Computational Topology for Point Data: Betti Numbers of α-Shapes

Vanessa Robins

Department of Applied Mathematics, Research School of Physical Sciences and Engineering, Australian National University, Canberra ACT 0200.
e-mail: vanessa.robins@anu.edu.au

Abstract. The problem considered below is that of determining information about the topology of a subset $X \subset \mathbb{R}^n$ given only a finite point approximation to X. The basic approach is to compute topological properties – such as the number of components and number of holes – at a sequence of resolutions, and then to extrapolate. Theoretical foundations for taking this limit come from the inverse limit systems of shape theory and Čech homology. Computer implementations involve constructions from discrete geometry such as alpha shapes and the minimal spanning tree.

1 Introduction

Two objects have the same topology if they are homeomorphic, i.e. when there is a continuous transformation from one to the other, with a continuous inverse. This means topological properties give a fundamental description of structure, and one that is independent of geometry. It is clear that objects can have the same topology and completely different geometry (a coffee cup and a donut are archetypal examples). However, the converse is also true: objects can have very similar geometric properties but a vastly different topology – the fractals in Sect. 5.2 are one example. Thus, in order to characterize spatial structures, we need both geometric and topological information.

Much attention has been given to the computation of geometric quantities from data, for example the Minkowski functionals [21, 22] and fractal dimensions [3, 26], but the field of 'computational topology' is relatively new [6]. The earliest work on extracting topological information from data targeted digital images. For black and white pixel or voxel images there are algorithms for labelling connected components [16], and for computing the Euler characteristic [19, 25] (a measure of connectivity that also appears as the zero-dimensional Minkowski functional). The technique of erosion and dilation is used in conjunction with these algorithms to probe structure at different length scales in digital images [2].

In this paper we discuss topological properties such as the number of connected components and number of k-dimensional holes, i.e., the Betti numbers, β_k. The Betti numbers are related to the Euler characteristic, χ, via the Euler-Poincaré formula:

$$\chi = \beta_0 - \beta_1 + \beta_2 - \ldots + \beta_d.$$

It follows that the Betti numbers give a more detailed description of topological structure than the Euler characteristic alone. Formal definitions of the Betti numbers are given in Sect. 2.

K.R. Mecke, D. Stoyan (Eds.): LNP 600, pp. 261–274, 2002.

We assume the data are given as a finite point pattern, $S \subset \mathbb{R}^d$. To give this finite set of points some non-trivial topological structure, we must first fatten or coarse-grain the set. This can be achieved, for example, by overlaying a digital mesh and colouring pixels if they contain data points, or by attaching spheres of radius α at each point. We focus on the latter objects, called alpha neighbourhoods or alpha parallel bodies, because they form a more flexible framework.

Once we choose the coarsening method, we have to set an appropriate level of coarse-graining. This might be physically motivated, or we may have complete freedom to choose. Without any a priori information, it is helpful to coarse-grain at a sequence of resolutions. This has the additional advantages of allowing us to detect fractal scaling, and if the data, S, approximate an underlying object, $X \subset \mathbb{R}^d$, we can extrapolate information about the topology of X. Theoretical underpinnings of this coarse-graining and extrapolation are described in Sect. 3.

To compute Betti numbers, we need a discrete complex (e.g., a triangulation) that has the same topology as the alpha neighbourhood. One solution comes from Edelsbrunner's alpha shapes [8] and their construction is described in Sect. 4. We use the publicly available alpha shape software [1] to compute Betti numbers of some example data in 2D and present the results in Sect. 5.

2 Homology Groups and Betti Numbers

Homology theory is a branch of topology that attempts to distinguish between objects by constructing algebraic invariants that reflect their connectivity properties. Readers should be warned that the homology groups do not completely determine the topology of an object. If two objects have different homology groups, then they certainly have different topologies. The converse, however, does not hold. For example, the set X shown in Fig. 3 has the same homology groups as a circle, but it is not homeomorphic to a circle because of the branch-point where the 'O' joins the 'C'.

In this section, we give a brief outline of the definitions for simplicial homology, following Munkres [24]. The basic building block is an *oriented k-simplex*, σ^k, the convex hull of $k+1$ geometrically independent points, $\sigma^k = [x_0, x_1, \ldots, x_k]$. Its orientation is defined by an arbitrary but fixed ordering of the vertices. Even permutations of this ordering give the same orientation and odd permutations reverse it. For example, a 0-simplex is just a point, a 1-simplex is a line segment, a 2-simplex a triangle, and a 3-simplex is a tetrahedron. A *simplicial complex*, $\mathcal{C}$, is a collection of oriented simplices with the property that the non-empty intersection of two simplices in $\mathcal{C}$ must itself be a simplex in $\mathcal{C}$. The set-theoretic union of all simplices from $\mathcal{C}$, when viewed as a subset of $\mathbb{R}^d$, is called a *polytope*. If a subset X of $\mathbb{R}^d$ is homeomorphic to a polytope, we say X is *triangulated* by $\mathcal{C}$.

The simplicial complex, $\mathcal{C}$, is given a group structure by defining the addition of k-simplices in a similar manner to addition in a vector space. The resulting free group is called the *chain group*, C_k. Its elements consist of *k-chains*, the sum of a finite number of oriented k-simplices: $c_k = \sum_i a_i \sigma_i^k$. The coefficients, a_i, are typically integers, but in general they can be elements of any abelian group, G. In this paper, we use the rational or real numbers for coefficients since this simplifies the homology group structure (we

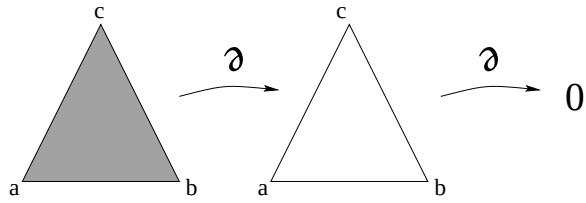

Fig. 1. The boundary operator, ∂, maps a 2-simplex onto the sum of three 1-simplices $\partial[a, b, c] = [b, c] - [a, c] + [a, b]$. The boundary of this 1-chain is empty. This property holds for any k-chain: $\partial\partial(c_k) = 0$.

