of surfaces).** Prove that the number of vertices in a finite triangulation of a boundaryless surface with Euler characteristic $\chi$ is at least

$$\left\lceil \frac{7 + \sqrt{49 - 24\chi}}{2} \right\rceil .$$

(You should be able to do this exercise without any knowledge of homology theory.)

**Exercise 2 (Non-homeomorphic spaces with equal Betti numbers).** Give an example of two simplicial complexes with equal Betti numbers, but with non-homeomorphic underlying spaces.

**Exercise 3 (Homology of connected graphs).** Let $G$ be a tree. Prove that $\beta_0(G, \mathbb{Q}) = 1$ and $\beta_1(G, \mathbb{Q}) = 0$ using the matrix of the boundary map. (*Hint:* Consider an enumeration of the vertices and oriented edges such that edge $e_i$ is directed from vertex $v_j$ to vertex $v_i$, with $j > i$.)

**Exercise 4 (Chain maps and chain homotopy).** Prove Propositions 2, 3 and 4.

**Exercise 5 (Cone construction and Betti numbers of spheres).** Let $L$ be a finite simplicial complex in $\mathbb{R}^n$, and regard $\mathbb{R}^n$ as the subspace of $\mathbb{R}^{n+1}$ with final coordinate zero. Let $v$ be a point in $\mathbb{R}^{n+1} \setminus \mathbb{R}^n$. If $\sigma$ is a $k$-simplex of $L$ with vertices $v_0, \ldots, v_k$, then the $(k+1)$-simplex with vertices $v, v_0, \ldots, v_k$ is called the *join* of $\sigma$ and $v$. The *cone of $L$ with apex $v$* is the simplicial complex consisting of the simplices of $L$, the join of each of these simplices and $v$, and the 0-simplex $\langle v \rangle$ itself. (One can check that these simplices form a simplicial complex.) Let $K$ be the cone of $L$.

1. Let the map $T_k \colon C_k(K, \mathbb{Q}) \to C_{k+1}(K, \mathbb{Q})$ be defined as follows: Let $\sigma = \langle v_0, \ldots, v_k \rangle$ be a $k$-simplex of $K$. If $\sigma$ is also a $k$-simplex of $L$, then $T_k(\sigma) = \langle v, v_0, \ldots, v_k \rangle$, otherwise $T_k(\sigma) = 0$. Prove that the sequence $\{T_k\}$ is a chain homotopy between the identity map and the zero map on the chain complex $C(K, \mathbb{Q})$.
2. Conclude that $H_k(K, \mathbb{Q}) = 0$, for $k > 0$. What is $H_0(K, \mathbb{Q})$?
3. Determine the Betti numbers of the $d$-dimensional disk, i.e., the space $\mathbb{B}^d = \{(x_1, \ldots, x_d) \in \mathbb{R}^d \mid x_1^2 + \cdots + x_d^2 \leq 1\}$. (*Hint:* Note that a disk is homeomorphic to a $d$-simplex.)
4. Use the previous result, and the incremental homology algorithm to determine the Betti numbers of the $d$-sphere.

**Exercise 6 (Homology of orientable surfaces).**

1. Prove that $\beta_0(K) = 1$ for every triangulation $K$ of an orientable surface of genus $g$ (a sphere with $g$ handles).
2. Let $K$ be a simplicial complex whose underlying space is the torus, and let all simplices of $K$ be oriented compatibly. Let $\alpha = \sum_\sigma \sigma$, where the sum ranges over all (oriented) simplices of $K$. Prove that $Z_2(K, \mathbb{Q}) = \mathbb{Q}\alpha$, and that $\beta_2(K, \mathbb{Q}) = 1$.
3. Use the same technique as in part 2 of this exercise to prove that $\beta_2(K, \mathbb{Q}) = 1$ for every triangulation $K$ of an orientable surface of genus $g$.
4. Let $L$ be the subcomplex of $K$ obtained by deleting an arbitrary 2-simplex. Use the incremental algorithm to prove that $\beta_2(L, \mathbb{Q}) = \beta_2(K, \mathbb{Q}) - 1$, and $\beta_i(L, \mathbb{Q}) = \beta_i(K, \mathbb{Q})$, for $i = 0, 1$.
5. Now let $K$ be the simplicial complex of Fig. 7.8. Prove that $L$ simplicially collapses onto the subcomplex $M$, the subgraph of $L$ consisting of the vertices $v_1, \ldots, v_5$ and the edges $v_1v_2$, $v_2v_3$, $v_3v_1$, $v_1v_4$, $v_4v_5$, and $v_5v_1$. Conclude that $\beta_1(K, \mathbb{Q}) = 2$, and $\beta_0(K, \mathbb{Q}) = 1$.
6. Try to generalize this exercise to an orientable surface of genus $g$.

**Exercise 7 (Morse Theory yields Betti numbers).**

1. Use Morse theory to compute the Betti numbers of the $d$-sphere $\mathbb{S}^d$.
2. Compute the Euler characteristic of a surface $M$ with $g$ handles by defining a suitable Morse function on it. Then compute the Betti numbers of this surface. (*Hint:* You may want to use the first and third result of Exercise 6).
3. For a Morse function $f$, let $s$ be a critical point with Morse index $i$. Consider the intersection $L^-(s)$ of the lower level set $f^{-1}((-\infty, f(s)])$ with a small sphere around $s$. Prove that the Euler characteristic of $L^-(s)$ equals $1 - (-1)^i$.

**Exercise 8 (The mountaineer's equation).** For a smooth Morse function on the 2-sphere $\mathbb{S}^2$, the number of peaks and pits (maxima and minima) exceeds the number of passes (saddles) by 2.

**Exercise 9 (Contour trees for bivariate Morse functions).** Show that, for a smooth Morse function on the 2-sphere $\mathbb{S}^2$, a saddle point will always generate a vertex of degree three in the Reeb graph. Use this observation and the previous exercise to prove that the Reeb graph is in fact a tree in this case.