

Alpha shapes in kernel density estimation

Erik Carlsson and John Carlsson

March 28, 2024

Abstract

A foundational problem in topology and data is to determine the topological type (e.g. the persistent homology groups) of the superlevel sets $\mathcal{L}(a) = f^{-1}[e^{-a}, \infty)$ of a sum of Gaussian kernels $f(x) = \sum_i a_i \exp(-\|x - x_i\|^2/2h^2)$ for $\{x_i\} \subset \mathbb{R}^d$. In this paper, we show that each $\mathcal{L}(a)$ coincides with the union of a certain power-shifted covering by balls, whose centers range over a closed subspace of the convex hull $\mathcal{S}(a) \subset \text{conv}(\{x_i\})$. We then present an explicit homotopy equivalence $p : \mathcal{L}(a) \rightarrow \mathcal{S}(a)$, realizing $\mathcal{S}(a)$ as a continuous version of the alpha shape. This leads to a prescription for modeling noisy point clouds by density-weighted alpha complexes which, in addition to computing persistent homology, give rise to refined geometric models. In order to compute alpha complexes in higher dimension, we used a recent algorithm due to the present authors based on the duality principle [10].

1 Introduction

Let $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ be a sum of Gaussian kernels with uniform covariance matrices, which after a linear change of coordinates takes the form

$$f(x) = \sum_{i=1}^N a_i \exp(-\|x - x_i\|^2/2h^2), \quad (1.1)$$

where $a_i > 0$, and $h > 0$ is the bandwidth, or scale parameter. We will denote the superlevel sets of f by

$$\mathcal{L}(a) = f^{-1}[e^{-a}, \infty) = \{x \in \mathbb{R}^d : f(x) \geq e^{-a}\}$$

using a in negative log coordinates so that the family of subspaces is increasing. The following problem is well-studied [16, 35, 9, 25]:

Problem 1. Determine the topological type (for instance, the persistent homology groups) of $\mathcal{L}(a)$.

One of the reasons to study this problem is that filtering by density gives a rigorous alternative to simply removing outliers or noise by a subjective criteria. For instance, the degree zero persistent homology can be taken as a precise definition of hierarchical clusters [17]. Another is that it would encode the landscape of a density function, which measures how the shapes change with a minimum density threshold.

However, it has been proved difficult to compute. To illustrate some of the difficulties that arise, consider the following simple algorithm:

Algorithm 1.1. Approximate the persistent homology groups of $\mathcal{L}(a)$.

1. Sample M points $\{\tilde{x}_1, \dots, \tilde{x}_M\}$ from the underlying distribution of $f(x)$, assumed to be sorted in increasing order of $-\log(f(\tilde{x}_i))$, so that $\tilde{x}_1$ is the densest.
2. Initialize $S = \emptyset$. For each $\tilde{x}_i$ in order, make the replacement $S \mapsto S \cup \{\tilde{x}_i\}$ if $\tilde{x}_i$ is at least distance ϵ to all existing points of S , for some predetermined choice of $\epsilon > 0$.
3. Construct a filtered family of simplicial complexes $X(a)$ using a general methods, such as weighted Vietoris-Rips, on the respective vertex sets given by $S \cap \mathcal{L}(a)$.
4. Calculate persistent homology.

Algorithm 1.1 seems intuitive, but it has undesirable properties. One is that while the homotopy type of the super-level sets $\mathcal{L}(a)$ is a function only of the pairwise distances $\|x_i - x_j\|$, the expected size of the vertex set S grows rapidly by simply adding coordinates of zeroes to the end of every x_i , as it is proportional to the covering number super-level set $\mathcal{L}(a)$ by ϵ -balls. Another issue is that the expected density of a sample point $f(\tilde{x}_i)$, which is used to sort the points and filter homology, becomes small and dominated by noise for $d \gg 0$.

An alternate approach to producing a finite vertex set is simply to discretize space, so that one can apply cubical homology, but this is clearly only

possible in very low dimensions. Another is to enumerate the critical points of $f(x)$ and compute the discrete Morse complex [26, 37], but this can in general lead to a combinatorial explosion in the number of points, including the case of sums of Gaussian kernels [21]. An intuitive fix is to use something similar to Algorithm 1.1, but using the data set $\{x_1, \dots, x_N\}$ itself in place of the samples $\{\tilde{x}_1, \dots, \tilde{x}_M\}$ in item 1, so that the final answer is technically independent of the dimension of the embedding. However, this is not a stable solution; as long as the underlying density that produced the $\{x_i\}$ is nonzero on all of $\mathbb{R}^d$, a large enough sample will eventually fill out space, so that we end up with the same poor scaling with the covering number as in the original algorithm.

Once a density-weighted vertex set S has been chosen, there are a number of persistence constructions that are robust with respect to noise, as well as statistically rigorous results about their output. Some of these methods include explicitly removing those points, and calculating distance-based complexes such as Vietoris-Rips or witness [39]; the aforementioned applications of zeroth-dimensional persistence to clustering [17], in which one scans density values over connecting edges; the distance to measure filtration, which is stable with respect to the 2-Wasserstein distance [15]; Other constructions which study rigorous properties of persistent homology, in the specific context of kernel density estimators [35, 7]; persistence landscapes, and studies of spaces of persistence diagrams [2, 9, 5]; and multidimensional persistence, which simultaneously filters by density and scale [12, 11, 33, 6].

1.1 Proposed method

Our method begins by replacing f by an modified function $\tilde{f}$, whose sublevel sets are topologically equivalent but better behaved. Let $f(x)$ be as in (1.1), and consider the transformed function $\tilde{f} : \mathbb{R}^d \rightarrow \mathbb{R}_+$ given by

$$\tilde{f}(y) = \inf_{x \in \mathbb{R}^d} f(x) \exp(\|x - y\|^2/2h^2). \quad (1.2)$$

By taking $x = y$, it is obvious that $\tilde{f}(y) \leq f(y)$, but it is not immediately clear that $\tilde{f}$ is even nonzero. However, we show (Proposition 2) that one can recover the original function by the formula

$$f(x) = \sup_{y \in \mathbb{R}^d} \tilde{f}(y) \exp(-\|x - y\|^2/2h^2) \iff$$

$$-\log(f(x)) = \inf_{y \in \mathcal{S}} (\|x - y\|^2/2h^2 + \alpha(y)) \quad (1.3)$$

where $\alpha(y) = -\log(\tilde{f}(y))$, and $\mathcal{S} \subset \mathbb{R}^d$ is the domain on which $\tilde{f}$ is nonzero. This is seen this by first showing that for any $f(x)$, the modified function $-h^2 \log(f(x)) + \|x\|^2/2$ is convex. Equation (1.3) then follows from the Fenchel–Moreau theorem, which states that any convex function satisfies $F = F^{**}$, where $F^* = \sup_{y \in \mathbb{R}^d} (x \cdot y - F(y))$ is the Legendre transform.

We now observe that the last expression in (1.3) expresses $\mathcal{L}(a)$ as the union of an infinite covering by closed balls,

$$\mathcal{L}(a) = \bigcup_{y \in \mathcal{S}(a)} B_r(y), \quad r = \sqrt{2h^2(a - \alpha(y))} \quad (1.4)$$

where $\mathcal{S}(a) = \alpha^{-1}(-\infty, a] \subset \mathcal{S}$ is the sublevel set. From this point of view, Algorithm 1.1 is backwards because it is selecting the vertices of a simplicial complex from the union of a cover, instead of selecting a subcover and taking the nerve. In the case of (1.4), a subcover associated to a finite subset $S \subset \mathcal{S}$ is known as a *power diagram* with power map $p = -2h^2\alpha$. Its nerve is the alpha complex, whose geometric realization the alpha shape [20, 22]. Alpha complexes have a number of theoretical advantages beyond persistent homology, namely that they are minimal in size, naturally embedded in space, and can be used to generate to beautiful geometric models [23].

Our main theorem states that the $\mathcal{S}(a)$ are continuous versions of alpha shapes. Theoretically speaking, one useful property of this shape is that $\mathcal{S}(a)$ is coordinate-free, in the sense that it is covariant under linear changes of coordinates (including extra coordinates); practically speaking, as our experiments will later demonstrate, this represents an attempt to address the curse of dimensionality by “projecting” noise introduced by the kernel density estimator back into the core “shape” of the original dataset.

