SPRINGER BRIEFS IN APPLIED SCIENCES AND TECHNOLOGY · MATHEMATICAL METHODS

Herbert Edelsbrunner

A Short Course in Computational Geometry and Topology

Springer

SpringerBriefs in Applied Sciences and Technology

Mathematical Methods

Series editor

Anna Marciniak-Czochra, Heidelberg, Germany

For further volumes:
http://www.springer.com/series/11219

Herbert Edelsbrunner

A Short Course in Computational Geometry and Topology

Springer

Herbert Edelsbrunner
Institute of Science and Technology Austria
Klosterneuburg
Austria

ISSN 2191-530X ISSN 2191-5318 (electronic)
ISBN 978-3-319-05956-3 ISBN 978-3-319-05957-0 (eBook)
DOI 10.1007/978-3-319-05957-0
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014936454

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This short book has been written with the purpose to bring the subject of *Computational Geometry and Topology* to a wider audience, namely to the scientists who deal with shapes in their work. To gain some focus, we have limited the scope to a small subset of topics, selected to form a coherent and convincing story. This includes fundamental topics and those that are relevant to applications—and we do not see any contradiction in these two requirements.

Coherence and relevance were also the selection guidelines when we developed the course at the Institute of Science and Technology Austria. Indeed, I have taught versions of the material three times: in 2010/2011 as an advanced course with Paul Bendich, in 2012 as a core course with Michael Kerber, and in 2013 as a core course by myself. I owe both, Paul and Michael, for their help in selecting topics and in finding innovative ways to teach the material. At IST Austria, we do not have departments, and core courses are taught to a mixed population of students: mathematicians, computer scientists, biologists, and neuroscientists. Similar to a typical course in the USA, the material is delivered in two lectures of 75 min each per week, and we have arranged the topics such that each section contains the material for a single lecture. A core course at IST Austria lasts for half a semester, which explains the small number of sections.

In conclusion, we recommend the prospective reader to approach this short book like a text in mathematics, and not like a novel. Read it slowly, pay close attention to detail, think about the concepts, and spend time to digest the material.

Klosterneuburg, Austria, 2013 Herbert Edelsbrunner

Contents

Part II Complexes

Part III Homology

Chapter 1
Roots of Geometry and Topology

Geometric questions have been pondered by people for thousands of years. In contrast, the abstraction to topological questions is only a few hundred years old. According to Galileo Galilei (1623), philosophy is written in a grand book—the Universe—which cannot be understood unless one first learns to comprehend its language, which is mathematics, and its characters, which are triangles, circles, and other geometric figures. These characters will figure prominently throughout this course.

1.1 Platonic Solids

A *convex polyhedron* is the intersection of finitely many closed half-spaces. As a practical exercise, we can build one by slicing off pieces of an apple with straight cuts of a knife. If we do this carefully, we can arrange that all faces are regular polygons of the same size and type, and that all vertices are endpoint of the same number of edges. Examples of such polyhedra with non-empty volume are the tetrahedron, the cube, the octahedron, the dodecahedron, and the icosahedron; see Fig. 1.1. There are only these five examples. They are referred to as *regular polytopes* or as *Platonic solids*, named after then Greek philosopher Plato,[1] who theorized that the elements were constructed from them.

Euclid gave a full description of the five Platonic solids in Book XIII of the Elements [1]. It is interesting that the face vector of the octahedron is the reverse of that of the cube. This is a result of the duality between them: we can map the vertices, edges, faces of the octahedron bijectively to the faces, edges, vertices of the cube so that incidences are preserved. Similarly, there is such a mapping from the dodecahedron, with face vector (20, 30, 12), to the icosahedron, with face vector (12, 30, 20). The vector of the tetrahedron is a palindrome, and there is an incidence preserving map that reverses dimensions to itself.

[1] Born in 428 BC, he was a student of Socrates, and became one of the most influential thinkers of all times.

H. Edelsbrunner, *A Short Course in Computational Geometry and Topology*, SpringerBriefs in Mathematical Methods, DOI: 10.1007/978-3-319-05957-0_1,

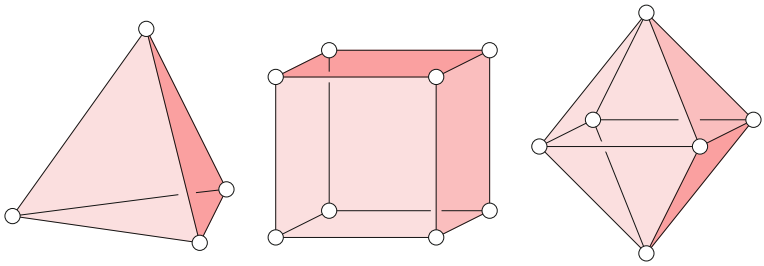

Fig. 1.1 Three of the five Platonic solids together with their face vectors. From *left* to *right*: the tetrahedron, the cube, and the octahedron

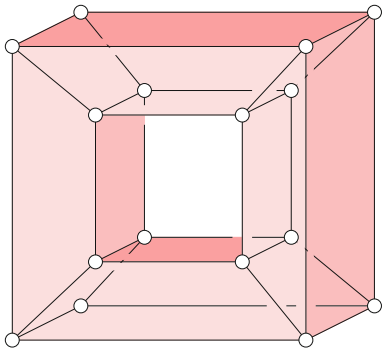

Fig. 1.2 The 'window-frame' polyhedron has a hole through the middle

1.2 Euler Formula

Another pattern we observe for the Platonic solids is that the alternating sum of the face numbers is always the same:

$$\#\text{vertices} - \#\text{edges} + \#\text{faces} = 2. \tag{1.1}$$

This equation hold more generally for bounded convex polyhedra. The relation is originally due to Leonhard Euler and is widely considered the starting point of topology: it is a global statement and it does not depend on the precise geometric shape. Indeed, it does not depend on the convexity of the polyhedron either, but it becomes false if we generalize the class of objects too far. For example, $\#\text{vertices} - \#\text{edges} + \#\text{faces} = 0$ for the 'window-frame' in Fig. 1.2. Indeed, it took more than a century to find a satisfying framework that includes Euler's original observation as a special case and elucidates why and when the relation holds [2]. This generalization is due to Henri Poincaré, which is the reason why the more general result is referred

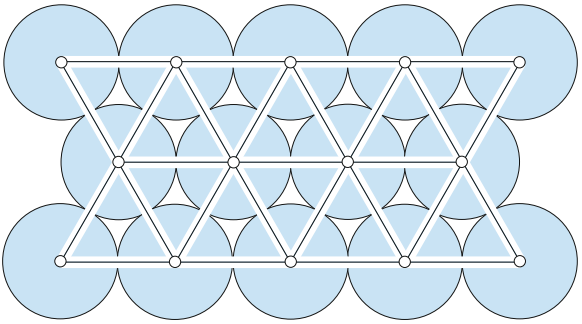

Fig. 1.3 Portion of a densest packing of disks in the plane

to as the Euler-Poincaré formula. It relates the alternating sums of face numbers and Betti numbers:

$$\sum_{i\geq 0}(-1)^i f_i = \sum_{i\geq 0}(-1)^i \beta_i, \tag{1.2}$$

where f_i is the number of i-dimensional faces, and β_i is the ith Betti number. It would be asking too much to explain the meaning of the Betti numbers now, but we will find out in due course. The common value of the two alternating sums is also known as the *Euler* or *Euler-Poincaré characteristic* of the polyhedron.

1.3 Disk Packings

Consider next the question of packing as many disks in a given area as possible. We assume the disks are all the same size, and they can touch each other but not overlap. Arranging pennies on a flat table is a good example. The question is generally difficult because the answer depends on the shape of the area, but it becomes easier when we pack disks in the infinite plane. It is intuitively obvious that the best packing is the arrangement in which the disk centers form the regular hexagonal grid; see Fig. 1.3, and this is also true.

To justify this assertion, we compute the *packing density*, which is the percentage of the plane covered by disks. Since the pattern is the same everywhere, we can just compute the covered fraction of an equilateral triangle spanned by the centers of three mutually touching disks. Assuming the distance between any two of the three points is 2, the area of the triangle is $\sqrt{3}$. There are three disks, each covering a

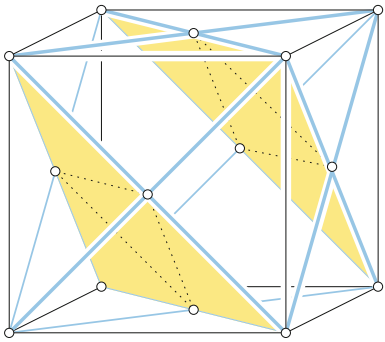

Fig. 1.4 We have a sphere centered at each vertex and at the center of each face of the cube

portion of the triangle with one sixth of its area. It follows the packing density of the hexagonal grid is

$$\varrho_2 = \frac{\pi}{2\sqrt{3}} \quad = \quad 90.68...\%. \tag{1.3}$$

Every other packing of disks in the plane covers at most that fraction of the plane.

1.4 Sphere Packings

The physically more relevant question is how well spheres can be packed in three dimensional space, which is significantly more difficulty than for disks in the plane. The answer has been known since Johannes Kepler, who worked on the problem and stated the conjecture in 1611, but a proof has been completed only recently [3]. The densest packing is perhaps obvious: arrange a layer of spheres in a hexagonal grid, as in Fig. 1.3, and put on top another such layer, resting in the crevices of the first layer. A third layer is added on top of the second layer, and so on. Drawing the layers as diagonal planes of the cube, we notice that one such configuration is obtained by placing the centers of the sphere at the corners and the face centers of a cube tiling; see Fig. 1.4. This gives us a short-cut to computing the packing density of the arrangement, which is the fraction of the cube covered by the spheres (or rather, balls). Assuming the spheres have unit radius, the edges of the cube have length $2\sqrt{2}$. The cube overlaps with one eight of each corner sphere and one half of each face center sphere. The total volume of these pieces is four times the volume of a unit sphere, which is $16\pi/3$. The cube itself has volume $16\sqrt{2}$. The packing density is therefore

$$\varrho_3 = \frac{\pi}{3\sqrt{2}} \quad = \quad 74.04...\%. \tag{1.4}$$

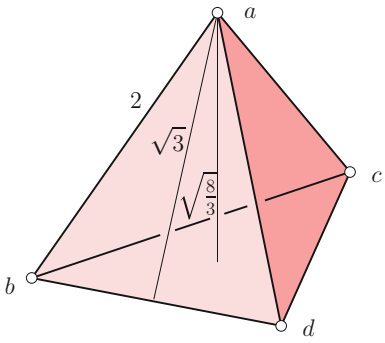

Fig. 1.5 A regular tetrahedron with edges of length 2 has height $\sqrt{8/3}$. Its four triangles have height $\sqrt{3}$ each

1.5 Space Filling

To illustrate the difficulty of the sphere packing question, we note a false belief that goes back to Aristotle,[2] namely that copies of the regular tetrahedron can be used to tile the 3-dimensional Euclidean space. If this were true, then we could compute the best packing density that can be achieved in $\mathbb{R}^3$ by calculating the percentage of the tetrahedron covered by the spheres centered at its four vertices. Take the tetrahedron with vertices $a = (\sqrt{2}, 0, 0, 0)$, $b = (0, \sqrt{2}, 0, 0)$, $c = (0, 0, \sqrt{2}, 0)$, $d = (0, 0, 0, \sqrt{2})$ in $\mathbb{R}^4$, and notice that its six edges all have length 2. The height of each triangular face is the distance between a and $\frac{1}{2}(b+c)$, which is $\sqrt{3}$, and the height of the tetrahedron is the distance between a and $\frac{1}{3}(b+c+d)$, which is $\sqrt{8/3}$; see Fig. 1.5. To compute the volume, we take the area of a triangle, which is $\sqrt{3}$, times the height over 3, which gives $\sqrt{8}/3 = 0.94...$. The dihedral angle between any two faces of the tetrahedron is $2\alpha = 2\arctan(1/\sqrt{2})$, which is about 70.52°. From this, we get the solid angle at any vertex as $6\alpha - \pi = 0.55...$; this is the surface area of the portion of the sphere inside the tetrahedron. The total volume of the four spheres inside the tetrahedron is 4/3 times as much. The fraction of the tetrahedron covered by the four spheres is therefore

$$\frac{4(6\arctan\frac{1}{\sqrt{2}} - \pi)}{\sqrt{8}} = 77.96...\%. \tag{1.5}$$

We see that this percentage is higher than ϱ_3 given in (1.4). It follows that the packing in layers is locally not optimal. This is the deeper reason why proving its global optimality has been so difficult.

[2] He was born in 384 BC, studied under Plato, and was the tutor of Alexander the Great.

References

1. Heath TL (eds) (1956) Euclid. The thirteen books of Euclid's elements, 2nd unabr edn. Dover, England
2. Lakatos I (1976) Proofs and Refutations: the logic of mathematical discovery. Cambridge University Press, Cambridge, England
3. Hales T (2005) A proof of the Kepler conjecture. Ann Math 162:1065–1185

Part I
Tessellations

In this part, we introduce Voronoi diagrams and their dual Delaunay triangulations. They both have myriads of applications in diverse disciplines. Indeed, there is now a conference series on "Voronoi Diagrams in Science and Engineering," which has its 10th meeting this year. While we explain both types of tessellations in two and three dimensions, they are defined in any dimension.

Chapter 2
Voronoi and Delaunay Diagrams

The goal of this book is the introduction of the Voronoi diagram and the Delaunay triangulation of a finite set of points in the plane, and an elucidation of their dual relationship. We begin with convex polygons, which can be constructed as the intersection of a finite number of half-planes or as the convex hull of a finite set of points.

2.1 Convex Polygons

A set X is *convex* if two points $x, y \in X$ imply that every point on the line segment connecting x and y belongs to X. For example, a disk in $\mathbb{R}^2$ is convex but a circle is not. Importantly, convexity is closed under taking intersections.

Lemma A *The intersection of convex sets is convex.*

Indeed, if points x and y belong to the intersection then every point of the line segment belongs to the intersection. A simple convex set is a closed *half-plane*, which consists of all points on or on one side of a line in $\mathbb{R}^2$. Assuming the line does not pass through the origin, we write the half-plane that contains the origin as the set of points $x = (x_1, x_2)$ that satisfy

$$\langle x, u \rangle = x_1 u_1 + x_2 u_2 \leq 1, \tag{2.1}$$

where $u = (u_1, u_2)$ is a non-zero vector. It is a normal vector of the line and its length is one over the distance of the origin from the line. We will use points and vectors interchangeably, so u is also a point in the plane, namely the endpoint of the vector. A *convex polygon* is the intersection of finitely many half-planes. By Lemma A, this intersection is indeed convex. Assuming all these half-planes can be written as in (2.1), the polygon contains the origin in its interior, as in Fig. 2.1.

Every bounded convex polygon can alternatively be constructed by taking the *convex hull* of its vertices. This is the intersection of all convex sets that contain all

H. Edelsbrunner, *A Short Course in Computational Geometry and Topology*,
SpringerBriefs in Mathematical Methods, DOI: 10.1007/978-3-319-05957-0_2,

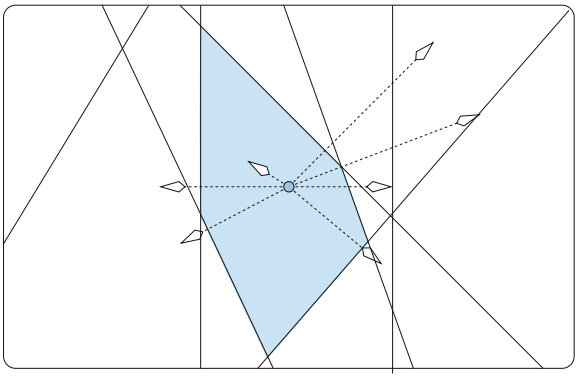

Fig. 2.1 The intersection of finitely many half-planes, each defined by its outward directed normal vector

vertices; in the plane you can visualize it as the region bounded by the rubber band that stretches around the given points.

2.2 Voronoi Diagrams

Let now S be a finite set of points in $\mathbb{R}^2$. We call the elements of S *sites* in order to distinguish them from other points in the plane. Recall that the *Euclidean distance* between a point $x = (x_1, x_2)$ and a site $s = (s_1, s_2)$ is $\|x - s\| = \langle x - s, x - s\rangle^{1/2}$, which is equal to $[(x_1 - s_1)^2 + (x_2 - s_2)^2]^{1/2}$. For each site, we are interested in the region of points that are at least as close as to any other site:

$$V_s = \{x \in \mathbb{R}^2 \mid \|x - s\| \leq \|x - t\|, \forall t \in S\}, \tag{2.2}$$

calling V_s the *Voronoi region* of s. It is named after the Ukrainian mathematician Georgy Voronoi [1, 2]. The set of points x that satisfy $\|x - s\| \leq \|x - t\|$ is a closed half-plane, so V_s is the intersection of finitely many half-planes and therefore a convex polygon. Any two Voronoi regions intersect at most along their boundaries, and together, the Voronoi regions cover the entire plane. The *Voronoi diagram* of S is the set of Voronoi regions, one for each site in S; see Fig. 2.2.

Suppose we pick a point x in the interior of the region V_s and draw a circle with center x and radius $\|x - s\|$. Then s lies on the circle and all other sites lie outside the circle. On the other hand, if x lies on the boundary of V_s then it also belongs to at least one other region, say to V_t. In this case, s and t both lie on the circle. Indeed, if we move x along the edge $V_s \cap V_t$, we get a pencil of circles, all passing through s and t. If x is an endpoint of that edge, then there is typically a third region, V_u, that contains x. In this case, s, t, u all lie on the circle around x and other other sites lie outside this circle.

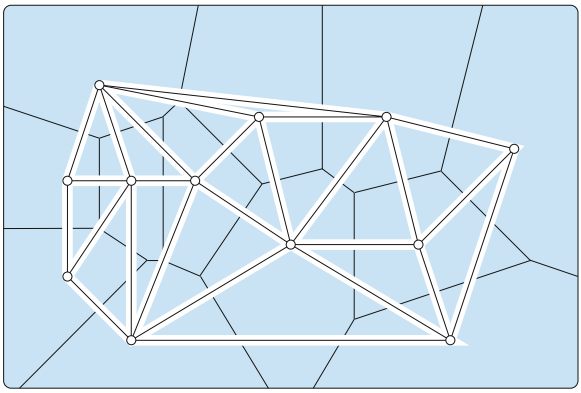

Fig. 2.2 The Voronoi diagram of a finite set of points in $\mathbb{R}^2$, and the corresponding Delaunay triangulation superimposed

2.3 Delaunay Triangulations

Given the Voronoi diagram of $S \subseteq \mathbb{R}^2$, we get the *Delaunay triangulation* by connecting two sites by a straight edge whenever the corresponding two Voronoi regions share an edge [3]. Generically, the intersection of any four or more Voronoi regions is empty. Three Voronoi regions may intersect in a point, which is then the endpoint of the three edges formed by taking pairwise intersections. Correspondingly, the three sites form a triangle in the Delaunay triangulation; see Fig. 2.2. We say the points in S are in *general position* if no four points lie on a common circle. It implies that no four Voronoi regions have a non-empty common intersection. A necessary condition for any notion of general position is that violations happen with probability zero, which is the case here. However, computers are finite and work with finite precision, so violations of such conditions become likely, in particular for large data sets. Still, arbitrary small perturbations can remove the violation, and such perturbations can often be simulated without being explicitly constructed.[1]

Delaunay triangulations have a number of interesting properties, some of which we now list. To explain them, we define the *star* of a site s in the Delaunay triangulation as the collection of edges and triangles that share s. The *link* of s is the collection of sites and edges in the boundary of the star that do not contain s.

Lemma B *Let S be a finite set of sites in general position in $\mathbb{R}^2$.*

(i) Sites $s, t, u \in S$ form a triangle in the Delaunay triangulation iff all other sites lie outside the unique circle that passes through s, t, u.
(ii) The triangles in the Delaunay triangulation decompose the convex hull of S.
(iii) The cyclic list of sites in the link of a site s is the same as the cyclic list of edges of V_s.

[1] Sect. 1.4 in [4]

A standard mistake is to assume that the link of every site bounds a convex polygon. This is not true.

2.4 Planar Graphs

Let n be the number of sites in S. To count the edges and vertices in the Voronoi diagram—as well as the edges and triangles in the Delaunay triangulation—we recall some basic facts about graphs that can be drawn without crossings in the plane. A *graph* consists of finitely many vertices and edges, each edge being a pair of vertices. We draw each vertex as a point and each edge as a curve connecting the points that represent its two vertices. The graph is *planar* if there exists such a drawing in which no two curves cross each other; that is, no two curves intersect other than possibly at shared endpoints. This drawing is called an *embedding* of the graph in the plane. An example is the 1-skeleton of the Delaunay triangulation, which has the extra property that each edge is drawn as a straight line segment. It is not entirely obvious but true that no two of these line segments cross.

Planar graphs cannot have substantially more edges than vertices. To make this precise, let v be the number of vertices, e the number of edges, and f the number of faces of an embedding. Similar to the Euler-Poincaré Formula for convex polyhedra, we have a linear relation provided the graph is connected.

Euler relation *A connected planar graphs satisfies* $v - e + f = 2$.

Proof Draw the vertices as points in $\mathbb{R}^2$. During the first phase of the construction, we add edges, one by one, always making sure that the two endpoints belong to separate components before we add the edge. After drawing $v - 1$ such edges, we reach a connected graph and therefore the end of the first phase. At this stage, we have $e = v - 1$ edges, and because there are no cycles, we have $f = 1$ face. Plugging these values into the alternating sum gives $v - (v - 1) + 1 = 2$, so the relation holds.

During the second phase, we draw the remaining edges, again one by one. Each edge decomposes a face of the current graph into two. Adding an edge thus increases e and f by 1 each. The two changes cancel, so the claimed relation is maintained. □

2.5 Maximally Planar Graphs

A planar graph is *maximal* if adding any new edge would render the graph non-planar. It is plausible that in an embedding of a maximal planar graph, every face is bounded by exactly three edges. This is also true, and because of this property, we have 3 edges per face and 2 faces per edge. Assuming $v \geq 3$, this implies $3f = 2e$. Plugging this equation into the Euler Relation, we get

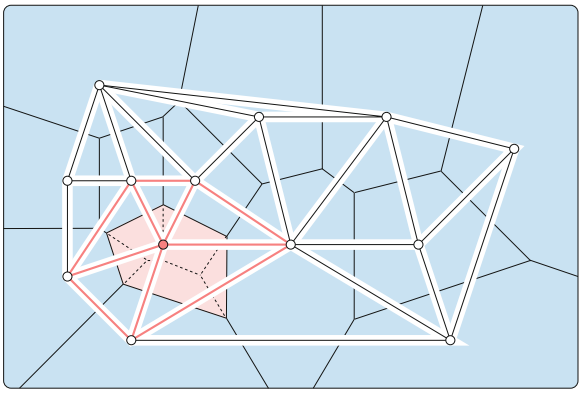

Fig. 2.3 Adding a new site to the Voronoi diagram and the corresponding Delaunay triangulation

$$e = 3v - 6, \tag{2.3}$$

$$f = 2v - 4. \tag{2.4}$$

The Delaunay triangulation can only have fewer edges and faces than the maximal planar graph for the same set of points. Writing n for the number of sites, we therefore get fewer than $3n$ edges and fewer than $2n$ triangles. Correspondingly, the Voronoi diagram of S has fewer than $3n$ edges and fewer than $2n$ vertices.

The *degree* of a vertex is the number of incident edges. Since every edge is incident to only two vertices, the sum of degrees, over all vertices, is less than $6n$. This implies that the average degree in the Delaunay triangulation is less than 6, and that there is at least one vertex of degree at most 5.

2.6 Incremental Construction

A popular algorithm for constructing the Voronoi diagram adds one site at a time. To add a site, s, we first find the Voronoi region that contains s, and second construct the Voronoi region of s by stealing pieces of the surrounding regions. The new Voronoi region is of course convex, which implies that its boundary is connected and can be computed by stepping from one neighboring region to the next, as illustrated in Fig. 2.3. The corresponding algorithm for the Delaunay triangulation removes and adds triangles. The triangles that are removed correspond to the Voronoi vertices that lie in the stolen region, which implies that their circumcircles enclose the new site. The new triangles are all incident to the new site, and they cover the same area in the plane. To implement this algorithm, it is convenient to add three dummy sites that span a triangle containing the entire set S. Starting with this triangle as the initial Delaunay triangulation, we add one site at a time:

Let $\mathcal{D}_0$ consist of the triangle xyz;
`for` $i = 1$ `to` n `do`
1 find the subset of triangles, $A \subseteq \mathcal{D}_{i-1}$,
whose circumcircles enclose the i-th site in S;
2 let B contain the triangles formed by joining
the i-th site to the boundary edges of $\bigcup A$;
3 $\mathcal{D}_i = (\mathcal{D}_{i-1} - A) \cup B$
`endfor`.

To implement this algorithm, we need a test that decides whether a point lies inside or outside the circle that passes through three other points. Such a test will be discussed during the next lecture.

