A Witness complex for density landscapes

Erik Carlsson[1] and John Carlsson[2]

[1]Department of Mathematics, UC Davis, 1 Shields ave. Davis, CA, 95618, 530-754-0274, ecarlsson@math.ucdavis.edu
[2]Department of Industrial Engineering, University of Southern California, jcarlsso@usc.edu

Abstract

We define a filtered simplicial complex associated to the superlevel sets of a sum of weighted Gaussians kernels $f(x)$, with uniform scale parameter $h > 0$. Regarding f as a kernel density estimator of a data set $\mathcal{D} \subset \mathbb{R}^m$, we obtain a method for filtering persistent homology by density in low or high dimensions. On the other hand, we also see that our construction can be highly visually descriptive.

1 Introduction

Let $f : \mathbb{R}^m \to \mathbb{R}_{>0}$ be a sum of Gaussian kernels with uniform scale parameter $h > 0$ and positive coefficients $a_i > 0$, centered at a collection of points $\{x_1, ..., x_N\} \subset \mathbb{R}^m$:

$$f(x) = \sum_{i=1}^{N} a_i K(\|x - x_i\|/h), \quad K(r) = \exp(-r^2/2).$$

When the coefficients are all constant, $f(x)$ is a standard kernel density estimator for the collection $\mathcal{D} = \{x_1, ..., x_N\}$, regarded as a point cloud in $\mathbb{R}^m$. In order to ensure that our proposed constructions ends up being finite, we restrict ourselves to a bounded domain such as $\mathcal{A} = f^{-1}[d_0, \infty)$, which is the set of points with density value greater than some chosen cutoff $d_0 > 0$. A two-dimensional example coming from a well-known species distribution data set is shown in Figure 2.1, which will be used as a running illustration.

We consider the problem of approximating the "landscape" of $f(x)$ by the data of a filtered simplicial complex. A filtered simplicial complex is a pair (X, w) consisting of a simplicial complex together with together with a weight function $w : X \to \mathbb{R}$ whose sublevel sets $X(a) = w^{-1}(-\infty, a]$ are subcomplexes. The function w is a discrete approximation of f, which is reparametrized by the direction-reversing function $a = -\log(d)$, so that higher weight corresponds to

lower density. Thus, each sublevel set $X(a)$ corresponds to the density superlevel set $f^{-1}[e^{-a}, \infty)$.

In terms of persistent homology, such a complex would be said to be "filtered by density" as opposed to other constructions such as Vietoris-Rips, which filter by distance. One advantage to filtering by density is that while distance based complexes such as Vietoris-Rips are stable under perturbation, they are sensitive to outliers. As a result, it is often necessary to remove low-density points as a preprocessing step. In density based methods, this would be unnecessary because the weight function is itself a measure of sparsity.

The primary issue with filtering by density, or sublevel sets in general is that the size of the complex tends to blow up combinatorially in higher dimensions. In the case of sums of Gaussians for instance, it is shown that the number of critical points, which are a natural candidate for the vertices of the complex, may be larger than the data set in higher dimensions [13] (the extra ones are called "ghost modes"). Thus, the problem of selecting vertices requires a well-suited criteria for selecting landmark points. Unlike many other persistent homology constructions, our landmark selection scheme is critical to the main construction, and we expect it will be of independent interest.

We summarize our proposed method, which proceeds in several steps:

1. Fix a parameter $0 < s < 1$ which controls the number of vertices, and a bounded domain $\mathcal{A}$, which will be called the reference set. In theory, $\mathcal{A}$ may be taken to be $f^{-1}[d_0, \infty)$ as above, but in practice it may be taken to be points in the data set itself, or some other finite set.

2. Using a particular algorithm, generate a function which is a *max* of Gaussians with the same scale parameter, but different weights and centers

$$g(x) = \max_i b_i K(\|x - z_i\|/h),$$

   which satisfies $g(x) \leq f(x) \leq s^{-1}g(x)$ for all $x \in \mathcal{A}$. The centers $\{z_i\}$ are the desired landmark points.

3. Realize $w(x) = -\log(g(x))$ as the weight function of a power diagram, also known as a shifted Voronoi diagram, with centers $\{z_i\}$, and powers given by $p_i = -2h^2 \log(b_i)$. Form the nerve of the corresponding covering, denoted by $X = \mathrm{DensAlpha}(f, \mathcal{A}, s)$.

4. Form the continuous map $\phi : |X| \to |\operatorname{Sd}(X)| \to \mathbb{R}^m$ which is linear on the barycenteric subdivision whose value on a vertex $[\sigma] \in \operatorname{Sd}(X)_0$ is the unique minimizer q_σ of w on any nonempty intersection in the power diagram, as defined in [5].

5. Define a filtered complex $Y = \mathrm{SubDens}(f, \mathcal{A}, s)$ so that $Y(a)$ is the maximal subcomplex of X with the property that $|Y(a)| \to |\operatorname{Sd}(Y)(a)| \to \mathbb{R}^m$ is entirely contained in $f^{-1}[e^{-a}, \infty)$. In practice, we will compute a closely related complex denoted $\mathrm{DensWit}(f, \mathcal{A}, s)$ using only the values of $-\log(f(x))$ at the barycenters q_σ.

Finding the function $g(x)$ from item 2 is clearly a critical step, which is described in Algorithm 1 below. While the algorithm itself is a straightforward search, it exploits a crucial fact given in Lemma 3.1 below: for any postively weighted sums of Gaussians $f(x)$ and any point $y \in \mathbb{R}^m$, there exists a unique weighted Gaussian $g_y(x) = bK(\|x - z\|/h)$ with the same scale parameter, satisfying

$$g_y(x) \leq f(x), \quad g_y(y) = f(y).$$

The center point z turns out to be an explicit convex combination of the points $\{x_i\}$, whose coefficients are determined by fitting $f(x)$ to first order at y. An example of the resulting function for the species distribution example is shown in Figure 3.2. The resulting complexes are shown in Figure 3.4.

The reader may also note that the computation of alpha complexes in higher dimension is not a straightforward computation. We have created our own algorithm based on dual quadratic programming which is mentioned in Section 2.4, and left for a separate paper.

Let $\mathcal{A} = f^{-1}[e^{-a}, \infty)$ be the superlevel set and let $Y = \text{SubDens}(f, \mathcal{A}, s)$ for some choice of s. Recall that we are reordering the filtration value according to the above filtration so that density is given by e^{-a}, and let us set $\epsilon = -\log(s)$. In Section 3.2 we prove the following theorem, which shows that the three spaces are interleaved above the minimum density cutoff, showing that their persistent homology groups approximate each other in a certain sense [7]:

Theorem A. *The above complexes fit into a sequence of continuous maps*

$$H_*(X(a)) \to H_*(Y(a)) \to H_*(f^{-1}[e^{-a}, \infty)) \to H_*(X(a + \epsilon))$$

for $a \leq -\log(d_0)$, which form an interleaving on their domain of definition.

In Section 4, we present examples of persistent homology computations, as well as graphical illustrations in low dimensional coordinate systems. Our higher dimensional examples include a data set of simulated states of the Ising model from statistical mechanics associated to certain graphs, and an analysis of local patches in the MNIST data set, in the same spirit as the study of natural images of [16]. In each case, a low-dimensional projection of the one-skeleta show visibly recognizable topological features, such as the low energy landscape in several instances of the Ising model, as well Klein bottle related shapes such as the Möbius strip appearing in the case of MNIST.

1.1 Acknowledgements

Both authors were supported by the Office of Naval Research, project (ONR) N00014-20-S-B001.

2 Notation and preliminaries

We summarize some preliminary definitions and notation, including the definition of the alpha and witness complexes.

2.1 Kernel density estimation

Let $\mathcal{D} \subset \mathbb{R}^m$ be a point cloud data set of N vectors in $\mathbb{R}^m$. A Gaussian kernel density estimator is a sum of the form

$$f(x) = \sum_{i=1}^{N} a_i \rho(x, x_i), \quad a_i > 0 \tag{1}$$

where

$$\rho(x, y) = K(\|x - y\|/h), \quad K(r) = \exp\left(-r^2/2\right).$$

We will be interested in the restriction of $f(x)$ to a bounded domain $\mathcal{A}$, such as the superlevel set regions $\mathcal{A} = f^{-1}[d_0, \infty)$ for some lower bound on density $d_0 > 0$. We will often use the notation $\mathcal{D}_{\geq d_0} = \mathcal{D} \cap f^{-1}[d_0, \infty)$ to denote the subset of the data set whose density is at least d_0. Unless otherwise specified, if $\mathcal{D} = \{x_i\}$ is a data set of size N, the weights a_i will be assumed to all be $1/N$. We are not including a standard volume normalizing term, because we would like the value of $f(x)$ to depend only on distance within the data set, and not directly on the embedding dimension.

For simplicity we will assume that the norm is always the usual L^2 metric, as other quadratic forms may be converted into this form by a change of coordinates. The following example will be referred to in several parts below.

Example 2.1. Figure 2.1 shows a well known species distribution data set from [21]. The data points on the left show locations of two South American mammals, namely 1536 instances of "Bradypus Variegatus," the Brown-throated Sloth, and 88 instances of "Microryzomys Minutus," the forest small rat. On the right is the corresponding kernel density estimator of the combined data set using a heat map on a log scale. We chose h to be one degree of latitude, or about 69 miles. While in some cases it may make sense to assign different weights to the different species, we took every coefficient to be equal $a_i = 1/1624$. The minimum density cutoff, which determines the boundary on the density estimator shown on the right, is set to be $d_0 = .005$.

2.2 Topological preliminaries and notation

We review some background and notation from computational topology. For more details, we refer the reader to [15].

By a simplicial complex on n vertices we will mean an abstract complex, which is a subset $X \subset \mathcal{P}(\{1, ..., n\})$ of the power set that is closed under taking subsets. All homology groups $H_*(X)$ are assumed to be taken over a field, which will be suppressed unless a particular one is of interest. We will denote by $|X| \subset \mathbb{R}^{n+1}$ the geometric realization, and by $|\sigma| \subset |X|$ the corresponding subspace for any simplex $\sigma \in X$. We denote by $\mathrm{Sd}(X)$ the barycentric subdivison, which has an identification $|X| = |\mathrm{Sd}(X)|$.