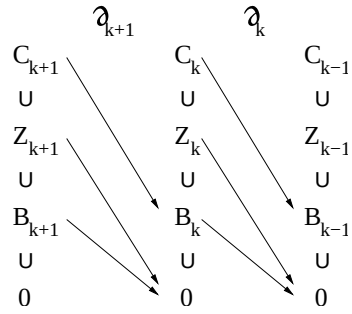

Fig. 2. The boundary operator, ∂_k, maps all k-chains onto the $k-1$-boundaries and the k-cycles to zero.

can ignore torsion). A negative coefficient is interpreted as reversing the orientation of the simplex: $-[x_0, x_1, \ldots, x_k] = [x_1, x_0, \ldots, x_k]$.

The next step towards defining the homology groups is to look at how the k-chains are related to the $k-1$-chains. This is done using the *boundary operator*, $\partial_k : C_k \to C_{k-1}$, a linear operator that maps a k-simplex onto the oriented sum of all $(k-1)$-simplices in its boundary

$$\partial_k[x_0, x_1, \ldots, x_k] = [x_1, x_2, \ldots, x_k] - [x_0, x_2, \ldots, x_k] + \ldots \\ +(-1)^k[x_0, x_1, \ldots, x_{k-1}].$$

For example, in Fig. 1 the boundary of the triangle is its three edges.

The action of the boundary operator on the chain groups leads to the definition of three more groups. Firstly, the image of ∂_k is a subgroup of C_{k-1} called the boundary group, and is denoted B_{k-1}. Secondly, the set of all k-chains that have empty boundary forms the group of k-cycles, Z_k (i.e. Z_k is the kernel or null space of ∂_k). These two groups are related by the fact that the boundary of a boundary is empty. This is the fundamental property of the boundary operator, $\partial_k\partial_{k+1} = 0$. It implies that B_k is a subgroup of Z_k. The relationships between the chain, cycle, and boundary groups are illustrated in Fig. 2.

The structures we are really interested in are the k-cycles that do not bound anything, since these indicate the presence of a k-dimensional 'hole'. It is this idea that leads to the definition of the homology groups as the quotient group $H_k = Z_k/B_k$. This means

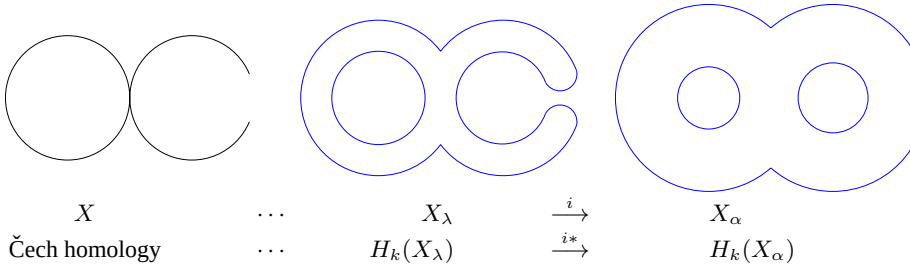

Fig. 3. The inclusion maps, $i : X_\lambda \to X_\alpha$, allow us to detect those holes which persist in the limit. In this example $\beta_1(\alpha) = 2$, but $\beta_1^\lambda(\alpha) = 1$.

that two k-cycles w_k and z_k belong to the same homology class if the k-chain formed by their difference is the boundary of some $(k+1)$-chain, $z_k - w_k = \partial v_{k+1}$.

The number of distinct equivalence classes of H_k is the kth Betti number β_k. The kth Betti number effectively counts the number of k-dimensional holes in X, so is exactly the type of information we seek. When $k = 0$, the Betti number counts the number of path-connected components of X. For subsets of $\mathbb{R}^3$, we can interpret β_1 as the number of independent tunnels, and β_2 as the number of enclosed voids. For example, the solid torus has $\beta_0 = 1$, $\beta_1 = 1$, and $\beta_2 = 0$ (the same as a simple circle), whereas the surface of a torus has $\beta_0 = 1$, $\beta_1 = 2$, and $\beta_2 = 1$.

Given a finite simplicial complex, the computability of Betti numbers using linear algebra techniques is well established [24]. However, these algorithms have run times at best cubic in the number of simplices. The development of fast algorithms for computational homology is thus an active area of research [5, 7, 13, 18]. We will describe one of these algorithms in Sect. 4.2.

3 A Theoretical Framework for Computational Homology

As mentioned in the introduction, we are interested in the situation where a finite number of points, S, approximate a subset $X \subset \mathbb{R}^d$. Our hope is to obtain accurate information about the topological structure of X by coarse-graining the data S at a sequence of resolutions, then extrapolating the limiting behaviour. There are two problems to be addressed here. First, is the extrapolation valid, even supposing we have perfect information about X? Second, what happens when we only have the finite approximation, S? The mathematical theory that helps answer these questions comes from Čech homology [15] and shape theory [20]. We give a brief overview of the concepts here – details can be found in [30, 31].