2.7 Expectations

Instead of giving a detailed analysis of the run time, we just count the triangles that are constructed during the course of the algorithm. It is possible to choose the sites, and their ordering, such that this number is a constant times n^2. However, this is not typical, which is a number that is linear in n. To substantiate this claim, we compute the average over the $n!$ orderings of the n sites. Note that there is no assumption on the configuration of the sites.

To understand this average, we count the triangles while we go backward in the ordering. In other words, we ask for the expected number of triangles constructed when we add the last, or n-th site. Every site is equally likely to be the last, so this is the average number of triangles in the star of a vertex, now averaged over all sites, which we know is less than 6. Let t_i be the expected number of triangles constructed (and possible already destructed) after adding the first i sites. Since expectations are additive, we have

$$t_n < t_{n-1} + 6. \tag{2.5}$$

We have $t_3 = 1$ and therefore $t_n < 6n$. In words, if we add the sites in a random order, then the expected number of triangles constructed during the course of the algorithm is less than $6n$. Less than $2n$ appear in the final triangulation, which implies that roughly two thirds of the triangles are wasted. Similarly, the expected number of edges constructed by the algorithm is less than $6n$, and about half of them are wasted.

References

1. Voronoi G (1907) Nouvelles applications des paramètres continus à la théorie des formes quadratiques. J Reine Angew Math 133:97–178
2. Voronoi G (1908) Nouvelles applications des paramètres continus à la théorie des formes quadratiques. J Reine Angew Math 134:198–287

3. Delaunay B (1934) Sur la sphère vide. Izv Akad Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk 7:793–800
4. Edelsbrunner H (2001) Geometry and topology for mesh generation. Cambridge University Press, Cambridge, England

Chapter 3
Weighted Diagrams

Every region in a 2-dimensional Voronoi diagram consists of all points for which the corresponding site minimizes the Euclidean distance. In this section, we modify the notion of distance by introducing weights. Particular attention will be paid to the case of subtracting the weight from the squared Euclidean distance because this gives convex regions, like for the unweighted Euclidean distance.

3.1 Apollonius Diagrams

Let S be a finite set of points or *sites* in the plane. For each $s \in S$, we let $w_s \in \mathbb{R}$ be its weight. Suppose we define the *weighted distance* of a point $x \in \mathbb{R}^2$ from s by subtracting the weight from the Euclidean distance:

$$d_A(s, x) \;=\; \|x - s\| - w_s. \tag{3.1}$$

Drawing the weighted site as a circle with center s and radius $|w_s|$, we can interpret this notion of distance geometrically. If $w_s \geq 0$ and $\|x - s\| \geq w_s$, then $d_A(x, s)$ is the Euclidean distance to the nearest point on the circle. The same interpretation works for $\|x - s\| < w_s$ except that $d_A(s, x)$ is now negative. Finally, if $w_s < 0$, then $d_A(s, x)$ is the Euclidean distance to the furthest point on the circle. To unify the three cases, we draw a vertical cone in $\mathbb{R}^3$, adding a third coordinate to the plane. Its axis of rotation is the vertical line passing through s, its apex is the point $(s, -w_s) \in \mathbb{R}^3$, and its opening angle is 90°; see Fig. 3.1. The weighted distance from x to s is then the vertical distance from x to the cone. In other words, (the surface bounding) the cone is the graph of the function that maps x to the weighted distance from s. Let now s and t be two weighted sites. The *bisector* consists of all points x with equal weighted distance from both: $d_A(s, x) = d_A(t, x)$ or, equivalently, $\|x - s\| - \|x - t\| = w_s - w_t$. This is the equation of a hyperbola; it is the vertical projection of the intersection of the two cones to $\mathbb{R}^2$. On one side of the bisector,

H. Edelsbrunner, *A Short Course in Computational Geometry and Topology*,
SpringerBriefs in Mathematical Methods, DOI: 10.1007/978-3-319-05957-0_3,

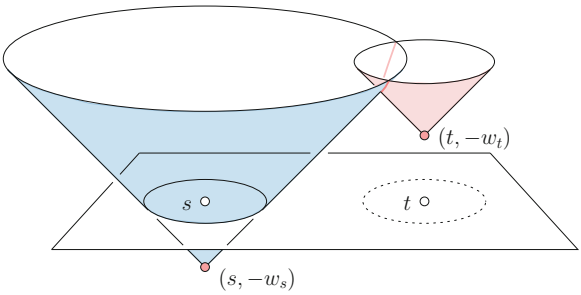

Fig. 3.1 The cones of two sites in the plane, one with positive and the other with negative weight

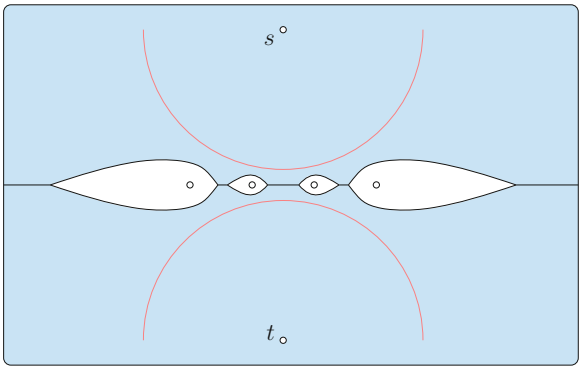

Fig. 3.2 The Apollonius diagram of six sites. The intersection of the regions of s and t consists of five segments

we have $d_A(s, x) < d_A(t, x)$, and on the other side we have the reverse inequality. Returning to the set of finitely many sites with real weights, the region in which s minimizes the weighted distance is

$$A_s = \{x \in \mathbb{R}^2 \mid d_A(s, x) \leq d_A(t, x), \quad \forall t \in S\}, \tag{3.2}$$

and the *Apollonius diagram* of S is the set of such regions. In contrast to the Voronoi diagram, the regions in the Apollonius diagram are not necessarily convex; only the region of the site with the smallest weight is guaranteed to be convex. Nevertheless, every non-empty region is connected. In contrast, the intersection of two regions is not necessarily connected; see Fig. 3.2. Recall how we defined the Delaunay triangulation by connecting sites whose Voronoi regions have a non-empty intersection. Because of the more complicated intersections, this construction is no longer as straightforward for the Apollonius diagram.

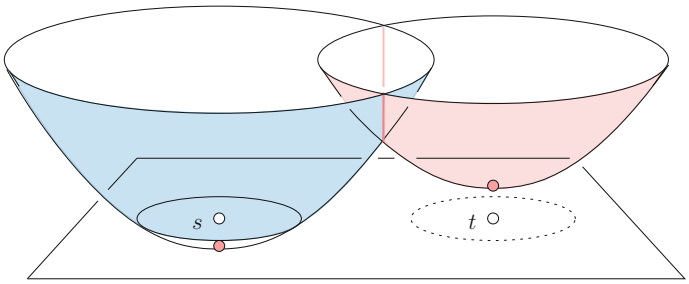

Fig. 3.3 The paraboloids of two sites in the plane, one with positive and the other with negative weight

3.2 Power Diagrams

Instead of subtracting the weight from the Euclidean distance, we now subtract it from the square of the Euclidean distance:

$$d_P(s,x) \;=\; \|x-s\|^2 - w_s. \tag{3.3}$$

Here, $d_P(s,x)$ is called the *power* or *power distance* of x from s. Recall that the bisector consists of all points that satisfy $d_P(s,x) = d_P(t,x)$. Writing the squared Euclidean distance as a scalar product, we get

$$\|x-s\|^2 \;=\; \langle x-s, x-s\rangle \;=\; \|x\|^2 - 2\langle x,s\rangle + \|s\|^2.$$

The equation of the bisector is therefore $2\langle x, t-s\rangle = \|t\|^2 - \|s\|^2 + w_s - w_t$, which is the equation of a line. We can see this geometrically, by drawing the graph of the function that maps a point to its power distance from s; it is a paraboloid with vertical axis of rotation and lowest point at $(s, -w_s)$. We also draw the graph of the function for t, which is a translate of the paraboloid for s; see Fig. 3.3. The intersection of the two paraboloids is a parabola that lies in a vertical plane; its projection to $\mathbb{R}^2$ is a line. Returning to the set of finitely many sites with weights, the region within which s minimizes the power distance is

$$P_s \;=\; \{x \in \mathbb{R}^2 \mid d_P(s,x) \le d_P(t,x), \quad \forall t \in S\}, \tag{3.4}$$

and the *power diagram* of S is the set of such regions. Since the bisectors are straight lines, each region is the intersection of half-planes and therefore convex. In contrast to the unweighted case, a site may have an empty region in the power diagram. Otherwise, the power diagrams are visually difficult to distinguish from Voronoi diagrams, which is perhaps the reason why they are sometimes referred to as *weighted Voronoi diagrams*.

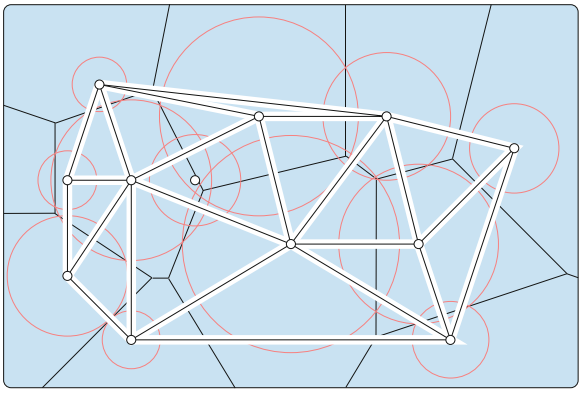

Fig. 3.4 The power diagram and the superimposed weighted Delaunay triangulation of twelve sites in the plane. Note that only eleven of the sites function as vertices of the triangulation

3.3 Weighted Delaunay Triangulations

Having convex regions, we can again draw the dual, connecting two sites with an edge whenever the corresponding two power regions share an edge; see Fig. 3.4. Similar to the unweighed case, this construction typically gives a triangulation, but unlike the unweighted case, not every site is necessarily also a vertex in that triangulation. To understand the construction, it will be useful to generalize the circumscribed circles that characterize the triangles in the unweighted Delaunay triangulation. To this end, we say a point x with weight w_x is *orthogonal* to s if

$$\|x - s\|^2 \;=\; w_s + w_x. \tag{3.5}$$

Drawing s as the circle with center s and radius $\sqrt{w_s}$, and similarly for x, we get two circles that meet at a right angle.[1] Given s, w_s, and x, we can always find w_x such that the two weighted points are orthogonal. In fact, the weight for which this is the case is unique. However, if in addition we are given t and w_t, we can find w_x such that x is orthogonal to s as well as to t only if x lines on the bisector of s and t. Finally, if in addition we are given u and w_u, then we can find w_x such that x is orthogonal to all three only if x lies on all three bisectors or, equivalently, it is the unique point at which the three regions in the power diagram of s, t, u meet.

We can now formulate a criterion for stu being a triangle in the weighted Delaunay triangulation of S. In the unweighted case, we said S is in general position if no four sites lie on a common circle. The appropriate notion in the weighted case is that no four unweighted sites have a common orthogonal circle.

[1] For this to be true, we have to assume that both weights are positive. A similar but geometrically not quite as compelling interpretation can also be found for the case in which one of the weights is non-positive.

Lemma A *Let S be a finite set of weighted sites in general position in $\mathbb{R}^2$. Then $s, t, u \in S$ form a triangle in the weighted Delaunay triangulation of S iff $\|x - v\|^2 \geq w_v + w_x$ for every $v \in S$, where x is the unique weighted point orthogonal to s, t, and u.*

3.4 Geometric Primitives

There are many algorithms for weighted and unweighted Delaunay triangulations, one being the incremental construction briefly mentioned at the end of last lecture. All these algorithms need to be able to decide whether the circle that passes through three sites is empty (see Lemma B (i) in Chap. 2) or, in the weighted case, whether the circle orthogonal to three weighted sites is further than orthogonal from all other sites (see Lemma A). We now study the details of these decisions, beginning with the unweighted case.

Let $a = (a_1, a_2)$, $b = (b_1, b_2)$, $c = (c_1, c_2)$, be three points in $\mathbb{R}^2$. If the points lie on a common line, then we can write $c = (1-\lambda)a + \lambda b$, assuming $a \neq b$. Hence, the determinant of

$$\Delta = \begin{bmatrix} 1 & a_1 & a_2 \\ 1 & b_1 & b_2 \\ 1 & c_1 & c_2 \end{bmatrix} \tag{3.6}$$

vanishes. Indeed, $\det \Delta = 0$ iff a, b, c are collinear. In addition, $\det \Delta > 0$ iff a, b, c form a left-turn. We can formulate a similar test for cocircularity. Lift a to the point $a^+ = (a_1, a_2, a_1^2 + a_2^2)$ in $\mathbb{R}^3$, and similarly for b and c. Furthermore, let $v = (v_1, v_2)$ be a fourth point, write $v^+ = (v_1, v_2, v_1^2 + v_2^2)$, and define

$$\Gamma = \begin{bmatrix} 1 & a_1 & a_2 & a_1^2 + a_2^2 \\ 1 & b_1 & b_2 & b_1^2 + b_2^2 \\ 1 & c_1 & c_2 & c_1^2 + c_2^2 \\ 1 & v_1 & v_2 & v_1^2 + v_2^2 \end{bmatrix}. \tag{3.7}$$

The crucial insight is that $\det \Gamma = 0$ iff v lies on the circle determined by a, b, c. To see this, we interpret the determinant as an expression of the points a^+, b^+, c^+, v^+ in $\mathbb{R}^3$. It vanishes iff the four points lie on a common plane. Now, we just need to verify that being coplanar in $\mathbb{R}^3$ corresponds to being cocircular in $\mathbb{R}^2$. To see this, we intersect the paraboloid given by $x_3 = x_1^2 + x_2^2$ with the plane given by $x_3 = 2Ax_1 + 2Bx_2 + C$. Eliminating x_3, we get $(x_1 - A)^2 + (x_2 - B)^2 - (A^2 + B^2 - C) = 0$, which is the equation of a circle; see Fig. 3.5.

Lemma B *Let a, b, c, v be points in $\mathbb{R}^2$ and Δ, Γ the matrices defined in (3.6) and (3.7). Then the point v belongs to the open disk bounded by the circle passing through a, b, c iff* $\det \Delta \cdot \det \Gamma < 0$.

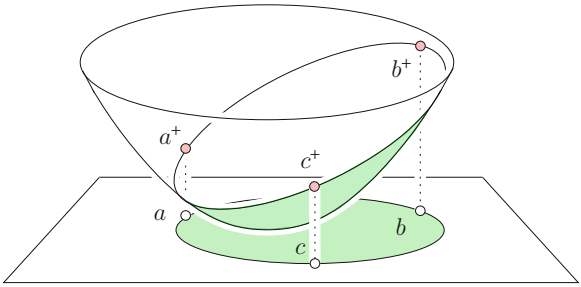

Fig. 3.5 Lifting a, b, c to the paraboloid, we get the circle that passes through them by projecting the intersection between the paraboloid and the plane passing through the lifted points

We omit the proof, which consists of two parts: verifying that v lies on the circle if $\det \Gamma = 0$, and making sure that the sign is negative inside and positive outside the circle. We note that the extra factor, $\det \Delta$, is necessary because switching two points of a, b, c changes the sign of the determinant without changing the geometric configuration. The test explained in Lemma B can be generalized to the weighted case by changing the matrix in (3.7) to

$$\Gamma_W = \begin{bmatrix} 1 & a_1 & a_2 & a_1^2 + a_2^2 - w_a \\ 1 & b_1 & b_2 & b_1^2 + b_2^2 - w_b \\ 1 & c_1 & c_2 & c_1^2 + c_2^2 - w_c \\ 1 & v_1 & v_2 & v_1^2 + v_2^2 - w_v \end{bmatrix}. \tag{3.8}$$

Indeed, if we project the intersection of the paraboloid with the plane that passes through the thus lifted points of a, b, c, then we get the unique orthogonal circle. We can now modify Lemma B.

Lemma C *Let a, b, c, v be weighted points in $\mathbb{R}^2$ and Δ, Γ_W the matrices defined in (3.6) and (3.8). Letting x with w_x be the weighted point orthogonal to a, b, c, we have $\|v - x\|^2 < w_v + w_x$ iff $\det \Delta \cdot \det \Gamma_W < 0$.*

Note that Lemma C agrees with Lemma B when the weights of a, b, c, v vanish.

Chapter 4
Three Dimensions

Voronoi diagrams and Delaunay triangulations in $\mathbb{R}^3$ are more interesting and more difficult to understand than in $\mathbb{R}^2$. In this section, we develop some intuition by considering these tessellations for a few symmetric point sets.

4.1 Lattices

The *lattice* generated by a set of d linearly independent vectors in $\mathbb{R}^d$ consists of all integer combinations: $S = \{\sum_{i=1}^{d} k_i \mathbf{u}_i \mid k_i \in \mathbb{Z} \text{ for all } i\}$. In the plane, we just need two vectors to define a lattice. The most common example is the *square lattice* generated by the vectors $\mathbf{u}_1 = (1, 0)$ and $\mathbf{u}_2 = (0, 1)$; see Fig. 4.1. The vectors are not unique. For example, the square lattice is also generated by the vectors $(2, 1)$ and $(1, 1)$. Every other lattice in the plane is the image of the square lattice under a linear map, but when we construct the Voronoi diagrams, we see significant differences.

Before we get there, we observe that every site in a lattice is locally the same. More precisely, for any two sites $a, b \in S$, the translation by the difference vector leaves the lattice invariant: $S = S + (b - a)$. This implies that the Voronoi regions are the same: $V_b = V_a + (b - a)$. Furthermore, the central reflection through the origin leaves the lattice invariant: $S = -S$. It follows that the Voronoi region of the origin is centrally symmetric. Since all other Voronoi regions are translates of this one, all Voronoi regions are centrally symmetric. In the plane, the only centrally symmetric convex polygons have $2k$ edges, for k an integer larger or equal to 2. Of these, only 4-gons and 6-gons can be used to tile the plane. We thus have really only two types of Voronoi diagrams, both exemplified in Fig. 4.1. The square lattice on the left is not in general position. Indeed, the vertices in the Voronoi diagram have degree 4, so when we take the dual, we get quadrangles instead of triangles. The Voronoi regions on the right are hexagons, and they meet in triples at vertices of the diagram. In this case, the dual is a triangulation in which every vertex has degree 6.

H. Edelsbrunner, *A Short Course in Computational Geometry and Topology*,
SpringerBriefs in Mathematical Methods, DOI: 10.1007/978-3-319-05957-0_4,

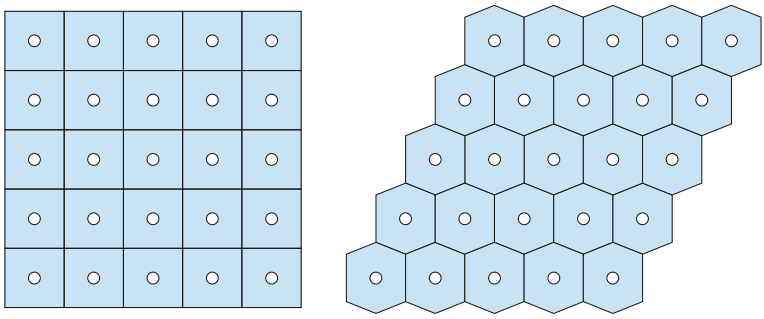

Fig. 4.1 The Voronoi diagram of the square lattice on the *left*, and that of another lattice on the *right*

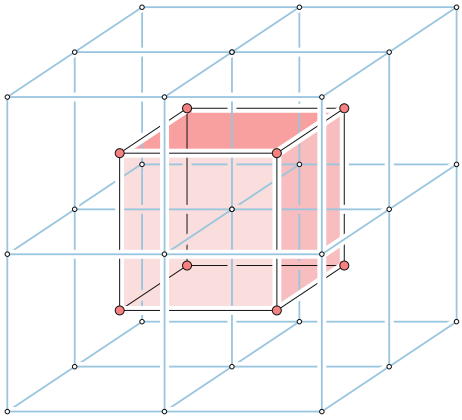

Fig. 4.2 A subset of 27 sites in the cube lattice, and the Voronoi region of the site at the center

4.2 Cube Lattice

We write $2\mathbb{Z}^3$ for the set of points with even integer coordinates in $\mathbb{R}^3$. It is a *3-dimensional lattice* since we can write $2\mathbb{Z}^3$ as the set of all integer combinations of three vectors: $(2, 0, 0)$, $(0, 2, 0)$, $(0, 0, 2)$. The Voronoi region of any site is a cube with sides of length 2; see Fig. 4.2. Similar to the square lattice, $2\mathbb{Z}^3$ is not in general position, which for Voronoi diagrams in $\mathbb{R}^3$ would require that no five points lie on a common sphere. For this reason, we do not get a triangulation when we dualize. Indeed, the Delaunay triangulation of $2\mathbb{Z}^3$ consists of cubes that tile $\mathbb{R}^3$; it is still called a triangulation out of convention. The only difference between the Delaunay triangulation and the Voronoi diagram is that cubes of the former have their vertices in $2\mathbb{Z}^3$, while the cubes of the latter have their centers in $2\mathbb{Z}^3$.

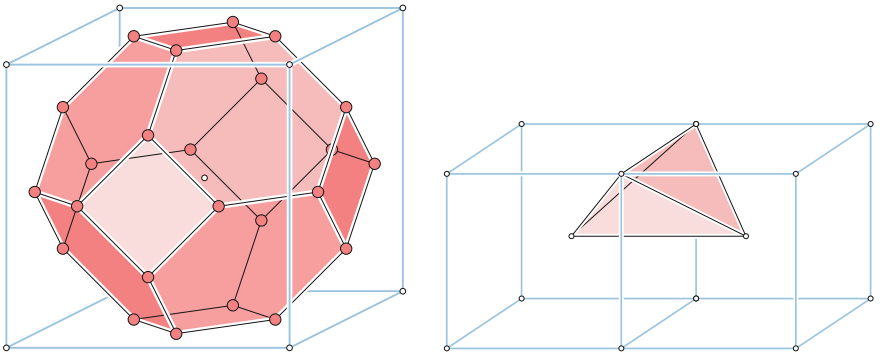

Fig. 4.3 *Left* the Voronoi region of a site in the BCC lattice. *Right* a tetrahedron in the Delaunay triangulation of the BCC lattice

4.3 BCC Lattice

We get the *body-centered cubic lattice* from $2\mathbb{Z}^3$ by adding the centers of the cubes to the set. Formally, we write $B = 2\mathbb{Z}^3 \cup \big(2\mathbb{Z}^3 + (1, 1, 1)\big)$. In other words, a point belongs to B iff all its coordinates are even integers or all its coordinates are odd integers. We call it a lattice since it is again generated by three vectors: $\mathbf{u}_1 = (1, 1, -1)$, $\mathbf{u}_2 = (1, -1, 1)$, $\mathbf{u}_3 = (-1, 1, 1)$. To see this, we note that $\mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3 = (1, 1, 1)$, and the pairwise sums give $(2, 0, 0)$, $(0, 2, 0)$, $(0, 0, 2)$. The latter generate $2\mathbb{Z}^3$, and adding $(1, 1, 1)$ gives the remaining sites. It is also clear that the vectors generate no points other than those in B. Recall that the Voronoi regions of the sites are translates of each other and that they are centrally symmetric. To draw the Voronoi region of the origin, we note that it is the intersection of its Voronoi region within $2\mathbb{Z}^3$, which is the cube $[-1, 1]^3$, and the Voronoi region within $2\mathbb{Z}^3 + (1, 1, 1)$, which is a regular octahedron. Intersecting the two gives the *truncated octahedron* drawn in Fig. 4.3.

It is also interesting to draw the Delaunay triangulation of the BCC lattice. We have *long* edges of length 2 and *short* edges of length $\sqrt{3}$. Every long edge connects the centers of two neighboring cubes, and every short edge connects the center of a cube with one of its corners. Every tetrahedron in the Delaunay triangulation consists of two long and four short edges; see Fig. 4.3. The dihedral angle at a long edge is $90°$, so that four tetrahedra fit around the edge. In contrast the dihedral angle at a short edge is $60°$, with six tetrahedra fitting around the edge. Indeed, these two types of edges correspond to the quadrangles and the hexagons in the dual Voronoi diagram.