Theorem A. *We have*

1. *The total space $\mathcal{S}$ is an open subset of the convex hull of the Gaussian centers $\mathcal{D} = \{x_i\}$, and each $\mathcal{S}(a)$ is a closed subset of $\mathcal{L}(a) \cap \mathcal{S}$.*
2. *We have a surjective map $p : \mathcal{L}(a) \rightarrow \mathcal{S}(a)$, which takes the form of an expectation:*

$$p(x) = \frac{\sum_i \exp(-\|x - x_i\|^2/2h^2) x_i}{\sum_i \exp(-\|x - x_i\|^2/2h^2)} \quad (1.5)$$

3. *The inclusion map $i : \mathcal{S}(a) \hookrightarrow \mathcal{L}(a)$ induces a homotopy equivalence with homotopy inverse p . If additionally the $\{x_i\}$ affinely span $\mathbb{R}^d$, then p is a homeomorphism.*

In Section 3.2, we use Theorem A to modify Algorithm 1.1 so that it selects vertices from $\mathcal{S}(a)$ instead of $\mathcal{L}(a)$, which by item 3 has the same homotopy type. Due to the explicit form of the inverse homotopy from item 2, we may generate a sample from $\mathcal{S}$ by sampling $\tilde{x}_i$ as usual, and setting $\tilde{y}_i = p(\tilde{x}_i)$. It follows from basic properties of Gaussians that the coefficients in (1.5) are independent of the embedding $\mathcal{D} \subset \mathbb{R}^d$, and therefore so is the sampling procedure. Now the potential number of vertices is only determined by the ϵ -covering number of $\mathcal{S}(a)$, which (unlike the covering number of $\mathcal{L}(a)$) has no dependence on the embedding dimension, because $\mathcal{S}(a)$ is contained in the convex hull by part 1. We then build the alpha complex of the corresponding shape instead of a general construction such as Vietoris-Rips. An illustration of the results of is shown in the case of spatial density estimation in Figure 1.1.1.

1.2 Examples

In Section 4, we present several constructions using this method, many using the Metropolis algorithm to generate point clouds. One example is an interesting energy landscape, whose persistent homology reveals a filtered version of a configuration space. Another uses connected components to reveal local basins in a loss functions in a nonlinear regression problem [38], using Bayesian sampling. Also within Bayesian models, we obtain a geometric representation of a simple singular model by two univariate Gaussian mixtures [40]. These two examples illustrate how a varying metric may be replaced by the Euclidean metric with more coordinates, as well as our assertions that our constructions are stable in higher dimension. The next example follows a similar setup as in as [28], in which the authors used the witness complex and persistent homology to detect subspaces of the Klein bottle in random 3×3 image patches from the van Hateren-Schaaf natural image data set [29]. Instead of calculating homology groups, we use alpha shapes to generate precise geometric models from a different type of patch taken from the MNIST data set of hand-drawn digits. In the last one we model energy landscapes of discrete simulations from the graphical Ising model, using spectral bases on the underlying graph to generating appropriate Euclidean embeddings.

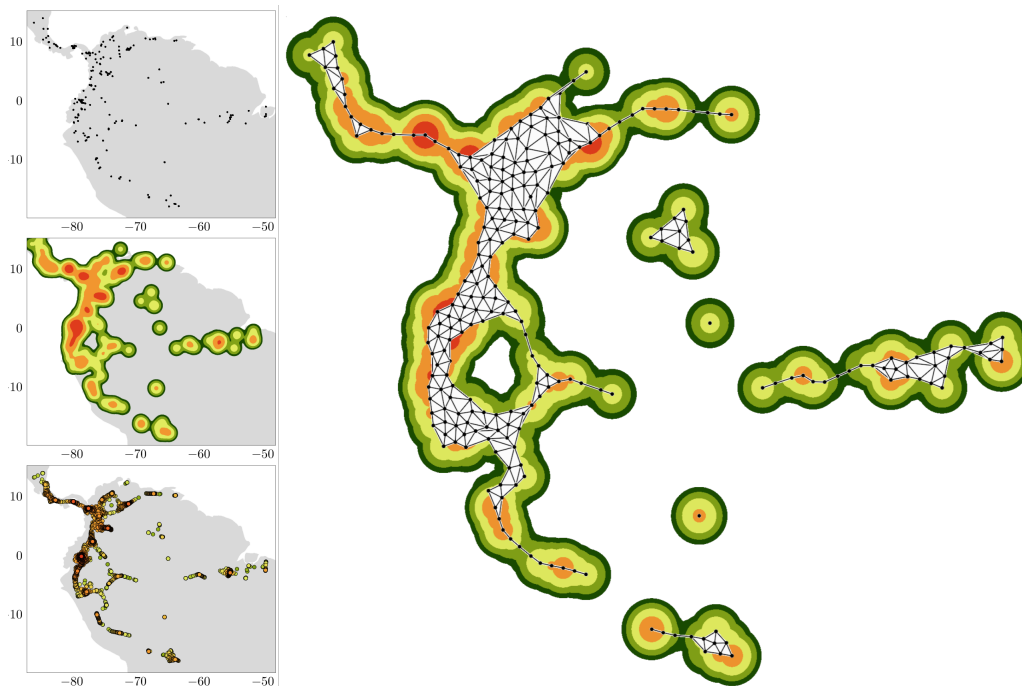


Figure 1.1.1: An illustration of the pipeline from Section 3.2. In the top left, we have a well-known geographical data set from [36]. Below that we have a heat map of a kernel density estimator with a certain scale parameter, and the result of sampling many points $\tilde{y}_i = p(\tilde{x}_i)$ from the shape, colored by the value of $\alpha(\tilde{y}_i)$. On the right, we have the resulting alpha complex and its shape, contained in the total space of a power diagram, which closely resembles the density superlevel set by the sampling algorithm.

In order to compute alpha complexes, we used a recent algorithm due to the present authors [10], based on the duality principle in optimization. Unlike other mainstream constructions, that algorithm is suitable in higher dimension because the dual variables are a function only of pairwise dot products, and so have no explicit dependence on the embedding dimension d . In every example in Section 4, computing the alpha complex took at most a few seconds. On the other hand, the higher dimensional examples could not have been calculated using standard methods, which almost always begin by computing the full (shifted) Delaunay triangulation, which can be astronomically large.

All our code for this paper was written in MAPLE, and is available on

the first author's website.

1.3 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 for kernel density estimation and computational topology, including the power diagram and alpha complexes.

2.1 Kernel density estimators

Let $\mathcal{D} = \{x_1, \dots, x_N\} \subset \mathbb{R}^d$ be a point cloud. A Gaussian kernel density estimator is a sum of the form

$$f(x) = \sum_{i=1}^N a_i K_h(x - y), \quad K_h(v) = \exp(-\|v\|^2/2), \quad (2.1)$$

for $a_i > 0$. We will be interested in the superlevel sets

$$\mathcal{L}(a) = f^{-1}[e^{-a}, \infty) = \{x : f(x) \geq e^{-a}\} \quad (2.2)$$

For simplicity, we consider only finite sums, but our results apply to the convolution of more general distributions by Gaussian kernels. We will assume the the norm is always the standard L^2 -norm, as any other quadratic form can be transformed in that way by a linear change of coordinates. Moving (anisotropic) metrics are an interesting extension, which we hope to study in future papers. For now, we remark that general Riemannian metrics can often be approximated by Euclidean ones in higher dimensions, for instance using spectral embeddings, which will be used in Section 4. A more general setup might involve replacing (2.1) by the convolution of a distribution on a Riemannian manifold by a heat kernel with respect to the metric.

2.2 Computational topology

By a simplicial complex on a vertex set S , we will mean a collection of nonempty subsets of S that is closed under taking nonempty subsets. If the vertex set comes equipped with a map to a vector space, for instance if it is described as an explicit subset $S = \{p_1, \dots, p_n\} \subset \mathbb{R}^d$, then we have its geometric realization defined by

$$|X| = \bigcup_{\sigma \in X} \text{conv}(\sigma), \quad (2.3)$$

which is the union of the convex hulls of the vertices of all simplices, i.e.

$$\text{conv}(\{x_i\}) = \left\{ \sum_i c_i x_i : c_i \geq 0, \sum_i c_i = 1 \right\} \quad (2.4)$$

The affine span, or affine hull $\text{aff}(\{x_i\})$ is the smallest affine subspace containing it, which is the same as the convex hull but without the condition that $c_i \geq 0$.