We can coarse-grain a set by taking its intersection with a digital mesh, or by applying opening and closing operations. In this paper we use the closed α-neighbourhood of a set (also known as the parallel set)

$$X_\alpha = \{y \text{ such that } \inf_{x \in X} |y - x| \leq \alpha\}$$

where $|.|$ is the Euclidean metric on $\mathbb{R}^d$. The following theory is easily adapted to the case of digital meshes and other standard methods of coarse-graining.

Our question now is: for what spaces do the Betti numbers of the α-neighbourhoods, X_α, converge to those of X? That is, does $\beta_k(X_\alpha) \to \beta_k(X)$ as $\alpha \to 0$? For the number of components, $\beta_0(X_\alpha)$, this holds for any compact set (i.e. closed and bounded subsets of $\mathbb{R}^d$) [15, 28]. This restriction to compact sets is not unreasonable, since we are primarily interested in objects that are well approximated on a computer by a finite number of points.

Higher-order Betti numbers present a more subtle problem, because fattening a set to its α-neighbourhood can introduce holes, as illustrated in Fig. 3, and also remove holes by filling them in. Mathematically, this problem is resolved by incorporating information about how a smaller neighbourhood maps inside a larger one. This leads to the definition of *persistent Betti number*. Suppose $0 \leq \lambda < \alpha$, so that $X_\lambda \subset X_\alpha$. The inclusion map $i : X_\lambda \to X_\alpha$ induces a homomorphism on the homology groups, $i* : H_k(\lambda) \to H_k(\alpha)$. We define the λ-persistent Betti number, $\beta_k^\lambda(\alpha)$, to be the rank of $i*$, i.e. the number of non-equivalent, non-bounding k-cycles in $H_k(\alpha)$ that are the image of a k-cycle from $H_k(\lambda)$. This leads to the formula

$$\beta_k^\lambda(\alpha) = \text{rank}[Z_k(\lambda)] - \text{rank}[i_*(Z_k(\lambda)) \cap B_k(\alpha)] \tag{1}$$

Geometrically, $\beta_k^\lambda(\alpha)$ is the number of holes in X_λ that do not get filled in by taking a fatter neighbourhood, X_α.

Cycles in X_α that genuinely come from a cycle in X are the 0-persistent Betti numbers, $\beta_k^0(\alpha) \leq \beta_k^\lambda(\alpha)$ for all $0 \leq \lambda \leq \alpha$. From the continuity of Čech homology, we know that the 0-persistent Betti numbers of the α-neighbourhoods converge to those of the original space: $\beta_k^0(\alpha) \to \beta_k(X)$, when X is compact. Thus, to find the regular Betti number of X, $\beta_k(X)$, from the persistent Betti numbers $\beta_k^\lambda(\alpha)$, we must first fix a sequence of α-values and let $\lambda \to 0$ to get the limits $\beta_k^\lambda(\alpha_n) \to \beta_k^0(\alpha_n)$. Then we can find $\beta_k(X)$ as the limit of this sequence: for $\alpha_n \to 0$, $\beta_k^0(\alpha_n) \to \beta_k(X)$.

The second problem we must address concerns the relationship between the neighbourhoods of the data, S_α, and those of the set they approximate, X_α. Intuitively speaking, the idea is that if the data and the underlying set are "close" then their α-neighbourhoods should have similar topological properties, provided α is sufficiently large. We measure the distance between the data, S, and the underlying space, X, using the Hausdorff metric on compact sets. This is defined as

$$d_H(S, X) = \min\{\alpha \mid X \subset S_\alpha \text{ and } S \subset X_\alpha\}. \tag{2}$$

If $\rho = d_H(S, X)$, it follows that $S_\alpha \subset X_{\alpha+\rho}$ and $X_\alpha \subset S_{\alpha+\rho}$ for all $\alpha > 0$. These inclusions allow us to derive a number of inequalities relating the persistent Betti numbers of X to those of S [31]. For example, we have that for $\alpha > \rho$, $\beta_k^0(X_{\alpha+\rho}) \leq \beta_k^\rho(S_\alpha)$ and $\beta_k^{\alpha-\rho}(S_{\alpha+\rho}) \leq \beta_k(X_\alpha)$. For the number of connected components, the persistent Betti number is the same as the regular Betti number and we find that $\beta_0(S_{\alpha-\rho}) \geq \beta_0(X_\alpha) \geq \beta_0(S_{\alpha+\rho})$. We refer to ρ as the *cutoff resolution*, since we cannot hope to get valid topological information about X from S_α unless $\alpha > \rho$.

4 Computer Implementation

We now describe some algorithms for computing Betti numbers from scattered point data at a sequence of α-values. We begin with the simple problem of counting the connected components of S_α.

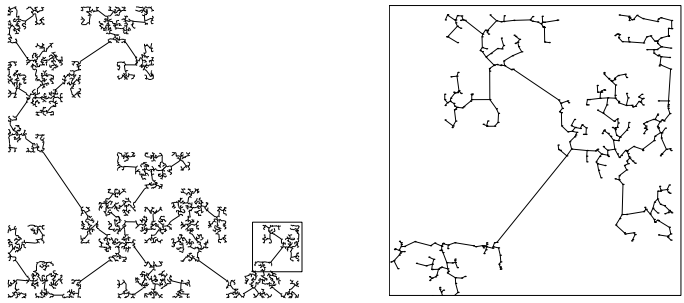

Fig. 4. The MST of 10^4 points on a Cantor set, and a close up of a small region.

4.1 Connected Components

For scattered point data, $S \subset \mathbb{R}^d$, the Euclidean minimal spanning tree (MST) encodes all the relevant information about connected components at all resolutions [29]. Recall that the MST is a tree (i.e. a graph with no cycles) that passes through every point in such a way as to minimize the total length. There are a number of algorithms for constructing the MST; the most easily implemented one is due Prim [4, 27]. This starts with any point and joins its nearest neighbour to create an initial subtree. The algorithm proceeds incrementally by adding the point which is closest to the existing subtree, finishing when all points have been added. The computational complexity is quadratic in the number of data points.