4.4 FCC Lattice

We get the *face-centered cubic lattice* by adding the centers of the faces to the cube lattice. Formally, we define F as the set of points with integer coordinates that add up to an even number. It is again a lattice, generated by $\mathbf{v}_1 = (1, 1, 0)$,

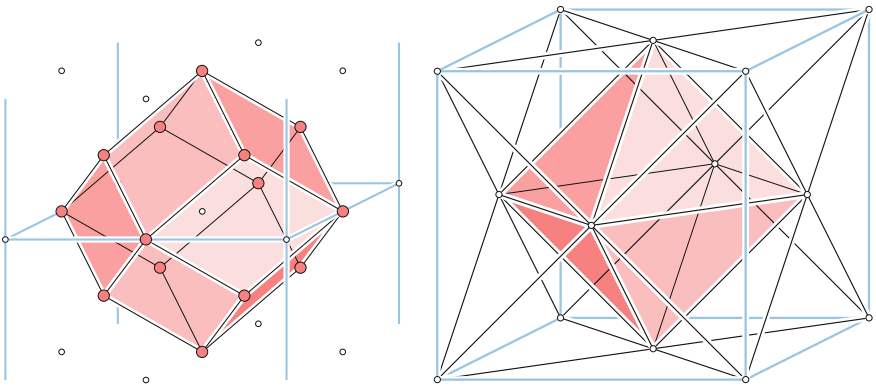

Fig. 4.4 *Left* the Voronoi region of a site in the FCC lattice. *Right* an octahedron and the eight neighboring tetrahedra in the Delaunay triangulation of the FCC lattice

$\mathbf{v}_2 = (1, 0, 1)$, $\mathbf{v}_3 = (0, 1, 1)$. To see that these vectors generate F, we note that $\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3 = (2, 0, 0)$, $\mathbf{v}_1 - \mathbf{v}_2 + \mathbf{v}_3 = (0, 2, 0)$, $-\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 = (0, 0, 2)$ generate $2\mathbb{Z}^3$, and that adding $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ gives the face centers. It is clear that the vectors generate no other points.

The Voronoi region of a site is show in Fig. 4.4. It is the intersection of a non-regular octahedron with a square cylinder. In contrast to the BCC lattice, the dual Delaunay triangulation does not just consist of tetrahedra. To see this, note that the Voronoi region has two kinds of vertices: of degree 3 and of degree 4. Every degree-3 vertex is shared by three other Voronoi regions, implying that it corresponds to a tetrahedron in the Delaunay triangulation. In contrast, every degree-4 vertex is shared by five other Voronoi regions, and the dual cell in the Delaunay triangulation is an octahedron. Both the tetrahedron and the octahedron are regular, as illustrated in Fig. 4.4.

4.5 Quadratic Example

In all the examples we have seen so far, the number of edges, faces, and cells in the Delaunay triangulation are at most some constant times the number of sites. This is not true in general. Indeed, if we place $n/2$ sites on a line, and another $n/2$ sites on a second, skew line, then we have more than $n^2/4$ edges in the Delaunay triangulation; see Fig. 4.5. This number can be further increased to $\binom{n}{2}$ by being more careful in how we place the points in $\mathbb{R}^3$. To describe how this is done, we define the *moment curve* as the set of points $M(t) = (t, t^2, t^3)$, with $t \in \mathbb{R}$. Suppose that S consists of n points chosen from the moment curve. We claim that the Delaunay triangulation of S contains all $\binom{n}{2}$ edges, which is clearly the maximum number possible. To see this, we observe that any sphere intersects the moment curve in at most four points.

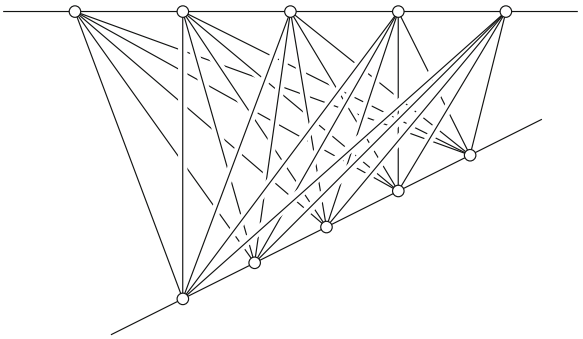

Fig. 4.5 Delaunay triangulation of sites that lie on two it skew lines

Indeed, we can write a sphere as the set of points that satisfy

$$(x_1 - A)^2 + (x_2 - B)^2 + (x_3 - C)^2 - R^2 = 0.$$

Substituting t for x_1, t^2 for x_2, and t^3 for x_3, we get a polynomial equation of degree 6:

$$\begin{aligned} 0 &= (t - A)^2 + (t^2 - B)^2 + (t^3 - C)^2 - R^2 \\ &= t^6 + t^4 - 2Ct^3 + (1 - 2B)t^2 - 2At + (A^2 + B^2 + C^2 - R^2). \end{aligned}$$

A theorem of Descartes asserts that the number of roots is bounded from above by the number of sign changes when we read the coefficients in sequence.[1] Here we have $1, 0, 1, -2C, 1-2B, -2A$, and $A^2+B^2+C^2-R^2$. No matter how we choose A, B, C, R, we cannot have more than four sign changes, which implies the claimed upper bound on the number of intersections.

Why does this imply that the Delaunay triangulation of S has $\binom{n}{2}$ edges? To prove this, let $t_1 < t_2 < \cdots < t_n$ be the parameters that define the n sites. For any four indices $1 \leq i < i+1 < j < j+1 \leq n$, we consider the sphere that passes through the four corresponding points. It intersects the moment curve in these four points and in no others. It follows that the segments between $M(t_i)$ and $M(t_{i+1})$ and between $M(t_j)$ and $M(t_{j+1})$ lie inside the sphere, and the other three segments along the moment curve lie outside the sphere. Hence, the tetrahedron spanned by the four points belongs to the Delaunay triangulation. Every one of the $\binom{n}{2}$ edges belongs to at least one such tetrahedron and therefore also to the Delaunay triangulation, as claimed.

[1] This is known as *Descarte's sign rule* and was first described in his *La Géométrie*, which was published in 1637 as an appendix to his *Discour de la méthod*, where he presents his method for obtaining clarity on any subject.

Part I
Tessellations

Exercises

Question 1. (20 points). Why are there only five Platonic solids? Did we miss some, or is there a reason there is not a sixth one?

Question 2. (20 points). No two edges of the Delaunay triangulation of a finite set of points in the plane cross each other. Give a convincing argument why this is the case.

Question 3. (20 points). All faces of an embedding of a maximal planar graph with three or more vertices in the plane are bounded by exactly tree edges. Give a convincing argument that this is the case.

HINT you might find Kuratowski's Theorem on planar graphs useful, which says, among other things, that the complete graph of 5 vertices is not planar.

Question 4. (20 points). Let S be a finite set of sites in $\mathbb{R}^2$, each with a real weight. Show that every region in the Apollonius diagram of S is connected.

Question 5. (20 points). Let $a = (a_1, a_2), b = (b_1, b_2), x = (x_1, x_2)$ be three points in $\mathbb{R}^2$. Prove that *a, b, x* form a left-turn iff $\det \begin{bmatrix} 1 & a_1 & a_2 \\ 1 & b_1 & b_2 \\ 1 & x_1 & x_2 \end{bmatrix} > 0$.

Question 6. ($20 = 10 + 10$ points). Consider the Delaunay triangulation of the BCC lattice in $\mathbb{R}^3$. Recall that it consists of congruent copies of only one tetrahedron, which has two long edges (of length 2) and four short edges (of length $\sqrt{3}$).

(a) Draw the four tetrahedra in the star of a long edge and highlight the link of the long edge.

(b) Draw the six tetrahedra in the star of a short edge and highlight the link of the short edge.

Part II
Complexes

Similar to the tessellations of Euclidean space introduced in Chap. 1 there are several equivalent ways to describe alpha shapes. This is perhaps the hallmark of an interesting concept, as it guarantees connections in multiple directions. Of major importance in this course are the bridges between geometry, topology, and applications in biology. We will see that alpha shapes provide substantial contributions to all three of these bridges.

Chapter 5
Alpha Complexes

The original motivation for the concept of alpha shapes was the desire to develop a concrete version of the intuitive notion of 'shape' of a finite point set. Starting from this idea, we explore connections to Voronoi diagrams and Delaunay triangulations.

5.1 Jarvis' Construction

In an early approach, Jarvis defines the shape of a point set procedurally, as the output of a generalized convex hull algorithm [1]. Letting S be a finite set of sites in $\mathbb{R}^2$, we can construct the convex hull by rotating a line about the set; see Fig. 5.1. Assuming general position, we begin by letting s be the leftmost site. Drawing a vertical line L through s and oriented downward, we see that all sites lie to the left of L, so s is indeed a vertex of the convex hull. Using s as a pivot, we rotate L in a ccw order until it hits another site, t. All sites other than s and t lie to the left of L, implying that st is an edge of the convex hull. We repeat this step now using t as the pivot. Each step gives a new edge of the convex hull, and the algorithm halts when it returns to the initial site, s.

What else we can do by wrapping the set? For example, we can construct non-convex shapes if we shorten the rotating line to a line segment of fixed length [2]. This works if the sites are nicely distributed, as in Fig. 5.1, but can get the line segment lost and return a sequence of sites that cannot be reasonably called its shape. Another shortcoming is that this algorithm necessarily constructs only one closed curve, while two closed curves might be more appropriate for the example at hand; see Fig. 5.2.

5.2 The Alpha Shape

A theoretically and practically more satisfying solution to constructing the shape of a finite set can be based on empty disks [3]. Letting $\alpha \geq 0$ be a fixed radius, we write $D_x(\alpha)$ for the closed disk with center $x \in \mathbb{R}^2$ and radius α. It is *empty* if it contains

H. Edelsbrunner, *A Short Course in Computational Geometry and Topology*,
SpringerBriefs in Mathematical Methods, DOI: 10.1007/978-3-319-05957-0_5,

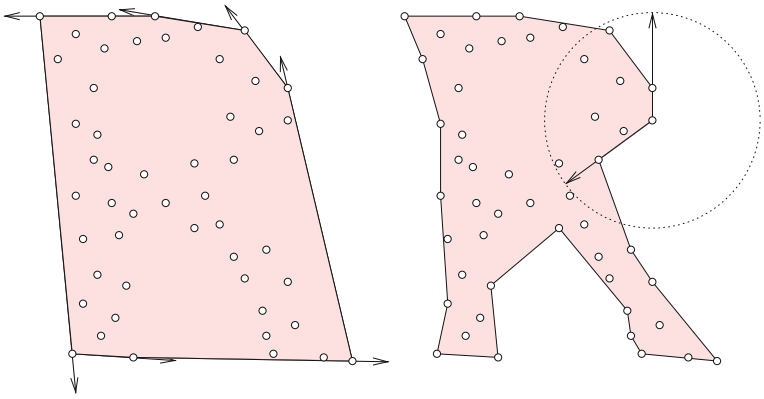

Fig. 5.1 *Left* illustration of Jarvis' convex hull algorithm. *Right* the generalization to constructing shapes obtained by shortening the line to a line segment

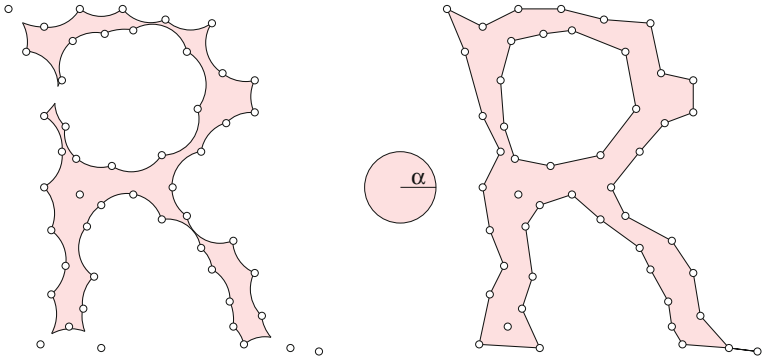

Fig. 5.2 A set of points sampling the letter 'R', with its α-hull on the *left* and its α-shape on the *right*

no site: $D_x(\alpha) \cap S = \emptyset$. The *$\alpha$-hull* of S is the complement of the union of empty disks of radius α; see Fig. 5.2. Setting α to zero, we get the set of points, and setting it to infinity, we get the convex hull.

The curved edges of the α-hull can sometimes be annoying. This motivates us to draw them straight, which results in the *α-shape* of S. We will give an alternative definition shortly, which will eliminate any remaining ambiguities. In contrast to the α-hull, the α-shape is a polyhedron in the general sense: it does not have to be convex, and it can have different intrinsic dimension at different places. For example, the α-shape in Fig. 5.2 is mostly 2-dimensional except it has a 1-dimensional extension at the right leg. If S contains a site that is further than distance 2α from any other site, then this point is isolated and forms a locally 0-dimensional portion of the α-shape.

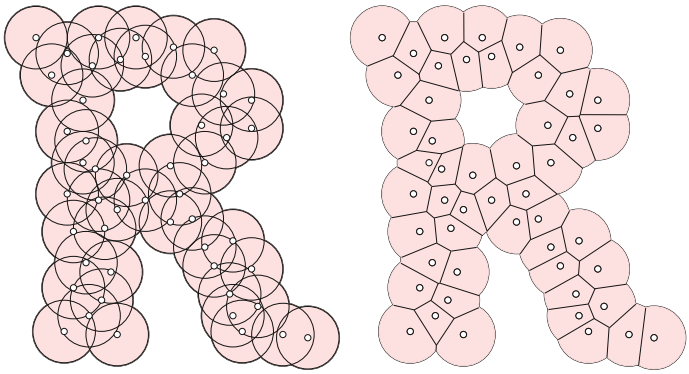

Fig. 5.3 *Left* the union of disks of radius α of the same points as in Fig. 5.2. *Right* the Voronoi decomposition of the union

5.3 Union of Disks

A point x is the center of an empty disk of radius α iff it is further than α from every site. To make this relationship concrete, we construct the union of disks of radius α centered at the sites:

$$\mathbb{U}_S(\alpha) = \bigcup_{s \in S} D_s(\alpha); \tag{5.1}$$

see Fig. 5.3. This union is the complement of the set of centers of the empty disks. We can therefore expect that the boundaries of the α-hull and the union of disks are related. Indeed, for each circular arc of $\mathbb{U}_S(\alpha)$, we have a vertex of the α-hull, and for each vertex of $\mathbb{U}_S(\alpha)$, we have a circular arc of the α-hull. In its details, this relationship is troubled by arcs in the boundary of the α-hull that intersect and partially or completely erase one another. An example are the two circular arcs that connect the right leg of the 'R' to its last site. Since they lie inside each other's empty disks, they do not belong to the α-hull.

5.4 Voronoi Decomposition

To get a cleaner relationship between the union of disks and the α-shape, we need an unambiguous definition of the latter. For this, we overlay the union of disks with the Voronoi diagram, effectively decomposing the union into convex regions; see Fig. 5.3. To formalize this idea, we write $R_s(\alpha) = V_s \cap D_s(\alpha)$ and note that this is a convex set because it is the intersection of convex sets. Furthermore, $\mathbb{U}_S(\alpha) = \bigcup_{s \in S} R_s(\alpha)$. In words, the regions R_s cover the union, but in contrast to the disks, which also cover the union, they do this without overlap. More specifically, the common intersection of the regions is limited to shared edges and vertices. Following the recipe for the

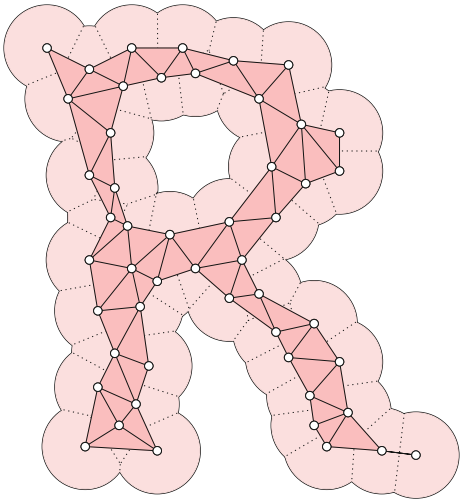

Fig. 5.4 The α-complex is superimposed on the union of disks, which is decomposed into convex regions by the Voronoi diagram

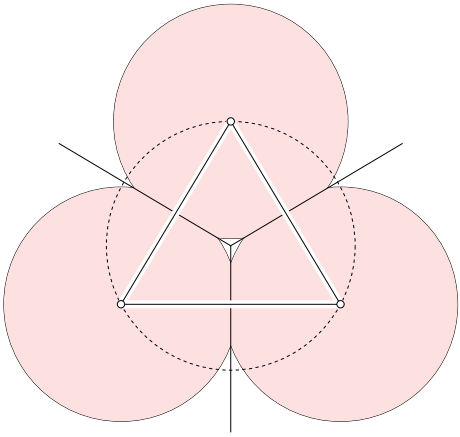

Fig. 5.5 While the intersection of any two regions is non-empty, the common intersection of all three regions is empty because α is smaller than the radius of the circumcircle. Accordingly, the α-complex contains the three edges but not the triangle spanned by the three sites

Delaunay triangulation, we construct the *α-complex* by drawing an edge between two sites if their regions intersect in a common edge, and by drawing a triangle between three sites if their regions intersect in a common point; see Fig. 5.4. For ease of reference, we denote the resulting complex by $A_S(\alpha)$, or by $A(\alpha)$ if the set of sites is understood. For example, $A(0)$ is the set of sites without any additional structure, and $A(\infty)$ is the Delaunay triangulation of S. We now formally define the *α-shape* as the union of simplices in the α-complex.

Note that the global connectivity of the union of disks in Fig. 5.4 is that same as that of the α-complex and of the α-shape: all three are connected and have a single hole. This is not a coincidence but rather a consequence of the Nerve Theorem, which will be discussed later in this course.

5.5 Filtration

Next we vary α and consider the complete range of possible values, which is from 0 to ∞. For $\alpha < \alpha'$, we have $D_s(\alpha) \subseteq D_s(\alpha')$ and therefore $R_s(\alpha) \subseteq R_s(\alpha')$. Recall that st is an edge in the α-complex iff $R_s(\alpha) \cap R_t(\alpha) \neq \emptyset$. Since the regions grow with the radius, this implies $R_s(\alpha') \cap R_t(\alpha') \neq \emptyset$, and therefore st is also an edge in the α'-complex. Similarly, every triangle in the α-complex belongs to the α'-complex. In summary, $A(\alpha) \subseteq A(\alpha')$ whenever $\alpha \leq \alpha'$. It thus makes sense to ask—for each vertex, edge, and triangle σ in the Delaunay triangulation—what the smallest value of α is for which σ belongs to $A(\alpha)$. Denoting this value by α_σ, we can construct the α-complex simply by collecting all vertices, edges, and triangles that have a value not larger than α:

$$A(\alpha) = \{\sigma \in K \mid \alpha_\sigma \leq \alpha\}, \tag{5.2}$$

where K is the Delaunay triangulation of S. Computing this value is easiest for the vertices, since we have $\alpha_s = 0$ for every $s \in S$. It is also easy for triangles, for which α_{stu} is the radius of the circumcircle; see Fig. 5.5. The computation of the smallest α-value is slightly more difficult for edges. Here, we distinguish between two cases. First, the edge st may intersect the dual Voronoi edge in its interior; as in Fig. 5.5. In this case, st belongs to the α-complex as soon as the two disks meet, which happens when the radius reaches half the distance between the sites: $\alpha_{st} = \frac{1}{2}\|s - t\|$. If st is shared by the triangles stu and stv in K, then the condition of st intersecting the dual Voronoi edge in its interior is equivalent to having acute angles at u and v. This leads us to the second case in which one of these two angles is obtuse, say the angle at u. Then it is not enough that the two disks meet; they also need to reach the Voronoi edge, which happens when the triangle stu enters the α-complex. Hence, $\alpha_{st} = \alpha_{stu}$. Now that we have the threshold value for every vertex, edge, and triangle in the Delaunay triangulation, we can sort them such that

$$\alpha_{\sigma_1} \leq \alpha_{\sigma_2} \leq \ldots \leq \alpha_{\sigma_n}. \tag{5.3}$$

The corresponding sequence of simplices is called a *filter*. Here, we make sure that every simplex is preceded by its faces. We get this from the formulas already, since the value of every edge is smaller than or equal to the values of the triangles it belongs to. However, in case the value of the edge is equal to that of an incident triangle, we make sure we order the edge before the triangle. With this, every prefix of the filter is a complex. Writing K_j for the collection of simplices σ_i with $i \leq j$, we get an increasing sequence of complexes,

$$\emptyset \;=\; K_0 \subseteq K_1 \subseteq \ldots \subseteq K_n \;=\; K. \tag{5.4}$$

We call such an increasing sequence as a *filtration*. The not necessarily contiguous subsequence of alpha complexes is sometimes referred to as the *alpha complex*

filtration of the set S. This and other filtrations will play an important role in later sections, when we talk about persistent homology.

5.6 Space-Filling Models of Proteins

A major application of alpha complexes are proteins and other molecules modeled as unions of balls [4]. One such model is the *van der Waals diagram* of a protein. It is based on the van der Waals force, which is weakly attractive at short distance between the atoms, and turns into a strongly repulsive force if we push the atoms closer together. The diagram is obtained by taking the union of the balls centered at the atoms in which the radii are chosen so that the atoms are at equilibrium when the balls touch.

Different types of atoms affect neighboring atoms differently, which leads to different radii. For example, hydrogen atoms are the smallest, with carbon, oxygen, and nitrogen atoms represented by somewhat larger balls. This motivates the concept of *weighted alpha complexes*, which are defined analogous to weighted Voronoi diagrams and weighted Delaunay triangulations. To be specific, we have a finite set of sites with real weights, and we recall that the bisector of two sites under the power distance is the set of points x that satisfy $\|x - s\|^2 - w_s = \|x - t\|^2 - w_t$. We have shown that the bisector is a straight line. If we add the same constant, α^2, to the weights, then the bisector stays the same. We therefore define $D_s(\alpha)$ as the disk with center s and radius $\sqrt{w_s + \alpha^2}$. We note that the radius depends on the weight as well as on α. We could therefore drop α and change all weights. We prefer to keep the weights fixed and modify the complex by varying the parameter, thus stressing that we use only one degree of freedom. Observe that the radius in the weighted case agrees with the definition in the unweighted case, when it is α. To construct the weighted α-complex of S, we take the union of the disks $D_s(\alpha)$, for all $s \in S$, we decompose the union into convex regions using the weighted Voronoi diagram, and we take the dual, as before.

If we choose the weights so that the radii agree with the van der Waals forces, we get the weighted 0-complex as the dual of the van der Waals diagram. There are reasons to also consider non-zero values of α. One is the *solvent accessible diagram* in which the van der Waals radii are increased by about 1.4 Angstrom, which is the radius used to approximate a water molecule. The diagram thus represents the interaction between the protein and solvent, which is water. Indeed, if we represent a water molecule by the ball with center x and radius 1.4 Angstrom, then this molecule is not yet repelled by the protein iff x lies outside the solvent accessible model. For other solvents, we would modify the radius by different amounts. A word or caution is in order: increasing all radii by the same amount relates to the Apollonius and not the power diagram. Indeed, only if we increase all squared weights by the same amount—which for reasons of compatibility with the unweighted case is denoted as α^2—we preserve the power diagram.

References

1. Jarvis RA (1973) On the identification of the convex hull of a finite set of points in the plane. Inform Process Lett 2:18–21
2. Jarvis RA (1977) Computing the shape hull of points in the plane. In: Proceedings of IEEE computer society conference pattern recognition and image processing, pp 231–241
3. Edelsbrunner H, Kirkpatrick DG, Seidel R (1983) On the shape of a set of points in the plane. IEEE Trans Inform Theory 29:551–559
4. Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. J Mol Biol 55:379–400

Chapter 6
Holes

Being possibly non-convex, alpha shapes can have many interesting features, some of which that are best classified as types of holes. We begin with the easy 2-dimensional case, noting that the types and structure of holes gets progressively more complicated as the dimension increases.

6.1 Holes in the Plane

There is really only one type of hole a connected α-shape in $\mathbb{R}^2$ can have, namely a bounded connected component of the complement. We refer to such a hole as a *void*. For example, most uppercase letters have no void, and only one has two voids:

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

If we allow for non-connected shapes, we can have a second type of hole: a *gap* between components. Among the lowercase letters, most a connected but two are not:

c f h k l m n r s t u v w x y z; a b d e o p q; g; i j.

In everyday language, we call things a hole even if they are not holes in the topological sense at all, such as for example the 'holes in the road'. (I am less sure about the 'holes in our memory'.) They are referred to as *cavities*, which are local depressions in the shape. Recall that teeth sometimes develop such, which can be painful. Their counterpart are *protrusions*, which are local extensions of the shape. In the planar case, we may define cavities by looking at the components of the convex hull minus the shape; see Fig. 6.1. Each cavity is a shape itself, and we can apply the same method, which would then identify cavities of the cavity, which are protrusions of the shape. Iterating this way, we get a hierarchical family of cavities and protrusions.

H. Edelsbrunner, *A Short Course in Computational Geometry and Topology*,
SpringerBriefs in Mathematical Methods, DOI: 10.1007/978-3-319-05957-0_6,

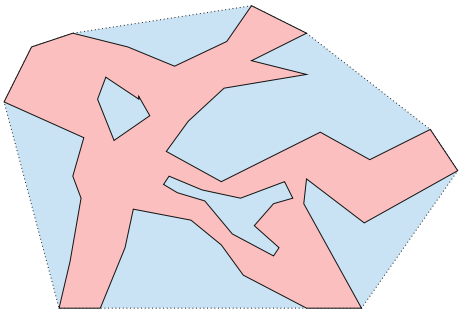

Fig. 6.1 A connected shape with two voids and five cavities

However, we do not advocate this classification since in three dimensions, the convex hull minus the shape can easily be connected, in which case this method fails to separate different cavities from each other.