If $\mathcal{U} = \{U_x : x \in S\}$ is a collection of (closed or open) subsets of $\mathbb{R}^d$, then the nerve of $\mathcal{U}$ is the complex

$$\text{Nrv}(\mathcal{U}) = \{\sigma \subset S : U_{\sigma_0} \cap \dots \cap U_{\sigma_k} \neq \emptyset\}. \quad (2.5)$$

Generally speaking, nerve theorems state that if $\mathcal{U}$ satisfies certain conditions, for instance if every k -fold union is contractible, then the nerve complex is homotopy equivalent to the union $\bigcup \mathcal{U} = \bigcup_{x \in S} U_x$. [32, 27, 8]. Different versions involve different realizations of the nerve as a topological space. In one version, suppose that each U_x is convex, and choose representatives $x_\sigma \in U_{\sigma_0} \cap \dots \cap U_{\sigma_k}$. Then as in [4], we have a linear map $\Gamma : |\text{Sd}(X)| \rightarrow \mathbb{R}^d$ on the barycentric subdivision of the nerve, whose value on each vertex σ is x_σ . By convexity, it is clear that its image is contained in $|\mathcal{U}|$. Theorem 3.1 from that reference states:

Proposition 1. *If $\mathcal{U}$ is convex then Γ is a homotopy equivalence, specifically the one from the nerve theorem.*

Beyond giving an explicit geometric realization of the nerve, this version gives an explicit form of the map that induces the nerve isomorphism.

In persistent homology, we will also have filtered families of simplicial complexes.

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

One source of filtered complexes comes from taking the nerve of a nested family of covers $\mathcal{U}(a) = \{U_i(a)\}$, where $U_i(a) \subset U_i(b)$ for $a \leq b$. For instance, the Čech and alpha complexes arise in this way.

For any filtered complex, we may compute its persistent homology groups [13, 11, 34, 24]. All barcode diagrams generated for this paper were calculated using `javaplex` [1].

2.3 Alpha complexes

For a reference on this section, we refer to [3, 20, 24]. Let $S = \{p_1, \dots, p_n\} \subset \mathbb{R}^d$ be a collection of points, and let $\pi : S \rightarrow \mathbb{R}$ be a function with values $\pi(p_i) = \pi_i$, called the powers. We have the weight map $w = w_{S,\pi} : \mathbb{R}^d \rightarrow \mathbb{R}$ defined by

$$w(x) = \min_{p_i \in S} w_i(x), \quad w_i(x) = \|x - p_i\|^2 - \pi_i. \quad (2.6)$$

Then we can define a family of covers $\mathcal{U} = \mathcal{U}_{S,\pi}$ by $\mathcal{U}(a) = \{U_i(a)\}$ for each $a \in \mathbb{R}$, where

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

Then

$$w^{-1}(-\infty, a] = \bigcup \mathcal{U}(a) \quad (2.7)$$

Definition 2. The power diagram $\mathcal{V} = \mathcal{V}_{S,\pi}$ associated to (S, π) is the collection of closed regions $\mathcal{V}(a) = \{V_i \cap U_i(a)\}$ where

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

When $\pi_i = 0$ for all i , the V_i are the cells of the usual Voronoi diagram. More generally, the regions are still determined by linear inequalities, in other words are separated by hyperplanes. In fact, for $\pi_i \geq 0$ they arise the intersection of true Voronoi diagrams with a linear subspace, with the π_i representing the negative squared normal distances.

Definition 3. The weighted alpha complex is the nerve of the power diagram $X(a) = \text{Nrv}(\mathcal{V}(a))$. The alpha shape is the geometric realization $|X(a)| \subset \mathbb{R}^d$, defined by identifying the vertices with $S \subset \mathbb{R}^d$.

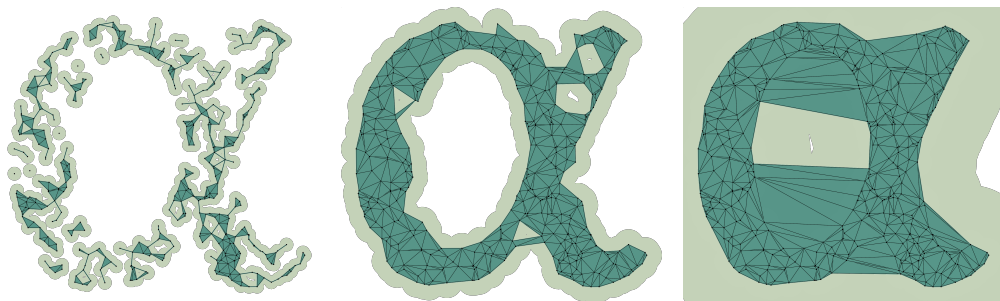


Figure 2.3.1: A sequence of power diagrams with their corresponding alpha shapes.

Written as a pair (X, w) , the full alpha complex X is the nerve of the shifted Voronoi diagram $\{V_i\}$, and

$$w(\sigma) = \inf_{x \in V_\sigma} w(x) \quad (2.8)$$

It follows from the nerve theorem that $X(a)$ is homotopy equivalent to $\bigcup \mathcal{V}(a)$. In [20], Edelsbrunner proved that $|X(a)| \subset \bigcup \mathcal{V}(a)$ is a homotopy equivalence with an explicit deformation retraction. Figure 2.3.1 illustrates the union of the cover and the corresponding shapes in the unweighted case.

3 Alpha shapes of Gaussian KDE's

We present our main theorem, and our proposed algorithm for sampling finite alpha complexes.

3.1 Main theorem

Let $\mathcal{D} = \{x_1, \dots, x_N\} \subset \mathbb{R}^d$, and consider a sum $f(x)$ of Gaussian kernels as in (2.1). In order to define our main construction, consider the following function:

$$\tilde{f}(y) = \inf_{x \in \mathbb{R}^d} f(x) K_h(x - y)^{-1} \quad (3.1)$$

While it is not obvious, this turns out to be a sort of “transform” of functions of the same form as f , and in fact we can recover f from $\tilde{f}$:

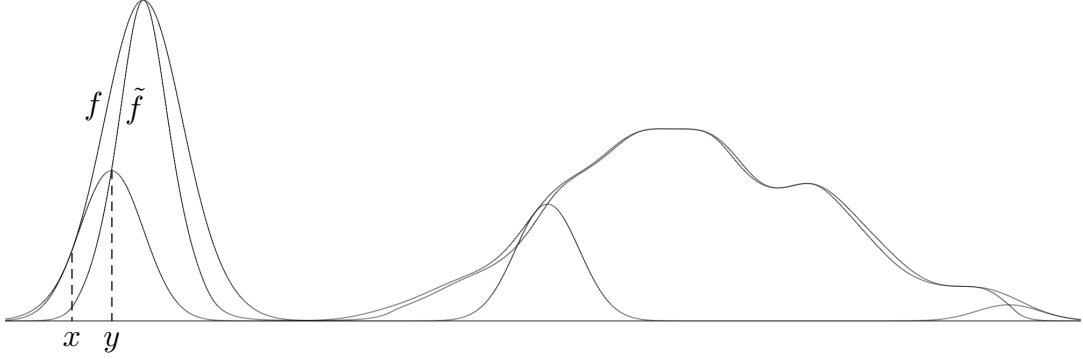


Figure 3.1.1: A univariate Gaussian KDE f shown as the top curve, with $\tilde{f}$ shown as the bottom curve. The Gaussians all have the same scale but different scalar multipliers and centers. Proposition 2 shows that f and $\tilde{f}$ determine each other, and that $y = p(x)$ for any pair (x, y) .

Proposition 2. *We have*

$$f(x) = \sup_{y \in \mathbb{R}^d} \tilde{f}(y) K_h(x - y). \quad (3.2)$$

Moreover, for each x , the supremum in (3.2) is obtained at a unique value $y = p(x)$ where

$$p(x) = \frac{\sum_{i=1}^N a_i K_h(x - x_i) x_i}{\sum_{i=1}^N a_i K_h(x - x_i)} \quad (3.3)$$

In other words, $p(x)$ is the mean of a probability distribution μ_x on $\mathbb{R}^d$ which is supported at the points of S , with normalized weights proportional to $a_i K_h(x - x_i)$. An illustration of the proposition is shown in Figure 3.1.1.

Proof. We may assume that $h = 1$. We show that the auxiliary function

$$F(x) = -\log(f(x)) + \|x\|^2/2$$

is convex. The first statement then easily follows from the fact that $F^{**} = F$, where

$$F^*(y) = \sup_x (x \cdot y - F(x))$$

is the Legendre transform.

To see the convexity, it suffices to prove the convexity of the one-dimensional function $F(\varphi(t))$ where $\varphi(t) = x + tv$ is a path with $\|v\| = 1$. But the restriction of any sum of Gaussians to an affine subspace is just another sum of Gaussians:

$$f(\varphi(t)) = \sum_{i=1}^N b_i \exp(-(t - t_i)^2/2) \quad (3.4)$$

where t_i is the coordinate of the orthogonal projection $\bar{x}_i = \varphi(t_i)$ of x_i , and $b_i = a_i \exp(-\|x_i - \bar{x}_i\|^2/2)$. Thus, we have reduced the problem to the one-dimensional case.