The property of connecting closest points gives the MST a natural correspondence with the connected components of the α-neighbourhood, S_α, [29]. In fact the number of components, $\beta_0(\alpha)$, is just one more than the number of MST edges that are longer than 2α:

$$\beta_0(\alpha) = 1 + \#\{\text{MST edges} > 2\alpha\}. \tag{3}$$

Note that we need only build the MST once to get component information at all scales.

4.2 Alpha Shapes and Homology

We now want to compute higher-order Betti numbers of the α-neighbourhoods S_α for a sequence of α values. This requires a sequence of simplicial complexes that reflect the topology of the α-neighbourhoods. An elegant solution to this problem comes from alpha shapes - a construction based on the Voronoi diagram and its dual, the Delaunay triangulation. Alpha shapes were defined by Edelsbrunner *et al.* for finite sets of points

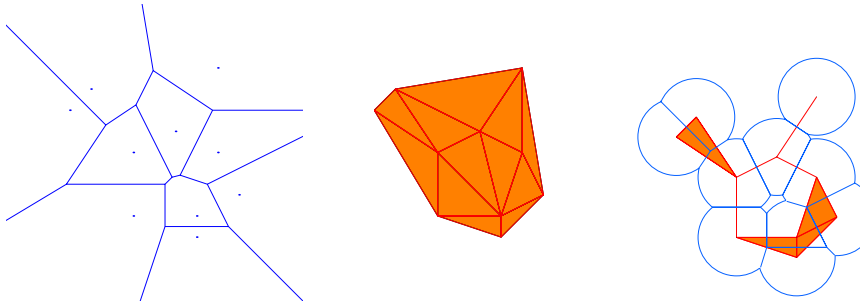

Fig. 5. Left: the Voronoi diagram of ten points. Middle: the corresponding Delaunay triangulation. Right: The alpha diagram is the intersection of the Voronoi diagram with the union of balls of radius α centred on the data points; it is shown together with its geometric dual, the alpha complex.

in the plane [9], then generalised to higher dimensions [11], and to points with weights [8]. Here we describe unweighted alpha shapes, and an algorithm to compute their Betti numbers.

Let $S = \{x_1, \ldots, x_n\}$ be a set of n points in $\mathbb{R}^d$ which satisfy the general position assumption that no four points simultaneously lie on the boundary of a circle. We write $V(x_i)$ for the Voronoi cell of x_i, i.e. the region that contains all points in $\mathbb{R}^d$ closer to x_i than to any other point in S. The dual tessellation of the Voronoi diagram is the Delaunay complex, $\mathcal{D}(S)$. This is obtained by associating a point with each Voronoi cell, an edge with two cells that share a $d-1$ face, a triangle with any triple of intersecting cells, and so on. See Fig. 5 for an example in the plane.

Edelsbrunner introduces a resolution parameter by taking the intersection of the Voronoi cell of each x_i with a ball of radius α centered at x_i, $B_\alpha(x_i)$. This gives us the *alpha diagram*:

$$\mathcal{Q}_\alpha(S) = \bigcup_{i=1}^{n} B_\alpha(x_i) \cap V(x_i). \tag{4}$$

The *alpha complex*, $\mathcal{C}_\alpha(S)$, is the dual of the alpha diagram. The *alpha shape* is the subset of $\mathbb{R}^d$ covered by the alpha complex. Figure 5 shows, for a simple example, the alpha diagram and its dual alpha complex. From the above definitions we have that $\mathcal{C}_\alpha(S) \subset \mathcal{D}(S)$ for all $\alpha > 0$. Also, if $\alpha_1 < \alpha_2$, then $\mathcal{Q}_{\alpha_1}(S) \subset \mathcal{Q}_{\alpha_2}(S)$, and this implies that $\mathcal{C}_{\alpha_1}(S) \subset \mathcal{C}_{\alpha_2}(S)$. Note that for sufficiently small α, the alpha complex consists of the vertices, $x_1, \ldots, x_n$, while at the other extreme, for large enough alpha, we get the Delaunay complex, $\mathcal{C}_\alpha(S) = \mathcal{D}(S)$.

The power of this construction comes from transforming the "continuous space" alpha diagrams into the discrete geometry of alpha complexes. The nerve theorem in topology [8, 15] guarantees that the alpha diagram is homotopy equivalent to the alpha shape. Thus, for a given alpha, each of these objects has the same Euler characteristic and Betti numbers.

Delfinado and Edelsbrunner's algorithm [5] for computing the Betti numbers of a simplicial complex relies on a sequential ordering of simplices, termed a filtration. The subset property $\mathcal{C}_{\alpha_1} \subset \mathcal{C}_{\alpha_2}$, for $\alpha_1 < \alpha_2$, means that every simplex σ has a threshold,

$\alpha_c(\sigma)$, such that $\sigma \in \mathcal{C}_\alpha$ for all $\alpha > \alpha_c(\sigma)$. The filtration orders the simplices according to their thresholds. Note that two simplices may have the same threshold, and in this situation they are ordered by adding the lower dimensional simplices first. The Betti numbers are computed incrementally as each simplex is added to the complex. This process depends on a test to determine whether the new simplex belongs to a k-cycle of the new complex. There are efficient algorithms for testing 1-cycles, and homology-cohomology duality theorems transform the $(d-1)$-cycles into 1-cocycles that are equally easy to test for. However, there is no fast test for other k-cycles, so Delfinado and Edelsbrunner's algorithm applies only to subcomplexes of $\mathbb{R}^2$ or $\mathbb{R}^3$.