6.2 Filtration by Thickening

Another approach to cavities makes use of the 1-parameter family we get by thickening the initial shape or point set. Let S be a finite set of sites, recall that $\mathbb{U}(\alpha) = \mathbb{U}_S(\alpha)$ is the union of disks of radius α centered at the sites, and note that $\mathbb{U}(\alpha) \subseteq \mathbb{U}(\alpha')$ whenever $\alpha \leq \alpha'$. We may grow the radius continuously and watch how the shape changes. For the regularly arranged points in Fig. 6.2, the shapes in this family are all geometrically different but most are topologically the same:

1. For small radii, the disks are disjoint, so $\mathbb{U}_S$ has as many components as there are sites;
2. When α reaches half the distance between contiguous sites, $\mathbb{U}_S$ becomes connected;
3. There is one big gap, and when α reaches half its size, then $\mathbb{U}_S$ closes up and forms a void;
4. When α reaches the radius of the circle on which the sites are placed, then the hole fills up, and $\mathbb{U}_S$ becomes a topological disk.

Writing $2c_1$ for the distance between two contiguous sites, $2c_2$ for the size of the gap on the right, and c_3 for the radius of the circle passing through the sites, we have $0 < \alpha_0 < c_1 < \alpha_1 < c_2 < \alpha_2 < c_3 < \alpha_3 < \infty$. Similar to the alpha complex filtration introduced in Chap. 5, we consider the growing sequence of shapes $\mathbb{U}(\alpha)$, for α from 0 to ∞.

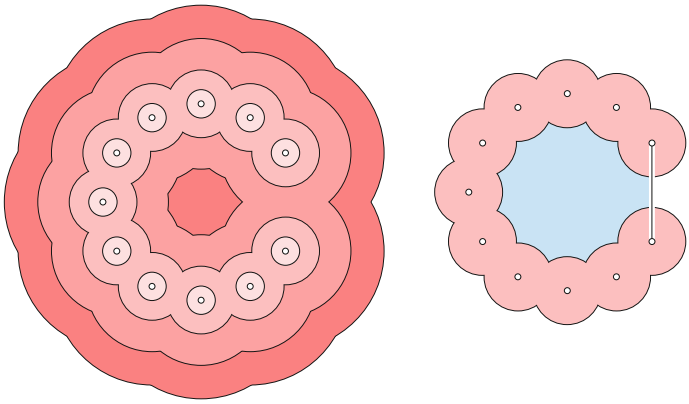

Fig. 6.2 Eleven points arranged in the shape of the letter C. *Left* apart from the points, we see four unions of disks, with radii $\alpha_0 < \alpha_1 < \alpha_2 < \alpha_3$. *Right* the second union, for radius α_1, and the pocket in *blue*

6.3 Pockets

This filtration enriches the shape description as we can now keep track of the holes as they appear and disappear during the thickening process. Indeed, the history of changes depends on the details of the original shape. Using this insight, we define a *pocket* of a shape as a subset of the complement that becomes a void before it disappears. To makes this into an unambiguous concept, we need a flow that identifies the subset of the complement that becomes the void. Using the flow normal to the boundary of the evolving shape, we get the pocket of $\mathbb{U}(\alpha_1)$ in part delimited by the straight edge connecting the centers whose disks meet to form the void; see Fig. 6.2. Indeed, this edge divides the flow, with points close to and to the left of the edge flowing into the void, while points close to and to the right of the edge recede to infinity. The same pocket already exists in $\mathbb{U}(\alpha)$ for $\alpha < \alpha_1$, or, more precisely, $\mathbb{U}(\alpha)$ has a pocket that contains the pocket of $\mathbb{U}(\alpha_1)$. We refer to [1] for more details, in particular about the more interesting 3-dimensional case.

Determining the existence of pockets, or counting them, is easier than drawing their precise shape. Specifically, each time we see a void appear in the filtration, we can read it as the witness of a pocket in preceding shapes. Tracking the changing topology in a filtration can be done with homology, as we will see later.

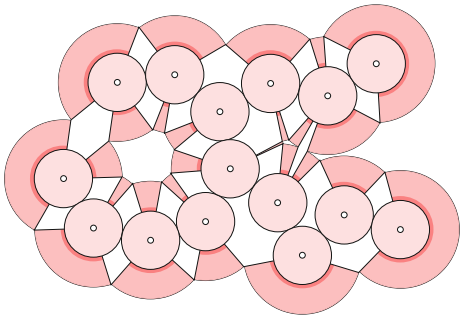

Fig. 6.3 Iterative construction of a branched peptide or a Brownian tree. The shaded arcs consist of the candidate points for the attachment of the next disk

6.4 Iterative Stochastic Construction

Let us look at materials modeled as configurations of atoms represented by spherical balls. For now we stay in two dimensions. A *branched peptide* is grown from an initial unit disk by randomly adding one unit disk at a time. The i-th disk is added such that

(i) It touches the union of the first $i - 1$ disks in precisely one point;
(ii) Every possible such point is equally likely to be chosen for the attachment.

The candidate points form open intervals on the boundary of the union of the first $i - 1$ disks. As illustrated in Fig. 6.3, we construct them by forming the union of disks with twice the radius and projecting the arcs of the boundary onto the circles bounding the corresponding smaller circles. Next consider a *Brownian tree*, which is constructed the same way, except that the probability of attachment is not the same for every candidate point. Specifically, the i-th unit disk follows a random path from far away until it hits one of the first $i - 1$ disks. For example, if there is a pocket whose opening is narrower than the diameter of the new disk, then the probability of attachment inside that pocket is zero. In Fig. 6.3, we have two such pockets corresponding to the voids in the union of larger disks.

6.5 Hierarchical Pocket Structure

By construction, both the branched peptides and the Brownian trees are topologically simple: they are connected and have no holes. Nevertheless, we will see that topology can be used to tell them apart. The most obvious difference between the two structures is the number and size of their pockets. It is plausible that Brownian trees have large pockets protected by narrow openings, while branched peptides have no large pockets. To see whether this intuition is correct, we consider two large structures

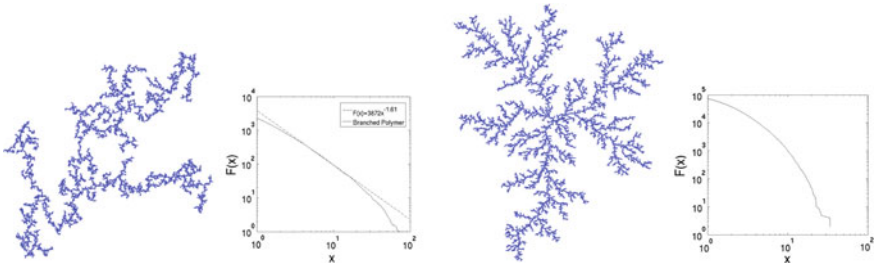

Fig. 6.4 A branched peptide on the *left* and a Brownian tree on the *right*. The diagrams show the number of pockets whose half-age exceeds a threshold x

and summarize their analysis, which can be found in [2]. To quantify their pocket structure, we count the voids in the filtration obtained by thickening. Specifically, we call $\frac{1}{2}(b+d)$ the *half-age* of a void born at the radius b and dying at the radius d. Let $F(x)$ be the number of voids with half-age at least x. As shown in Fig. 6.4, the branched peptide has more voids with large half-age and fewer voids with small half-age, which seems to contradict our intuition. Does it?

A closer look at the functions, drawn in log–log scale, shows that the graph for the branched peptide can be approximated by a straight line. The approximation fails for small and large mean-age, which is to be expected since we are dealing with a finite size example. In contrast, the graph for the Brownian tree has no good approximation by a straight line. This implies that the structure of the Brownian tree is unlikely to be self-similar on any scale, even if only in a statistical sense.

6.6 Tunnels

Let us venture one dimension up, to 3-dimensional space. An alpha shape in $\mathbb{R}^3$ can have voids and gaps, but it can also have *tunnels*, like open doors and windows, or prison escape routes. Curiously, tunnels are much more difficult to count than voids and components. What is, for example, the number of tunnels formed by the edge skeleton of the cube? We will later argue that the answer is 5. Indeed, if we flatten the skeleton and draw it in the plane, we get 5 holes. A possibly more convincing argument why the answer should be 5 will be furnished by the homology groups to be introduced in Chap. 10.

In the case of a long and possibly narrow tunnel, we can use pockets to give it a geometric identity. Indeed, if we thicken the shape, the tunnel may be cut off the outside at two or more places, defining the space enclosed by these openings as a pocket with more than one mouth. Pockets have been defined and implemented as part of the 3-dimensional alpha shape software [3], and they have found applications in the study of protein structure [4]. Using this software, we can construct α-shapes of zeolites, which are periodic molecular structures used in oil refinery. On the left

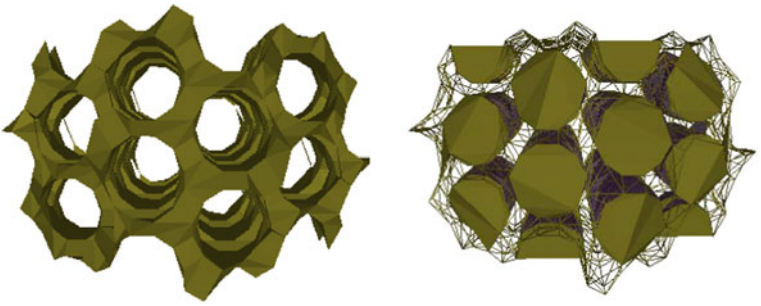

Fig. 6.5 *Left* the alpha shape of a zeolite. *Right* the pockets of the zeolite

in Fig. 6.5, we see tunnels going in parallel through the entire structure. On the right, we see the pockets of the same structure. Many of them have two mouths and can thus be identified as geometric representations of tunnels. In later sections, we will use the ranks of homology groups to count tunnels or, more precisely, to measure the size of a basis that generates all possible ways to pass through the shape following tunnels.

References

1. Edelsbrunner H, Facello MA, Liang J (1998) On the definition and the construction of pockets in macromolecules. Discrete Appl Math 88:83–102
2. MacPherson R, Schweinhart B (2010) Measuring shape with topology. arXiv:1011.2258
3. Edelsbrunner H, Mücke EP (1994) Three-dimensional alpha shapes. ACM Trans Graphics 13:43–72
4. Liang J, Edelsbrunner H, Woodward C (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci 7:1884–1897

Chapter 7
Area Formulas

Perhaps surprisingly, α-complexes are useful in measuring a union of balls in two and higher dimensions. In the study of proteins and other biomolecules, this connection is the basis of the fastest precise algorithms for the volume, surface area, and derivatives of space-filling models of molecular structures.

7.1 Space-Filling Models

The central dogma of biology asserts that strings of DNA get transcribed into RNA molecules, which then get translated into proteins, the work-horses of living organisms. Each protein consists of a backbone, which is decorated by a sequence of amino acids—we can think of it as a word over a 20-letter alphabet. Upon creation, the protein folds up to a tightly packed geometric structure. Importantly, the process is repeatable: the same sequence of amino acids folds up into the same structure. Since we know the DNA of many organisms in great detail, we know the chemical structure of many proteins, but only for a small fraction do we also know the geometric structure. Similarly, we have very incomplete knowledge of which proteins interact with each other, which depends on their shapes. The tasks of predicting the geometric from the chemical structure, and of predicting interaction from geometric structure are known as *protein folding* and *protein docking*. Both are important challenges in structural molecular biology.

A common representation of a protein, but also of RNA and DNA, is the *space-filling model*; see Fig. 7.1. Each atom is represented by a ball, with the radius chosen to approximate the range of the weakly attractive van der Walls force, or 1.4 Angstrom larger to model the interaction with water molecules; see also Chap. 5. Measurements of space-filling models are used as ingredients in the force that drives the simulation of the folding process. For example, some amino acids are hydrophobic while others are hydrophilic. We are therefore interested in the surface area, weighted by hydrophobicity factors, and the derivative of the same under continuous motion.

H. Edelsbrunner, *A Short Course in Computational Geometry and Topology*,
SpringerBriefs in Mathematical Methods, DOI: 10.1007/978-3-319-05957-0_7,

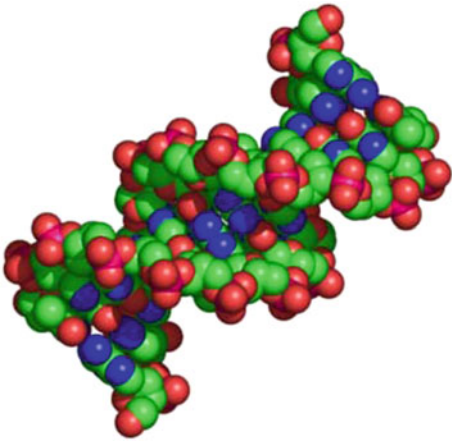

Fig. 7.1 The space-filling representation of a short piece of DNA. Note how the helix forms two grooves on opposite sides of the ribbon

We will not discuss the area derivative formulas, which are more advanced [1] but otherwise based on the same principles to be presented here.

7.2 Principle of Inclusion-Exclusion

To simplify the exposition, we limit ourselves to two dimensions, where the equivalent of volume of a space filling model is the area of a union of disks. Let S be a finite set of sites in $\mathbb{R}^2$, and consider the union of disks of radius α centered at the sites. As before, we write $\mathbb{U}(\alpha) = \bigcup_{s \in S} D_s(\alpha)$ for the union, and we write area$(\mathbb{U}(\alpha))$ for its area. Suppose we have a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{U}(\alpha), \\ 0 & \text{if } x \notin \mathbb{U}(\alpha), \end{cases}$$

which in mathematical jargon is called the *indicator function* of the union. Then we can compute the area by integration:

$$\text{area}(\mathbb{U}(\alpha)) = \int_{x \in \mathbb{R}^2} f(x)\, \mathrm{d}x. \tag{7.1}$$

To construct such a function, let $S_x \subset S$ contain all sites for which $x \in D_s(\alpha)$. Suppose S_x consists of m sites. Then it has $\binom{m}{k}$ subsets of size $0 \le k \le m$. The alternating sum gives $(1-1)^m = \sum_{k=0}^{m} (-1)^k \binom{m}{k}$, which is zero. Hence,

$$f(x) = m - \binom{m}{2} + \cdots \pm \binom{m}{m} \;=\; 1. \tag{7.2}$$

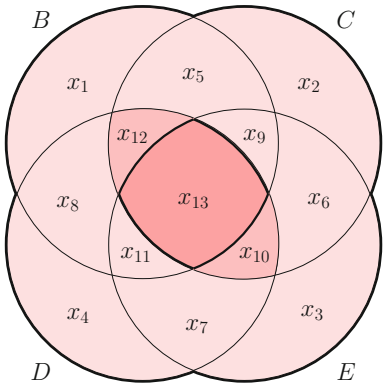

Fig. 7.2 The union of four disks, with the intersection of two and of all four disks highlighted

In words, if we count the disks that contain x, subtract the pairs whose intersections contain x, add the triplets whose common intersections contain x, etc., then we get 1, as desired. Integrating over the plane, the contribution of a disk is its area, the contribution of a pair is minus the area of its intersection, etc. Writing this as a formula, we get the area of $\mathbb{U}(\alpha)$ as the integral of $f(x)$:

Long PIE Formula *The area of the union of disks of radius α centered at the sites in S is*

$$\text{area}(\mathbb{U}(\alpha)) = \sum_{\emptyset \neq T \subseteq S} (-1)^{|T|-1} \cdot \text{area}\left(\bigcap_{s \in T} D_s(\alpha) \right).$$

This formula holds for much more general sets than disks and is known as the *principle of inclusion-exclusion*.

7.3 Independent Disks

The Long PIE Formula has two shortcomings that make it difficult to use. First, each term can be complicated to compute itself, and second there are exponentially many terms, possibly all non-zero. Fortunately, it is possible to shorten the formula without sacrificing its correctness. To explain this, we first identify redundant terms in the formula for the four disks in Fig. 7.2. Writing $b = \text{area}(B)$, $bc = \text{area}(B \cap C)$, etc., the area of $B \cup C \cup D \cup E$ is

$$b + c + d + e - bc - bd - be - cd - ce - de + bcd + bce + bde + cde - bcde.$$

From Fig. 7.2, we see that $-cd + bcd + cde - bcde = 0$, so we can remove the four terms and get a shorter formula. The reason for the existence of redundant terms that can be cancelled is the dependence of the disks in the last term.

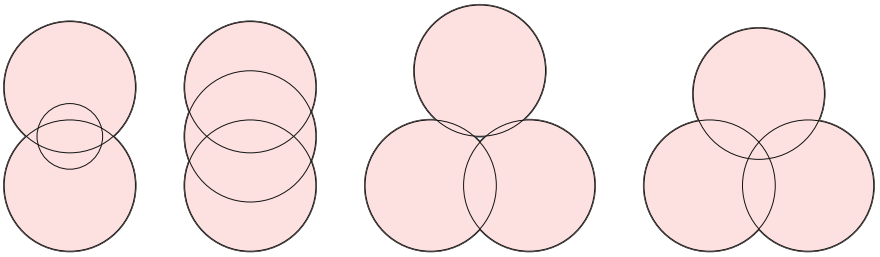

Fig. 7.3 The four triplets of disks whose pairs are independent. Except for the first, each configuration exists for equal-sized disks

We call a collection U *independent* if for every subset $V \subseteq U$ there exists a point x that is contained in all disks in V but in none of the disks in $U - V$. A collection consisting of only one disk is always independent: we can choose x inside or outside the disk. A collection of two disks is independent if they have a non-empty common intersection but neither disk contains the other. Clearly, a collection of three disks can only be independent if every subcollection of two disks is independent. There are four triplets whose pairs are independent, all drawn in Fig. 7.3. Only the last of the four triplets is independent. The first is not because one disk is contained in the union of the other two, the second is not because one disk contains the intersection of the other two, and the third is not because the common intersection of the three disks is empty.

7.4 Substitution

A collection of four or more disks is never independent, as a straightforward counting argument shows. A collection of n disks has 2^n subsets, so when we draw the bounding circles, we need 2^n regions, each contained in a different maximal subset. Drawing three circles, we can indeed get $2^3 = 8$ regions, namely for the configurations in Fig. 7.3. Adding a fourth circle, it intersects the first three in at most 6 points, which decompose the circle into 6 arcs. Each arc decomposes a region into two. Starting with 8 regions, we thus reach at most $8 + 6 = 14$, which is too small for independence.

Whenever we have a dependent collection of disks, we can replace the corresponding term in the inclusion-exclusion formula by terms involving fewer disks [2]. In the example of four disks depicted in Fig. 7.2, we replace $bcde$ by $bcd + cde - cd$. To see why this is possible, we write the areas of the intersections in terms of the areas of the regions in the arrangement of circles:

$$\begin{aligned} b &= x_1 + x_5 + x_8 + x_9 + x_{11} + x_{12} + x_{13}, \\ &\cdots \\ bc &= x_5 + x_9 + x_{12} + x_{13}, \\ &\cdots \\ &\cdots \\ bcde &= x_{13}. \end{aligned}$$

In the dependent case, we have fewer variables than equations. In the above case, we have 15 equations in 13 variables. It follows that some rows of the matrix that represents this system of equations are equal to linear combinations of other rows.

7.5 Full Subcomplexes

Instead of studying the details of the substitution process, we give a description of the final result in terms of the simplices in the α-complex. Assume the sites in S are in general position and write $A = A(\alpha)$ for the α-complex of S. Let x be any point in the plane, and let $S_x \subseteq S$ be the subset of sites whose disks of radius α contain x, as before. Note that S_x is a subset of the vertices of A, and we get the corresponding *full subcomplex*, A_x, by adding the edges and triangles in A whose vertices belong to S_x. It is dual to the collection of Voronoi regions generated by the sites in S_x. As illustrated in Fig. 7.4, the union of the disks with centers in S_x is star-convex. More precisely, for each point y in the union, the entire line segment connecting x to y is contained in the union. Similarly, the union of the Voronoi regions generated by the sites in S_x is star-convex, and so is the intersection of this union with the union of disks. The full subcomplex, A_x, is the nerve of the decomposition of this intersection into restricted Voronoi regions. Since the intersection is contractible, the Nerve Theorem implies that A_x is contractible. It follows that the Euler characteristic (the number of vertices minus the number of edges plus the number of triangles) of the full subcomplex is

$$\chi(A_x) = \begin{cases} 1 & \text{if } S_x \neq \emptyset, \\ 0 & \text{if } S_x = \emptyset. \end{cases} \tag{7.3}$$

Since this is true for every point in the plane, we can set $f(x) = \chi(A_x)$ and compute the area of the union of disks by integration, as before. To collect the contributions, we iterate over all vertices, edges, and triangles of A:

Short PIE Formula *The area of the union of disks of radius α centered at the sites in S is*

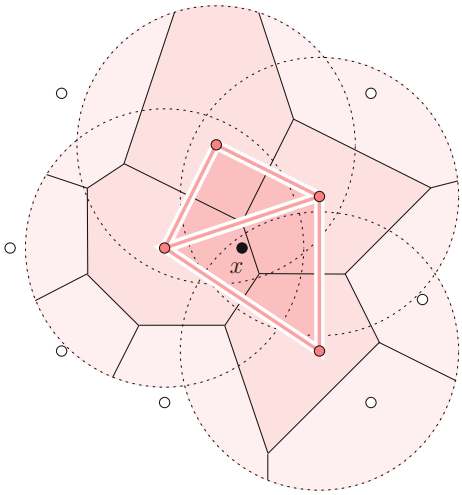

Fig. 7.4 The Voronoi diagram of 11 sites and the full subcomplex of the alpha complex defined by the four sites whose disks contain the point x

$$\text{area}\mathbb{U}(\alpha) = \sum_{\sigma \in A} (-1)^{\dim \sigma} \cdot \text{area}\left(\bigcap_{s \in \sigma} D_s(\alpha) \right).$$

In the plane, A has at most a constant times $n = |S|$ simplices, and each simplex corresponds to an independent collection of disks. We thus have a formula with few and simple terms. The formula generalizes to higher dimensions, and to collections of balls with different radii [3]. The summation can also be done over all simplices of the Delaunay triangulation, and while this sum contains redundant terms, the result is still correct [4].

References

1. Bryant R, Edelsbrunner H, Koehl P, Levitt M (2004) The area derivative of a space-filling diagram. Discrete Comput Geom 32:293–308
2. Kratky KW (1978) The area of intersection of n equal circular disks. J Phys A: Math Gen 11:1017–1024
3. Edelsbrunner H (1995) The union of balls and its dual shape. Discrete Comput Geom 13:415–440
4. Naiman DQ, Wynn HP (1992) Inclusion-exclusion Bonferroni identities and inequalities for discrete tube-like problems via Euler characteristics. Ann Statist 20:43–76

Part II
Complexes

Exercises

Question 1. ($20 = 10 + 10$ points). Let S be a finite set of points in $\mathbb{R}^2$, s and t two points in S, and α positive real number. Recall that $D_t(\alpha)$ is the α-disk centered at $s \in S$, V_s is the Voronoi region of s, and $R_S(\alpha) = D_s(\alpha) \cap V_S$.

(a) Prove that every point $x \in V_s \cap D_t(\alpha)$ is contained in $D_S(\alpha)$,
(b) Prove $\cup_{s \in S} D_s(\alpha) = \cup_{s \in S} R_s(\alpha)$.

Question 2. ($20 = 10 + 10$ points). Let S be a finite set of sites in general position in $\mathbb{R}^2$. Let *stu* and *stv* be two triangles in the Delaunay triangulation, and denote the angles at u and v inside these triangles by φ and ψ. Recall that an *acute* angle is smaller than 90° and an *obtuse* angle is larger than 90°.

(a) Show that at most one of φ and ψ is obtuse.
(b) Is it possible that both φ and ψ are obtuse if the sites are weighted and we are talking about the weighted Delaunay triangulation?

Question 3. ($20 = 10 + 10$ points). Consider the 26 uppercase letters of the Roman alphabet, each drawn in a simple font, without any fancy extras.

(a) Classify each uppercase letter in terms of its symmetries.
HINT a *symmetry* is a transformation (here: rotation, or reflection) that leaves the letter unchanged. For example, "A" has a vertical symmetry axis, while "C" has a horizontal symmetry axis, and "S" has neither but is centrally symmetric.
(b) Which of the uppercase letters have no hole but develop holes as they get thickened?

Question 4. ($20 = 10 + 10$ points). Let S be a finite set of points in $\mathbb{R}^2$, consider the union of disks with radius α centered at the points, and write $L(\alpha)$ for the total length of the boundary.

(a) Prove $L(2\alpha) \leq 2L(\alpha)$.
(b) Is it true that $L(\alpha) \leq L(2\alpha)$ for all choices of S and α?