For the one-dimensional case, it suffices to show that the second derivative is nonnegative. We check

$$f'' = \left(\sum_i a_i (x - x_i)^2 e^{-(x - x_i)^2/2} \right) - f.$$

Now write

$$F'' = \frac{f''}{f} - \left(\frac{f'}{f} \right)^2 + 1 = E_\rho[(X_i - x)^2] - (E_\rho[X_i - x])^2$$

where the expectations are over the random variables $X_i - x$ defined on the finite probability distribution ρ from (3.3). Since this is the expression variance, we obtain the desired nonnegativity.

For the second statement, suppose y is any value that obtains the supremum in (3.2) for a given x , and let $g(z) = c \exp(-\|z - y\|^2/2)$ be the corresponding Gaussian centered at y with $c = \tilde{f}(y)$. Then we have that g agrees with f to first order at x :

$$g(x) = f(x), \quad \nabla g(x) = \nabla f(x). \quad (3.5)$$

Dividing the second second expression by the first and solving for y , we obtain (3.3), which also establishes the uniqueness. $\square$

Thanks to Greg Kuperberg, who proposed using the Legendre transform to simplify an earlier version of this argument.

We then make the following definition, which actually applies to the convolution of any distribution with a Gaussian kernel, not just discrete ones:

Definition 4. The alpha shape of the kernel density estimator $f(x)$ is the filtered family of regions

$$\mathcal{S}(a) = \left\{ x : \tilde{f}(x) > e^{-a} \right\} \quad (3.6)$$

In other words, $\mathcal{S}(a) = \alpha^{-1}(-\infty, a]$ where $\mathcal{S}$ consists of all x for which $\tilde{f}(x) \neq 0$, and $\alpha : \mathcal{S} \rightarrow \mathbb{R}$ is the function $\alpha(x) = -\log(\tilde{f}(x))$.

We now have our main theorem, which roughly speaking says that $\mathcal{S}(a)$ is a continuous version of the alpha shape.

Theorem 1. *Let $\mathcal{L}(a)$ be the superlevel set of a sum of Gaussians $f(x)$ centered at $\mathcal{D} = \{x_i\}$ as in (2.1). Let $(\mathcal{S}, \alpha)$ and $p : \mathbb{R}^d \rightarrow \mathcal{S}$ be as above, with $\mathcal{S}(a) = \alpha^{-1}(-\infty, a]$.*

1. *We have that $\mathcal{L}(a) = \bigcup \mathcal{U}_{\mathcal{S}, \pi}(2h^2a)$, where $\mathcal{U}_{\mathcal{S}, \pi}$ is the filtered covering by weighted balls with vertex set $\mathcal{S}$ and power map $\pi = -2h^2\alpha$. In other words, $-2h^2\log(f) = w_{\mathcal{S}, \pi}$, using an infimum in (2.6).*
2. *The shape is an open subset of the convex hull $\mathcal{S} \subset \text{conv}(\mathcal{D})$ on which α is a continuous map. Each $\mathcal{S}(a)$ is a (closed) subset of $\mathcal{L}(a) \cap \mathcal{S}$, and the restriction $p|_{\mathcal{L}(a)} : \mathcal{L}(a) \rightarrow \mathcal{S}(a)$ is surjective.*
3. *For any a we have $\mathcal{S}(a) \subset \mathcal{L}(a)$, and the inclusion map induces a homotopy equivalence with homotopy inverse $p|_{\mathcal{L}(a)}$. If $\text{aff}(\mathcal{D})$ is all of $\mathbb{R}^d$, then p is a homeomorphism.*

We start with the final statement.

Lemma 1. *Suppose the affine span of $\mathcal{D}$ is all of $\mathbb{R}^d$. Then the restriction of p to the level set gives a homeomorphism $\mathcal{L}(a) \sim \mathcal{S}(a)$.*

Proof. Proposition 2 implies that p is surjective, so we must check that it is also injective, and that its inverse is continuous. We may again assume that $h = 1$.

Starting with the one-dimensional case, it suffices to check that the derivative $p'(x)$ is positive, so that p is increasing and has a continuous inverse by the inverse function theorem. Using the quotient rule, we find that the numerator in the expression for p' is

$$\sum_{1 \leq i < j \leq N} a_i a_j (x_i - x_j)^2 \exp((x - x_i)(x - x_j) - (x_i - x_j)^2/2)$$

which is positive. The denominator is f^2 which is also positive.

In the general case, suppose that $p(x) = p(x')$ for $x \neq x' \in \mathbb{R}^d$. Then the restriction of f to the line $\varphi(t) = x + tv$ for $v = (x' - x)/\|x - x'\|$ is the one-dimensional sum of Gaussian kernels in (3.4). Then we have

$$(p(x + tv) - p(x)) \cdot v = \frac{\sum_i b_i t_i \exp(-(t - t_i)^2/2)}{\sum_i b_i \exp(-(t - t_i)^2/2)} \quad (3.7)$$

The t_i must also affinely span $\mathbb{R}$ (which just means they are not all the same point), so taking $t = \|x' - x\|$ contradicts the one-dimensional case.

Therefore p is bijective, and it remains to show that its inverse is continuous. For this, we have that $v^t J_p(x) v$ is the derivative of the right hand side of (3.7) at $t = 0$, where $J_p(x)$ is the Jacobian matrix. Then using the first paragraph, we find that the Jacobian is positive definite for all x , so p is locally invertible by a continuous function by the inverse function theorem. Since p is globally invertible, its inverse must agree with each local inverse, so p^{-1} is continuous. $\square$

We can now prove Theorem 1.

Proof. Part 1 follows immediately from Proposition 2.

Suppose that the affine span of $\mathcal{D}$ is a lower-dimensional affine subspace $\text{aff}(\mathcal{D}) \subset \mathbb{R}^d$. Then the superlevel sets of the KDE associated to the restriction of f to $\text{aff}(\mathcal{D})$ are homotopy equivalent to those of f . Moreover, p factors as the orthogonal projection onto $\text{aff}(\mathcal{D})$ composed with a map p' on $\text{aff}(\mathcal{D})$ that takes the same form as p . Thus, the remaining statements are reduced to the case in which $\mathcal{D}$ affinely spans all of $\mathbb{R}^d$.

In the case that $\mathcal{D}$ spans, we have that p is a homeomorphism by Lemma 1. The statement that $\mathcal{S} \subset \text{conv}(\mathcal{D})$ follows from the explicit form of p . The statement that α is continuous follows from the expression

$$\alpha(y) = -\log(f(p^{-1}(y))) + \|y - p^{-1}(y)\|^2/2h^2.$$

The statement that $\mathcal{S}$ is open in the convex hull follows since it is the union of the open sublevel sets.

The only remaining statement is that the inclusion map $i : \mathcal{S}(a) \subset \mathcal{L}(a)$ is homotopic to p^{-1} , or equivalently that $ip : \mathcal{L}(a) \rightarrow \mathcal{L}(a)$ is homotopic to the identity map. We define a family $h_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ by

$$h_t(x) = tx + (1 - t)p(x).$$

Since $p(x)$ sends x to the center of a closed ball $U_y \subset \mathcal{S}(a)$ contacting x along the boundary, it follows from part 1 that h_t carries each $\mathcal{L}(a)$ into itself, and therefore gives the desired homotopy. $\square$

Remark 1. In [20], Edelsbrunner gave an explicit deformation retraction of the union of balls onto the shape. However, while $\mathcal{S}(a) \subset \mathcal{L}(a)$ induces a homotopy equivalence, the reverse map p does not act as the identity on $\mathcal{S}(a)$, so it is not a deformation retraction.

3.2 Subsampling and finite alpha complexes

We now explain our method for choosing finite subsets $S \subset \mathcal{S}$, corresponding to finite subcovers. Taking the nerve, we obtain finite alpha complexes filtered by the negative log of density.

Algorithm 3.1. Generate a finite alpha complex from $\mathcal{S}$.

1. Sample $\{\tilde{x}_1, \dots, \tilde{x}_M\}$ from the underlying distribution of $f(x)$.
2. Let $\tilde{y}_i = p(\tilde{x}_i)$, where $p : \mathbb{R}^d \rightarrow \mathcal{S}$ is the map from Theorem 1, and suppose $\{\tilde{y}_1, \dots, \tilde{y}_M\}$ are sorted in increasing order of $\alpha(\tilde{y}_i)$.
3. Generate a vertex set as follows:
 - (a) Initialize $S = \emptyset$. Fix $0 < s < 1$, and let $\epsilon = -2h^2 \log(s)$.
 - (b) For each $\tilde{y}_i$ in order, make the addition $S \mapsto S \cup \{\tilde{y}_i\}$ if the squared distance of $\tilde{y}_i$ from all existing points in S is at least ϵ .
4. Return the weighted alpha complex with vertex set S , and power map $\pi = -2h^2 \alpha$.