The computational complexity involved in building the Delaunay triangulation in $\mathbb{R}^3$ is quadratic in the number of points, and the incremental algorithm for computing Betti numbers is barely super-linear in the number of simplices. The NCSA ftp site provides software that implements the above alpha shape constructions in $\mathbb{R}^2$ and $\mathbb{R}^3$ [1]. The Betti number data for the examples in the following section were generated using this software.

The numerical computation of persistent Betti numbers is possible using linear algebra techniques but is not yet implemented. Recently, Edelsbrunner [10] has made a similar definition of persistence specific to alpha shapes and has developed an incremental algorithm for this quantity.

5 Example Data

The random point clouds we study tend to have three regimes determined by the cutoff resolution ρ discussed in Sect. 3. For $\alpha \gg \rho$ we recover the coarse-scale topology of the approximated region X. At the other extreme, $\alpha \ll \rho$ we see purely the statistical effects of a point process. When α is on the same order as ρ, there is transitional behaviour as the finite data effects begin to dominate the underlying topological structure.

In the following two subsections we analyse the Betti numbers for data from a binomial point process and a family of self-similar fractals. For a finite domain, the coarse scale topology of the binomial point pattern is extremely simple, so the interesting Betti number behaviour occurs only for $\alpha \leq \rho$. The fractal examples have identical fractal dimension, but greatly differing topological structure. The Betti number data at coarse resolutions distinguishes between the different topological types. For $\alpha < \rho$, however, the fractals show essentially the same behaviour, presumably dictated by the iterated function system technique of generating the data points.

5.1 Binomial Point Process

We begin by analysing N points that are randomly distributed in the unit square according to a binomial point process. The coordinates of each point are assigned using the uniform distribution on $[0, 1]$ and the point pattern therefore constitutes a finite-domain approximation to a Poisson process with intensity $\lambda = N/1$. We describe the behaviour of a typical single realisation of N points, and then present results from a numerical investigation of the mean values of β_0 and β_1 as functions of the radius α.

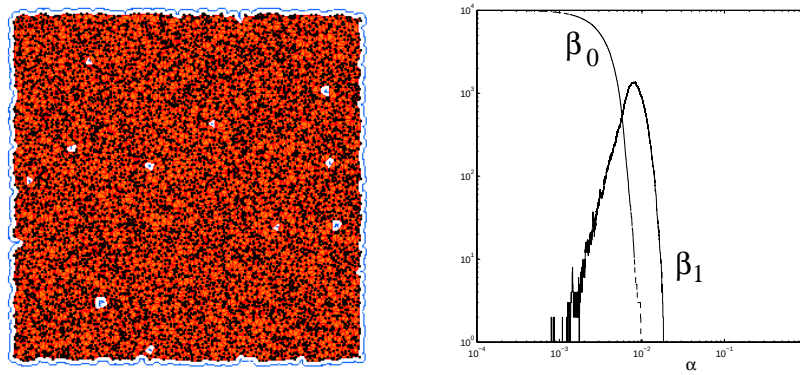

Fig. 6. Left: An example alpha neighbourhood (faint outline) and alpha shape (shaded region) of 10^4 points sampled from the uniform distribution on the unit square. On the right we plot the Betti number data on logarithmic axes; $--$ the number of components β_0, and – number of holes β_1.

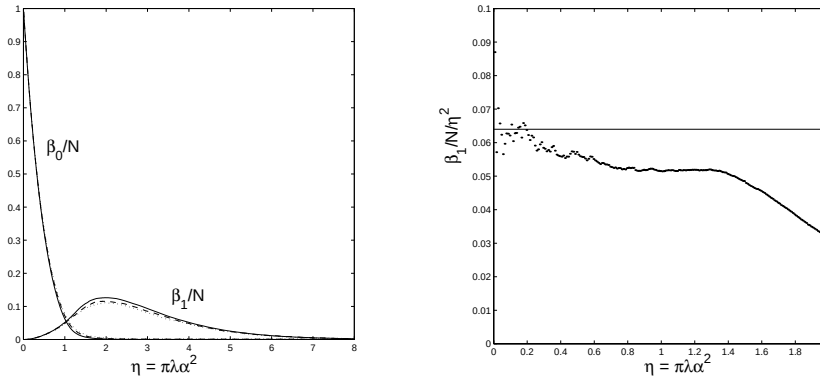

Fig. 7. On the left we show that the Betti number data rescale to functions of $\eta = \pi\lambda\alpha^2$. We compute sample mean values of β_0/N and β_1/N for $N = 500$ ($\cdots$), $N = 1000$ (- -), and $N = 5000$ (—).

Figure 6 shows for an example the α-shape of $N = 10^4$ points. The points form an approximation to the unit square, a connected region that has extremely simple topology and geometry. For α above the cutoff resolution we expect to see a single connected component with no holes, and this is reflected by the data presented in Fig. 6 for $\alpha > 0.04$. For sufficiently small values of α, each point is a single component, so there are no holes and $\beta_0(\alpha) = N$, the number of points. As α increases, edges are added to the alpha complex, decreasing β_0 and possibly increasing β_1. When a triangle is added to the complex, β_1 may decrease; this is the reason for the noisy non-monotonic behaviour in the number of holes. As alpha approaches the cutoff resolution there is a sharp drop in the number of components until the point when the alpha neighbourhood forms a single connected component. At around the same α-value, the number of holes reaches a maximum then starts to decrease with increasing α as the holes are filled in.