Question 5. (20 points). Let S be a finite set of points in $\mathbb{R}^2$ and α positive real number. Prove that the common intersection of the disks with radius α centered at the points in S is nonempty iff there exists a disk of radius α (not necessarily centered at a point in S) that contains all points in S.

Part III
Homology

Gently shifting from geometry to topology, we introduce simplicial complexes as a general representation of shapes and spaces. The transition comes with an increased emphasis on connectivity. Contrary to a widely held belief, this does not imply that topology is not useful to distinguish intricate shapes with simple global connectivity, such as human faces and root systems of plants. Evidence to support this claim will be given in the next part, whose topic is persistent homology. This part focuses on the homology of individual topological spaces, which is a classic topic in algebraic topology.

Chapter 8
Topological Spaces

In geometry, we can decide which points are near and which are far by computing distance. In topology, no such notion is available and distance is replaced by the weaker concept of neighborhoods. A full treatment of this idea is beyond the scope of this course. Indeed, we are motivated to side-step the technical difficulties by restricting ourselves to spaces in which connectivity can be defined by elementary means. The primary tool to achieve this goal are simplicial complexes.

8.1 Topology and Topology Equivalence

Let $\mathbb{X}$ be a set of points. A *topology* of $\mathbb{X}$ is a collection of subsets, called *open sets*, such that

(i) $\mathbb{X}$ is open and the empty set is open;
(ii) the intersection of any two open sets is open;
(iii) the union of any family of open sets is open.

The set $\mathbb{X}$ together with the topology is called a *topological space*. By a *neighborhood* of a point we mean an open set that contains that point. For example, the plane together with the topology generated by the Euclidean metric is a topological space. To construct it, we call the set of points at distance less than $r > 0$ from a point $x \in \mathbb{R}^2$ an *open disk*. Taking finite intersections and arbitrary unions of open disks, we get a collection of open sets that satisfies the above three conditions. It is usually referred to as the *Euclidean topology* of the plane. Similarly, we can construct topologies for subsets of the plane. Consider for example the unit disk, $\mathbb{B}^2 = \{x \in \mathbb{R}^2 \mid \|x\| \leq 1\}$. The topology *inherited* from the topology of the plane consists of all intersections of open sets with $\mathbb{B}^2$. Thus, $\mathbb{B}^2$ together with the topology inherited from $\mathbb{R}^2$ is a topological space.

A function from one topological space to another is *continuous* if the preimage of every open set is open. This is derived from the familiar notion of continuity in calculus. A *homeomorphism* between two topological spaces is a bijective function

H. Edelsbrunner, *A Short Course in Computational Geometry and Topology*, SpringerBriefs in Mathematical Methods, DOI: 10.1007/978-3-319-05957-0_8,

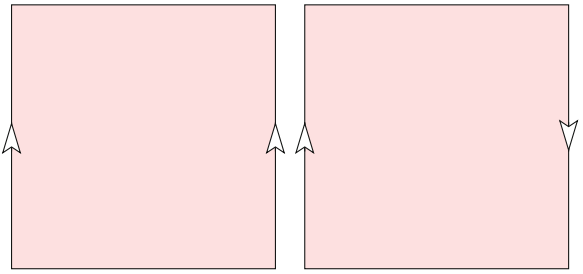

Fig. 8.1 *Left* the cylinder. *Right* the Möbius strip

$f : \mathbb{X} \to \mathbb{Y}$ such that f and f^{-1} are both continuous. If such a function exists, then $\mathbb{X}$ and $\mathbb{Y}$ are said to be *homeomorphic*, or *topologically equivalent*, or they have the same *topology type*, and this is denoted by writing $\mathbb{X} \approx \mathbb{Y}$. An *embedding* is a function $g : \mathbb{X} \to \mathbb{Y}$ whose restriction to the image, $g(\mathbb{X}) \subseteq \mathbb{Y}$, is a homeomorphism. For example, the open disk is homeomorphic to the plane. To prove this, we introduce $\mathbb{S}^1 = \{x \in \mathbb{R}^2 \mid \|x\| = 1\}$ and note that $\mathbb{B}^2 - \mathbb{S}^1$ is the prototypical open disk. The function $f : \mathbb{B}^2 - \mathbb{S}^1 \to \mathbb{R}^2$ defined by $f(x) = \frac{x}{1-\|x\|}$ is bijective and bicontinuous implying that $\mathbb{B}^2 - \mathbb{S}^1 \approx \mathbb{R}^2$. While f is a homeomorphism between the open disk and the plane, the identity defined by $g(x) = x$ is an embedding of the open disk in the plane.

The basic question in topology is to classify spaces up to topology type. For example, most coffee cups have the same type as the solid torus, which is a ball with a handle. However, this is a difficult undertaking in general, with known answers restricted to very limited situations.

8.2 2-Manifolds

Perhaps the best known family for which a complete classification into topological types is known are the compact surfaces. We define a 2-*manifold* (without boundary) as a topological space $\mathbb{X}$ for which every point $x \in \mathbb{X}$ has an open neighborhood homeomorphic to $\mathbb{R}^2$. As mentioned earlier, this is equivalent to saying that x has an open disk neighborhood. If the 2-manifold is connected and compact, then we sometimes call it a *compact surface*. Similarly, a 2-*manifold with* (possibly empty) *boundary* is a topological space $\mathbb{Y}$ for which every point $y \in \mathbb{Y}$ has a neighborhood homeomorphic to $\mathbb{R}^2$ or to $\mathbb{H}^2 = \{x = (x_1, x_2) \mid x_1 \geq 0\}$. The boundary of a 2-manifold with boundary is necessarily a 1-manifold, that is: a collection of closed curves. For example, the *cylinder* is a 2-manifold with boundary, and its boundary consists of two closed curves. As illustrated in Fig. 8.1, it can be constructed from a square by gluing the left edge to the right edge.

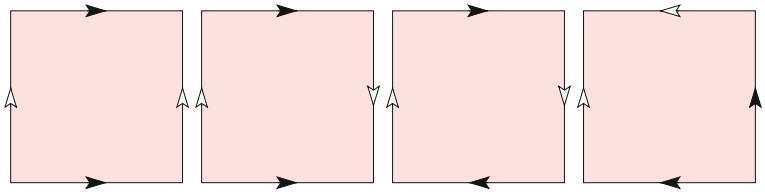

Fig. 8.2 From *left* to *right*: the torus, the Klein bottle, the projective plane, and the sphere

If we glue the same two edges with reversed orientation, as shown in Fig. 8.1 on the right, then we get the *Möbius strip*. In contrast to the cylinder, it is *non-orientable*, by which we mean that it is not possible to assign an unambiguous orientation (cw or ccw) to every point. Indeed, we can walk from a point x—standing on one side of the Möbius strip—to the same point x—standing on the other side of surface. A cw orientation observed at the first visit appears as a ccw orientation at the second visit. This implies that the surface has really only one side, and the two sides make sense only locally.

We can also glue the top to the bottom edge so that all boundary is removed and we get a compact surface. Doing this the obvious way, as shown in Fig. 8.2 on the left, we get the *torus*, and with one reversed orientation, as in the second picture, we get the *Klein bottle*. If we reverse the direction for both pairs of glued edges, as in the third picture, we get the *projective plane*. Finally, if we glue the top edge to the left, and the bottom edge to the right, as in Fig. 8.2 on the right, we get the *sphere*. The torus and the sphere can be embedded in $\mathbb{R}^3$, so we are quite familiar with their curved appearance, but the Klein bottle and the projective plane cannot, which is perhaps the reason why they are more difficult to imagine. The projective plane seems most difficult to imagine of all, so we offer an alternative construction. Starting with the sphere, $\mathbb{S}^2 = \{x \in \mathbb{R}^3 \mid \|x\| = 1\}$, we call points x and $-x$ *antipodal*. Gluing the antipodal points in pairs gives the projective plane. This recipe glues the northern hemisphere to the southern hemisphere, and it glues the equator to itself, like wrapping a hair-band twice around a pony-tail.

8.3 Classification of Compact Surfaces

We use the Euler characteristic and the orientability of a surface for classification. To define the former, we decompose the surface into triangles, making sure that any two meet in a shared edge, or a shared vertex, or not at all. Then we compute the *Euler characteristic* as the alternating sum of simplices: number of vertices minus number of edges plus number of triangles. The four triangulations shown in Fig. 8.3 all have 27 edges and 18 triangles, but they differ in the number of vertices, which from left to right is 9, 9, 10, 11. It follows that $\chi = 0, 0, 1, 2$ for the torus, the Klein bottle, the projective plane, and the sphere. Remembering that the torus and the sphere are orientable, and the Klein bottle and the projective plane are not, we use these two

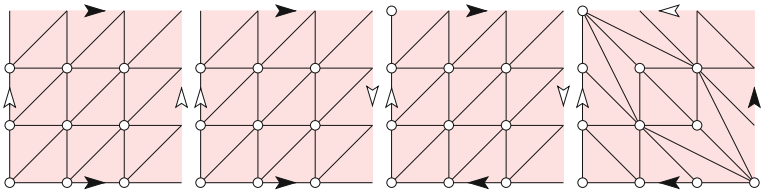

Fig. 8.3 From *left* to *right*: triangulations of the torus, the Klein bottle, the projective plane, and the sphere. Only one copy of each vertex and one copy of each edge is shown

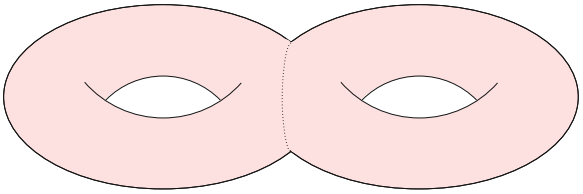

Fig. 8.4 The double torus obtained by forming the connected sum of two tori

pieces of information to tell the four surfaces apart. As observed about a century ago in [1, 2], the method extends to all compact surfaces. It is worth mentioning that such a classification is out of reach for 4-manifolds since it is undecidable whether two triangulated 4-manifolds are homeomorphic [3]. To explain this extension, we write $\mathbb{T}^2$ for the torus, $\mathbb{K}^2$ for the Klein bottle, $\mathbb{P}^2$ for the projective plane, and $\mathbb{S}^2$ for the sphere. Assuming triangulations, we can form the *connected sum* of two by removing a triangle in each and gluing their boundaries to each other; see Fig. 8.4.

Denoting this operation by #, we observe that forming the connected sum with $\mathbb{S}^2$ does not change the topological type. However, forming the connected sum with any of the other three surfaces does change the type. In all three cases, it changes the Euler characteristic, and if we start with an orientable surface and form the connected sum with the projective plane or the Klein bottle, it changes the surface from orientable to non-orientable. Define

$$T_n = \mathbb{T}^2 \# \mathbb{T}^2 \# \ldots \# \mathbb{T}^2, \tag{8.1}$$

$$P_n = \mathbb{P}^2 \# \mathbb{P}^2 \# \ldots \# \mathbb{P}^2 \tag{8.2}$$

by iterating the connected sum $n - 1 \geq 0$ times each. The Euler characteristic of the resulting spaces is $\chi(T_n) = 2 - 2n$ and $\chi(P_n) = 2 - n$. Note that all T_n are orientable and all P_n are non-orientable.

Classification Theorem *Two connected, compact 2-manifolds (without boundary) are homeomorphic iff they have the same Euler characteristic and they are both orientable or both non-orientable.*

This result amounts to saying that the sphere together with the two infinite families defined in (8.1) and (8.2) exhaust all compact surfaces. Note that it also says that $\mathbb{K}^2 \approx \mathbb{P}^2 \# \mathbb{P}^2$.

8.4 Simplicial Complexes

We are now more formal about the decompositions of surfaces into triangles. This is also a good moment to drop the restriction to two dimensions. A set of $k+1$ points, $u_0, u_1, \ldots, u_k$, is *affinely independent* if the k vectors, $u_1 - u_0, u_2 - u_0, \ldots, u_k - u_0$, are linearly independent. A *k-simplex* is the convex hull of $k+1$ affinely independent points. Writing σ for the k-simplex, we call $k = \dim \sigma$ its *dimension*, and u_0 to u_k its *vertices*. Simplices of dimension 0, 1, 2, 3 are usually referred to as *vertices*, *edges*, *triangles*, *tetrahedra*. A *face* of σ is a simplex spanned by a subset of the vertices of σ. Since a set of $k+1$ elements has $\binom{k+1}{\ell+1}$ subsets of size $\ell+1$, σ has this number of ℓ-faces, for $0 \leq \ell \leq k$. The total number of faces is therefore

$$\sum_{\ell=0}^{k} \binom{k+1}{\ell+1} = 2^{k+1} - 1, \tag{8.3}$$

the number of subsets minus 1 because we do not count the empty set. Sometimes it is convenient to call $\emptyset$ a face of σ, namely the unique (-1)-face, but we will refrain from this practice. When we triangulate a surface, we choose the triangles such that they have only proper intersections. Similarly, we define a *simplicial complex* as a finite collection of simplices, K, such that

(i) for every simplex $\sigma \in K$, every face of σ is in K;
(ii) for every two simplices $\sigma, \tau \in K$, the intersection, $\sigma \cap \tau$, is either empty or a face of both simplices.

If the intersection of two simplices is a common face, then (i) implies that it is a simplex in K. The *dimension* of K is the largest dimension of any simplex in K. A *subcomplex* is a subset of the simplices that is itself a simplicial complex. The *Euler characteristic* of K is the alternating sum of simplex numbers: $\chi(K) = s_0 - s_1 + s_2 - \ldots \pm s_k$, where k is the dimension of K and s_i is the number of i-simplices, for $0 \leq i \leq k$.

8.5 Triangulations

Until now, we have avoided any mention of the space in which the simplicial complex lives. Clearly, $k+1$ points in $\mathbb{R}^d$ can be affinely independent only if $k \leq d$. Suppose K is a simplicial complex in $\mathbb{R}^d$. The *underlying space* of K is the union of the simplices,

$$|K| = \bigcup_{\sigma \in K} \sigma, \tag{8.4}$$

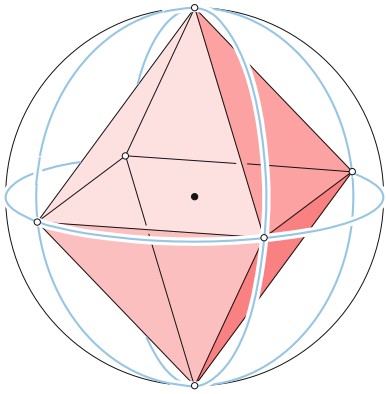

Fig. 8.5 Triangulation of the sphere. Each triangle of the octahedron projects centrally to a spherical triangle that covers an eighth of the sphere

together with the topology inherited from the Euclidean topology of $\mathbb{R}^d$. We recall that this means that the open sets in $|K|$ are the intersections of the open sets in $\mathbb{R}^d$ with the underlying space. Note that the underlying space is a set of points, while the simplicial complex is a set of simplices. Sometimes, this difference is important but more often than not, it is convenient to ignore it. A *triangulation* of a space $\mathbb{X}$ is a simplicial complex, K, whose underlying space is homeomorphic to $\mathbb{X}$.

Many times we talk about a topological space, we do so in reference to a triangulation of the space. It therefore pays off to elaborate on the concept by showing a few examples of spaces and triangulations. The simplicial complexes consisting of 18 triangles and their edges and vertices depicted in Fig. 8.3 are triangulations of the torus, the Klein bottle, the projective plane, and the sphere. We note a few pitfalls that need to be avoided when constructing a triangulation. For example, the triangulation of the square used for the torus on the left does not work for the sphere, because the identification of boundary edges would render some of the diagonal edges invalid, namely the ones for which the endpoints are the same. Similarly, the identification would render some edges the same, which is not allowed since the simplicial complex is a set and not a multi-set. Another popular method for constructing a triangulation is by projection. Take for example any convex polytope in $\mathbb{R}^3$ whose faces are all triangles, such as the octahedron. Fixing a point in the interior of the polytope as the origin, the central projection from the boundary of the polytope to the sphere is a homeomorphism; see Fig. 8.5.

References

1. Brahana HR (1922) Systems of circuits on two-dimensional manifolds. Ann Math 23:144–168
2. Dehn M, Heegard P (1907) Analysis situ. Enzykl Math Wiss III AB 3:153–220
3. Markov AA (1958) Insolubility of the problem of homeomorphy. In: Proceedings of International Congress of Mathematicians, pp 14–21

Chapter 9
Homology Groups

Given a topological space, its homology is a formal, algebraic way to talk about its connectivity. Better known than the homology groups are their ranks, which are the Betti numbers of the space. In contrast to most other topological formalisms that capture connectivity, homology groups have fast algorithms.

9.1 Little Creatures

We motivate the sequence definitions needed to construct homology groups with a simple example: the 2-dimensional torus. Embedding it in $\mathbb{R}^3$, it looks like a tire filled with air; see Fig. 9.1. Imagine the torus finely triangulated, too fine to show the vertices and edges. There is a large population of little creatures that explore the surfacethrough traveling. After some time, they all deliver a report describing in great detail the traveled path. Some never get back to where they started, and we find their reports not very useful in determining anything about the global connectivity of the surface. Many return to their original position, and from each we get a closed path as the record of the trip. There is a large number of records, some similar and some different. To bring order into the collection, we first select the closed paths that surround a piece of real estate. A closed path that does not fall into this category has the property that the real estate on its left is mysteriously connected to the real estate on its right. To find out, we consult the open path records we have prematurely labeled as useless for our exploration. But there are too many closed paths in the second category, so we further simplify by grouping closed paths that together surround a piece of real estate.

After this elaborate classification effort, we get the closed paths separated into four categories. In the first, all paths surround real estate, and in each of the other three, every pair of closed paths surrounds real estate. Strangely, there is also a relationship between the latter three categories: picking one closed path from each, the triplet again surrounds real estate, as shown in Fig. 9.1.

H. Edelsbrunner, *A Short Course in Computational Geometry and Topology*,
SpringerBriefs in Mathematical Methods, DOI: 10.1007/978-3-319-05957-0_9,

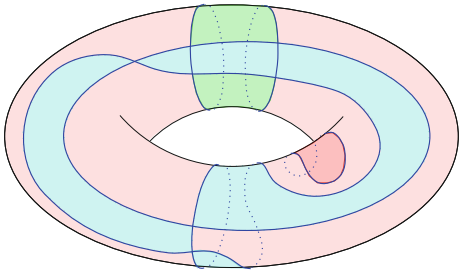

Fig. 9.1 The torus embedded in $\mathbb{R}^3$. Of the six paths, one surrounds real estate by itself, two form a pair with a ring-shaped piece of real estate between them, and three collaborate to surround a complicated piece of real estate that reaches all the way around the torus

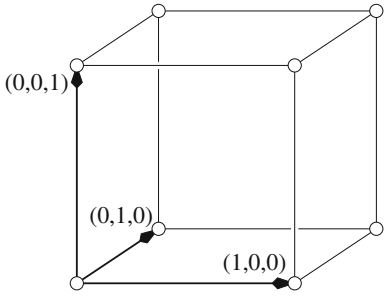

Fig. 9.2 Vector space interpretation of the chain group

9.2 Chain Groups

To formalize the exploration effort, we now sort the records into an algebraic structure. First, we specify what exactly these records are. Assuming a simplicial complex, K, a *p-chain* is a subset of the p-simplices in K. To relate this concept to our motivating example, we set $p = 1$, but we no longer insist that a chain is connected. Indeed, we can think of a chain as the travel record of a team of little creatures. This is useful because we may want to combine teams and their records to form new records. Formally, the *sum* of two p-chains is the symmetric difference between the two sets of p-simplices, which is again a p-chain. We write $(\mathsf{C}_p, +)$ for the thus constructed group of p-chains, or usually just C_p.

Before we continue, we point out that C_p may be large but it has the simple structure of a vector space. This vector space is generated by the s_p p-simplices in K, which we draw as the unit coordinate vectors of this s_p-dimensional space. Each coordinate takes only two values: 0 or 1, so the vector space consists of the vertices of the s_p-dimensional unit cube; see Fig. 9.2.

9.3 Chain Complex

To prepare the next step, we define the *boundary* of a p-simplex as the set of $(p-1)$-faces. There are $p+1$ of them. The boundary of a p-chain is the sum of the boundaries of its p-simplices: $\partial_p c = \sum_{\sigma \in c} \partial_p \sigma$. Note that the boundary is a $(p-1)$-chain. Hence, we can formalize this operation as a map between vector spaces:

$$\partial_p : \mathsf{C}_p \to \mathsf{C}_{p-1}, \tag{9.1}$$

called the *p-th boundary map* or *boundary homomorphism*. Indeed, it has the property of a homomorphism, namely that the map commutes with the group operation: $\partial_p(c+c') = \partial_p c + \partial_p c'$. If we now line up the groups and the maps, we get the *chain complex* of K:

$$\ldots \xrightarrow{\partial_{p+2}} \mathsf{C}_{p+1} \xrightarrow{\partial_{p+1}} \mathsf{C}_p \xrightarrow{\partial_p} \mathsf{C}_{p-1} \xrightarrow{\partial_{p-1}} \ldots. \tag{9.2}$$

Only the groups of dimension $0 \leq p \leq \dim K$ are non-trivial, but it is convenient to add trivial groups at both ends so we do not have to worry about where the line ends. Doing so requires that the entire group of 0-chains maps to 0, but this makes sense because the boundary of a vertex is empty in any case. Similarly, we will usually drop the index and not distinguish between maps in different dimensions.

Note that for $p \geq 2$, every $(p-2)$-face of a p-simplex, σ, belongs to exactly two $(p-1)$-faces. Hence, $\partial\partial\sigma = 0$, and this also holds for $p < 2$. It follows that applying the boundary map twice to a chain gives 0. This may be obvious but it is important so we state it explicitly.

Fundamental Lemma $\partial\partial c = 0$.

9.4 Cycles and Boundaries

Next, we focus on chains that close up, like the closed paths traveled by our little creatures. A *p-cycle* is a p-chain whose boundary is empty. Since the boundary map commutes with the group operation, the sum of two cycles is again a cycle: $\partial(z+z') = \partial z + \partial z' = 0$. We use this as a reminder that a cycle does not need to be connected, and its components may cross one another. The p-cycles thus form a subgroup of the p-chains, denoted as Z_p, for every dimension p. Recall that the *kernel* of a map is everything in the domain that maps to 0, so $\mathsf{Z}_p = \ker \partial_p$; see Fig. 9.3. Recall that for $p = 0$, the entire chain group maps to zero. This implies $\mathsf{Z}_0 = \ker \partial_0 = \mathsf{C}_0$.

A special type of cycle is obtained by taking the boundary of a higher-dimensional chain. A *p-boundary* is the boundary of a $(p+1)$-chain: $b = \partial_{p+1} c$, and since $\partial b = \partial\partial c = 0$, the p-boundary is indeed a p-cycle. Similarly, if $b' = \partial_{p+1} c'$ then

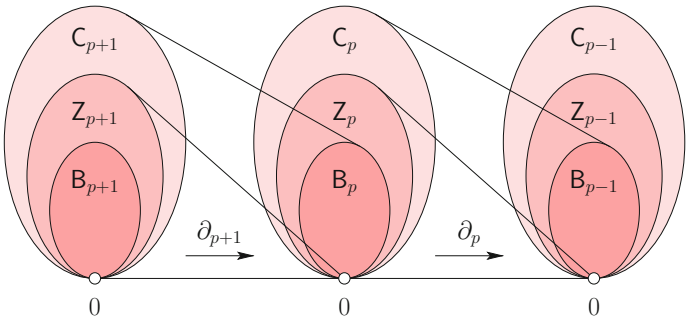

Fig. 9.3 Three consecutive groups in the chain complex. The cycle and boundary subgroups are shown as kernels and images of the boundary maps

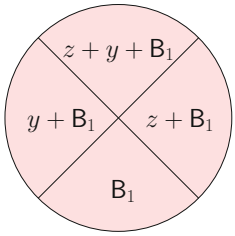

Fig. 9.4 Partition of the first cycle group of the torus into homology classes

$b + b' = \partial_{p+1}(c + c')$, so the sum of two p-boundaries is again a p-boundary. In other words, the p-boundaries form a subgroup of the p-cycles, as indicated in Fig. 9.3. We denote this subgroup by B_p. Recall that the *image* of a map consists of all elements in the range reached by elements in the domain. Hence, $\mathsf{B}_p = \operatorname{im} \partial_{p+1}$.

9.5 Homology

The boundary group collects all records that—in the language of our motivational example—surround real estate. We still have to categorize the cycles that do not bound. The idea is to not distinguish between two such cycles if they differ only by a boundary. In other words, we say $z, z' \in \mathsf{Z}_p$ are equivalent if there exists $b \in \mathsf{B}_p$ such that $z = z' + b$. Since this implies the existence of $c \in \mathsf{C}_{p+1}$ with $\partial c = b$, we conclude that $z - z' = z + z'$ is the boundary of a $(p + 1)$-chain. In this case, we write $z \sim z'$ and say that z and z' are *homologous*.