Definition 5. We denote the alpha complex resulting from running Algorithm 3.1 by $X = \mathcal{X}(f, s)$, separately specifying the number of sampled points M . We will let $\mathcal{X}(f, s, d_0)$ be the result of stopping upon reaching a minimum density cutoff of $f(\tilde{y}_i) \geq d_0$ in step 3.

We have chosen to represent our minimum separation parameter ϵ in terms of exponential coordinates $s = \exp(-\epsilon/2h^2)$, in other words as a minimum ratio in density coordinates. We find this more useful since s is unitless, and independent of the situation and scale. For instance, a value of $s = .9$

will always result in a relatively dense vertex set, whereas $s = .5$ will be relatively spread out. Similarly, it often makes more sense to refer to the minimum density d_0 in terms of the fraction of the data that falls under that range. A minimum density d_0 corresponding to %80 of the data would mean the value at which %80 of the points satisfy $f(x_i) \geq d_0$.

Notice that the union of the corresponding power diagram $\bigcup \mathcal{V}_{S,\pi}(a)$ is always contained in $\mathcal{L}(a)$. On the other hand, if we have chosen enough samples, then we would also have that the union $\bigcup \mathcal{V}(a + \epsilon)$ contains $\mathcal{S}(a)$. This relies on the fact that we are proceeding in increasing order of α , so that the shifted squared distance of any new point to an existing site $p_i \in S$ is bounded above by the unshifted squared distance. In this case, we would have

$$\mathcal{S}(a) \subset \bigcup_{p_i \in S} V_i(a + \epsilon) \subset \mathcal{L}(a + \epsilon) \sim \mathcal{S}(a + \epsilon). \quad (3.8)$$

Algorithm 3.1 has several crucial scalability properties. First, the shape is covariant under linear coordinate changes, so its covering number by balls of radius ϵ is also unchanged, meaning we do not have the poor scaling associated with sampling from $\mathcal{L}(a)$ from the introduction. Furthermore, it follows from properties of Gaussian samples and the form of p as an expectation that the sampling procedure itself, in other words the resulting distribution on $\mathcal{S}$ from which the $\tilde{y}_i$ are sampled, is independent of the embedding as well. To illustrate, notice that increasing the dimension by adding extra zeros to the end of each x_i has no effect: the values along the new coordinates will be sampled from independent Gaussians, and p will simply map those coordinates back to zero, as it factors through the projection map to the affine span of $\mathcal{D}$. In particular, unlike the values of $f(\tilde{x}_i)$, the resulting real distribution coming from the values of $\tilde{f}(\tilde{y}_i)$ or $\alpha(\tilde{y}_i)$ is stable, and indeed it depends only on the pairwise distances $\|x_i - x_j\|$. We also point out that evaluating $\alpha(\tilde{y}_i)$ creates no extra computational cost because the the infimum defining (3.1) is attained by the sample $\tilde{x}_i$ itself.

We will use the following variant of $\mathcal{X}(f, s)$. The alpha complex (or Vietoris-Rips) necessarily creates “noise” in zeroth persistent betti number when the vertices are well-spaced, due to the gap between a new point and the main component. A tempting fix is to add “slack” by shifting the persistence of higher-dimensional simplices back by a specified amount, similar to what the Rips or lazy witness complex does to fill in higher simplices, but this would add parameters and compromise objectivity. A far better solution

is to modify the weight of each simplex according to the following definition:

Definition 6. Let $\mathcal{X}^\circ(f, s)$ be the result of replacing

$$w(\sigma) \mapsto \max(w(\sigma), \alpha(\sigma_0) + \epsilon, \dots, \alpha(\sigma_k) + \epsilon) \quad (3.9)$$

for all simplices $\sigma \in X = \mathcal{X}(f, s)$, where $\epsilon = -2h^2 \log(s)$ is the spacing from Algorithm 3.1.

This is equivalent to requiring that each element of the power diagram appears only once its squared-radius crosses is at least ϵ . This maintains the desirable property that the resulting complex is the nerve of a cover whose union is contained in $\mathcal{L}(a)$, and also preserves (3.8) for sufficiently dense samples.

4 Experiments

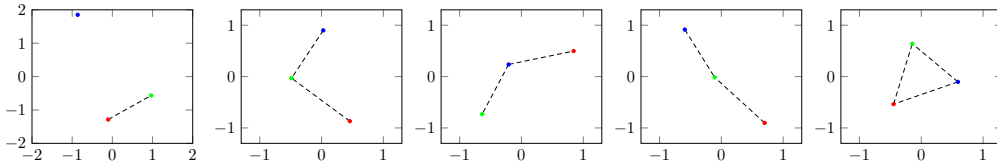
We construct the filtered complexes from Section 3.2 in several examples. To compute alpha complexes in higher dimension we used a recent algorithm based on dual programming to compute the corresponding alpha complexes [10], which in all cases required only a few seconds to complete. In the sampling step we usually used $M = 10000$, which took a couple of minutes, coming from evaluating f at that many points.

4.1 Interesting energy landscapes

Consider the energy function for particle interaction for three points in the plane $(p_1, p_2, p_3) \in \mathbb{R}^2$, using the Lennard-Jones potential:

$$H(p_1, p_2, p_3) = \sum_{i < j} V(\|p_i - p_j\|), \quad V(r) = 4 \left(\frac{1}{r^{12}} - \frac{1}{r^6} \right). \quad (4.1)$$

The Lennard-Jones potential rewards pairs of particles that are approximately distance 1 apart, but strongly penalizes particles that are much closer than that, and is neutral for far away points. Some typical points can be represented as



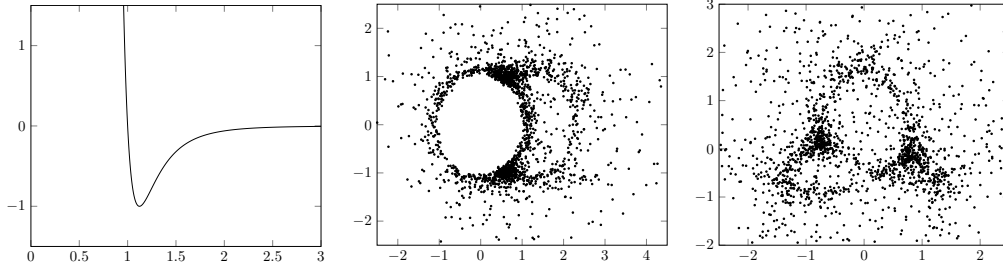


Figure 4.1.1: On the left, a graph of the Lennard-Jones potential. In the middle and right, representations of a point cloud of samples using the two-point normalization and Hopf map respectively.

where we have shown a dotted line when points are roughly distance one apart. We estimate the sublevel set persistent homology of H by applying the density-based complex to a mean-centered point cloud of samples, revealing an interesting the homology groups of a configuration space.

We sampled 100000 points from the distribution

$$\rho(p_1, p_2, p_3) = \exp \left(-\beta \left(H(p_1, p_2, p_3) + \sum_{i=1}^3 \|p_i\|^2 / 2R^2 \right) \right) \quad (4.2)$$

thought of as a density in $\mathbb{R}^6$, using the Metropolis algorithm. Here the temperature parameter β was set to 3.0, and a radius of $R = 3.0$ was used to keep particles from wandering off to ∞ .