Readers familiar with the field of stochastic geometry will recognize that the α-neighbourhoods of a random point process are analogous to a Boolean model with circular grains of radius α. Most geometric quantities associated with such a circular-grain Boolean model are really functions of $\eta = \pi\lambda\alpha^2$ [32]. For example, the mean Euler characteristic per unit area is $e = \lambda(1-\eta)e^{-\eta}$ [21]. Unsurprisingly, the expected value of the mean Betti numbers per unit area (written b_0 and b_1) are also functions of this quantity. The graphs in Fig. 7 show the mean specific Betti number curves from 500, 1000, and 5000 points in the unit square.

Hopes for deriving analytic formulas for the mean Betti numbers are limited by their connection with percolation. (The percolation transition is signalled by a singularity in the second moment of the cluster size distribution and β_0 is the total number of clusters.) However, an expression for the expected number of components per unit area in the low-intensity limit, $\lambda \to 0$, is known [14]:

$$b_0(\eta) = \lambda(1 - 2\eta + \tfrac{2}{3}(4 - 3\sqrt{3}/\pi)\eta^2) + O(\eta^3).$$

Assuming $e = b_0 - b_1$ [23], we can find the limiting behaviour of $b_1(\eta)$ as $\eta \to 0$. Specifically

$$b_1(\eta) = \lambda(\tfrac{7}{6} - 2\sqrt{3}/\pi)\eta^2 + O(\eta^3) \approx 0.0640\lambda\eta^2 + O(\eta^3).$$

Data from the alpha shapes closely match these expressions, see Fig. 7.

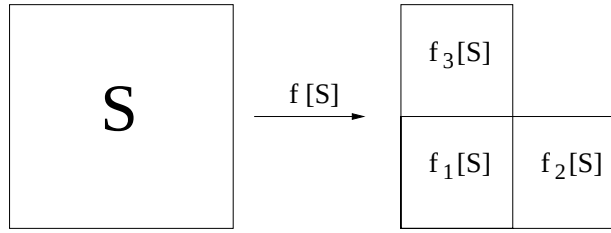

Fig. 8. Template for the iterated function system that generates the Sierpinski triangle relatives.

5.2 Sierpinski Triangle Relatives

The fractals we study here are a family of iterated function systems [12], illustrated by the template in Fig. 8:

$$S = f[S] = f_1[S] \cup f_2[S] \cup f_3[S].$$

The functions f_i are similarity transformations of the unit square with contraction ratio $\frac{1}{2}$, i.e. $|f_i(x) - f_i(y)| = \frac{1}{2}|x - y|$. Those that generate the Sierpinski triangle, shown in Fig. 9, are simple contractions followed by a translation; the generators of the other examples involve additional rotation or reflection symmetries of the square. There are 232 different fractals in this family [26]. Their topology ranges from simply connected

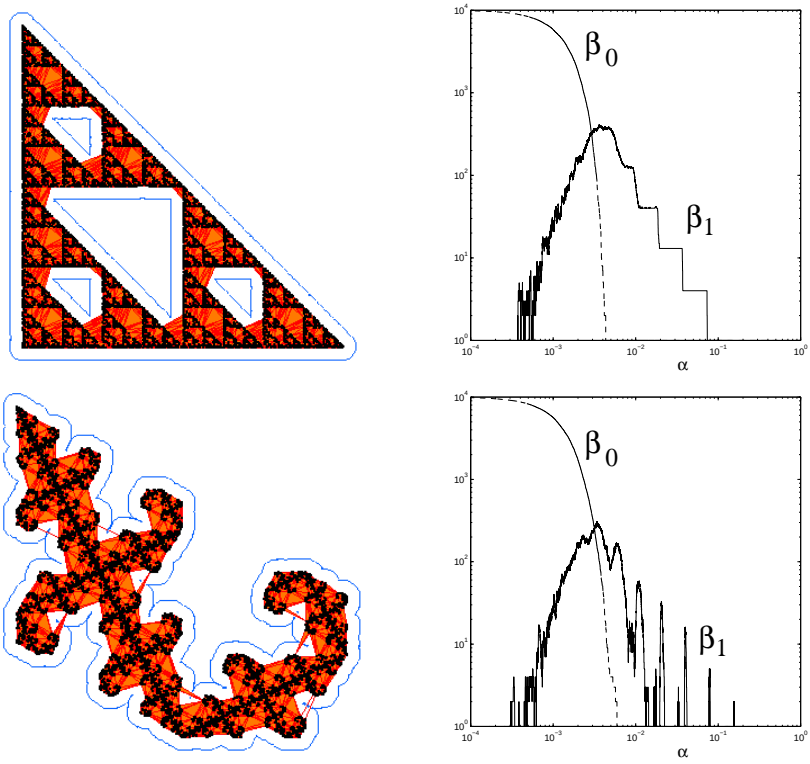

Fig. 9. Left column: an example alpha neighbourhood (outline) and alpha shape (shaded) of 10^4 points on two fractals. Right column: Betti number data on logarithmic axes. The Sierpinski triangle (top row) is a connected set with infinitely many holes. The Betti numbers of its alpha neighbourhoods are $\beta_0(\alpha) = 1$ for $\alpha \geq 0$, and $\beta_1(\alpha_1/2^n) = \frac{1}{2}(3^{n+1}-1)$, where $n = 0, 1, 2, \ldots$ and α_1 is the radius of the circle inside the largest triangular hole. This behaviour is accurately reflected by data from the finite point approximation for $\alpha \geq 0.005$. The other fractal is simply connected, so in principle $\beta_0 = 1$ and $\beta_1 = 0$. There are spikes in the hole data however, and these are due to the geometry of the fractal – the way it hooks around creates non-persistent holes in the α-neighbourhoods.