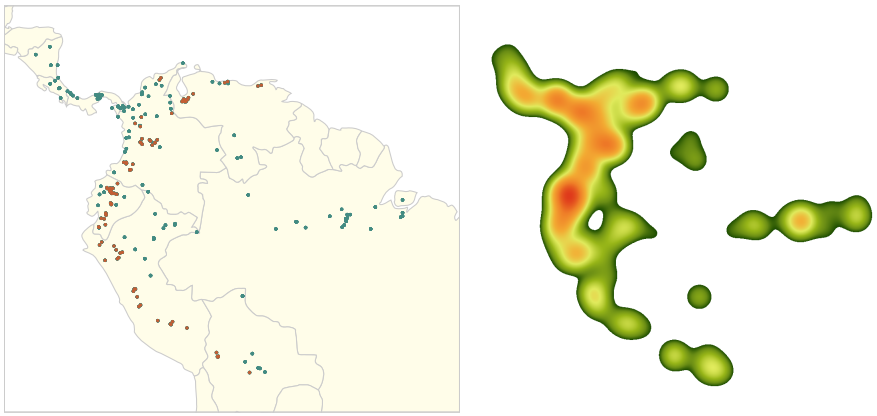


Figure 2.1: On the left: the data set of species distribution of Bradypus Variegatus (green) and Microryzomys Minutus (brown) from Phillips et al. [21]. On the right: a corresponding kernel density estimator of the whole data set shown as a heat map in log coordinates. The boundary is determined by the minimum density cutoff.

Definition 2.1. A filtered simplicial complex is a pair (X, w) where X is a simplicial complex, and $w : X \to \mathbb{R}$ is a function with the property that $X(a) = w^{-1}(-\infty, a]$ is a subcomplex of X for all a.

For $a \leq b$ the persistence maps

$$i_*^{a,b} : H_k(X(a)) \to H_k(X(b)) \tag{2}$$

are the ones induced by the inclusion maps $i^{a,b} : X(a) \to X(b)$.

Definition 2.2. Given $a \leq b$, the kth persistent homology group of a filtered complex X, denoted $H_k^{a,b}(X)$, is the image of $i_*^{a,b} : H_k(X(a)) \to H_k(X(b))$.

In this paper persistent homology groups will be represented by barcode diagrams generated by `javaplex` [1].

If $\mathcal{U} = \{U_i\}$ is a collection of n closed or open sets in $\mathbb{R}^m$, then the nerve $X = \text{Nerve}(\mathcal{U})$ is the complex

$$X = \{\sigma \subset \{1, ..., n\} : U_{\sigma_0} \cap \cdots \cap U_{\sigma_k} \neq \emptyset\}. \tag{3}$$

The nerve theorem states that whenever $\mathcal{U}$ is a good cover, for instance one whose members are convex sets $U_i \subset \mathbb{R}^m$, then the resulting complex is homotopy equivalent to the union, denoted $|\mathcal{U}| = \bigcup_i U_i$.

Suppose that each U_i is convex, and choose representatives $q_\sigma \in U_\sigma$, where U_σ denotes the intersection in (3). Then we have a linear map $\Gamma : |\operatorname{Sd}(X)| \to \mathbb{R}^m$ on the barycentric subdivision of the nerve, whose value on each vertex σ is q_σ, as in [5]. By convexity, it is clear that its image is contained in $|\mathcal{U}|$. We will denote by

$$\phi : |X| \to |\operatorname{Sd}(X)| \to |\mathcal{U}| \tag{4}$$

the corresponding piecewise linear map on $|X|$. The following proposition is Theorem 3.1 from [5]:

Proposition 2.1. *If $\mathcal{U}$ is convex then Γ (and therefore ϕ) is a homotopy equivalence, specifically the one from the nerve theorem.*

They determine an explicit homotopy inverse, Φ determined by a partition of unity subordinate to the cover associated to $|\operatorname{Sd}(X)|$. An illustration of Proposition 2.1 is shown below in the case when the cover comes from a power diagram.

Remark 2.1. The proposition would not necessarily hold if $\mathcal{U}$ is only a good cover. On the other hand, if $\mathcal{U}$ were a covering by balls, not just convex sets, then we would have a map $|X| \to U$ induced by the centers, which is linear on $|X|$, not just $|\operatorname{Sd}(X)|$.

2.3 Alpha and witness complexes

Let $\mathcal{L} = \{z_1, ..., z_n\} \subset \mathbb{R}^m$ be a collection of points, and let $p : \mathcal{L} \to \mathbb{R}$ be a function. The values $p_i = p(z_i)$ are called the powers. We have a function $w : \mathbb{R}^m \to \mathbb{R}$ defined by

$$w(x) = \min_i w_i(x), \quad w_i(x) = \|x - z_i\|^2 - p_i. \tag{5}$$

Definition 2.3. The weighted ball cover $\operatorname{BallCov}(\mathcal{L}, p)$ is the filtered family of collections $\mathcal{U}(a) = \{U_i(a)\}$ for each $a \in \mathbb{R}$, where

$$U_i(a) = \{x : w_i(x) \leq a\}.$$

Definition 2.4. The power diagram (see [3, 12]) denoted $\operatorname{PowDiag}(\mathcal{L}, p)$ is the collection of closed regions $\mathcal{V} = \{V_i\}$ where

$$V_i = \{x \in \mathbb{R}^m : w_i(x) \leq w_j(x) \text{ for all } j\}.$$

The intersections will be denoted by $V_\sigma = V_{\sigma_0} \cap \cdots \cap V_{\sigma_k}$. It is filtered by $\mathcal{V}(a) = \{V_i(a)\}$, where $V_i(a) = V_i \cap U_i(a)$ is the intersection of the covering with the corresponding balls. We will also denote by $V_\sigma(a)$ the intersection of the $V_i(a)$.

When $p_i = 0$ for all i, we obtain the usual Voronoi diagram. More generally, the regions are all determined by linear inequalities, in other words are separated by hyperplanes. In fact, they come from the intersection of true Voronoi diagrams with a linear subspace, with the powers p_i representing the negative squared normal distances. For more on this topic, we refer to [15].

The (weighted) Čech and alpha complexes are the filtered complexes defined in terms of the corresponding nerves.

Definition 2.5. The weighted Čech complex $X = \operatorname{Cech}(\mathcal{L}, p)$ is the filtered complex determined by $X(a) = \operatorname{Nerve}(\mathcal{U}(a))$, where $\mathcal{U} = \operatorname{BallCov}(\mathcal{L}, p)$.

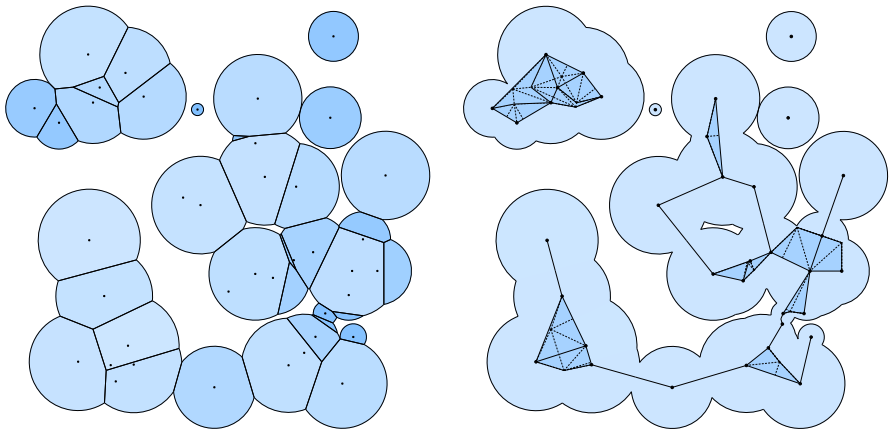


Figure 2.2: A randomly generated power diagram $\mathcal{V}(a)$, and its alpha complex mapped into $|\mathcal{V}(a)|$ via ϕ, taking $\{q_\sigma\}$ to the minimizers of the weight function. We have used dashed lines to illustrate the map of 2-simplices. The resulting map is a homotopy equivalence by Proposition 2.1.

Definition 2.6. The weighted alpha complex $X = \text{Alpha}(\mathcal{A}, p)$ is given by $X(a) = \text{Nerve}(\mathcal{V}(a))$, where $\mathcal{V} = \text{PowDiag}(\mathcal{L}, p)$.

Said another way, the weight function on the alpha complex is given on a simplex $\sigma = (\sigma_0, ..., \sigma_k)$ by

$$w(\sigma) = \min_{x \in V_\sigma} w_{\sigma_i}(x) \tag{6}$$

for any choice of $i \in \{0, ..., k\}$, all answers being equivalent. We will denote by q_σ the unique corresponding argmin in (6), which as described above defines a map $\phi : |\,\text{Alpha}(\mathcal{L}, p)| \to \mathbb{R}^m$ associated to any alpha complex. An example is shown for a random power diagram in Figure 2.2.

Since $|\mathcal{U}(a)| = |\mathcal{V}(a)|$, the nerve theorem implies that the Čech and alpha complexes are homotopy equivalent.

The witness complex [16] can be viewed as a refinement of a power diagram.

Definition 2.7. Let $\mathcal{W} \subset \mathbb{R}^m$ be any subset, finite or infinite, and let $(\mathcal{L}, p)$ be as above. The strong (weighted) witness complex is given by

$$Y(a) = \{\sigma : \mathcal{W} \cap V_\sigma(a) \neq \emptyset\},$$

where V_σ is the nonempty intersection following Definition 2.4.

In other words, it is the nerve of the power diagram intersected with $\mathcal{W}$. An element $x \in \mathcal{W} \cap V_\sigma(a)$ is called a strong witness for σ, since its existence confirms the existence of the corresponding simplex $\sigma \in Y(a)$. In this case the collection of the $Y(a)$ form a filtered complex, but other filtrations are common, such as ones that add slack to the condition of being in V_σ.

We have the definition of weak witnesses:

Definition 2.8. A an element $x \in \mathcal{W}$ a weak witness for σ if

$$w_i(x) \leq w_j(x) \text{ for all } i \in \sigma,\ j \in \mathcal{L} - \sigma. \tag{7}$$

Then we have a complex $Y(a)$ for which $\sigma \in Y(a)$ if for any face $\tau \leq \sigma$ there exists an element $x \in \mathcal{W} \cap |\mathcal{U}(a)|$ which is a weak witness for τ, where $\mathcal{U} = \text{BallCov}(\mathcal{L}, p)$.

A strong witness is the same as a weak witness if we also have equality in (7) whenever $i, j \in \sigma$, from which it follows that the strong witness complex is contained in the weak witness complex. On the other hand, de Silva's theorem [10] shows that we have equality in the case of $\mathcal{W} = \mathbb{R}^m$. Unlike the strong witness complex, the weak witness complex has the advantage that each cell has nonzero measure, so if a data set $\mathcal{W} = \mathcal{D}$ is sufficiently dense, one may reasonably expect that each cell will contain a witness.