We can now categorize the cycles into equivalence classes, called *homology classes*; see Fig. 9.4. One such class is B_p, another is $z + \mathsf{B}_p$, where z is a p-cycle not in B_p. Note that $z' \in z + \mathsf{B}_p$ iff z and z' are homologous. Hence, $z \in z' + \mathsf{B}_p$, so it really does not matter whether we use z or z' to represent the class. The homology

classes that partition Z_p are the elements of the *p-th homology group*, and we write $\mathsf{H}_p = \mathsf{Z}_p/\mathsf{B}_p$, where taking the quotient is the algebraic way of saying that we partition into equivalence classes. Homology classes can also be added, simply by choosing representatives and adding them: $(z + \mathsf{B}_p) + (y + \mathsf{B}_p) = (z + y) + \mathsf{B}_p$. Indeed, if z' and y' are different representatives of the same classes, then $z' = z + b_1$ and $y' = y + b_2$, and therefore

$$(z' + y') + \mathsf{B}_p = (z + y + b_1 + b_2) + \mathsf{B}_p \tag{9.3}$$

$$= (z + y) + \mathsf{B}_p. \tag{9.4}$$

In other words, H_p together with the addition operation forms again a group, and in the particular case in which we add chains by taking the symmetric difference, it is a vector space. The rank of the p-th homology group is called the *p-th Betti number* of K: $\beta_p = \operatorname{rank} \mathsf{H}_p$. Recall that the elements in H_p are classes of size 2^r, where $r = \operatorname{rank} \mathsf{B}_p$. Since the classes partition Z_p, we have

$$2^{\operatorname{rank} \mathsf{Z}_p} = 2^{\operatorname{rank} \mathsf{B}_p} \cdot 2^{\operatorname{rank} \mathsf{H}_p}, \tag{9.5}$$

and therefore $\beta_p = \operatorname{rank} \mathsf{Z}_p - \operatorname{rank} \mathsf{B}_p$. This rank is a measure of the difference between the p-th cycle group and the p-th boundary group.

9.6 Euler-Poincaré Formula

Recall that the Euler characteristic of the simplicial complex is $\chi = s_0 - s_1 + s_2 - \cdots \pm s_k$, where $k = \dim K$. Recall also that $s_p = \operatorname{rank} \mathsf{C}_p$. We may partition the chain group by calling two chains equivalent if they differ by a cycle. The resulting quotient is known as the *coimage* of the map: $\operatorname{coi} \partial_p = \mathsf{C}_p/\mathsf{Z}_p$. All chains in an equivalence class map to the same $(p-1)$-boundary, which implies $\operatorname{rank} \operatorname{coi} \partial_p = \operatorname{rank} \mathsf{B}_{p-1}$ and therefore $\operatorname{rank} \mathsf{C}_p = \operatorname{rank} \mathsf{Z}_p + \operatorname{rank} \mathsf{B}_{p-1}$. We use this equation to rewrite the Euler characteristic:

$$\chi = \sum_{p=0}^{k} (-1)^p \left(\operatorname{rank} \mathsf{Z}_p + \operatorname{rank} \mathsf{B}_{p-1} \right) \tag{9.6}$$

$$= \sum_{p=0}^{k} (-1)^p \left(\operatorname{rank} \mathsf{Z}_p - \operatorname{rank} \mathsf{B}_p \right) \tag{9.7}$$

$$= \sum_{p=0}^{k} (-1)^p \beta_p. \tag{9.8}$$

In words, the Euler characteristic is equal to the alternating sum of Betti numbers. This result is known as the Euler-Poincaré Formula. It was established by Henri Poincaré, who thus achieved a vast generalization of Euler's relation, which was originally perceived for convex polytopes in $\mathbb{R}^3$. See [1] for a discussion of the mathematical effort that culminated in Poincaré's proof.

It is important to note that the Betti numbers do not depend on the triangulation but only on the triangulated space. While this is not obvious, it is true and justifies the practice to speak of the Betti numbers of the space and the Euler characteristic of the space, rather than of the triangulation. Similarly, we can talk about the homology groups of the space. However, here we need to mention that the groups depend on the triangulation; after all, the simplices are the elements of the cycles, which are the elements of the classes. Nonetheless, the homology groups of two different triangulations, K and L, of the same space are *isomorphic* to each other: $\mathsf{H}_p(K) \simeq \mathsf{H}_p(L)$. By this we mean that there is a bijection between the classes that commutes with the group operations.

9.7 Some Computations

To comprehend the meaning of the Betti numbers, it is helpful to know what they are for a few spaces we understand. We begin with a general statement about the meaning of the zeroth Betti number.

Claim A *β_0 is the number of connected components.*

Proof Recall that $\mathsf{H}_0 = \mathsf{Z}_0/\mathsf{B}_0$ and assume first that there is only one component. Every set of vertices is a cycle, hence the size of Z_0 is 2^{s_0}. Every even number of vertices is a boundary, and every odd set is not, which implies that the size of B_0 is 2^{s_0-1}. Finally, $\beta_0 = s_0 - (s_0 - 1) = 1$, as claimed. The argument for $j > 1$ the cycle group. The size of the cycle group is still 2^{s_0}. For a set of vertices to be a boundary, it needs to have an even number of vertices in each component. Writing n_i for the number of vertices in the i-th component, the number of boundaries is therefore

$$2^{n_1-1} \cdot 2^{n_2-1} \cdot \ldots \cdot 2^{n_j-1} = 2^{s_0-j}. \tag{9.9}$$

The rank of the zeroth homology group is therefore $\beta_0 = s_0 - (s_0 - j) = j$, as claimed. □

Next, we consider two special spaces.

Claim B *The non-zero Betti numbers of the torus and the Klein bottle are $\beta_0 = 1$, $\beta_1 = 2$, and $\beta_2 = 1$.*

The proof is not very difficult, in particular for $p = 0, 2$. For $p = 1$, we use the help of our little creatures, and we put them to work on the Klein bottle for which they generate the same result. Claim B illustrates the limitations of the discriminative power of homology groups. In particular, using modulo 2 arithmetic, homology

cannot distinguish between the torus and the Klein bottle. This is not necessarily a drawback. However, it is worth mentioning that the standard coefficients for homology are integers, and with them the first homology groups are different: isomorphic to $\mathbb{Z} \oplus \mathbb{Z}$ for the torus and to $\mathbb{Z} \oplus \mathbb{Z}_2$ for the Klein bottle. We chose $\mathbb{Z}^2$ coefficients for two reasons: they simplify the exposition by not requiring the simplices to be oriented, and they have homology groups that are vector spaces, which will be important when we discuss persistent homology groups.

Reference

1. Lakatos I (1976) Proofs and refutations: the logic of mathematical discovery. Cambridge University Press, Cambridge, England

Chapter 10
Complex Construction

There are not many ways to automatically construct interesting topological spaces, and by 'interesting' we mean spaces that go beyond the designed ones, such as the ball, and the sphere. Taking the nerve of a collection of sets is one such method, and we have seen examples: the Delaunay triangulation and the alpha complex. There, we did not have to worry about the embedding in Euclidean space, but in general we do.

10.1 Abstract Simplicial Complexes

It is often convenient to talk about a simplicial complex abstractly, without reference to its realization in a Euclidean space. Suppose we have a finite collection of elements, which can be anything but we call them *vertices*. An *abstract simplicial complex* is a system of subcollection, A, such that $\alpha \in A$ and $\beta \subseteq \alpha$ imply $\beta \in A$. This is the abstract equivalent of Condition (i) for simplicial complexes. There is no abstract equivalent of Condition (ii). The sets α are called *abstract simplices*. The *dimension* of an abstract simplex is one less than its cardinality, and the *dimension* of A is the maximum dimension of any of its abstract simplices. We can draw A in Euclidean space by mapping each vertex to a point in $\mathbb{R}^d$, and mapping each abstract k-simplex to the convex hull of the $k+1$ corresponding points. If this drawing satisfies Conditions (i) and (ii) of a simplicial complex, then we call it a *geometric realization* of A. For example, a 1-dimensional abstract simplicial complex is a graph, and a geometric realization is a straight-line embedding of the graph. Given an abstract simplicial complex, we ask whether it has a geometric realization in $\mathbb{R}^d$. The answer is in the affirmative if d is large enough.

Geometric Realization Theorem *Any abstract simplicial complex of dimension k has a geometric realization in $\mathbb{R}^{2k+1}$.*

For $k = 1$, the theorem says that every graph can be geometrically realized in $\mathbb{R}^3$. Indeed, if we place the vertices in general position, with no four points lie on a common plane, then no two edges can cross. Sometimes, $2k + 1$ is best possible.

H. Edelsbrunner, *A Short Course in Computational Geometry and Topology*, SpringerBriefs in Mathematical Methods, DOI: 10.1007/978-3-319-05957-0_10,

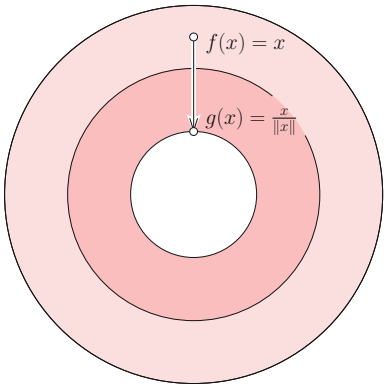

Fig. 10.1 With time, the homotopy moves $f(x)$ to $g(x)$. The half-way function, $\frac{1}{2}f + \frac{1}{2}g$, shrinks the annulus to half its width

For example, the graph with 5 vertices and all 10 edges cannot be drawn without crossing in $\mathbb{R}^2$. Hence, $2k+1 = 3$ dimensions are necessary to geometrically realize this graph. A generalization of this example to a k-dimensional abstract simplicial complex that has no geometric realization in $\mathbb{R}^{2k}$ can be found in [1].

10.2 Homotopy

We may compare two spaces by deforming one to the other, without gluing or cutting. We begin with deforming maps. Let $f, g : \mathbb{X} \to \mathbb{Y}$ be continuous maps from one topological spaces to another. A *homotopy* between f and g is a continuous map $H : \mathbb{X} \times [0, 1] \to \mathbb{Y}$ for which $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$ for all $x \in \mathbb{X}$. The two functions are *homotopic* if a homotopy exists. It is reasonable to think of the second variable as time, going from 0 to 1, and to consider the 1-parameter family of functions $f_t : \mathbb{X} \to \mathbb{Y}$ defined by $f_t(x) = H(x, t)$, which interpolates between $f_0 = f$ and $f_1 = g$.

As example, we consider two functions from the annulus to itself. Let $\mathbb{X} = \{x \in \mathbb{R}^2 \mid 1 \leq \|x\| \leq 3\}$, and define $f, g : \mathbb{X} \to \mathbb{X}$ by setting $f(x) = x$ and $g(x) = x/\|x\|$; see Fig. 10.1. To show that f and g are homotopic, we construct $H : \mathbb{X} \times [0, 1] \to \mathbb{X}$ by setting

$$H(x, t) = (1 - t) f(x) + t g(x). \tag{10.1}$$

We have $f_0 = f$ and $f_1 = g$ by construction. It is also clear that H is well defined, and that it is continuous.

10.3 Homotopy Equivalence

To compare two spaces, we construct two homotopies, both involving the identity map. Specifically, $\mathbb{X}$ and $\mathbb{Y}$ are *homotopy equivalent* if there exist continuous functions $a : \mathbb{X} \to \mathbb{Y}$ and $b : \mathbb{Y} \to \mathbb{X}$ such that $b \circ a$ is homotopic to $\mathrm{id}_{\mathbb{X}} : \mathbb{X} \to \mathbb{X}$ and $a \circ b$ is homotopic to $\mathrm{id}_{\mathbb{Y}} : \mathbb{Y} \to \mathbb{Y}$. In this case, we also say that $\mathbb{X}$ and $\mathbb{Y}$ have the same *homotopy type*, and we write $\mathbb{X} \simeq \mathbb{Y}$. We already did most of the work to show that the annulus, $\mathbb{X}$, has the same homotopy type as the circle, $\mathbb{Y} = \mathbb{S}^1$. Let a map the point $x \in \mathbb{X}$ to the point $a(x) = x/\|x\|$ in $\mathbb{Y}$, and let b be the canonical embedding of the circle in the annulus, that is: $b(y) = y \in \mathbb{X}$. We proved earlier that $g = b \circ a$ and $\mathrm{id}_{\mathbb{X}}$ are homotopic. To see the other direction, we note that $f = a \circ b$ is the identity on the circle: $f = \mathrm{id}_{\mathbb{Y}}$, which is stronger than f and $\mathrm{id}_{\mathbb{Y}}$ being homotopic. It follows that the annulus has the homotopy type of the circle. Even simpler than the circle is the point, and a space that has the homotopy type of the point is said to be *contractible*. For example, the disk is contractible, and so is every tree.

We note that being homotopy equivalent is weaker than being homeomorphic. Indeed, the annulus and the circle have the same homotopy type but not the same topology type. The two notions are different because the former permits a local change of dimension while the latter does not. On the other hand, having isomorphic homology groups is yet weaker than having the same homotopy type:

$$\mathbb{X} \approx \mathbb{Y} \Longrightarrow \mathbb{X} \simeq \mathbb{Y} \implies \mathsf{H}_p(\mathbb{X}) \simeq \mathsf{H}_p(\mathbb{Y}), \tag{10.2}$$

for all dimensions p. To compute the Betti numbers of $\mathbb{X}$, we may therefore find a homotopy equivalent space $\mathbb{Y}$ and compute its Betti numbers.

10.4 Nerves

We are now ready to introduce the announced construction of simplicial complexes from arbitrary collections of sets. We prefer finite, so let X be a finite collection of sets. The *nerve* of X is the system of subcollections of X whose sets have a non-empty common intersection:

$$\mathrm{Nrv}\, X = \left\{ \emptyset \neq V \subseteq X \mid \bigcap V \neq \emptyset \right\}. \tag{10.3}$$

Here, we use a short-form for taking the intersection of all sets in a collection: $\bigcap V = \bigcap_{v \in V} v$. We note that the nerve is an abstract simplicial complex because $V \in \mathrm{Nrv}\, X$ and $U \subseteq V$ implies $U \in \mathrm{Nrv}\, X$. To make this more clear, we call the singleton sets in the nerve abstract vertices, the pairs abstract edges, etc.

As mentioned earlier, the Delaunay triangulation of a point set is the nerve of the collection of Voronoi regions. More precisely, the Delaunay triangulation is the geometric realization of the nerve obtained by mapping each Voronoi region to the point

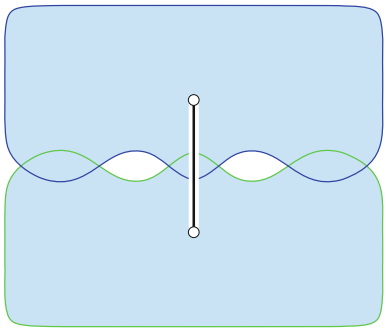

Fig. 10.2 Two contractible regions whose union is a disk with two holes. The nerve consists of two vertices connected by an edge, which has a different homotopy type

that generates the region. Here it is important that the points are in general position, else the dimension of the nerve can be higher than that of the ambient Euclidean space. A geometric realization in this space would then be impossible. However, we can always go to higher-dimensional Euclidean spaces to construct geometric realizations. Independent of the realization, the nerve has the same homotopy type as the union.

Nerve Theorem *If all sets in X are closed and triangulable, and all non-empty common intersections of the sets are contractible, then* $\mathrm{Nrv}\, X$ *and* $\bigcup X$ *have the same homotopy type.*

If all sets in X are convex, then their non-empty common intersections are convex and therefore contractible. This is the situation for restricted Voronoi regions, so the Nerve Theorem implies that the alpha complex has the same homotopy type as the union of balls it represents. It follows that it has isomorphic homology groups and therefore the same Betti numbers. If the sets are not convex then be aware that the contractibility of the sets themselves is not sufficient to conclude homotopy equivalence. To get an example, it suffices to have two contractible regions overlap in multiple locations, as in Fig. 10.2. We note that the example is similar to Fig. 3.2, which shows two regions in an Apollonius diagram.

10.5 Čech Complexes

Suppose we simplify the construction of the alpha complex by intersecting the disks directly, without first restricting them to the corresponding Voronoi regions. To describe this more formally, let S be a finite set of points in $\mathbb{R}^2$, let $r \geq 0$ be a real number, and write $D_s(r)$ for the disk with center s and radius r, as before. The *Čech complex* is isomorphic to the nerve of the disks:

$$\check{\mathrm{C}}\mathrm{ech}(r) = \{\emptyset \neq T \subseteq S \mid \bigcap_{s \in T} D_s(r) \neq \emptyset\}. \tag{10.4}$$

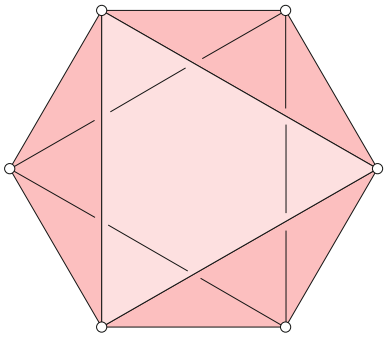

Fig. 10.3 The Vietoris-Rips complex of six points equally spaced on the unit circle and a parameter $\sqrt{3}/2 < r < 1$. It consists of eight triangles connected like the faces of an octahedron

Even if the points are in general position, the radius may be large enough so that the complex has dimension larger than 2. For this to happen, it suffices that four disks have a non-empty common intersection, such as in Fig. 7.2. Nevertheless, the Nerve Theorem implies that the Čech complex has the same homotopy type as the union of disks. Similarly, the Nerve Theorem implies that the alpha complex for radius r, denoted as $A(r)$, has the same homotopy type as the union of disks. Since homotopy equivalence is transitive—every equivalence relation is—this implies $|A(r)| \simeq |\check{\text{C}}\text{ech}(r)|$. Note also that $A(r)$ is isomorphic to a subcomplex of $\check{\text{C}}\text{ech}(r)$, simply because the restricted disks are subsets of the unrestricted disks.

It might be interesting to study the structure of the extra simplices in the Čech complex. How does it relate to the substitution method that reduces a long Pie formula to a short Pie formula; see Chap. 7.

10.6 Vietoris-Rips Complexes

To construct the Čech complex, we need to test whether a collection of disks has a non-empty intersection, which can be difficult or, in some metric spaces, impossible. We now define a complex that needs only the distances between the points in S for its construction. Letting $r \geq 0$ be a real number, the *Vietoris-Rips complex* of S and r, denoted as Vietoris-Rips(r), consists of all abstract simplices in 2^S whose vertices are at most a distance $2r$ from one another. In other words, we connect any two vertices at distance at most $2r$ from each other by an edge, and we add a triangle or higher-dimensional simplex to the complex if all its edges are in the complex; see Fig. 10.3.

While the Vietoris-Rips complex is easy to construct, it generally does not have the homotopy type of the union of disks of radius r. Indeed, for the 6 points in Fig. 10.3, the disks with radius r form an annulus, which has $\beta_0 = \beta_1 = 1$ and $\beta_p = 0$ for all $p \neq 0, 1$. In contrast, the Vietoris-Rips complex triangulates a 2-sphere, which has $\beta_0 = \beta_2 = 1$ and $\beta_1 = 0$. This example suggests that Vietoris-Rips complexes can have topological artifacts that do not show up in the data. While this is true, the artifacts are limited.

Vietoris-Rips Lemma *Let S be a finite set in $\mathbb{R}^2$. Then $\hat{C}ech(r) \subseteq Vietoris\text{-}Rips(r) \subseteq \check{C}ech(r')$, where r' is $\frac{2\sqrt{3}}{3} = 1.154\ldots$ times r.*

Proof The first containment is obvious because Čech(r) and Vietoris-Rips(r) have the same edges. For the second containment, we note that the equilateral triangle with edges of length $2r$ has a circumcircle of radius $r' = \frac{2\sqrt{3}r}{3}$. If three points lin on this circle, then their disks of radius r' have a non-empty common intersection, so the triangle belongs to Čech(r'). The longest of the three edges connecting the points has length at least $2r$, implying that all triangles in Vietoris-Rips(r) have a circumcircle of radius at most r'. This implies the second claimed containment. □

Reference

1. Flores A (1933/34) Über n-dimensionale Komplexe die in $\mathbb{R}_n$ selbstverschlungen sind. Ergeb Math Koll 6:4–7

Part III
Homology

Exercises

Question 1. ($20 = 10 + 10$ points). Recall the conditions (i) and (ii) in the definition of a simplicial complex.

(a) Give a small collection of simplices that violates (i) but not (ii).
(b) Give a small collection of simplices that violates (ii) but not (i).

Question 2. ($20 = 10 + 10$ points). Let K be a triangulation of a compact surface, and let L be another triangulation obtained by refining K. In other words, each simplex in L is contained in a simplex in K.

(a) Determine the Euler characteristic of a simplicial complex that triangulates $\mathbb{B}^2$.
(b) Use the result in (a) to prove that K and L have the same Euler characteristic.

Question 3. ($20 = 10 + 10$ points). Consider the one-dimensional simplicial complex with vertices A, B, C, D, E and edges AB, AC, AD, BE, CE, DE.

(a) What are the ranks of $\mathrm{C}_1, \mathrm{Z}_1$ and B_0?
(b) How many 1-chains have the same boundary? In other words, how many elements of C_1 map to the same element in B_0?

Question 4. ($20 = 10 + 10$ points). Recall the four 2-manifolds depicted in Figs. 9.2 and 9.3.

(a) What are the Betti numbers of the projective plane?
(b) What are the Betti numbers of the double torus, $T_2 = \mathbb{T}^2 \# \mathbb{T}^2$?

Question 5. ($20 = 7 + 7 + 6$ points). Let A_n^2 be a two-dimensional abstract simplicial complex with n vertices and $\binom{n}{3}$ triangles. Determine the largest value of n such that A_n^2 has a geometric realization

(a) in $\mathbb{R}^2$;
(b) in $\mathbb{R}^3$;
(c) in $\mathbb{R}^5$.

Question 6. ($20 = 6 + 7 + 7$ points). Let A be an abstract simplicial complex, and v, w vertices not in A. The *suspension* of A consists of all simplices of A together with $\{v\}$, $\{w\}$, and all simplices of the form $\alpha \cup \{v\}$ and $\alpha \cup \{w\}$, for $\alpha \in A$.

(a) Verify that the suspension of A is an abstract simplicial complex.
(b) Describe the space triangulated by the suspension of $\mathbb{S}^1$.
(c) What are the Betti numbers of the suspension of the letter B?

Part IV
Persistence

Persistent homology addresses a major weakness of homology, namely its instability under small changes. Reconstructing a tire by scanning its surface and connecting the millions of points into a triangulation, we expect a torus, but if we miss only one percent of one percent of the triangles, the first Betti number will be in the hundreds rather than equal to two. The theory of persistence is inspired by Morse theory and related to spectral sequences, both topics studied in topology. It adds a quantitative component to the Betti numbers that is stable under perturbations. Perhaps most importantly, it captures the various scales under which a dataset can be interpreted in one consistent formalism.

Chapter 11
Filtrations

Complicated shapes and data sets—in particular those that occur in nature—have different appearance depending on the resolution of the observation. Accordingly, persistence measures holes in terms of the range of scales at which they appear [1]. To begin, we need a representation of the data at different scale levels.

11.1 Alpha Complex Filtration

A *filtration* is an increasing sequence of topological spaces, each contained in the next. An example is the filtration already introduced in Chap. 5. Let S be a finite set of points in general position in $\mathbb{R}^3$, perhaps the centers of the atoms making up the Gramicidin protein, and write K for its Delaunay triangulation. We recall the basic steps in the construction of the filtration of alpha complexes, four of which are displayed in Fig. 11.1. For each simplex $\sigma \in K$, there is a real number α_σ such that σ belongs to the α-complex of S iff $\alpha_\sigma \leq \alpha$. We index the $n = s_0 + s_1 + s_2 + s_3$ simplices such that $\alpha_{\sigma_1} \leq \alpha_{\sigma_2} \leq \cdots \leq \alpha_{\sigma_n}$, calling the corresponding sequence of simplices a *filter* if every simplex is preceded by its faces. To achieve this for the Delaunay triangulation, we only need to make sure that ties in the ordering are broken such that lower-dimensional simplices precede higher-dimensional simplices. Assuming a filter, we let K_j be the collection of the first j simplices, noting that it is a simplicial complex for every j. The increasing sequence of complexes,

$$\emptyset \;=\; K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n \;=\; K \tag{11.1}$$

is called a *flat* filtration because any two contiguous complexes differ by only one simplex. Every α-complex belongs to the flat filtration, but not every complex in (11.1) is an α-complex. In other words, the alpha complex filtration is a subsequence of (11.1) and it is generally not flat.