(Fig. 9 bottom) to connected (Fig. 9 top) to totally disconnected (Fig. 10 top) to a class of examples with infinitely many connected components of non-zero diameter (Fig. 10 bottom). This range of topological structure makes them ideal test cases for our techniques [28, 29, 31].

We generate a finite number of points on a fractal using Barnsley's chaos game technique [3]. This records the trajectory of an initial point under the iteration $x_{n+1} = f_{i_n}(x_n)$, setting $i_n = 1, 2$ or 3 with equal probability at each step.

We present data for the number of components, $\beta_0(\alpha)$, and the number of holes, $\beta_1(\alpha)$, for a fractal of each topological type in Fig. 9 and Fig. 10. The Betti numbers are plotted on a logarithmic scale to show the self-similar fractal scaling.

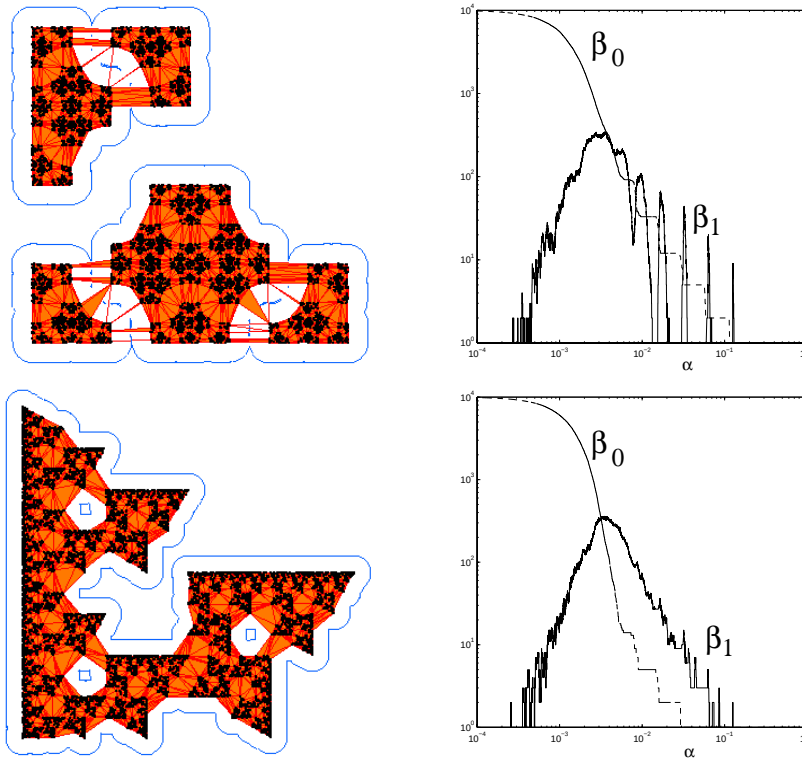

Fig. 10. Left column: an example alpha neighbourhood (outline) and alpha shape (shaded) of 10^4 points on two fractals. Right column: Betti number data on logarithmic axes. The upper fractal is a Cantor set so it is totally disconnected and perfect. We find [28] $\beta_0(\alpha_0/2^n) \to 3^n$ as $n \to \infty$ (where α_0 is the width of the largest gap) and $\beta_1(\alpha) = 0$. Although topologically the Cantor set does not contain any holes, the graph of $\beta_1(\alpha)$ has regularly spaced spikes. The reason for these is seen in the α-neighbourhood picture. The edges bridging the gaps cause holes in the triangulation that disappear for slightly smaller values of α when the edges are removed from the alpha shape. The lower fractal is disconnected, and consists of infinitely many line segments. Again, there are no topological holes in this fractal. It is the geometry of the set, i.e., the arrangement of the line segments in the plane, that creates holes in the α-neighbourhoods. Again, $\beta_0(\alpha_0/2^n) \to 3^n$. The regular Betti number shows fractal growth $\beta_1(\alpha_1/2^n) = 3^n$, while the 0-persistent Betti numbers are $\beta_1^0(\alpha) = 0$.

Each of these fractals has a cutoff resolution $\rho \approx 0.005$. For $\alpha < \rho$, the behaviour of the Betti number data is quantitatively similar for each fractal and qualitatively similar to that of the point process in Sect. 5.1. Preliminary analysis suggests that for each of these fractals, the Betti numbers are really functions of the rescaled quantity $\eta = \pi N \alpha^{d_f}$, where $d_f = \log 3 / \log 2$ is the fractal dimension. In particular we find $\beta_0(\eta)/N \to 1 - \eta$ and $\beta_1(\eta) = O(\eta^{d_f})$ as $\eta \to 0$.

It is in the region $\alpha > \rho$ that we see the effects of the different topology of each fractal. The difference between the connected and disconnected fractals shows up clearly

in the number of components. The connected sets in Fig. 9 have $\beta_0(\alpha) = 1$ for $\alpha > \rho$. The difference between the Sierpinski triangle and the simply connected fractal is seen the behaviour of β_1. The Sierpinski triangle has regular staircase growth in the number of holes, as expected, while the simply connected fractal has no persistent holes.

The disconnected fractals show a stair-step decrease in the number of components due to their self-similar structure. The slope is equal to the fractal dimension, $\log 3/\log 2$, in both cases. In order to distinguish between the totally disconnected Cantor set, and the other disconnected fractal, we need information about the size of the connected components [28, 29]. (For a totally disconnected set, the diameters go to zero as $\alpha \to 0$.) The disconnected fractal at the bottom of Fig. 10 illustrates the need to compute the persistent Betti numbers to obtain an accurate diagnosis of the underlying topology.