2.4 Practical implementations

Generally, computations related to power diagrams and their associated alpha complexes are determined by a family of quadratic programs, which are computationally expensive. An efficient algorithm in dimension 3 is given in [14]. To compute these complexes in higher dimension we have formulated an algorithm based on *dual* quadratic programming. This is in part because there are a large number of potential simplices, but most can be ruled out. In dual programming, early termination is built in to the setup. Our implementation of this in MAPLE using a highly flexible and elegant recent active set algorithm due to Ärnstrom, Bemporad and Axehill [2], which will be explained in a separate paper.

3 Density filtered complex

We define a filtered complex associated to a density estimator $f(x)$ and prove our main result, which is the interleaving property described in the introduction.

3.1 Landmark selection

Let $\mathcal{D} = \{x_1, ..., x_N\} \subset \mathbb{R}^m$, and consider a sum $f(x)$ of Gaussian kernels as in (1). We begin with a landmark selection scheme which determines the vertices, encoded by a certain approximation $g(x) \leq f(x)$ by a max of weighted Gaussians, rather than a sum.

We first show that for every point $y \in \mathbb{R}^m$, we can "fit" f to first order at y by a single weighted Gaussian kernel of the form $g(x) = b\rho(z, x)$ using the formula

$$z = c^{-1} \sum_i a_i \rho(y, x_i) x_i, \quad b = \rho(z, y)^{-1} c, \quad c = \sum_i a_i \rho(y, x_i). \tag{8}$$

Said another way, z is the expectation of the random variable x_i with respect to the probability measure on $\{1, ..., N\}$, with density $i \mapsto a_i \rho(y, x_i)$, normalized so that the sum is 1. We denote the resulting function by $g = \text{GaussFit}(f, y)$.

Lemma 3.1. *We have that $g(x)$ is the unique function of the form $b\rho(z, x)$ satisfying $g(x) \leq f(x)$ for all $x \in \mathbb{R}^m$, with equality at $x = y$.*

Proof. First we have that $g(x)$ agrees with $f(x)$ to first order at y:

$$g(y) = f(y), \quad \nabla_y g = \nabla_y f. \tag{9}$$

To check this, substitute (8) into (9), and divide the second part by the first by the first to obtain z, and then solve for b. Since (9) must hold for any function $g(x)$ satisfying the properties of the lemma, the uniqueness statement is clear.

For simplicity, suppose that $h = 1$, all coefficients are one, and the fitting is performed at the origin $y = 0$. To see that $g(x)$ satisfies the desired properties, we first consider the case of dimension $m = 1$ and only two points, which gives:

$$f(x) = e^{-(x-x_1)^2/2} + e^{-(x-x_2)^2/2}, \quad g(x) = be^{-(x-z)^2/2}.$$

Consider the function

$$q(x) = f(x)/g(x) =$$
$$b^{-1}e^{(x-z)^2/2-(x-x_1)^2/2} + b^{-1}e^{(x-z)^2/2-(x-x_1)^2/2} =$$
$$b^{-1}e^{z^2/2}\left(t_1 e^{(x_1-z)x} + t_2 e^{(x_2-z)x}\right)$$

where $t_i = c^{-1}a_i\rho(y, x_i)$ is the coeffficient of x_i in the construction of z, from (8). Since $t_i > 0$ by construction, we find that q is a strictly convex function of x, so its global minimizer x^* (if one exists) occurs if and only if $q'(x^*) = 0$. We see that indeed $q'(0) = 0$, because

$$\begin{aligned}\frac{d}{dx}q(x)\bigg|_{x=0} &= t_1(x_1 - z)e^{(x_1-z)x} + t_2(x_2 - z)e^{(x_2-z)x}\bigg|_{x=0}, \\ &= t_1(x_1 - z) + t_2(x_2 - z) = t_1x_1 + t_2x_2 - z = 0,\end{aligned}$$

where the the last step uses the fact that $t_1 + t_2 = 1$ and the the definition of z in (8).

Next, we have the case of two points in dimension $m > 1$. In this case it is enough to notice that $f(x)/g(x)$ is independent of the orthogonal direction to the line through x_1, x_2, as both functions scale by a common factor.

Finally, suppose we have N points in $\mathbb{R}^m$, and proceed by induction on N. In the base case we obviously recover the original function $g = f$. For $N > 1$, write $f = f_1 + f_2$ where

$$f_1(x) = e^{-(x-x_1)^2/2} + \cdots + e^{-(x-x_{N-1})^2/2}, \quad f_2(x) = e^{-(x-x_N)^2/2}$$

and let $g_i = \text{GaussFit}(f_i, 0)$. Then we have

$$g = \text{GaussFit}(f, 0) = \text{GaussFit}(f_1 + f_2, 0) =$$

$$\text{GaussFit}(g_1 + f_2, 0) \leq g_1 + f_2 \leq f_1 + f_2 = f$$

completing the proof. □

Lemma 3.1 implies a method for approximating $f(x)$ from below by functions which are maxes of Gaussian kernels, as follows. First, fix a number $0 < s < 1$ and a bounded domain $\mathcal{A} \subset \mathbb{R}^m$. We successively make the replacement $g(x) \mapsto \max(g(x), \text{GaussFit}(f, y))$ each time using a point $y \in \mathcal{A}$ for which $g(y) < sf(y)$ until there are none, described explicitly in Algorithm 1. The resulting center points z will correspond to the vertices in our main construction.

Algorithm 1 Max of Gaussians

input $(f, \mathcal{A}, s)$
$\mathcal{A}_0 \leftarrow \mathcal{A}$
$g_0 \leftarrow (x \mapsto 0)$
$n \leftarrow 0$
while $\mathcal{A}_n \neq \emptyset$ **do**
  $n \leftarrow n + 1$
  $y_n \leftarrow \text{argmax}_{\mathcal{A}_{n-1}}(f)$
  $b_n \rho(\_, z_n) \leftarrow \text{GaussFit}(f, y_n)$
  $g_n \leftarrow \max(g_{n-1}, b_n \rho(\_, z_n))$
  $\mathcal{A}_n \leftarrow \{y \in \mathcal{A} : g(y) < sf(y)\}$
end while

We now have the following lemma:

Lemma 3.2. *Let $\mathcal{A} \subset \mathbb{R}^m$ be a bounded region, and let $0 < s < 1$. Then we have that Algorithm 1 terminates, and the resulting function $g(x) = g_n(x)$ satisfies*

$$g(x) \leq f(x) \leq s^{-1} g(x) \tag{10}$$

for all $x \in \mathcal{A}$.

Proof. If the algorithm terminates, equation (10) is clear. To prove it terminates, we use a Lipschitz argument: at the end of iteration n of Algorithm 1, there are n locations y_i where $g_n(y_i) = f(y_i)$. We will show that there exists $\delta > 0$, depending only on f, $\mathcal{A}$, and s, such that $\|x - y_i\| \leq \delta \implies f(x) \leq s^{-1} g(x)$ for $x \in \mathcal{A}$. This will suffice to prove the lemma, because it establishes that the locations y_i must all be a distance δ apart from one another, so the number of iterations is bounded above by (for instance) the packing number of $\mathcal{A}$ for balls of radius $\delta/2$.

As $f(x)$ is a mixture of Gaussians and $\mathcal{A}$ is bounded, there exist constants c, C_1 such that $0 < c \leq f(x)$ and $\|\nabla f(x)\| \leq C_1$ for all $x \in \mathcal{A}$. Since $\mathcal{A} \times \mathcal{A}$ is also bounded, there also exists a constant C_2 such that for all $x, x' \in \mathcal{A}$, we have $\|\nabla h(x)\| \leq C_2$, where $h = \text{GaussFit}(f, x')$, so that C_1 and C_2 are Lipschitz constants for f and all possible mixtures h:

$$|f(x) - f(x')| \leq C_1 \|x - x'\|$$
$$|h(x) - h(x')| \leq C_2 \|x - x'\| .$$

For notational compactness, let $h_n = b_n\rho(\_, z_n) = \text{GaussFit}(f, y_n)$ be the contribution to g in the n-th iteration, and let $C := \max\{C_1, C_2\}$. We claim that if $\delta = \frac{(1-s)c}{(1+s)C}$, then $\|x - y_n\| \leq \delta \implies h_n(x)/f(x) \geq s$, which will complete the proof. For any $x \in \mathcal{A}$, we have

$$\frac{h_n(x)}{f(x)} \geq \frac{h_n(y_n) - C\|y_n - x\|}{f(y_n) + C\|y_n - x\|} = \frac{f(y_n) - C\|y_n - x\|}{f(y_n) + C\|y_n - x\|},$$

and so if $\|x - y_n\| \leq \delta$, we have

$$\frac{f(y_n) - C\|y_n - x\|}{f(y_n) + C\|y_n - x\|} \geq \frac{f(y_n) - C\delta}{f(y_n) + C\delta} \geq \frac{c - C\delta}{c + C\delta} = \frac{c - C\left[\frac{(1-s)c}{(1+s)C}\right]}{c + C\left[\frac{(1-s)c}{(1+s)C}\right]} = s$$

as desired. □

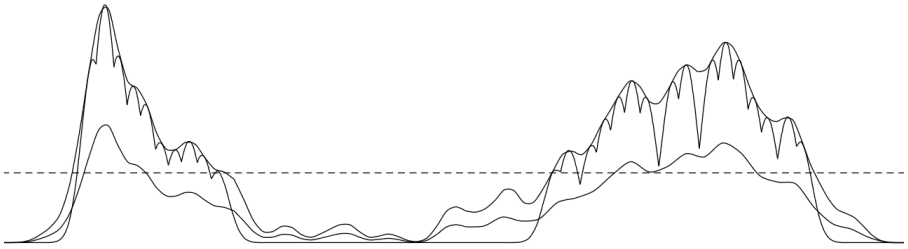
Remark 3.1. The proof of Lemma 3.2 does not depend on the criteria for determining the choice of the next reference point, which is given by $y_n = \text{argmax}_{\mathcal{A}_{n-1}}(f)$ in Algorithm 1. We experimented with other criteria that seem to give comparable results, another natural choice being the "greedy" one of $\text{argmin}_{\mathcal{A}_{n-1}}(g/f)$. The reason we choose the densest remaining point is that it leads to the useful property that $g_{\mathcal{A}_{\geq d}} \leq g_{\mathcal{A}_{\geq d'}}$ for $d \geq d'$, using the subscript to denote the resulting function for different reference sets (assuming we have chosen a consistent criteria for breaking ties). This results in a desirable compatibility property of the filtered complexes defined in Section 3.