Some projections of point clouds are shown in Figure 4.1.1. In order to visualize the point clouds, we used two different prescriptions to lower the dimension. In the middle frame on the lower row, we reduced the dimension to 2 by choosing a standard reference frame in which p_1 is at the origin, and p_2 is on the x -axis, using a translation followed by a rotation. We then normalized the sum of the norm squares, and plotted the remaining point p_3 , which moves around a figure 8 shape. In the frame on the lower right, we instead mean-center each sample $(p_1, p_2, p_3) \in \mathbb{R}^6$, so as to have a point in $\mathbb{R}^4$. We then normalized the rotational angle using the map

$$\varphi : \mathbb{R}^4 \rightarrow \mathbb{R}^3, \quad \varphi(v) = \|v\| \pi(v/\|v\|) \quad (4.3)$$

where $\pi : S^3 \rightarrow S^2$ is a version of the Hopf fibration that respects conformal structure, which is equivalent to the map $SU(2) \rightarrow SU(2)/T$ where $T \cong$

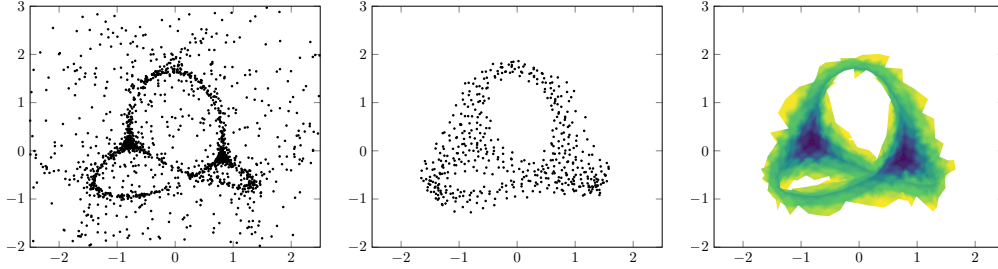


Figure 4.1.2: In the first frame, several samples $\{\tilde{y}_i\}$ from the shape using Algorithm 3.1. In the second, the sampled landmark set S with $s = .8$. In the final frame, the resulting alpha complex, using a heatmap to describe the weights.

$S^1 \subset SU(2)$ is the torus. Unlike the two-point version, this preserves the metric structure on the quotient, seen by the $\pi/3$ rotational symmetry in the figure.

We then defined a kernel density estimator f on both the mean centered original point cloud $\mathcal{D} \subset \mathbb{R}^4$, and its image $\mathcal{D}' \subset \mathbb{R}^3$ under (4.3). In both cases, we chose a value of $h = .3$, and computed $\mathcal{X}(f, s, d_0)$ using density cutoff d_0 corresponding to the %60 of the data. We created resulting filtered complex from $\mathcal{D}'$ with $s = .8$, which had sizes $(|X_0|, |X_1|, |X_2|) = (339, 1287, 1449)$. The outputs of Algorithm 3.1 are shown in Figure 4.1.2, using color values to represent the simplex weights.

We then applied the same procedure to the full dataset $\mathcal{D}$, this time using $s = .7$, and computing up to the 3-simplices, resulting in the sizes $(|X_0|, |X_1|, |X_2|, |X_3|) = (2298, 24582, 64896, 64674)$. The persistent homology groups are consistent with a disjoint union of two circles turning into a configuration space of 3 ordered points in the plane. This is because the densest points are the ones forming an equilateral triangle, which come in two types corresponding to the rotationally inequivalent permutations of the labels. The points with two connections form the shape of a configuration space, which has betti numbers of $(\beta_0, \beta_1, \beta_2) = (1, 3, 2)$ corresponding to the arrows in the diagram [18].

The `jvavplex` output of the persistent homology groups is shown in Figure 4.1.3.

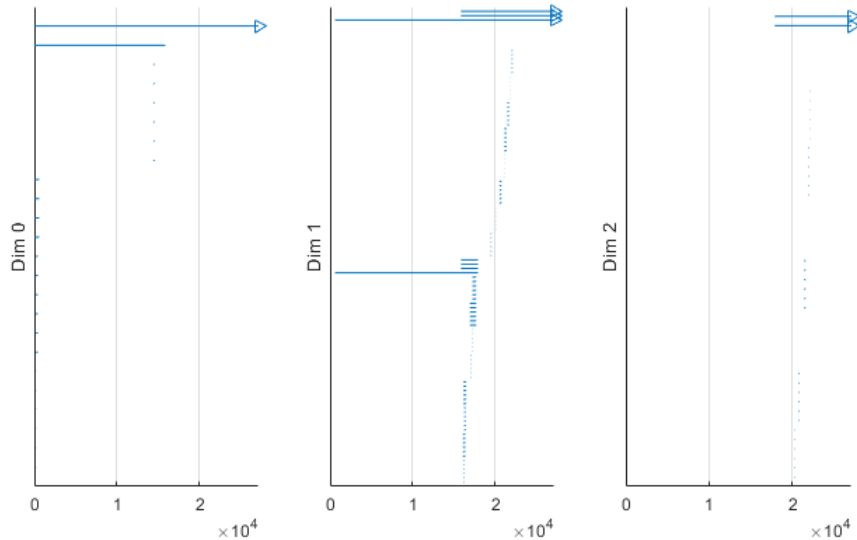


Figure 4.1.3: The barcode diagrams of the full density landscape computed with `javaplex`. Before the threshold around 1.5 we see two connected components and two betti 1 features. Beyond that threshold, we see the homology of the ordered configuration space of 3-points in the plane, $(\beta_0, \beta_1, \beta_2) = (1, 3, 2)$.

4.2 Multiple optimizers in nonlinear regressions

We next use alpha complexes to reveal local basins in the loss function of a nonlinear regression, using the Metropolis algorithm and maximum likelihood estimation. In order to produce a point cloud, we map each collection of parameter values into the Euclidean vector consisting of predicted values at each training point. We find that this is a good candidate for density estimation and the alpha complex, and that there is a natural choice of the scale h in terms of the temperature parameter of the model β .

We follow Example 2.6 from [38], which deals with the catalytic isometrization of n -pentane to i -pentane in the presence of hydronium, based on an original study by Carr [14]. The training data consists of 24 experimental runs, with four columns labeled x_1, x_2, x_3, r , measuring the partial pressures of hydrogen, n -pentane, and i -pentane, and the corresponding reaction rate r . The

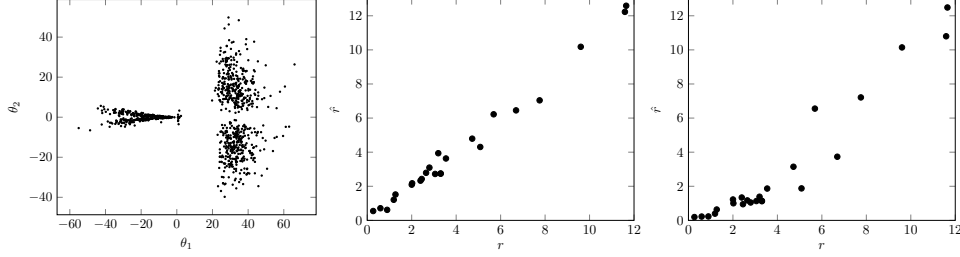


Figure 4.2.1: Scatter plot of the first two model parameters (θ_1, θ_2) from modeling the Carr data. In the middle/right frames we have the plot of predicted value versus true value for a typical value from the positive/negative values of θ_1 groups respectively. The first group tend to be better fits.

modeling problem is to predict the last column using the model

$$r \sim \frac{\theta_1 \theta_3 (x_2 - x_3 / 1.632)}{1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3}. \quad (4.4)$$

We simulated 10000 samples of the θ -parameters using the Metropolis algorithm, and a maximum likelihood with squared residual losses,

$$p(\theta) = \exp(-\beta L(\theta)), \quad L(\theta) = \sum_{i=1}^{24} (r_i - \hat{r}_i(\theta))^2 \quad (4.5)$$

and temperature of $\beta = 3.0$, resulting in a point cloud $\mathcal{D} \subset \mathbb{R}^4$. One finds that there was a near-symmetry in simultaneously sending $\theta_i \mapsto -\theta_i$ for $i = \{2, 3, 4\}$, due to the fact that the parameters tended to dominate the leading 1 in the denominator. Additionally, there were two local basins of solutions corresponding to the sign of θ_1 , with the positive values having better predictions. The results are shown in Figure 4.2.1.

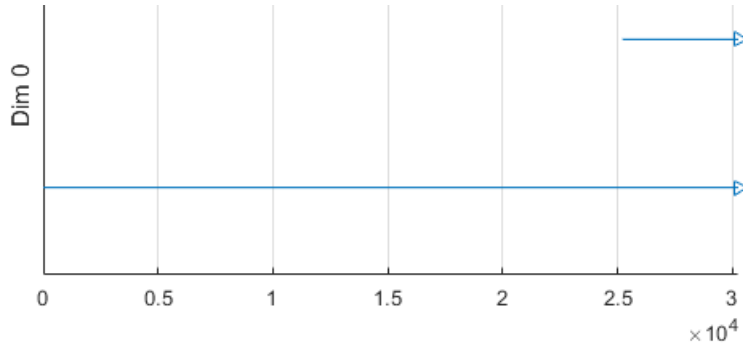
A naïve next step would be to apply kernel density estimation to the point cloud $\mathcal{D} \subset \mathbb{R}^4$, with some choice of scale h . However, this would not be meaningful, as it is depends on the parametrization of the model. A more appropriate one is to define $f(x)$ using a new data set $\mathcal{D}' \subset \mathbb{R}^{24}$ by mapping each θ to the corresponding vector of predictions

$$\theta \mapsto (\hat{r}_1(\theta), \dots, \hat{r}_{24}(\theta)). \quad (4.6)$$

We then have a natural choice of the scale parameter, $h = (\beta/2)^{-1/2} \sim .816$.