H. Edelsbrunner, *A Short Course in Computational Geometry and Topology*,
SpringerBriefs in Mathematical Methods, DOI: 10.1007/978-3-319-05957-0_11,

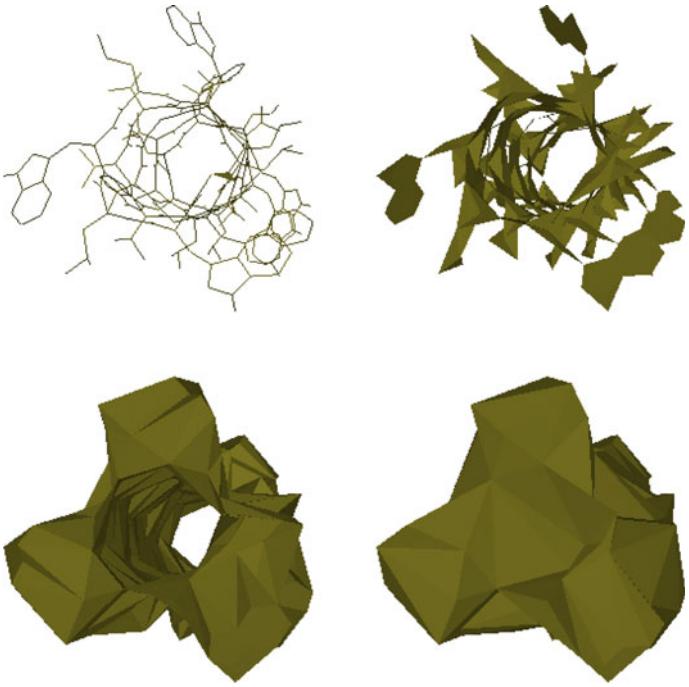

Fig. 11.1 Four alpha complexes of the Gramicidin protein. It functions as an ion channel embedded in a cell membrane

11.2 Incremental Algorithm

Since consecutive complexes in (11.1) are very similar, it is not surprising that it is easy to compute the Betti numbers of K_{j+1} if we know the Betti numbers of K_j. Indeed, there is only one additional simplex, σ_{j+1}, and we just need to determine how the addition of that simplex affects the connectivity of K_j. Starting with the empty complex, we compute the Betti numbers by adding one simplex at a time. Since our complexes are geometrically realized in $\mathbb{R}^3$, only the first three Betti numbers are possibly non-zero.

$\beta_0 = \beta_1 = \beta_2 = 0$;
`for` $j = 0$ `to` $n - 1$ `do`
  $p = \dim \sigma_{j+1}$;
  `if` $\sigma_{j+1} \in z \in \mathsf{Z}_p(K_{j+1})$ `then` $\beta_p = \beta_p + 1$
    `else` $\beta_{p-1} = \beta_{p-1} - 1$
  `endif`
`endfor`.

In the original publication [2], this algorithm was used for flat filtrations in $\mathbb{R}^3$, where p-simplices that belong to p-cycles can be recognized in almost constant time. More

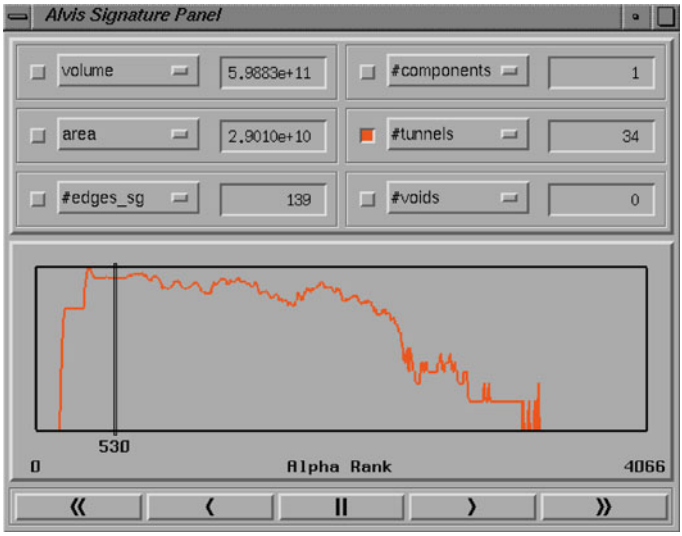

Fig. 11.2 Part of the interface to the alpha shape software showing $\log_2(\beta_1 + 1)$ as a function of the index along the alpha complex filtration of the Gramicidin protein

precisely, the time for a decision is the inverse of Ackermann's function of n, which is at most 6 unless n is unimaginably large. The connection is that each decision can be reduced to a few operations in a UNION-FIND data structure; see [3]. This is most easily seen for $p = 1$, when σ_{j+1} is an edge. It belongs to a 1-cycle in K_{j+1} iff its two endpoints belong to the same component in K_j. Storing the components as sets of vertices, we can decide whether or not this is the case by two FIND operations. If the two endpoints belong to different components, we follow up with a UNION operation that merges them into one.

11.3 Topological Noise

Having a fast algorithm for computing the Betti numbers of all complexes in a filtration, we try it out on the points marking the centers of the atoms of the Gramicidin protein displayed in Fig. 11.1. While we might have expected a single tunnel for a good fraction of the complexes, we get a few tens of tunnels almost always. The graph in Fig. 11.2 shows the first Betti number as a function of the index in the filtration. At the marked position, we have $\beta_1(K_{530}) = 34$.

It seems daunting to extract the one tunnel that is responsible for the functioning as an ion channel. The missing information is when a tunnel is born and when it dies. Adding this information, we produce β_1 as a function of the index and the index persistence, which is defined as the absolute difference between the index at the

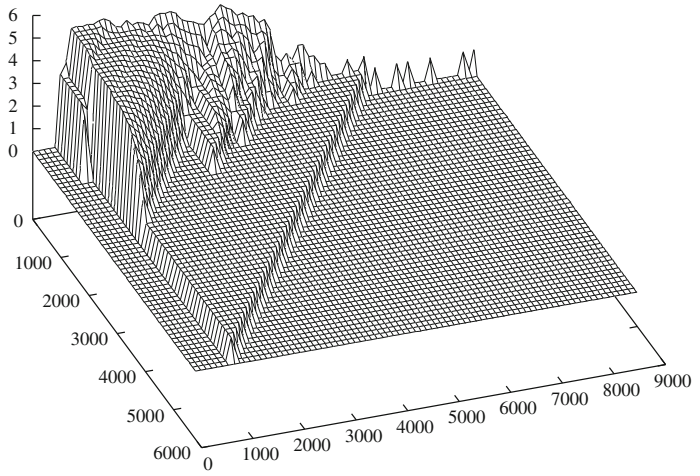

Fig. 11.3 Graph of $\log_2(\beta_1 + 1)$, the number of persistent tunnels in log-scale for the Gramicidin protein. The index in the filtration varies from *left* to *right* and the persistence from *back* to *front*

birth and the index at the death; see Fig. 11.3. We see a single triangular plateau in the graph of the function that reaches significantly further to the front than anything else; it corresponds to the channel that passes long-side through the protein.

11.4 Birth and Death

We now address the eternal question of birth and death. The starting point is a sequence of homomorphisms connecting the homology groups of the complexes in the filtration. For $i \leq j$, K_i is a subcomplex of K_j, which can be written as an injective map $f^{i,j} : K_i \hookrightarrow K_j$. It maps a simplex in K_i to the same simplex in K_j, which is why it is called the *inclusion map*. It carries over to an inclusion map on the p-cycles, $f_p^{i,j} : \mathsf{Z}_p(K_i) \hookrightarrow \mathsf{Z}_p(K_j)$. This in turn *induces* a map on homology:

$$\varphi_p^{i,j} : \mathsf{H}_p(K_i) \to \mathsf{H}_p(K_j), \tag{11.2}$$

which generally is not an inclusion map. Specifically, if γ is a class in $\mathsf{H}_p(K_i)$, and $z \in \mathsf{Z}_p(K_i)$ is a representative cycle, we let $\varphi_p^{i,j}(\gamma)$ be the class in $\mathsf{H}_p(K_j)$ that contains $f_p^{i,j}(z)$. It should be clear that the definition of $\varphi_p^{i,j}$ does not depend on the choice of representative. It takes a p-cycle in K_i and *pushes* it *forward* to K_j. For example, if γ surrounds a hole in K_i that fills up at the time we reach K_j, then $\varphi_p^{i,j}$ maps γ to $0 \in \mathsf{H}_p(K_j)$. The image of $\varphi_p^{i,j}$ is called a *persistent homology group*. It contains all p-dimensional homology classes that have representatives already

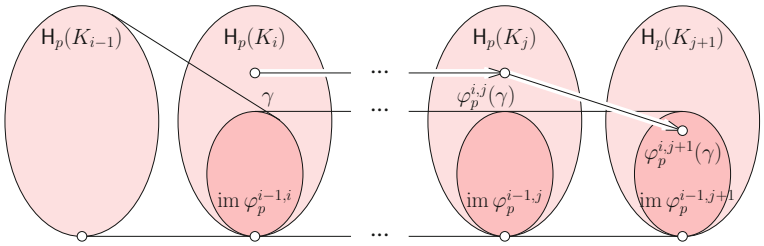

Fig. 11.4 The class γ is born at K_i and dies entering K_{j+1}

present in K_i. The corresponding *persistent Betti number* is the rank of the persistent homology group. It counts the p-dimensional holes that exist all the way from K_i to K_j. For a particular class, we are interested in the smallest index i and the largest index j such that the class is non-trivial within the entire interval from K_i to K_j. This motivates the following definitions, which are illustrated in Fig. 11.4. We say a class $\gamma \in \mathsf{H}_p(K_i)$ is *born* at K_i if γ is not in the image of $\varphi_p^{i-1,i}$, and a class born at K_i *dies entering* K_{j+1} if $\varphi_p^{i,j}(\gamma)$ is not in the image of $\varphi_p^{i-1,j}$ but $\varphi_p^{i,j+1}(\gamma)$ is in the image of $\varphi_p^{i-1,j+1}$. The *index persistence* of γ is $j - i + 1$. In most applications, we have a function that governs the construction of the filtration, and we call the difference between the function values at the birth and the death the *persistence* of the class. In an alpha complex filtration, this would be the function that maps each simplex, σ, to the threshold, α_σ, beyond which the simplex belongs to the α-complex.

11.5 Barcodes

It is time for a concrete example. Consider the construction of the simplicial complex consisting of a single tetrahedron and its faces, as illustrated in Fig. 11.5. Note the three Betti numbers given at each stage of the construction. From each stage to the next, exactly one Betti number changes, and it changes by ± 1. This is in agreement with the incremental algorithm, whose case analysis distinguishes between births and deaths. The missing piece of information is the pairing of births and deaths, as provided by the definition of persistent homology groups. We illustrate this by drawing the 16 events from left to right in Fig. 11.6. Pairing up the births and deaths, as described, we get a set of intervals or *bars*, which is called the *barcode* of the filtration. Each bar represents a class in one of the homology groups and thus has a definite dimension. One bar in our example does not end. It represents the component that is born at the beginning and does not die.

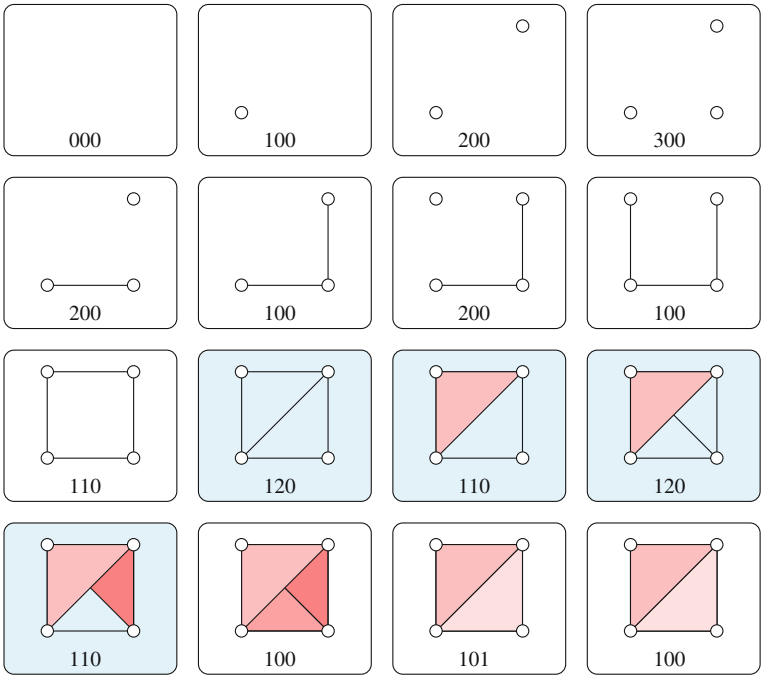

Fig. 11.5 Constructing a tetrahedron one simplex at a time. The three numbers are β_0, β_1, β_2, in this sequence. Two homology classes exists during the interval of four highlighted steps

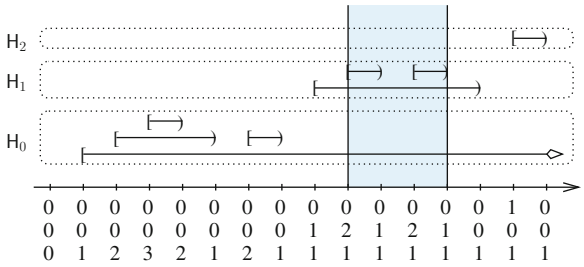

Fig. 11.6 The barcode of the filtration in Fig. 11.5. Two bars span the shaded interval from $i = 9$ to $j = 12$

11.6 Persistence Diagrams

The number of bars tends to get large with more complicated shapes and filtrations. For these more complicated cases, a representation as points in the plane is more convenient. For example, we can replace each bar by the top vertex of the right-angled, isosceles triangle above the bar, as in Fig. 11.7. Usually, we draw all points in one diagram, labeling them by the dimension of the classes they represent.

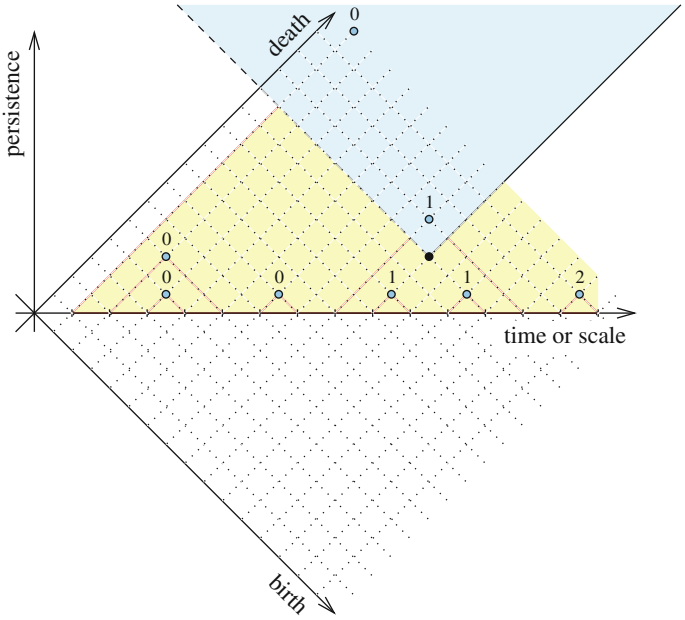

Fig. 11.7 The persistence diagram of the filtration in Fig. 11.5. One point is at infinity, and together with one other point, it lies inside the V-shaped region defined by $i = 9$ and $j = 12$

Given a persistence diagram, we ask how we can read off the Betti numbers of the complexes in the filtration or, more generally, the persistent Betti numbers of the images of the maps between the homology groups. We describe how this works for the more general case. The rank of the image of $\mathsf{H}_p(K_i)$ in $\mathsf{H}_p(K_j)$ is the number of classes that are born at or before K_i and that die entering K_{j+1} or later. Drawing the half-open V-shaped region whose lowest point is $(i + j, j - i)$, we can count the points labeled with dimension p to get the desired rank.

References

1. Edelsbrunner H, Letscher D, Zomorodian A (2002) Topological persistence and simplification. Discrete Comput Geom 28:511–533
2. Delfinado CJA, Edelsbrunner H (1995) An incremental algorithm for Betti numbers of simplicial complexes on the 3-sphere. Comput Aided Geom Design 12:771–784
3. Tarjan RE (1983) Data structures and network algorithms. SIAM, Philadelphia, Pennsylvania

Chapter 12
PL Functions

Other than from alpha, Čech, and Vietoris-Rips complexes, we get filtrations from real-valued functions on topological spaces. The computationally most convenient form are the lower star filtrations of piecewise linear maps, which we introduce in this section. We will use them to formulate an important property of persistence diagrams, namely their stability under small perturbations. We begin with standard material on piecewise linear functions.

12.1 Piecewise Linear Interpolation

Functions on a topological space are conveniently constructed using a triangulation. To explain how this works, we recall that the k-simplex spanned by points $u_1, u_2, \ldots, u_{k+1}$ consists of all points $x = \sum_{i=1}^{k+1} \lambda_i u_i$, with $\sum_{i=1}^{k+1} \lambda_i = 1$ and $\lambda_i \geq 0$ for all i. The λ_i are the *barycentric coordinates* of x; see Fig. 12.1. Suppose the k-simplex belongs to a simplicial complex, K, with vertices u_i, for $1 \leq i \leq n$, and that we have a function value, $f(u_i)$, for each. To extend these values to the points x of the k-simplex, we set $\lambda_i = 0$, for $k+2 \leq i \leq n$, and define

$$f(x) = \sum_{i=1}^{n} \lambda_i f(u_i). \tag{12.1}$$

Doing this for all points $x \in |K|$, we get a continuous, piecewise linear function $f : |K| \to \mathbb{R}$. We call f a *PL function*, tacitly assuming that it is linear—and not just piecewise linear—on each simplex.

H. Edelsbrunner, *A Short Course in Computational Geometry and Topology*,
SpringerBriefs in Mathematical Methods, DOI: 10.1007/978-3-319-05957-0_12,

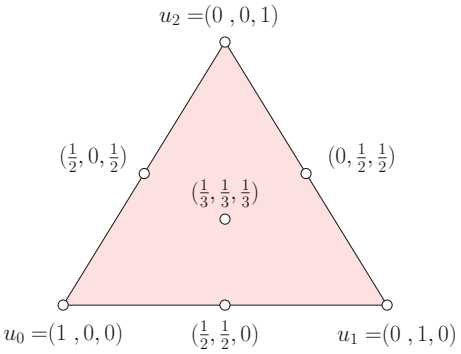

Fig. 12.1 A triangle with 7 of its points labeled by their barycentric coordinates: the 3 vertices, the midpoints of the 3 edges, and the centroid or barycenter

12.2 Sublevel Sets

For the remainder of this section, we assume that K is a simplicial complex, and $f : |K| \to \mathbb{R}$ is a PL function. For convenience, we also assume that the vertices have distinct function values, so we can sort them as $f(u_1) < f(u_2) < \cdots < f(u_n)$. Given a value $v \in \mathbb{R}$, the *level set*, $f^{-1}(v)$, consists of all points $x \in |K|$ that map to that value. Similarly, the *sublevel set* and the *superlevel set* are $f^{-1}(-\infty, v]$ and $f^{-1}[v, \infty)$. For the time being, we are primarily interested in the sublevel sets, which form a filtration:

$$\emptyset \subseteq f^{-1}(-\infty, v] \subseteq f^{-1}(-\infty, w] \subseteq |K|, \tag{12.2}$$

for any $v \leq w$. It is not difficult to see that the sublevel set can change its homotopy type only at the function value of a vertex. To unify the different cases of this claim, we set $f(u_0) = -\infty$ and $f(u_{n+1}) = \infty$.

Lemma A *Let $v, w \in \mathbb{R}$ and $0 \leq i \leq n$ such that $f(u_i) \leq v \leq w < f(u_{i+1})$. Then the sublevel sets have the same homotopy type*: $f^{-1}(-\infty, v] \simeq f^{-1}(-\infty, w]$.

We omit the proof, which is similar to that of the next claim.

12.3 Full Subcomplexes

To formulate it, let K_i be the full subcomplex of K that contains the vertices u_1 to u_i but not u_{i+1} to u_n. The K_i form again a filtration. In contrast to (12.2), the filtration of the full subcomplexes is finite. Nevertheless, the two are closely related.

Lemma B *Let $v \in \mathbb{R}$ and $0 \leq i \leq n$ such that $f(u_i) \leq v < f(u_{i+1})$. Then the full subcomplex and the sublevel set have the same homotopy type*: $|K_i| \simeq f^{-1}(-\infty, v]$.

Proof A good picture says it all. We therefore draw a simplicial complex in the plane, a sublevel set, and the full subcomplex defined by the vertices in this sublevel

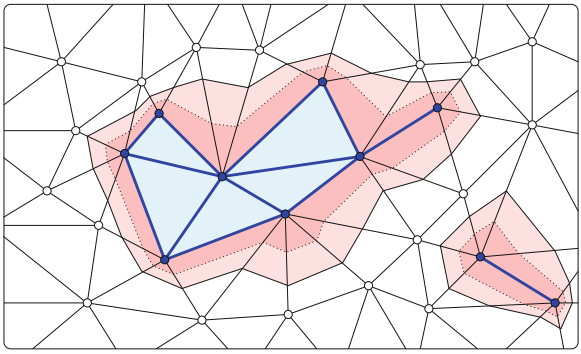

Fig. 12.2 From inside out: the full subcomplex, the remainder of the sublevel set, and the remainder of the simplicial complex. The division of the sublevel set into two zones illustrates the deformation used to prove the homotopy equivalence

set; see Fig. 12.2. The portion of the sublevel set that is not contained in $|K_i|$ can be partitioned into line segments that connect the boundaries of the two. Writing $\mathbb{X} = f^{-1}(-\infty, v]$, we introduce $f_t : \mathbb{X} \to \mathbb{X}$ that shrinks each of these line segments to $1 - t$ times its original length, namely the piece of this length that contains the endpoint in $|K_i|$. Naturally, the restriction of f_t to $|K_i|$ is the identity.

It is now easy to prove that $\mathbb{X}$ and $|K_i|$ have the same homotopy type. Set $a : \mathbb{X} \to |K_i|$ by defining $a(x) = f_1(x)$, for all $x \in \mathbb{X}$, and let $b : |K_i| \to \mathbb{X}$ be the inclusion map. Then $b \circ a$ is homotopic to $\mathrm{id}_{\mathbb{X}}$, as shown above, and $a \circ b$ is equal to the identity on $|K_i|$. □

The result of Lemma B motivates the following construction. Recall that the star of a vertex, u, contains all simplices that share u. The *lower star* of u is the subset of simplices in the star whose vertices have function values smaller than or equal to $f(u)$. Since we assumed that all vertices have distinct function values, each simplex belongs to a unique lower star, namely to that of the vertex with largest function value. It follows that K_i is the disjoint union of the lower stars of the vertices u_1 to u_i, and we can construct K_{i+1} by adding the simplices in the lower star of u_{i+1} to K_i. This is the reason we call the sequence of the K_i the *lower star filtration* of f.

12.4 PL 2-Manifolds

For the time being, let us assume that K triangulates a 2-manifold (without boundary). In this case, the star of a vertex, u, has the topology of an open disk. Recall that the *link* of u is the collection of faces of the star that do not belong to the star themselves; it has the topology of a circle. Similarly, we define the *lower link* of u as the full subcomplex of the link whose vertices satisfy $f(v) < f(u)$. In other words, the lower link contains all simplices in the link whose points all have function value less than $f(u)$. We call u a *regular vertex* if the lower link is non-empty and connected.

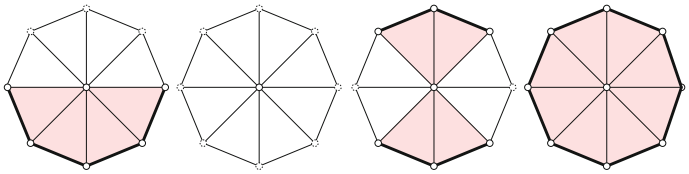

Fig. 12.3 From *left* to *right*: the lower star and lower link of a regular vertex, a minimum, a 1-fold saddle, a maximum

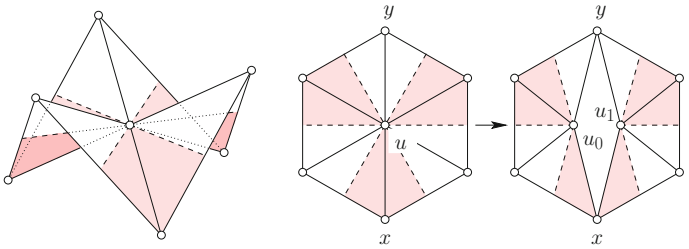

Fig. 12.4 Splitting a monkey or 2-fold saddle into two simple saddles

Otherwise, u is a *critical vertex*, and $f(u)$ is a *critical value*. Among the critical vertices, we distinguish between a *minimum* (the lower link is empty), a *maximum* (the lower link is the entire circle), and a *k-fold saddle* (the lower link has $k+1 \geq 2$ components); see Fig. 12.3.

If u is a k-fold saddle, for $k \geq 2$, we can split it into k *simple* or 1-fold saddles. As illustrated in Fig. 12.4, we take points x and y in the link such that $f(x) < f(u) < f(y)$ and at least one component of the lower link lies between x and y. We replace u by two vertices, u_0 and u_1, and we connect u_0 with the ccw arc from x to y and u_1 with the ccw arc from y to x. The function values at the new vertices are chosen close to $f(u)$ but slightly different from each other. The result is a j-fold saddle, u_0, and an ℓ-fold saddle, u_1, with $j + \ell = k$ and $j, \ell \geq 1$. Repeating this operation, we eventually get k simple saddles.