In practice, we will take $\mathcal{A}$ to be a finite set, so that the algorithm obviously terminates. Because of Lemma 3.2, we have that choosing a very dense reference set does not result unboundedly large numbers of landmarks n, as illustrated in one dimension in Example 3.1 below. In higher dimensions, dense subsets of $f^{-1}[a, \infty)$ are usually too large to consider. An alternative in this case that we will use in Section 4 is to take the data set itself intersected with the minimum density cutoff as the reference set, $\mathcal{A} = \mathcal{D}_{\geq d_0}$.

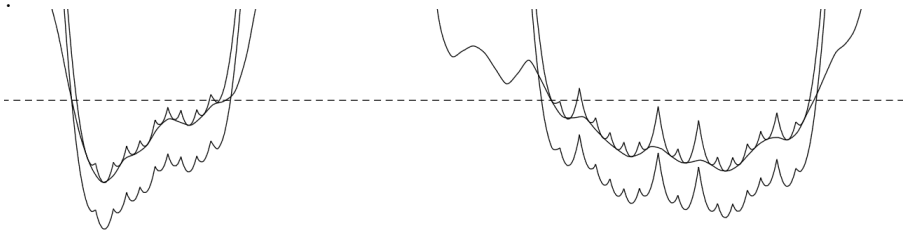
Example 3.1. The Old Faithful geyser data set consists of 272 measurements of eruptions from the Old Faithful Geyser in Yellowstone [4]. We considered the one-dimensional density estimator of the data set of eruption times with $h = .05$ seconds, equal weights $a_i = 1/272$, and minimum density of $d_0 = .03$. We applied Algorithm 1 to a dense finite set of 10000 points in the region $f^{-1}[d_0, \infty)$. The results are shown in Figure 3.1. We see that $g(x)$ is a max of 26 weighted Gaussians, and this number will not increase with more samples, according to Lemma 3.2 (and also Remark 3.1).

3.2 An associated power diagram

We now make a critical observation, which is that the function $g(x)$ resulting from Algorithm 1 is piecewise Gaussian, and in fact determines the cells of a power diagram. This is stated in the following lemma whose proof is clear:



(a) The inequality $.5f(x) \leq g(x) \leq f(x)$



(b) Equivalent inequality $g(x) \leq f(x) \leq 2g(x)$ on a log scale.

Figure 3.1: The result of applying Algorithm 1 to eruption times from the Old Faithful Geyser data set. Here $f(x)$ is density estimator with scale parameter $h = .05$ seconds, the domain $\mathcal{A}$ is a set of 10000 points with $f(x) \geq d_0 = .03$, and $s = .5$. The horizontal range is 1.4 to 5.3 seconds, and the density cutoff is the dashed line.

Lemma 3.3. *Let* $g(x) = \max_i b_i \rho(x, z_i)$ *as in Algorithm 1. Then* $w(x) = -\log(g(x))$ *is the weight function of the power diagram with*

$$\mathcal{L} = \{z_1, ..., z_n\}, \quad p_i = -2h^2 \log(b_i) \tag{11}$$

Definition 3.1. The power diagram corresponding to the output of Algorithm 1 and its associated alpha complex will be denoted $\text{PowDiag}(f, \mathcal{A}, s)$ and $\text{DensAlpha}(f, \mathcal{A}, s)$ respectively.

Example 3.2. The resulting function from applying Algorithm 1 to the density estimator from Example 2.1 and its associated power diagram are shown in Figure 3.2, using a value of $s = .8$, and the previously used values of h, d_0. We have taken $\mathcal{A}$ to be a densely sampled region from $f^{-1}[d_0, \infty)$. Notice that we always have the same number of cells as landmarks at the density cutoff, as each cell necessarily contains a unique reference point y_i, making it nonempty. However, there are many landmarks z_i that are not contained in their own cell.

3.3 Density based complexes

We can now define our main constructions.

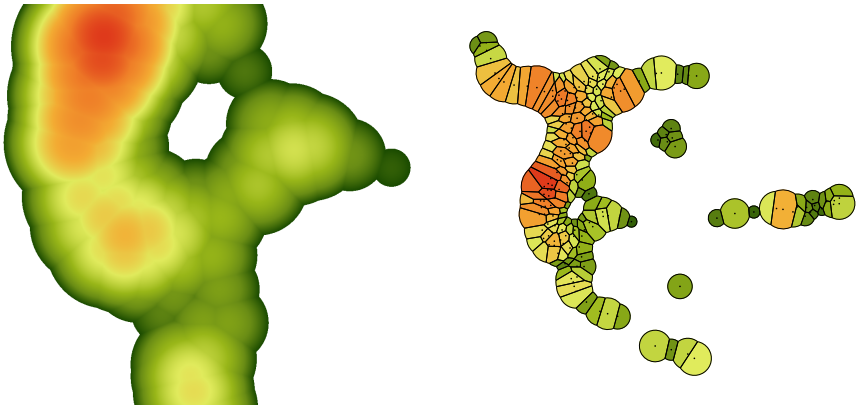


Figure 3.2: The max of Gaussians $g(x)$ associated to the density estimator $f(x)$ from the species distribution data set, according to Algorithm 1 using $s = .8$, and the resulting power diagram.

Definition 3.2. Let $Y = \text{SubDens}(f, \mathcal{A}, s)$ be the filtered complex with total space $X = \text{DensAlpha}(f, \mathcal{A}, s)$, and weight function

$$w(\sigma) = \max_{x \in |\sigma|} -\log(f(\phi(x))). \tag{12}$$

In other words, $Y(a)$ is the maximal subcomplex of $X = X(-\infty)$ with the property that the image of $\phi : |Y(a)| \to \mathbb{R}^m$ is completely contained in $f^{-1}[e^{-a}, \infty)$.

We now have the following theorem which says that the three filtered homology groups $H_*(X(a)), H_*(Y(a)), H_*(f^{-1}[a, \infty))$ are interleaved up to the minimum density cutoff, see [7] Definition 4.2. It is shown that when persistence modules are strongly interleaved, their persistent homology groups approximate one another according to a certain metric called the bottleneck distance.

Theorem 3.1. *Let $\mathcal{A} = f^{-1}[e^{-a_0}, \infty)$, fix some $0 < s < 1$, and set $\epsilon = -\log(s)$. Then $X = \text{DensAlpha}(f, \mathcal{A}, s)$ and $Y = \text{SubDens}(f, \mathcal{A}, s)$ define finite filtered complexes. Moreover, we have a family of maps*

$$H_*(X(a)) \to H_*(Y(a)) \to H_*(f^{-1}[e^{-a}, \infty)) \to H_*(X(a + \epsilon)) \tag{13}$$

defined for any $a \leq a_0$, which commute with the persistence maps $i_*^{a,b}$ from (2). The composition is equal to $i_*^{a,a+\epsilon} : H_*(X(a)) \to H_*(X(a + \epsilon))$, and similarly for the compositions $H_*(Y(a - \epsilon)) \to H_*(Y(a))$ and $H_*(f^{-1}[e^{-a+\epsilon}, \infty)) \to H_*(f^{-1}[e^{-a}, \infty))$.

Proof. By Lemma 3.2, we have that Algorithm 1 terminates so that X and Y are well-defined and finite. Since $g(x) \leq f(x)$, we have that $X(a) \subset Y(a)$. The map $|Y(a)| \to f^{-1}[e^{-a}, \infty)$ is induced by the restriction of $\phi : |X| \to \mathbb{R}^m$ to $|Y(a)|$, whose image is contained completely in $f^{-1}[e^{-a}, \infty)$ by the definition

of Y. By (10), we have that $f^{-1}[e^{-a},\infty) \subset |\mathcal{V}(a+\epsilon)| = g^{-1}[e^{-a-\epsilon},\infty)$, where $\mathcal{V} = \mathrm{DensPow}(f,\mathcal{A},s)$. We thus have a sequence of continuous maps

$$|X(a)| \to |Y(a)| \to f^{-1}[e^{-a},\infty) \to |\mathcal{V}(a+\epsilon)|, \tag{14}$$

which define the maps from (13) by taking homology, and applying the nerve isomorphism $H_*(|\mathcal{V}(a+\epsilon)|) \cong H_*(X(a+\epsilon))$.

Next, we have that (14) is compatible with $i^{a,b}$ as continuous maps. Then we find that (13) is compatible with the persistence maps by taking homology of the resulting diagram, and adding an extra square on the right

$$\begin{array}{ccccc} \cdots & \longrightarrow & H_*(|\mathcal{V}(a+\epsilon)|) & \longrightarrow & H_*(X(a+\epsilon)) \\ & & \downarrow & & \downarrow \\ \cdots & \longrightarrow & H_*(|\mathcal{V}(b+\epsilon)|) & \longrightarrow & H_*(X(b+\epsilon)) \end{array}$$

coming from the functoriality of the nerve isomorphism.

Finally we check that the composition $H_*(X(a)) \to H_*(X(a+\epsilon))$ is equal to $i_*^{a,a+\epsilon}$, and similarly for $Y(a-\epsilon), f^{-1}[e^{-a+\epsilon},\infty)$. First, the map $|X(a)| \to f^{-1}[e^{-a},\infty)$ from (14) factors as $\phi : |X(a)| \to |\mathcal{V}(a)|$ composed with the inclusion $|\mathcal{V}(a)| = g^{-1}[e^{-a},\infty) \subset f^{-1}[e^{-a},\infty)$. We then find that the induced maps from taking homology in

$$\cdots \subset f^{-1}[e^{-a+\epsilon},\infty) \subset |\mathcal{V}(a)| \subset f^{-1}[e^{-a},\infty) \subset |\mathcal{V}(a+\epsilon)|$$

agree with the maps in (13) after extending cyclically to the left, and taking the composition over $H_*(Y(a))$. We thus obtain the first and third compositions.

In the remaining case we have $Y(a-\epsilon) \subset X(a) \subset Y(a)$ as subcomplexes of X, since taking the max over $|\sigma|$ respects (10). It suffices to check that the first induced map $H_*(Y(a-\epsilon)) \to H_*(X(a))$ is the corresponding composition from (13) at $a = a-\epsilon$. To check this, take homology of the diagram

$$\begin{array}{ccccc} |Y(a-\epsilon)| & \longrightarrow & |X(a)| & & \\ \downarrow & & \downarrow & & \\ \phi(|Y(a-\epsilon)|) & \longrightarrow & \phi(|X(a)|) & \longrightarrow & |\mathcal{V}(a)| \end{array}$$

Then the composition $|X(a)| \to |\mathcal{V}(a)|$ induces the nerve isomorphism, whereas the one from the upper left to lower right is the composition from (14).