(a) Alpha complex projected to 2 dimensions



(b) Zero-dimensional persistence barcodes

Figure 4.2.2: Alpha complex $\mathcal{X}(f, s, d_0)$ built from the data set of prediction vectors in $\mathbb{R}^{24}$ with $s = .6$ and d_0 corresponding to %80 of the data, projected into 2 dimensions using a PCA on the vertex set. The denser component on the left represents the $\theta_1 > 0$ group. In the second frame, the barcode diagram indicating the local basins

We built the alpha complex using a value of $s = .6$, with density cutoff d_0 corresponding to 80% of the data, shown in Figure 4.2.2. The map to $\mathbb{R}^{24}$ effectively collapses the symmetry arising from the sign changes, leaving only the two connected components. The component corresponding to positive values of θ_1 gives better predictions, resulting in the denser component on the left. The others lead to local basin which wider but less dense, corresponding to higher values of the loss function. These are reflected in the betti-zero persistence barcodes, also shown.

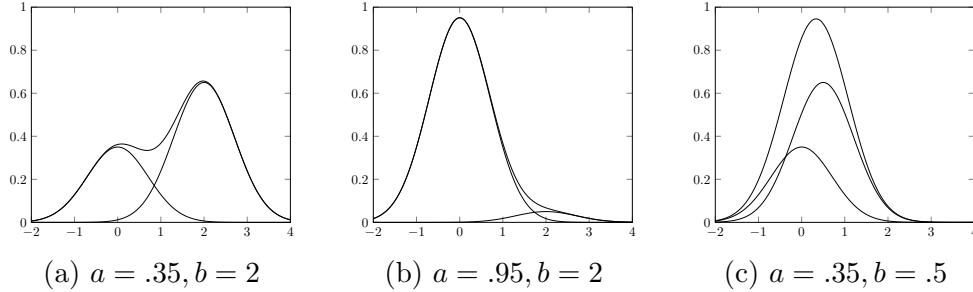


Figure 4.3.1: Simple singular statistical model with two Gaussians. The singularity arises because the second two frames correspond to different parameter values, but result in nearly equal distributions.

4.3 A simple singular learning model

We apply a similar method from the previous example to a singular statistical model, which is Example 1.2 from [40], with thanks to Dan Murfet for the suggestion. We see that the alpha complex exhibits interesting behavior as a certain Riemannian metric related to the Fisher information matrix becomes degenerate near the singularity.

Consider a simple one-variable Gaussian mixture model

$$p(x|a, b) = ae^{-x^2/2} + (1 - a)e^{-(x-b)^2/2}, \quad 0 \leq a \leq 1, \quad (4.7)$$

consisting of a weighted sum of Gaussian distributions with standard deviation 1, and a varying mean in the second one. An illustration is shown in Figure 4.3.1.

For any point cloud in $\mathbb{R}$, we can use the Metropolis algorithm and Bayes' rule to sample from a density proportional to

$$p(a, b|t) = \frac{p(t|a, b)p(a, b)}{p(t)},$$

starting with the uniform measure $p(a, b) = 1$. As usual, we do not need to know $p(t)$. If our point cloud is sampled from $p(t|a, b)$ for a particular choice of (a, b) , we would expect the resulting distribution to be supported near the ones we started with. This will indeed happen as expected if we chose values giving a bimodal distribution as in Figure 4.3.1a. However, interesting things happen if our point cloud comes from a single Gaussian centered at the

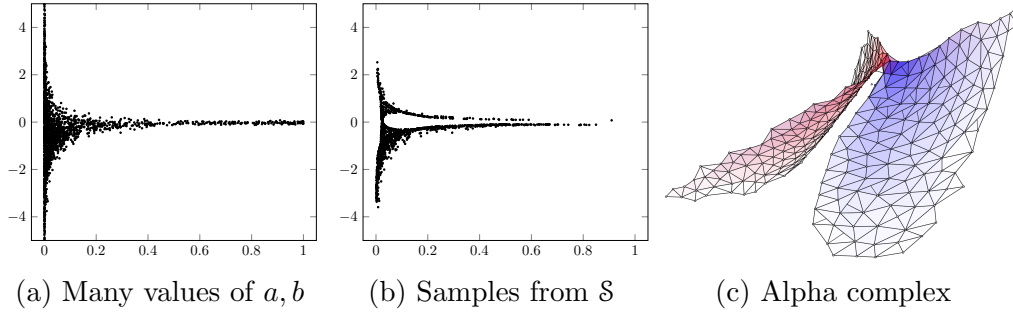


Figure 4.3.2: On the left, scatter plot of samples of the model parameters (a, b) using Markov chain Monte Carlo (MCMC) with respect to the distribution $e^{-t^2/2}$. In the middle, samples from the shape displayed in $\mathbb{R}^2$ using the coefficients of $p(\tilde{x}_i)$ and the original samples (a_i, b_i) . On the right, the corresponding alpha complex $\mathcal{X}^o(f, s, d_0)$ for $s = .8$ and d_0 corresponding to %80 of the data. The red/blue colors correspond to positive/negative values of b , respectively.

origin, because many values of (a, b) correspond to that same distribution, as shown in Figures 4.3.1b and 4.3.1c.

We generated a scatter plot of 10000 values of the parameters (a, b) associated to the singular Gaussian using 1000 training points in $\mathbb{R}$. The results, shown in Figure 4.3.2a, reveal the expected behavior at the singularity at the origin $(a, b) = (0, 0)$. As in the previous section, building an alpha complex on the resulting point cloud in $\mathbb{R}^2$ would be arbitrary. We no longer have a vector of predictions as in (4.6), and instead we use the vector of values of the loss function

$$(a, b) \mapsto (-\log(p(t_1|a, b)), \dots, -\log(p(t_{1000}|a, b))). \quad (4.8)$$

This map has the property that the pullback Riemannian metric on $\mathbb{R}^2$ is a version of the Fisher information matrix near the true distribution. We then reduced the dimension down to 50 using a PCA, in part to speed up the sampling, but more importantly because rounding errors become a factor when evaluating $y = p(\tilde{x}_i)$ at a sample. The second issue can easily be fixed by representing components of the calculation in log coordinates. We then considered the density estimator $f : \mathbb{R}^{50} \rightarrow \mathbb{R}_+$ corresponding to the resulting point cloud $\mathcal{D} \subset \mathbb{R}^{50}$, and chose a (this time arbitrary) value of the scale parameter of $h = .1$.

We then sampled from the shape using Algorithm 3.1, using the values of $s = .8$, cutting off at a value corresponding to %80 of the data. By taking the coefficients determining the expression $p(\tilde{x}_i) = \sum_j c_{i,j} x_j$ as a convex combination of the x_i and plotting the points $\sum_j c_{i,j} (a_i, b_i) \in \mathbb{R}^2$ instead of $\tilde{y}_i = \sum_j c_{i,j} x_j \in \mathbb{R}^{50}$, we obtain a description of the shape in the original 2-dimensional plane. Instead of corresponding to the original singularity, the points on the line $b = 0$ were all mapped to the same point near the origin, creating the interesting pattern shown in Figure 4.3.2b. This is because all those values become very close together when mapped to $\mathbb{R}^{50}$, resulting in a degenerate metric at the origin. We then generated $\mathcal{X}(f, s)$ with $s = .8$, and plotted the projection onto a unitary subspace in $\mathbb{R}^{50}$, showing the interesting shape in Figure 4.3.2c.

4.4 Local patches in the MNIST data set

In [28], 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 [29, 31]. They gave quantitative evidence using the witness complex that those patches lie along a sublocus of a parametrized Klein bottle, called the three-circle model. Using a similar setup, we apply our construction to data sets of coordinate patches taken from 28×28 images of handwritten digits from the MNIST data set [19]. Instead of using small 3×3 patches, we project onto discrete versions of the Hermite polynomials up to quadratic order. Using the alpha complexes, we obtain surprisingly descriptive geometric models corresponding to different regions in the Klein bottle as one varies the digit.