12.5 Alternating Sum

Suppose now that we have split all multiple saddles so that the only critical vertices are minima, simple saddles, and maxima. With this simplification, we are ready to follow the lead of traditional Morse theory [1] and define the *index* of a critical vertex:

$$\mathrm{ix}(u) = \begin{cases} 0 & \text{if } u \text{ is a minimum,} \\ 1 & \text{if } u \text{ is a simple saddle,} \\ 2 & \text{if } u \text{ is a maximum.} \end{cases} \tag{12.3}$$

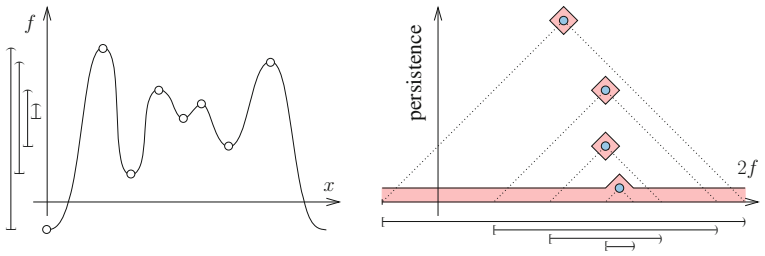

Fig. 12.5 *Left* the graph of a function on the circle. *Right* the corresponding persistence diagram, with the diamonds specifying a neighborhood of the diagram

It is easy to prove that the alternating sum of the critical vertices is the same as the alternating sum of the Betti numbers, namely the Euler characteristic. Let c_p be the number of critical vertices of index p.

Lemma C $\chi(K) = \sum_{p \geq 0} (-1)^p c_p$.

Proof We sort the vertices in the order of increasing function values, and construct K by adding the lower stars of the vertices in this sequence. Let K_i be the union of the lower stars of u_1 through u_i, as before. If u_{i+1} is a regular vertex, then its lower star has one vertex, ℓ edges, and $\ell - 1$ triangles. Adding this lower star does not change the Euler characteristic: $\chi(K_{i+1}) = \chi(K_i)$. For the three types of critical vertices, we have $\chi(K_{i+1}) = \chi(K_i) + (-1)^{\text{ix}(u_{i+1})}$, as required. Indeed, for a minimum we have just one vertex, for a simple saddle we have one vertex and two more edges than triangles, and for a maximum we have one vertex and the same number of edges and triangles. □

We note that the claimed equation is one of the Morse inequalities that relate the number of critical points of different indices with the Betti numbers of different dimensions. These inequalities can be found in any textbook on Morse Theory, as well as in [2].

12.6 Bottleneck Distance

We now return to the concept of persistence and, in particular, to persistence diagrams and their stability under perturbations. Let $f : |K| \to \mathbb{R}$ be a PL function and write $F = \mathsf{Dgm}(f)$ for the persistence diagram of the filtration of sublevel sets. Suppose we have a second PL function, $g : |K| \to \mathbb{R}$, with persistence diagram $G = \mathsf{Dgm}(g)$. We measure the distance between the functions by taking the maximum difference between function values:

$$\|f - g\|_\infty = \max_{x \in \mathbb{X}} |f(x) - g(x)|. \tag{12.4}$$

We would like to say that if g is near f, then the two have similar shape. This is not true if we characterize shape in terms of the critical points since they can move a far distance by a small perturbation, in particular if they are located on relatively flat regions of the function. The critical values are better suited as stable descriptors, by they can appear and disappear. More interestingly, the critical values form a stable shape description even after pairing them to give the points in the persistence diagram. To make this more concrete, we need the *bottleneck distance* between two diagrams, which is the length of the longest edge in a perfect matching:

$$W_\infty(F, G) = \inf_{\mu: F \to G} \max_{A \in F} \|A - \mu(A)\|_\infty, \tag{12.5}$$

where the infimum is over all bijections from F to G. Here, we are allowed to borrow points from the horizontal axis to form bijections if the two diagrams do not have the same number of points, or when this can shorten the longest edge. Note that we measure the distance between two points, A and $\mu(A)$, by the maximum absolute difference between their birth-values and their death-values. To justify the name, we observe that the bottleneck distance satisfies the axioms of a metric:

1. $W_\infty(F, G) = 0$ iff $F = G$;
2. $W_\infty(F, G) = W_\infty(G, F)$;
3. $W_\infty(F, H) \leq W_\infty(F, G) + W_\infty(G, H)$,

in which H is the persistence diagram of a third function on the same domain.

12.7 Stability

We are now ready to state the claim of stability for PL functions.

Stability Theorem *Let f and g be PL functions with persistence diagrams F and G. Then $W_\infty(F, G) \leq \|f - g\|_\infty$.*

In the original publication [3], this result is formulated for the more general class of *tame functions*, whose sublevel sets have finite rank homology groups and which have only finitely many homological critical values. A more elementary proof can be found in [4].

We remark that the inequality in the Stability Theorem is true if we substitute the Hausdorff distance for the bottleneck distance between diagrams, but this would be a substantially weaker result. Note also that the upper bound on the bottleneck distance cannot be complemented by a similar lower bound. Indeed, completely different functions can have identical persistence diagrams. This should not be surprising since topologically different spaces can have isomorphic homology groups.

References

1. Milnor J (1963) Morse theory. Princeton University Press, Princeton, New Jersey
2. Edelsbrunner H, Harer JL (2010) Computational Topology. An Introduction. Amer Math Soc, Providence, Rhode Island
3. Cohen-Steiner D, Edelsbrunner H, Harer J (2007) Stability of persistence diagrams. Discrete Comput Geom 37:103–120
4. Cohen-Steiner D, Edelsbrunner H, Morozov D (2006) Vines and vineyards by updating persistence in linear time. In: Proceedings 22nd annual symposium on computational geometry, pp 119–126

Chapter 13
Matrix Reduction

We have seen how mathematical reasoning can be used to compute persistent homology for simple filtrations. In this section, we present a more generally applicable computer algorithm that constructs persistence diagrams through matrix reduction. It can be viewed as a common extension of the incremental algorithm, discussed two sections ago, and the classic normal form algorithm, which can also be found in standard texts on algebraic topology [1].

13.1 Boundary Matrices

The input to the algorithm are the boundary maps represented by binary matrices. Recall that the p- and $(p-1)$-simplices form bases of the vector spaces of p- and $(p-1)$-chains. The p-th boundary map is fully specified by the p-th *boundary matrix*,

$$\partial_p = \partial_p[1 \ldots s_{p-1}, 1 \ldots s_p], \tag{13.1}$$

whose columns correspond to the p-simplices, whose rows correspond to the $(p-1)$-simplices, and whose entries record the face relation, with $\partial_p[i, j] = 1$ if the i-th $(p-1)$-simplex is a face of the j-th p-simplex, and $\partial_p[i, j] = 0$, otherwise. For example, the four boundary matrices of the tetrahedron are

$$\partial_0 = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}, \tag{13.2}$$

$$\partial_1 = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}, \tag{13.3}$$

H. Edelsbrunner, *A Short Course in Computational Geometry and Topology*, SpringerBriefs in Mathematical Methods, DOI: 10.1007/978-3-319-05957-0_13,

$$\partial_2 = \begin{bmatrix} 0&0&1&1 \\ 0&1&0&1 \\ 1&0&1&0 \\ 1&1&0&0 \\ 1&0&0&1 \\ 0&1&1&0 \end{bmatrix}, \tag{13.4}$$

$$\partial_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}. \tag{13.5}$$

We use standard row and column operations to reduce each matrix to *normal form*. By this we mean a matrix that is entirely zero, except for an initial segment of the diagonal, which consists of 1s. Since we use modulo 2 arithmetic, it is easy enough to get the matrices into this form. For the four boundary matrices above, we get

$$R_0 = \begin{bmatrix} 0&0&0&0 \end{bmatrix}. \tag{13.6}$$

$$R_1 = \begin{bmatrix} 1&0&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&0&0&0 \end{bmatrix}, \tag{13.7}$$

$$R_2 = \begin{bmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \end{bmatrix}, \tag{13.8}$$

$$R_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{13.9}$$

The dimensions of the null-spaces are the ranks of the cycle groups, the number of 1s in the diagonals are the ranks of the boundary groups in one lower dimension, and we get the Betti numbers by subtraction Table 13.1:

Table 13.1 The ranks of the cycle and boundary groups and the Betti numbers of the tetrahedron

	$p=0$	$p=1$	$p=2$	$p=3$
rank Z_p	4	3	1	0
rank B_p	3	3	1	0
β_p	1	0	0	0

13.2 Reduction Algorithm

To compute the Betti numbers, we need no particular ordering on the simplices, but for computing persistence, the ordering is essential. We therefore modify the classic algorithm in two ways, storing all boundary maps in one big matrix, and using only left-to-right column additions and no row operations. These are not enough to transform the matrix into normal form, but we will see that they suffice to give the needed rank information.

Let σ_1 to σ_n be the filter, and write $\partial = \partial[1 \dots n, 1 \dots n]$ for the big boundary matrix, with $\partial[i, j] = 1$ if $\dim \sigma_i = \dim \sigma_j - 1$ and σ_i is a face of σ_j, and $\partial[i, j] = 0$, otherwise. The algorithm reduces the matrix by iterating through the columns from left to right. It zeroes out a column j from bottom to top, to the extent possible. To explain this, we set $\text{pivot}(j)$ equal to the maximum row index for which column j has a non-zero entry. We set $\text{pivot}(j) = 0$ is the entire column is zero.

```
R = ∂;
for j = 1 to n do
  while ∃k < j with pivot(k) = pivot(j) ≠ 0 do
    add column k to column j
  endwhile
endfor.
```

We call the resulting matrix *reduced* because each row contains at most one pivot. We shortly discuss why the pivots are significant.

While the above pseudo-code states only two loops explicitly, there are really three. The innermost loop adds two columns of length n in time proportional to n. The middle loop searches for a column to the left of column j that has the pivot in the same row. It iterates until no such column can be found, and at each step, the pivot of column j moves up by at least one position. It follows that we iterate fewer than n times. Finally, the outer loop iterates over n columns. The total running time of the algorithm is therefore at most cubic in the number of simplices. All fast implementations of this algorithm use sparse matrix representations of ∂ and R. While ∂ is necessarily sparse, it is not clear that R inherits this property, but very often it does. Already the original publication on persistent homology [2] used a sparse matrix representation in which the columns are stored as lists that store only the non-zero entries. The availability of fast software is indeed an important factor in the success of the mathematical framework within applied communities.

13.3 Translation to Barcode

We illustrate the algorithm with an example, first turning the filtration in Fig. 11.1 into a single boundary matrix. We recall that we have a sequence of 15 simplices that correspond to the rows and the columns of the boundary matrix; see Fig. 13.1.

$$\partial = \begin{array}{c} \begin{array}{cccccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & \end{array} \\ \left[\begin{array}{cccccccccccccccc} 0 & 0 & 0 & \boxed{1} & 0 & 0 & \boxed{1} & 0 & \boxed{1} & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 & \boxed{1} & \boxed{1} & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & \boxed{1} & \boxed{1} & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & \boxed{1} & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 0 & \boxed{1} & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & \boxed{1} & 0 & 0 & \boxed{1} & 0 & 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 & \boxed{1} & 0 & 0 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 0 & \boxed{1} & 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 & 0 & \boxed{1} & 0 & 9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & \boxed{1} & 0 & 0 & 11 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 12 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 13 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 14 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 15 \end{array}\right] \end{array}$$

Fig. 13.1 The big boundary matrix of the filtration of the tetrahedron in Fig. 11.5

$$R = \begin{array}{c} \begin{array}{cccccccccccccccc} & & & & 4 & & & 4 & 4 & & & & & 12 & & \\ & & & & 5 & & & 5 & 5 & & 4 & & & 13 & & \\ & & & & 7 & & & 8 & 9 & & 7 & & 12 & 10 & & \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & \end{array} \\ \left[\begin{array}{cccccccccccccccc} 0 & 0 & 0 & \boxed{1} & \boxed{1} & 0 & \boxed{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & \boxed{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & \boxed{1} & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 & \boxed{1} & 0 & 0 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 0 & \boxed{1} & \boxed{1} & 0 & 0 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 & 0 & 0 & 0 & 9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 & 0 & 11 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 12 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 13 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & 14 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 15 \end{array}\right] \end{array}$$

Fig. 13.2 The reduced boundary matrix with lowest non-zero entries highlighted. The stacks of column indices at the *top* indicate the left-to-right column additions used to reduce the matrix

It should be clear that even just storing this matrix is a waste of resources because most entries are zero. However, the algorithm is easiest to explain for the full matrix, and we worry about speeding up the algorithm later.

The above matrix is small enough that we can run the algorithm by hand, displaying the result in Fig. 13.2. Each zero column is a cycle, adding to the rank of the corresponding cycle group. Together with the pivots, they encode all the information we need. The meaning of a pivot in row i and column $j+1$ is that the addition of σ_i creates a class that the addition of σ_{j+1} destroys. In other words, there is a homology class born at K_i that dies entering K_{j+1}. The dimension of the class is the dimension of σ_i. To translate this information into the barcode, we draw an interval from i to $j+1$ for every pivot in row i and column $j+1$, and we draw an interval from i to infinity if column i is zero but row i does not contain any pivot.

Translating the reduced matrix in Fig. 13.2 gives the barcode in Fig. 11.2. This is perhaps easiest to see by translating the reduced matrix into the persistence diagram in Fig. 11.3. To do this, we rotate the matrix by 45° in a ccw order so that its diagonal lies horizontal. The pivots become the points in the diagram, and we just need to add points at infinity for zero columns that do not correspond to pivots.

13.4 Uniqueness of Pivots

The number of pivots is the rank of the reduced matrix. Since column operations do not alter the rank, this number is also the rank of the initial boundary matrix. More than the number, we now show that the pivots themselves are unique and do not depend on the sequence of column operations used to reduce the matrix. To prove this, we consider lower-left submatrices ∂_i^j obtained from ∂ by deleting the first $i-1$ rows and the last $n-j$ columns. For example, $\partial = \partial_1^n$. Taking the alternating sum of the ranks of four submatrices gives

$$r_\partial(i, j+1) \quad = \quad \operatorname{rank} \partial_i^{j+1} + \operatorname{rank} \partial_{i+1}^{j} - \operatorname{rank} \partial_i^j - \operatorname{rank} \partial_{i+1}^{j+1}.$$

Substituting R for ∂, we get again submatrices and alternating sums of ranks. Since left-to-right column additions do not alter the ranks of any of these submatrices, $\operatorname{rank} \partial_i^j = \operatorname{rank} R_i^j$. Writing

$$\begin{aligned} A &= \operatorname{rank} R_i^{j+1}, & B &= \operatorname{rank} R_i^j, \\ C &= \operatorname{rank} R_{i+1}^{j+1}, & D &= \operatorname{rank} R_{i+1}^j, \end{aligned}$$

we therefore have $r_\partial(i, j+1) = r_R(i, j+1) = A - B + C - D$; see Fig. 13.3.

Pivot Uniqueness Lemma *An entry $R[i, j+1]$ is the pivot of column $j+1$ in the reduced boundary matrix iff $r_\partial(i, j+1) = 1$.*

Proof Since $r_\partial = r_R$, it suffices to show $\operatorname{pivot}(j+1) = i$ iff $r_R(i, j+1) = 1$. We prove this in a case analysis, making use of the fact that the non-zero columns in R and any of its lower-left submatrices are linearly independent.

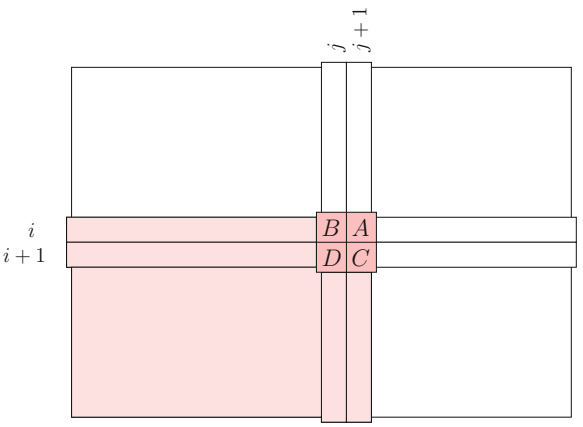

Fig. 13.3 Four submatrices of ∂ or of R, and their ranks

First, assume $\text{pivot}(j+1) = i$. Since the last column is non-zero, we have $A - B = 1$. If we now delete the top row from R_i^{j+1}, then the last column is zero, implying $C - D = 0$. Hence, $r_R(i, j+1) = 1$, as required. Second, assume $\text{pivot}(j+1) \neq i$. If $\text{pivot}(j+1) < i$, then the last column is zero, and we get $A - B = C - D = 0$. Else if $\text{pivot}(j+1) > i$, then the last column is non-zero, even after deleting the top row, and we again get $A - B = C - D = 0$. The claimed result follows. □

An off-shot of this analysis is an explanation why the reduction algorithm does not continue the elimination of non-zero entries above the pivots.

13.5 Alternative Algorithm

The Pivot Uniqueness Lemma can also be used to justify variants of the algorithm. For example, we could use the current column to reduce future columns, as opposed to using past columns to reduce the current column.

```
R = ∂;
for j = 1 to n do
  i = pivot(j);
  for ℓ = j + 1 to n do
    if R[i, ℓ] = 1 then add column j to column ℓ endif
  endfor
endfor.
```

This works provided we stipulate that $R[0, \ell] = 0$ for all columns ℓ. The running time is similar to the earlier algorithm, namely at most cubic in the number of simplices. While the resulting reduced matrices computed by the two algorithms

may be different, the Pivot Uniqueness Lemma implies that the sets of pivots are the same. This is comforting because it confirms that the persistence diagram is uniquely defined by the filtration and does not depend on the algorithm that computes it.

References

1. Munkres JR (1984) Elements of algebraic topology. Perseus Publications, Cambridge, Massachusetts
2. Edelsbrunner H, Letscher D, Zomorodian A (2002) Topological persistence and simplification. Discrete Comput Geom 28:511–533

Part IV
Persistence

Exercises

Question 1. ($20 = 6 + 7 + 7$ points). Let S be the set of points $A = (-1, 0, 0)$, $B = (1, 0, 0)$, $C = (0, -1, 0)$, $D = (0, 1, 0)$, $E = (0, 0, -1)$, $F(0, 0, 1)$ in $\mathbb{R}^3$. Note that the convex hull of S is a regular octahedron.

(a) Draw the simplicial complex, K, whose vertices are the six points in S, that contain the edge connecting (± 1, 0, 0), and whose underlying space is the convex hull of S?

(b) List the simplices of K such that σ precedes τ either because $\dim \sigma < \dim \tau$ or because the dimensions are equal and the ordering of the vertices, sorted by name, is lexicographically smaller for σ.

(c) Draw the barcode and the persistence diagram of the filtration defined by the filter in (b).

Question 2. ($20 = 6 + 7 + 7$ points). Let S be a set of n points in general position in $\mathbb{R}^2$.

(a) Show that the number of different alpha complexes of S is less than $5n$.

(b) Show that the number of different Vietoris–Rips complexes of S is less than n^2.

(c) Show that the number of different Čech complexes of S is less than n^3.

Question 3. ($20 = 10 + 10$ points). The *dunce cap* is a two-dimensional topological space obtained by gluing the three edges of a triangle, as shown in Fig. 1.

(a) Draw a triangulation of the dunce cap (making sure that the identification of vertices and edges on the boundary of the triangle does not violate basic properties of a simplicial complex).

(b) Compute the Betti numbers of the dunce cap.

Question 4. ($20 = 6 + 7 + 7$ points). Let A, B, C three vertices of a triangle, Pick two numbers S, $T \in [0, 1]$, independently and uniformly at random from the unit

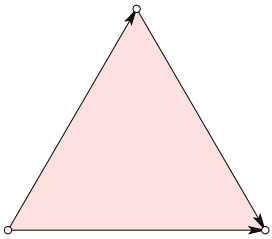

Fig. 1 The dunce cap. Gluing the *left* to the *bottom* edge, we get a cone with a rim and a seam. Gluing the rim to the seam, we get the dunce cap, which is perhaps similar to a snail house

interval, and set $s = \min\{\mathrm{S}, \mathrm{T}\}$ and $\mathrm{t} = \max\{\mathrm{S}, \mathrm{T}\}$. Finally, define $\alpha = s$, $\beta = t - s$, $\gamma = 1 - t$, interpreting them as the barycentric coordinates of the point $x = \alpha A + \beta B + \gamma C$.

(a) What is the probability that S and T are both smaller than or equal to $\frac{1}{2}$? Draw the corresponding region of points in the triangle?

(b) What are the expected values of α, β, γ?

(c) Prove that x is chosen uniformly at random from the triangle.

Question 5. ($20 = 10 + 10$ points). Let $X = (b, d)$ and $X' = (b', d')$ be points in the plane. The L_1 -and L_∞ -distances between them are $\| X - X' \|_1 = | b - b' | + | d - d' |$; $\| X - X' \|_\infty = \max\{| b - b' |, | d - d' |\}$.

(a) Draw the unit disk (the points at distance at most 1 from the origin) under both definitions of distance.

(b) Define $A = (b + d, d - b)$ and $A' = (b' + d', d' - b')$. Prove $\| A - A' \|_1 = 2\| X - X' \|_\infty$.

Question 6. (20 points). Recall the two matrix reduction algorithms for persistent homology given in Chap. 13. Construct a boundary matrix such that the two algorithms produce different reduced matrices.

Epilogue

We hope the reader has enjoyed the material arranged in this short book. There are two chapters in **Discrete and Computational Geometry** and two chapters in **Computational Algebraic Topology**. There is no clear border separating these fields, and we make an effort to further blur the difference and convey that there are three complementing ingredients: *geometry*, *topology*, and *algorithms*, and leaving any one out is to the detriment of the others.

For the reader who asks herself where to go from here, we mention a few texts that develop the threads we started in various directions. There are several established texts in **Computational Geometry**, and we recommend [1] because it is currently most representative of the field, and [2] because it focuses on Voronoi and Delaunay diagrams. The field of **Computational Topology** is considerably younger, and we mention [3] because of its focus on algebraic algorithms for homology, and [4] as a general text in the area. The field of **Algorithms** is a core subject within computer science, and we recommend [5] for a broad representation of the field, and [6] for an in-depth treatment of the data structures which are instrumental to obtain fast algorithms. Within **Discrete Geometry**, we recommend [7] for a modern treatment of convex polytopes, and [8] for its focus on circles, spheres, and transformations. The mathematical forerunners of the topological material in this book can be found in **Morse Theory** [9] and [10], in **Differential Topology** [11] and [12], and most importantly in **Algebraic Topology** [13] and [14]. In this context, we mention [15] which makes an effort to introduce the novice to the beauty of algebraic topology. This text is recommended if you are not yet acquainted to the way of thinking in topology, which at first may seem strange but soon becomes so natural that is seems strange it was strange at first.

H. Edelsbrunner, *A Short Course in Computational Geometry and Topology*,
SpringerBriefs in Mathematical Methods, DOI: 10.1007/978-3-319-05957-0,

References

1. de Berg M, Cheong O, van Krefeld M, Overmars M (2008) Computational geometry. Algorithms and applications, 3rd edn. Springer, Berlin, Germany
2. Edelsbrunner H (2001) Geometry and topology for mesh generation. Cambridge University Press, Cambridge, England
3. Kaczynski T, Mischaikow K, Mrozek M (2004) Computational homology. Springer, New York
4. Edelsbrunner H, Harer JL (2010) Computational Topology. An introduction. American mathematical society, Providence, Rhode Island
5. Cormen TH, Leiserson CE, Rivest RL, Stein C (2001) Introduction to algorithms, 2nd edn. MIT Press, Cambridge, Massachusetts
6. Tarjan RE (1983) Data structures and network algorithms. SIAM, Philadelphia, Pennsylvania
7. Ziegler GM (1995) Lectures on polytopes. Springer, New York
8. Pedoe D (1970) Geometry. A comprehensive course. Dover, New York
9. Milnor J (1963) Morse theory. Princeton University Press, Princeton, New Jersey
10. Matsumoto Y (2002) An introduction to Morse theory. Translated from (Japanese by Hudson K, Saito M), American mathematical society, Providence, Rhode Island
11. Guillemin V, Pollack A (1974) Differential topology. Prentice Hall, Englewood Cliffs, New Jersey
12. Milnor J (1965) Topology from the differential viewpoint, Revised edn. Princeton University Press, Princeton, New Jersey
13. Munkres JR (1984) Elements of algebraic topology. Perseus, Cambridge, Massachusetts
14. Hatcher A (2002) Algebraic topology. Cambridge University Press, Cambridge, England
15. Giblin PJ (1981) Graphs, Surfaces and homology. Chapman and Hall, London, England