□

Example 3.3. Figure 3.3 illustrates equation (14) from the proof of Theorem 3.1 using the Geyser data set of Example 3.1. The four subspaces given by $X = \mathrm{DensAlpha}(f,\mathcal{A},s)$, $Y = \mathrm{SubDens}(f,\mathcal{A},s)$, $f^{-1}[a,\infty)$, and $|\mathcal{V}(a+\epsilon)|$ of $\mathbb{R}$ appear as the sublevel sets of the four graphs shown as solid lines. They appear in order from top to bottom, whenever the persistence value of a, shown on the y-axis, is below the dashed line $a = a_0$.

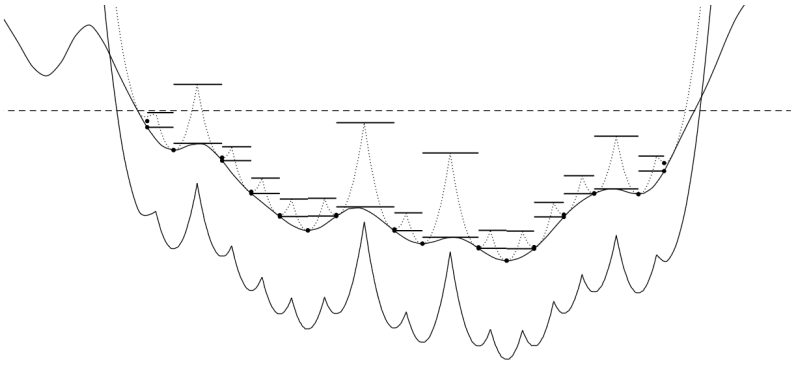


Figure 3.3: The four filtered subspaces of $\mathbb{R}$ from the proof of the main theorem, given by $|X(a)|, |Y(a)|, f^{-1}[e^{-a}, \infty), |\mathcal{V}(a+\epsilon)|$ for the Old Faithful Geyser example, zoomed in on part of Figure 3.1b, shown as solid lines. The y-axis is the persistence value of a.

In practice, we will use the following complex, which replaces the max in (12) with a max over the just the barycenters of $|\sigma|$. Recall that q_σ are defined for any alpha complex as the minimizers of the weight function (6).

Definition 3.3. We define $Y = \text{DensWit}(f, \mathcal{A}, s)$ to have the same total space as X as in Definition 3.2, but using the modified weight function

$$w(\sigma) = \max_{\tau \subset \sigma} -\log(f(q_\tau)). \tag{15}$$

Remark 3.2. In light of our terminology, the reader may wonder in what sense Y is a witness complex. While Y is not technically a witness complex, a related weight function which replaces the max over $\tau \subset \sigma$ with a min over $\tau \supset \sigma$ exhibits similar properties, and would yield a filtered family of witness complexes $Y(a)$ with witnesses $\mathcal{W} = \{q_\sigma : f(\sigma) \geq e^{-a}\}$. Moreover, the given weight function in (15) would actually be an example of a weak witness complex with the same choice of $\mathcal{W}$, if we replaced the inequality in (7) with strict inequality.

Example 3.4. We show the sequence of complexes in the case of Examples 2.1 and 3.2 in Figure 3.4, using the same values of $(f, \mathcal{A}, s)$, and the complex $Y = \text{DensWit}(f, \mathcal{A}, s)$. In the upper left we have the part of the subcomplex of $|X|$ that is contained in $|\mathcal{V}(a_0 + \epsilon)|$, which contains all witness that could possibly appear in Y, assuming that $\mathcal{A}$ is sufficiently dense. The subfigure in the upper right illustrates the last two containments in (14), showing the 1-skeleton of Y instead of $\text{SubDens}(f, \mathcal{A}, s)$, the middle one being the same boundary of $f^{-1}[a_0, \infty)$ shown in Figure 2.1. The containment shown is not actually guaranteed in this case, first because the reference set $\mathcal{A}$ is not actually the entire superlevel set, second because we have used the approximation of

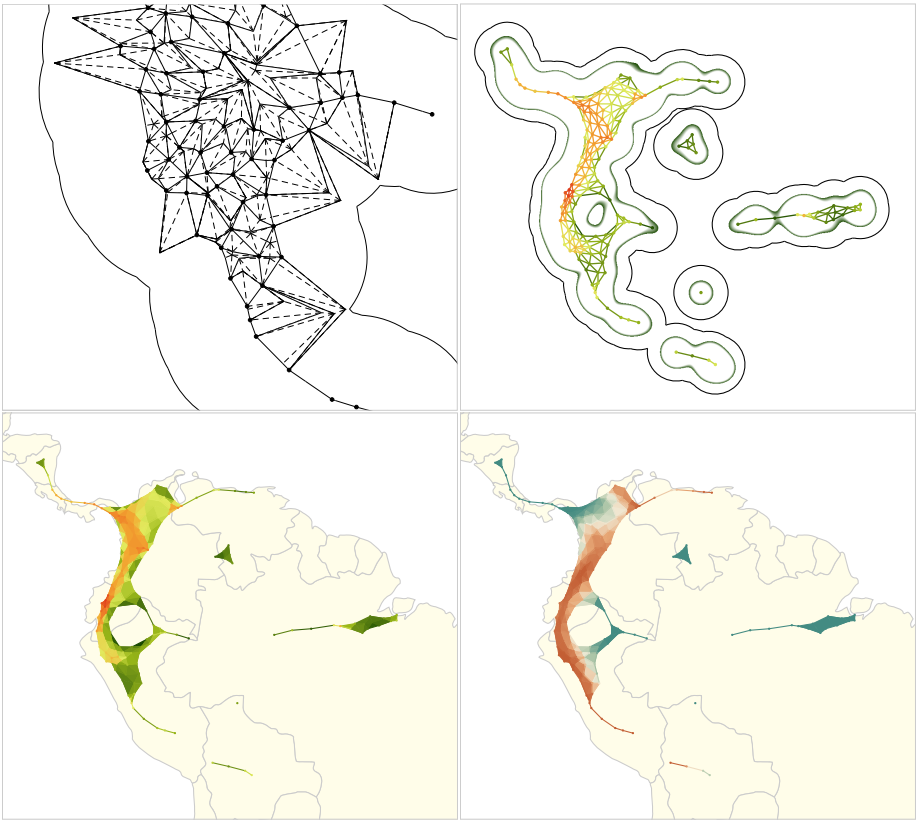


Figure 3.4: Illustration of the containments from Theorem 3.1 in the context of the species distribution example, using the previously used specifications. Top left: the map ϕ applied to the part of $X = \text{DensAlpha}(f, \mathcal{A}, s)$ contained in $|\mathcal{V}(a_0 + \epsilon)|$ for $a_0 = -\log(d_0)$. Top right: the one-skeleton of $Y = \text{DensWit}(f, \mathcal{A}, .8)$, the boundary of $f^{-1}[d_0, \infty)$, and $|\mathcal{V}(a_0 + \epsilon)|$. Bottom left: a heat map of the persistence values of the full two skeleton. Bottom right: another coloring of Y by relative density values of each of the two species, evaluated at the barycenters.

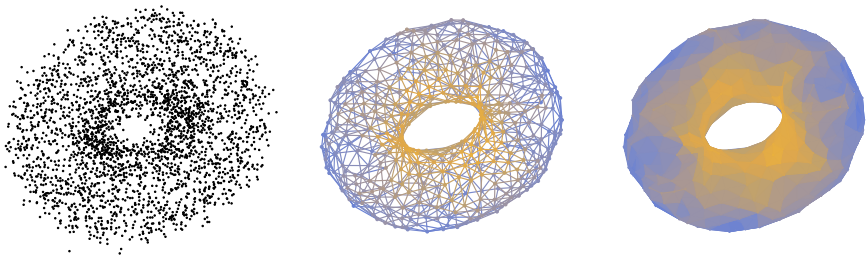


Figure 4.1: A data set of a torus $\mathcal{D} \subset \mathbb{R}^3$ with Gaussian noise, and the 1 and 2-skeleta of $\text{DensWit}(f, \mathcal{D}_{\geq .005}, .6)$ with $h = .3$.

$\text{DensWit}(f, \mathcal{A}, s)$ instead of $\text{SubDens}(f, \mathcal{A}, s)$, and last because we have used the values of the landmarks z_i to map the complex two $\mathbb{R}^2$ instead of ϕ. Nevertheless, we do still see the containment, indicating that all three approximations are reasonable in this example. In the lower left we have the full complex Y using the heat map from Figure 2.1. In the lower right, we overlayed a different color map onto the total complex of $Y(a_0)$, by restricting separate density estimators for the two different species, using their values at the vertices of ϕ in the upper left. Notice that Y retains one-dimensional features of the data, due to the congregation along various parts of the Amazon river for instance, while $f(x)$ does not. This happens because the centers z_i are convex combinations of the elements of $\mathcal{D}$ by (8).

4 Higher dimensional examples

We demonstrate the performance of the complex in higher dimensions using some persistent homology computations, and by viewing the results using a low dimensional projection.

4.1 Persistent homology computations

We start with two persistent homology computations using a torus data set with noise, and simulated points from an ordered configuration space.

Example 4.1. We formed a data set of points on the two-dimensional torus embedded in $\mathbb{R}^3$ by sampling

$$(\cos(\theta_1)(1 + .5\cos(\theta_2)), \sin(\theta_1)(1 + .5\cos(\theta_2)), .5\sin(\theta_2))$$

using 3000 random values of θ_1, θ_2. We then added noise using Gaussian samples, and considered the density estimator $f(x)$ using the scale parameter $h = .3$. We computed $Y = \text{DensWit}(f, \mathcal{D}_{\geq .005}, .6)$, shown in Figure 4.1. We found sizes $(|Y_0|, |Y_1|, |Y_2|, |Y_3|) = (466, 1581, 1308, 199)$, a reasonable number of simplices for a data set that is centered around a two-dimensional manifold.