Acknowledgements

This paper is based on part of my dissertation [30]. I would like to thank my thesis advisors Professors J.D. Meiss and E. Bradley for many stimulating discussions in relation to this work. I am also grateful to Dr Klaus Mecke for inviting me to attend the Second Conference on Spatial Statistics and Statistical Physics.

References

1. Software available from NCSA via anonymous ftp from ftp.ncsa.uiuc.edu/Visualization/Alpha-shape/.
2. Arns, C.H., M.A. Knackstedt, W.V. Pinczewski, and K. Mecke (2001): 'Characterisation of irregular spatial structures by parallel sets and integral geometric measures'. *Applied Mathematics*, preprint. (RSPhysSE, ANU, Canberra)
3. Barnsley, M.F. (1993): *Fractals Everywhere*, second edition. (Academic Press, Boston)
4. Cormen, T.H., C.E. Leiserson, and R.L. Rivest (1990): *Introduction to Algorithms*. (MIT Press, Cambridge, MA)
5. Delfinado, C.J.A., and H. Edelsbrunner (1995): 'An incremental algorithm for Betti numbers of simplicial complexes on the 3-sphere'. *Computer Aided Geometric Design* **12**, pp. 771–784.
6. Dey, T.K., H. Edelsbrunner, and S. Guha. (1999): 'Computational topology'. In: *Advances in Discrete and Computational Geometry*, volume 223 of *Contemporary Mathematics*. ed. by B. Chazelle, J.E. Goodman, and R. Pollack (American Mathematical Society)
7. Dey, T.K., and S. Guha. (1998): 'Computing homology groups of simplicial complexes in $\mathbf{R}^3$'. *Journal of the ACM* **45** pp. 266–287.
8. Edelsbrunner, H. (1995): 'The union of balls and its dual shape'. *Discrete and Computational Geometry* **13**, pp. 415–440.
9. Edelsbrunner, H., D.G. Kirkpatrick, and R. Seidel (1983): 'On the shape of a set of points in the plane'. *IEEE Transactions on Information Theory* **29**, pp. 551–559.
10. Edelsbrunner, H., D. Letscher, and A. Zomorodian (2000): 'Topological persistence and simplification'. *41st Annual Symposium on Foundations of Computer Science*.
11. Edelsbrunner, H., and E.P. Mücke (1994): 'Three-dimensional alpha shapes'. *ACM Transactions on Graphics* **13**, pp. 43–72.
12. Falconer, K. (1990): *Fractal Geometry: Mathematical Foundations and Applications*. (Wiley, Chichester)

13. Friedman, J. (1998): 'Computing Betti numbers via combinatorial Laplacians'. *Algorithmica* **21**, pp. 331–346.
14. Hall, P. (1988): *Introduction to the Theory of Coverage Processes*. (Wiley, Chichester)
15. Hocking, J.G., and G.S. Young (1961): *Topology*. (Addison-Wesley)
16. Hoshen, J. and Kopelman, R. (1976): 'Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm'. *Physical Review B* **14**, pp. 3438–3445.
17. Hutchinson, J.E. (1981): 'Fractals and self similarity'. *Indiana University Mathematics Journal* **30**, pp. 713–747.
18. Kalies, W.D., K. Mischaikow, and G. Watson (1999): 'Cubical approximation and computation of homology'. *Banach Center Publications* **47**, pp. 115–131.
19. Lee, C.-N., T. Posten, and A. Rosenfeld (1991): 'Winding and Euler numbers for 2D and 3D digital images'. *CVGIP: Graphical Models and Image Processing* **53**, pp. 522-537.
20. Mardešić, S. and J. Segal (1982): *Shape Theory*. (North-Holland)
21. Mecke, K. (1998): 'Integral geometry and statistical physics'. *International Journal of Modern Physics B* **12**, pp. 861–899.
22. Mecke, K. and D. Stoyan (Eds.) (2000): *Statistical Physics and Spatial Statistics – The Art of Analyzing and Modeling Spatial Structures and Pattern Formation* Lecture Notes in Physics **554** (Springer-Verlag, Berlin)
23. Mecke, J., and Stoyan, D. (2001): 'The specific connectivity number of random networks'. *Advances in Applied Probability* **33**, pp. 576–583.
24. Munkres, J.R. (1984): *Elements of Algebraic Topology*. (Benjamin Cummings)
25. Nagel, W., J. Ohser, and K. Pischang (2000): 'An integral-geometric approach for the Euler-Poincaré characteristic of spatial images'. *Journal of Microscopy* **198**, pp. 54–62.
26. Peitgen, H.-O., S. Jürgens, and D. Saupe. (1992): *Chaos and Fractals: New Frontiers of Science*. (Springer-Verlag, Berlin)
27. Prim, R.S. (1957): 'Shortest connection networks and some generalizations'. *The Bell System Technical Journal* November 1957, pp. 1389–1401.
28. Robins, V., J.D. Meiss, and E. Bradley (1998): 'Computing connectedness: An exercise in computational topology'. *Nonlinearity* **11**, pp. 913–922.
29. Robins, V., J.D. Meiss, and E. Bradley (2000): 'Computing connectedness: Disconnectedness and discreteness'. *Physica D* **139**, pp. 276–300.
30. Robins, V. (2000): *Computational topology at multiple resolutions*. PhD Thesis, Department of Applied Mathematics, University of Colorado, Boulder.
31. Robins, V. (1999): 'Towards computing homology from finite approximations'. *Topology Proceedings* **24**, pp. 503–532.
32. Stoyan, D., W.S. Kendall, and J. Mecke (1987): *Stochastic Geometry and its Applications*. (Wiley, Chichester)