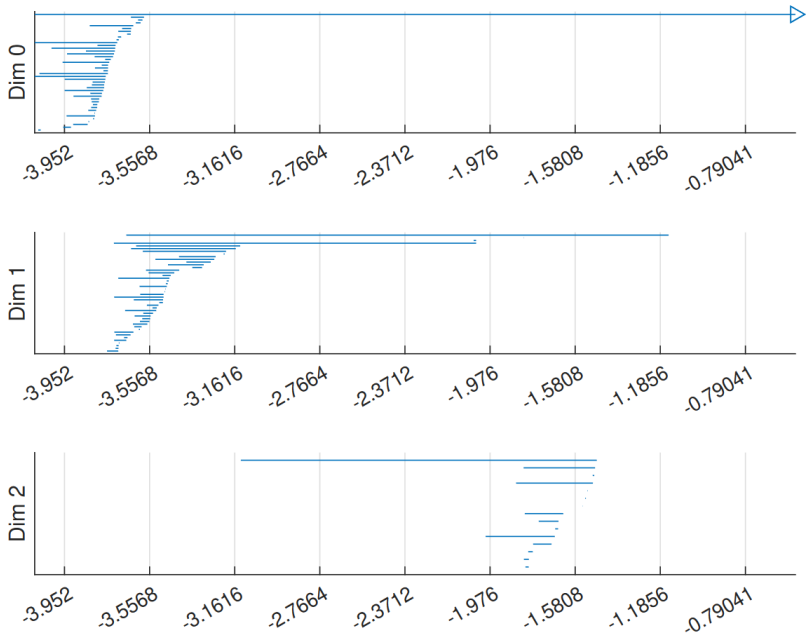


Figure 4.2: The persistent homology groups for the torus with noise with rescaled weights.

We then computed the persistence barcodes, which captured the β_1-cycle near the center, but failed to capture the remaining one, as well as the $\beta_2 = 1$ feature due to the fact that the data is insufficiently dense around the periphery. One might expect to capture the remaining features by using different values of the scale parameter in different points of the data set. The problem of combining complexes coming from power diagrams associated to different metric is indeed an interesting one. However, we have a simpler available option in this case, which is to simply rescale the weights a_i. We computed the complex for the result of scaling the weights of $f(x)$ by distance from the origin, $a_i \mapsto \|x_i\| a_i$, thus increasing the density around the periphery by a factor of about 3 relative to the inner circle. The resulting persistent homology groups capture all the desired betti numbers, shown in Figure 4.2.

Example 4.2. We consider a space with more sophisticated homology groups, namely the ordered configuration space $\mathrm{Conf}_3(\mathbb{R}^2)$ of 3 ordered points in the plane. The homology groups of this space are well-known in greater generality [9]. We use a homotopy equivalent variant that is more suitable for density estimation, in which the points must be a distance at least one apart, two of those distances being equal to one.

To form a data set, we sampled pairs of points $\theta_1, \theta_2 \in [0, 2\pi)$ uniformly at random, and computed the unique triple $(p_1, p_2, p_3) \in (\mathbb{R}^2)^3 = \mathbb{R}^6$ whose mean

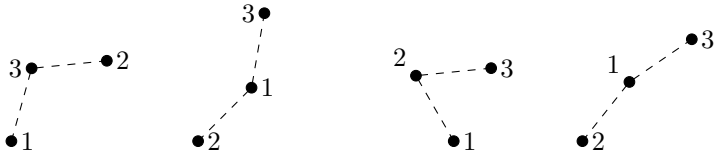


Figure 4.3: Some points in configuration space.

is at the origin, satisfying

$$p_2 - p_1 = (\cos(\theta_1), \sin(\theta_1)),\ p_3 - p_2 = (\cos(\theta_2), \sin(\theta_2)).$$

We then discarded any triples with $\|p_3 - p_1\| < 1$, and permuted the order via a random element of S_3. We sampled 20000 instances to obtain a data set of points $\mathcal{D} = \{(p_1, p_2, p_3)\} \subset \mathbb{R}^6$ with the property that the pairwise distances are all at least 1, and all but one are equal to 1. Some typical elements are pictured in Figure 4.3.

We then chose the corresponding density estimator $f(x)$ with values of $h = .25$ in the L^2 metric, which does not represent a uniform measure on the underlying space in any sense. We computed $\mathrm{DensWit}(f, \mathcal{D}_{\geq .006}, .4)$ up to the 3-simplices, and obtained a 3-dimensional filtered complex with sizes (271,1264,1558,631). The persistent homology groups are shown in Figure 4.4. The bars which extend indefinitely correspond to the desired betti numbers of $\mathrm{Conf}_3(\mathbb{R}^2)$ given by $(\beta_0, \beta_1, \beta_2) = (1, 3, 2)$. We also see that in the medium density range, we have a value of $\beta_0 = 2$ and $\beta_1 = 2$, whose bars are too long to be due to noise. This is due to the fact that the triples which nearly form an equilateral triangle tend to be denser that triples which are colinear. As a result, we obtain two disconnected loops in the medium density range, corresponding to the two rotationally inequivalent permutations of the labels, when the points lie on an equilateral triangle.

4.2 Local patches from the MNIST data set

In [16], the authors studied the topology of a certain space of local 3×3 high intensity patches of the van Hateren data set of natural images, which was investigated earlier by Lee, Mumford, and Pederson [17, 19]. They gave quantitative evidence using the witness complex that those patches lie along a two-dimensional locus parametrized by the Klein bottle.

We illustrate our complex on a parallel construction of high intensity local patches in the MNIST data set of 28×28 images of hand-drawn digits [11]. Because digits tend to have lines in the middle of blank space, but rarely the reverse, we expect those patches to lie in a sublocus of the Klein bottle homeomorphic to the Möbius strip. Some points in the MNIST data set and a parametrization of this Möbius strip are shown in the first two rows of Figure 4.5. The circle that traces around the boundary of the Möbius strip through the top and bottom row is called the primary circle, and it usually contains

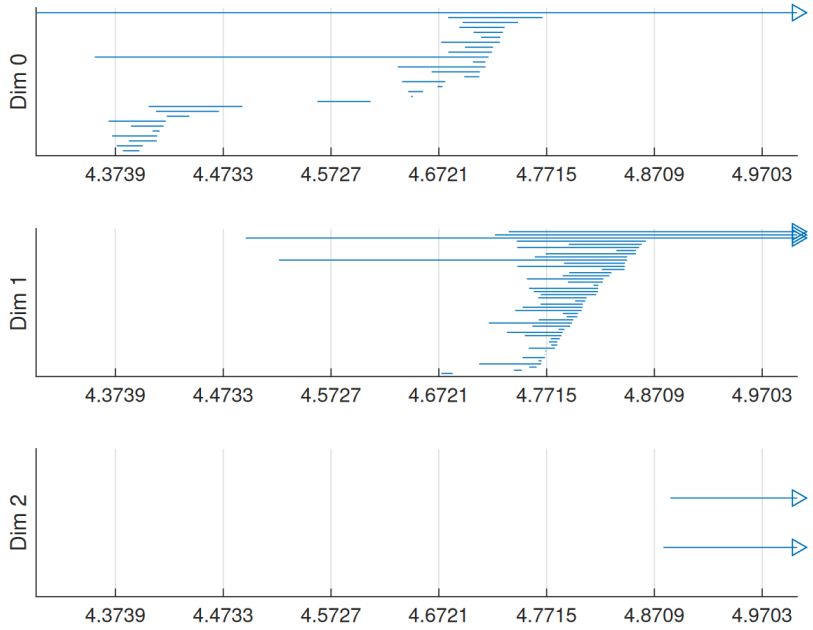


Figure 4.4: Persistent homology groups of the Configuration space data set.

points of higher density than those in the middle row. The circle consisting of the middle row is called the secondary circle, and it is more difficult to detect. We will see that the density based complex exhibits descriptive visual models of both primary and secondary circle features, with clear differences between different digits.

To form a parametrization of image patches, we considered an inner product on the l^2-dimensional vector space $\mathrm{Mat}(l,l)$ of $l \times l$ image patches, given by

$$(A,B) = \frac{1}{2^{2(l-1)}} \sum_{i=1}^{l} \sum_{j=1}^{l} \binom{l-1}{i-1} \binom{l-1}{j-1} A_{i,j} B_{i,j} \tag{16}$$

We then consider an orthonormal basis given by $H^l_{a,b} = H^l_a \otimes H^l_b$, where the $H_a \in \mathbb{R}^l$ are a discrete form of the Hermite polynomials, given by applying the Gram-Schmidt algorithm to the vectors of polynomial functions $V^a = (i^a)_{i=1}^l$, using the one-dimensional form of (16). For several reasons, we find the inner product to be more robust than the usual dot product, which would lead to products of Legendre polynomials. The bottom row of Figure 4.5 shows an example of two such vectors, and the result of projecting an image patch onto the span of the 6 Hermite polynomials up to quadratic order for $l = 7$.

We then constructed a data set as follows. For 50 instances each digit, we sampled all $l \times l$ patches for various choice of l, and projected those patches onto $\mathrm{span}(H^l_{1,0}, H^l_{0,1}, H^l_{2,0}, H^l_{1,1}, H^l_{0,2})$, the span of nonconstant terms up to quadratic

order, to obtain a data set of size $50 \cdot 28 \cdot 28 = 39200$ in $\mathbb{R}^5$ (assuming that patches are extended by zero outside the domain of the image). We then chose only those images whose norm is above a fixed number of $r = .3$, resulting in a subset of roughly 10% of the original size. This is a version of the choice of "high intensity patches" from [16]. We then normalized the resulting points to arrive at a data set $\mathcal{D}(k) \subset S^4$ of size 3000-5000 for each digit $k \in \{0, ..., 9\}$.

We show the results of DensWit$(f, \mathcal{A}, s)$ using the values of $l = 11$, $h = .16$, $\mathcal{A} = \mathcal{D}(k)_{\geq d_0}$, $s = .5$, with varying choices of d_0 in the top row of Figure 4.6. We see that primary circle features are dominant, corresponding to the dense regions around the periphery, with secondary circle features connecting them. The coordinate system is the one spanned by $H^{11}_{1,0}, H^{11}_{0,1}$. In order to highlight the secondary circle features, we produced a variant on the above data set, in which high intensity is determined only by the norm of the quadratic terms, and in which we normalize by that value. This has the effect of dimensionally reducing just the second degree terms, thereby accentuating the secondary circle. We then mapped the resulting complexes into low dimension using a similar parametrization of the Klein bottle to the one given in [16]. In the case of the digit 0, we see a very clear Möbius strip. In the case of the digit 2, we lowered the density threshold, revealing a primary circle feature encircling the second order ones.

4.3 The Ising model on a graph

In our final example, we consider density estimation on a simulated data set consisting of trials of the Ising model [18] on a graph with m vertices, thought of as a collection of real-valued vectors in $\mathbb{R}^m$. In order to obtain a viable density estimator on this data set, we use the Laplacian operator associated to the graph, which has the effect of replacing each trial by something resembling a continuous function.

Let $G = (V, E)$ be an $m \times m$ graph such as the ones shown in Figure 4.7, represented by a symmetric adjacency matrix J, diagonal entries being zero. For every discrete vector of "spins" $\sigma \in \{1, -1\}^m$, we have the Hamiltonian energy

$$H_G(\sigma) = -\sum_{i,j} J_{i,j}\sigma_i\sigma_j = H_{\min} + 2|\{(i,j) \in E : \sigma_i \neq \sigma_j\}|. \tag{17}$$

Those points $(i, j) \in E$ for which $\sigma_i \neq \sigma_j$ are called transitions. For each choice of $\beta > 0$, called the temperature parameter, one seeks to sample from the Boltzmann distribution on $\{1, -1\}^m$ given by

$$P_\beta(\sigma) = \frac{1}{Z_\beta} e^{-\beta H(\sigma)}, \quad Z_\beta = \sum_\sigma e^{-\beta H(\sigma)} \tag{18}$$

which is done using the single-flip Metropolis algorithm.

A collection of N trials can be interpreted as a data set

$$\mathcal{D}(G, \beta) = \{\sigma^1, ..., \sigma^N\} \subset \mathbb{R}^m,$$

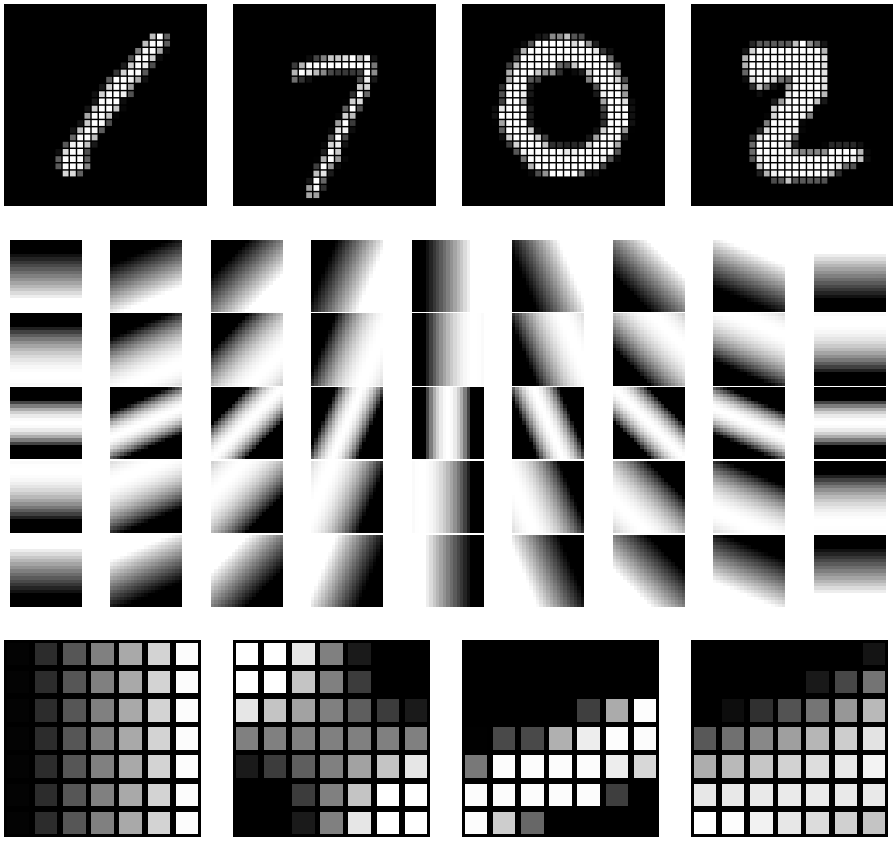
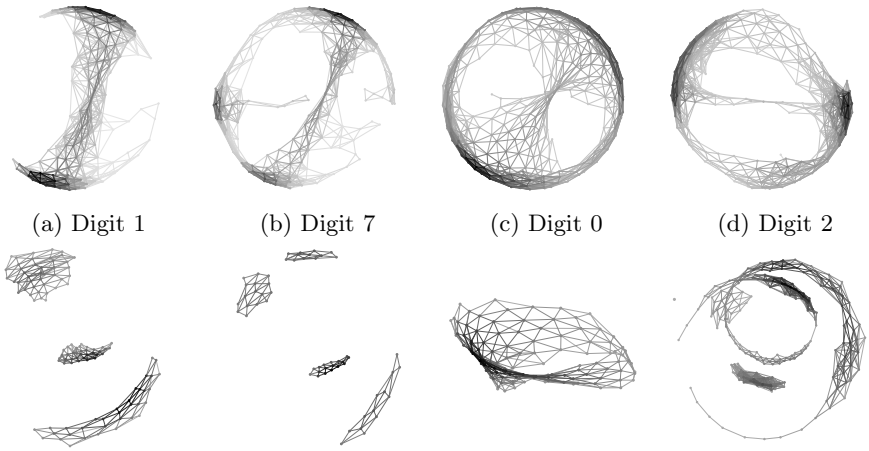


Figure 4.5: Top row: some elements from the MNIST data set of 28×28 grayscale images of hand drawn digits. Middle row: a Möbius strip in the space of image patches. Bottom row: on the left are two instances of the Hermite polynomials $H^7_{0,1}$ and $H^7_{1,1}$. In the second from the right, a typical 7×7 patch from the data set, and its projection onto the span of the Hermite polynomials up to quadratic order on the far right.



(a) Digit 1 (b) Digit 7 (c) Digit 0 (d) Digit 2

(e) Digit 1 (f) Digit 7 (g) Digit 0 (h) Digit 2

Figure 4.6: Top row: The complex DensWit$(f, \mathcal{A}, s)$ for the digits $\{1, 7, 0, 2\}$. All have the same parameters $h = .16$, $s = .5$, $r = .3$, and varying choices of d_0. Bottom row: a similar construction but defining intensity using only the second order features.

by viewing the spin states σ as real vectors in $\mathbb{R}^m$. We generated $\mathcal{D}(G, \beta)$ for the three different types of graph G shown in Figure 4.7, but with different numbers of vertices. Specifically, we took an interval consisting of 30 sites, a circle with 30 sites, and a graph with three flares of length 14 each coming from the center, for a total of 43 vertices. For every one we chose $\beta = 1.5$, and $N = 20000$. In the case of the interval, the distribution of the energy values is given by

$$(a_k) = (4907, 7035, 4942, 2193, 709, 167, 41, 6, 0, ...)$$

where a_k is the number of states $\sigma \in \mathcal{D}$ with k transitions. For instance, we would have 4907 instances in which all spins are the same, and 6 instances in which there are 7 transitions. These numbers are consistent with the predicted values, which by a simple combinatorial argument are proportional to

$$a_k \sim \binom{m-1}{k} / 2^{m-1} e^{-2\beta k}$$

for the given parameters.

Applying density estimation to these spaces directly would be subject to the curse of dimensionality, and would not produce useful results. Instead of a dimensional reduction, we will consider a blended form of the data set using the left-normalized Laplacian operator $I - D^{-1}A$. Here A is the adjacency matrix of G, normalized so that the diagonal entry $A_{i,i}$ is the degree of v_i, and D is the row-sum of A. There are several reasons for this choice of diagonal in A, for instance a troublesome dependence of the eigenvalues of L^t on the parity of m

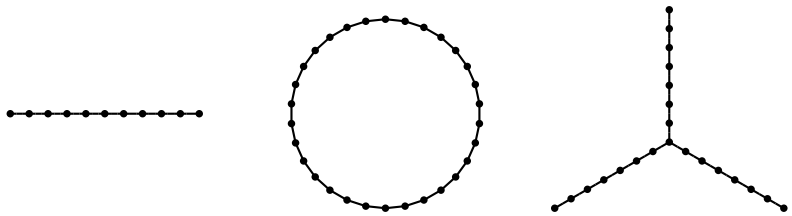


Figure 4.7: Top row: three different graphs G used to simulate the Ising model, denoted INT(11), CIRC(30), FLARES(22). The index corresponds to the number of vertices.

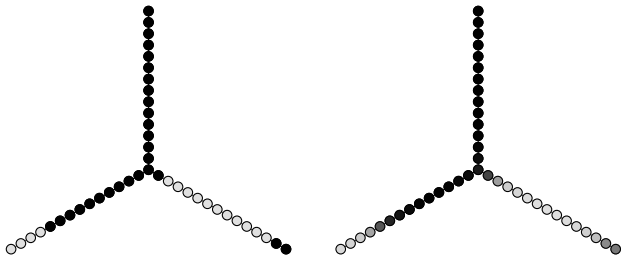


Figure 4.8: On the left: a typical data point in $\mathcal{D}(G, \beta)$ with three transitions for $(G, \beta) = (\text{FLARES}(43), 1.5)$. The spins ± 1 are represented by black and white. On the right: the same point after diffusion, i.e. right multiplication by $\exp(-tL^t)$ for $t = 10$.

when the diagonal entries are zero. In the case of the interval with 30 vertices, the first few values of these eigenvalues are

$$\Lambda = (1.000, .997, .988, .974, .954, .928, .898, .863, .824, .781, ...).$$

We then make the replacement $\mathcal{D}(G, \beta) \mapsto \mathcal{D}(G, \beta) \exp(-tL^t)$ with the value of $t = 10$, viewing the data set as an $N \times m$ matrix, to produce a continuous form of each data point as shown in Figure 4.8.

We then chose the density estimator $f(x)$ with scale parameter $h = 2.0$, and computed $Y = \text{DensWit}(f, \mathcal{D}(\beta, J)_{\geq .001}, .4)$ for each graph G up to the 3-simplices. In the case of the $G = \text{INT}(30)$, we obtain a filtered complex with sizes $(|Y_0|, |Y_1|, |Y_2|, |Y_3|) = (210, 2213, 6500, 8570)$. The persistent homology groups, shown in Figure 4.9, show the betti numbers of low energy states, which are the ones of higher density. For instance, the β_0 barcodes show two connected components at high density, corresponding to the two states in which all spins are the same.

Not surprisingly, if we then consider a smaller data of only those states σ with low energy, our data set becomes concentrated around a smaller-dimensional space, resulting in a smaller complex. This does not considerably affect the barcode diagram, as the current one is already essentially noiseless up to the chosen

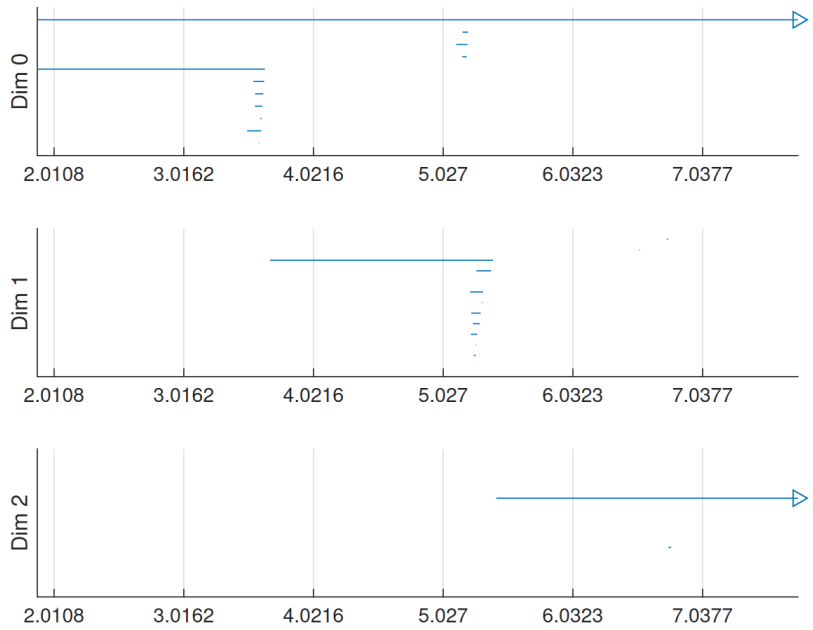


Figure 4.9: Persistent homology groups of DensWit$(f, \mathcal{D}(\text{INT}(30), 1.5)_{\geq .001}, .4)$, with scale parameter $h = 2.0$.

cutoff, but it does result in a smaller computation. Perhaps more importantly, the resulting complexes are more suitable for visual purposes. We computed the complex on the smaller data sets on states of up to 2 transitions, leading to a complex with sizes $(|Y_0|, |Y_1|, |Y_2|, |Y_3|) = (106, 554, 825, 491)$ in the case of the interval. We then projected the resulting one-skeleta onto $\mathbb{R}^3$ using the first three eigenvectors of the transposed Laplacian matrix, which are orthonormal with respect to the dot product weighted by the diagonal elements of D, followed by a random projection onto a 2-dimensional subspace. The results for all three types of graph, shown in Figure 4.10, show the geometry of the space of low energy configurations, with not necessarily obvious results. For instance, the edges of the cube in the case of the flares graph correspond to 6 dense states with exactly one transition, and 6 less dense states with two transitions, with one of them neighboring the center point of the graph.

4.4 Running time

In most examples, our complex was computed in a few seconds, and in all cases the largest time cost was evaluating the kernel density estimator, either at the reference set $\mathcal{A}$, or the witness set $\{q_\sigma\}$, whichever was bigger. In particular, it took longer than the computation of the alpha complex and its representatives. The cost of computing $f(x)$ could be decreased by choosing a covering of the

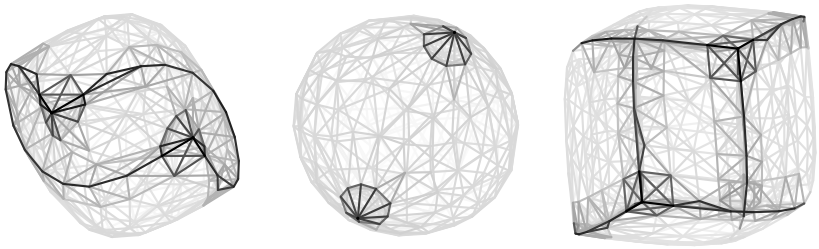


Figure 4.10: Low dimensional projections of of DensWit$(f, \mathcal{A}, s)$ using the graphs INT(30), CIRC(30), and FLARES(43), and only states with at most two transitions.

data set and ignoring the contribution from far away points $\|x_i - x\| > r$. We also note that both this step and the computation of the alpha complex may be done in parallel.

The most time consuming example was that of the Ising model from Section 4.3, which took several minutes, due to the larger size of the data set, the fact that we computed up to the 3-simplices, and because the data was embedded in $\mathbb{R}^{30}$. Because our algorithm for computing the alpha complex is based on dual programming and therefore takes only the dot products $z_i \cdot z_j$ and powers p_i as input, its running time is not directly affected by the higher dimensionality. However, the running time of evaluating $f(x)$ scales with dimension simply because it requires more operations compute the distance.

5 Conclusions and future directions

In this paper we defined a filtered simplicial complex associated to a Gaussian kernel density estimator, which we illustrated through persistent homology calculations and by viewing the complex in low dimensions in several examples. We conclude with a some potential extensions and future directions.

- Zeroth degree persistent homology group were shown to be a valuable way of viewing clustering in [8]. In clustering applications, our algorithm would not need to solve any quadratic programs, only to test when the (power-shifted) midpoints between two landmark points have those points as their nearest neighbors. This also results in a considerably smaller number of points on which to evaluate $f(x)$.

- We expect that the resulting clustering algorithm would be a strong candidate for the Mapper Algorithm [22], which combines the outputs of a given clustering algorithm over multiple overlapping intervals or other types of covering, through a filter function. One reason is that the density-based approach is not sensitive to outliers or small changes in the data. Another is that other than perhaps the scale parameter h, our construction has no

tuning parameters which could require different choices for different intervals. Enlarging the other parameters $(\mathcal{A}, s)$ leads to a finer approximation, but not a different target.

- It is often necessary to consider data sets with varying metric. If $\mathcal{D}$ is partitioned into groups associated to different quadratic forms, one may use a combined complex by solving a quadratic program over the intersection of cells defined in different power diagrams. In another direction, recall from Section 2.1 that our density estimators do not included a volume normalizing factor of $(1/\sqrt{2\pi})^m$. As a result, we have that $f(x) \leq f'(x)$ when f, f' are density estimators for the same data set with scale parameters $h \leq h'$, leading to a multidimensional persistence setting [6, 20, 23].

- It would be interesting to determine to what extent Lemma 3.1 applies to diffusion-based density estimators in manifolds M other than $\mathbb{R}^m$, such as the examples $\mathcal{D} \subset S^4$ of Section 4.2. At a minimum, it is clear that the conclusions of the Lemma would hold when M is the product of a vector space and a compact torus, by interpreting a data set $\mathcal{D} \subset \mathbb{R}^m \times (S^1)^{m'}$ as a periodically repeating one in $\mathbb{R}^{m+m'}$. In the case of a circle, $f(x)$ would take the form of an infinite periodic sum of Gaussian kernels on $\mathbb{R}$, or equivalently as a finite sum of Jacobi theta functions defined on the circle.

- In a further extension, our construction applies to the convolution of any distribution by the diffusion process, not just a discrete one coming from a data set. It would be interesting to consider a distribution defined in terms of Fourier modes on the torus $(S^1)^m$, in which the diffusion operator acts diagonally. It then becomes a calculus problem to represent the Gaussian fit function in that basis.

References

[1] Henry Adams, Andrew Tausz, and Mikael Vejdemo-Johansson. JavaPlex: A research software package for persistent (co) homology. In *International congress on mathematical software*, pages 129–136. Springer, 2014.

[2] Daniel Arnström, Alberto Bemporad, and Daniel Axehill. A dual active-set solver for embedded quadratic programming using recursive ldl^T updates. *IEEE Transactions on Automatic Control*, 67(8):4362–4369, 2022.

[3] F. Aurenhammer and H. Edelsbrunner. An optimal algorithm for constructing the weighted voronoi diagram in the plane. *Pattern Recognition*, 17(2):251–257, 1984.

[4] A. Azzalini and A. W. Bowman. A look at some data on the old faithful geyser. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 39(3):357–365, 1990.

[5] U. Bauer, M. Kerber, F. Roll, and A. Rolle. A unified view on the functorial nerve theorem and its variations. *arXiv preprint arXiv:2203.03571*, 2022.

[6] Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete and Computational Geometry*, 42:71–93, 06 2007.

[7] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas Guibas, and Steve Oudot. Proximity of persistence modules and their diagrams. *Proc. 25th ACM Sympos. Comput. Geom.*, 12 2008.

[8] Frédéric Chazal, Leonidas Guibas, Steve Oudot, and Primoz Skraba. Persistence-based clustering in riemannian manifolds. *Journal of the ACM*, 60, 06 2011.

[9] F.R. Cohen. On configuration spaces, their homology, and lie algebras. *Journal of Pure and Applied Algebra*, 100(1):19–42, 1995.

[10] Vin de Silva. A weak characterisation of the delaunay triangulation. geom. dedic. 135(1), 39-64. *Geom. Dedicata*, 135:39–64, 08 2008.

[11] Li Deng. The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.

[12] H. Edelsbrunner. The union of balls and its dual shape. *Discrete and Computational Geometry*, pages 415–440, 1995.

[13] H. Edelsbrunner, B.T. Fasy, and Rote. Add isotropic gaussian kernels at own risk: More and more resilient modes in higher dimensions. *G. Discrete Comput Geom*, pages 797–822, 2013.

[14] H. Edelsbrunner and E. Mücke. Three-dimensional alpha shapes. *ACM Trans. Graph.*, 13(1), 1994.

[15] Herbert Edelsbrunner and John Harer. *Computational Topology - an Introduction.* American Mathematical Society, 2010.

[16] Carlsson G., T. Ishkhanov, V. de Silva, and A. Zomorodian. On the local behavior of spaces of natural images. *Int J Comput Vis*, 76:1–12, 2008.

[17] J. H. van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings: Biological Sciences*, 265(1394):359–366, Mar 1998.

[18] E. Ising. Beitrag zur theorie des ferromagnetismus. *Z. Physik*, 31:253–258, 1925.

[19] A.B. Lee, K.S. Pedersen, and D. Mumford. The nonlinear statistics of high-contrast patches in natural images. *International Journal of Computer Vision*, pages 83–103, 2003.

[20] Michael Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Foundations of Computational Mathematics*, 15, 06 2011.

[21] Steven J. Phillips, Robert P. Anderson, and Robert E. Schapire. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3):231–259, 2006.

[22] Gurjeet Singh, Facundo Memoli, and Gunnar Carlsson. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007.

[23] The RIVET Developers. Rivet, 2020.