A “local image patch” will mean an $l \times l$ subimage of a larger one, which in our case will be taken the MNIST data set of handwritten digits. We will view local image patches as elements of the vector space $V = \text{Mat}(l, l)$ using its grayscale intensity value. We have a scalar product on V given by

$$(A, B) = \frac{1}{2^{2(l-1)}} \sum_{i=1}^l \sum_{j=1}^l \binom{l-1}{i-1} \binom{l-1}{j-1} A_{i,j} B_{i,j} \quad (4.9)$$

This inner product has a number of advantages over the usual L^2 product in that the weights fall off gradually near the image border. It also has the property of being nearly rotationally invariant for larger values of l , as the binomial coefficient approximates the Gaussian. There is an orthonormal

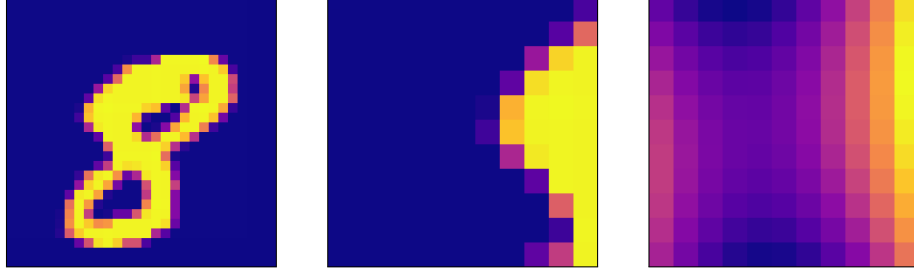


Figure 4.4.1: A typical digit, a random 11×11 image patch, and its orthogonal projection onto $V_1 \oplus V_2$.

basis given by $H_{a,b} = H_a \otimes H_b$, where the $H_a \in \mathbb{R}^l$ are a discrete form of the Hermite polynomials. They can be obtained by applying the Gram-Schmidt algorithm to the vectors of polynomial functions $v_a = (i^a)_{i=1}^l$, with respect to the one-dimensional form of (4.9).

We have a decomposition

$$V = V_0 \oplus V_1 \oplus \cdots, \quad V_i = \text{span}\{H_{a,b} : a + b = i\}$$

as well as orthogonal projections $\pi_i : V \rightarrow V_i$. We will be interested in the images of image patches under the map $\pi_{1,2} : V \rightarrow V_1 \oplus V_2 \cong \mathbb{R}^5$, which analogous to projecting onto low-frequency modes in Fourier analysis. One such projection is shown in Figure 4.4.1. For any image patch we have its norm squared $r^2 = r_1^2 + r_2^2$ where r_i is the norm of its image in V_i . Image patches centered on the points of the digit would tend to have relatively high r_2 values, while points on the boundary would have higher r_1 values, due to the gradient. The patch in Figure 4.4.1 would both have relatively high r_1 and r_2 terms.

For 50 instances of each digit, we sampled all $l \times l$ patches using the choice of $l = 11$, and projected those patches onto their linear and quadratic components $V_1 \oplus V_2$, to obtain a point cloud of size $50 \cdot (28 - l + 1)^2 = 16200$ in $\mathbb{R}^5$. We then chose only those images whose L^2 -norm is above a fixed number of $r \geq .3$, resulting in a subset of around 20% of the original size. This is analogous to the step of selecting “high intensity patches” from [28]. We then divided the remaining points by r to arrive at a point cloud $\mathcal{D}_k \subset S^4$ of size a few thousand for each digit $k \in \{0, \dots, 9\}$. We defined a kernel density estimator $f : \mathbb{R}^5 \rightarrow \mathbb{R}_+$ using $h = .15$, and built $\mathcal{X}^\circ(f, s, d_0)$ using the same choices of $s = .5$, and d_0 corresponding to about %60 of the data for each digit.

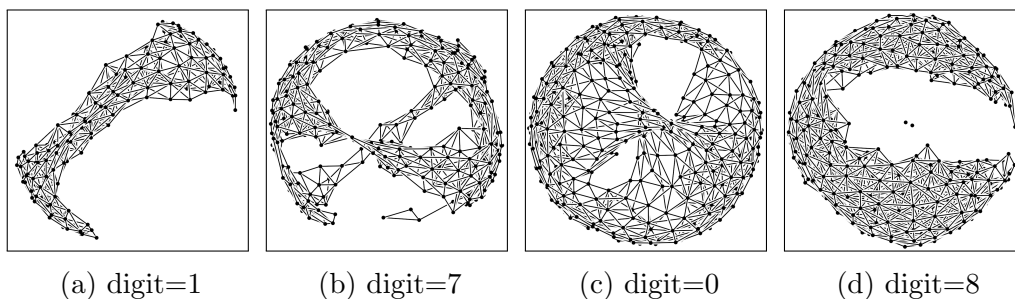


Figure 4.4.2: Alpha complexes build out of the high-intensity image patches.

The results, shown for the numbers $\{1, 7, 0, 8\}$ in Figure 4.4.2, exhibited distinctive features, which can be understood in terms of the Klein bottle model. Starting with the digit 1, we see two arcs connected by a connecting region in the center. Analyzing the images associated to each point shows that the arcs are the regions on either side of the digit, which have large magnitude in linear terms r_1 . The strip lives on the digit, which has relatively high quadratic norm r_2 . The digit 7 has a similar explanation but with two different components. These linear terms fill out the entire periphery of a circle in the case of the digit 0, coming from both the interior and exterior. In the center of the figure, we see the high r_2 points twist and connect across points on the digit, which are only dense enough on the left and right sides because of the oval shape. The digit 8 shows no points dominated by second order, except a small disconnected region corresponding to the two voids.

We then ran the experiment again, this time using r_2 to determine intensity, and dividing by r_2 in place of r . This has the effect of making patches with higher r_2 denser, thereby accentuating the second order features. This time we cut off at an intensity value of $r_2 \geq .125$, and chose d_0 to correspond to only %30 of the data, keeping the values of $h = .15$ and $s = .5$.

The results are shown in Figure 4.4.3. In the digit 1, we see three connected components, corresponding to points on the digit itself, and two others which are not on the boundary, but slightly away from it on either side. For instance the patch in Figure 4.4.1 would be such a point. Again, digit 7 has the same explanation, with 6 components instead of 3. In the digit zero, we have a complete Möbius strip as one traverses halfway around the digit itself. The two additional components have the same meaning as with the digit 1, appearing only on the sides because the digit is not a perfect circle. In the digit 8, we actually see 5 connected components. The main one consists of

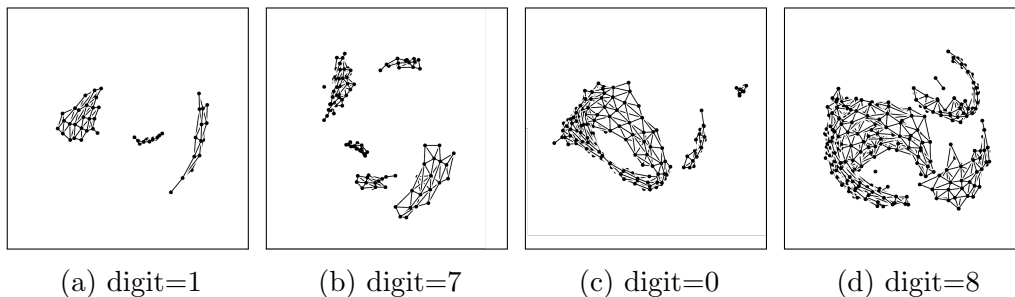


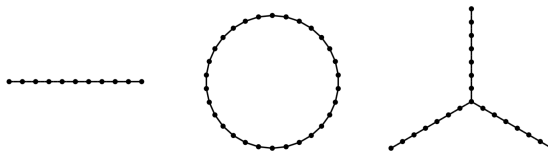
Figure 4.4.3: Complexes built on the corresponding digits, normalizing only the second-order coefficients.

points on the digit itself, whereas the second largest ones are points on either side, as in the other digits. One of the remaining small clusters comes from the voids inside either loop, while the other represents the crossing point in the center.

4.5 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 [30] on a graph with d vertices, thought of as a collection of real-valued vectors in $\{\pm 1\}^d \subset \mathbb{R}^d$. Attempting to apply kernel density estimation on the resulting point cloud directly would not yield good results, and we would not even expect to be able to correlate kernel based density at a particular state with the theoretical density determined by the energy function. We show that we can create geometric models of the density landscape as we did in Section 4.1, by using Laplacian operator L of the underlying graph to obtain smooth versions of the spin vectors.

Let $G = (V, E)$ be a graph with n vertices, represented by a symmetric adjacency matrix J , with diagonal entries being zero. In our example we will use



denoted $\text{int}(n)$, $\text{circ}(n)$, $\text{flares}(n)$, where n is the number of vertices. For every

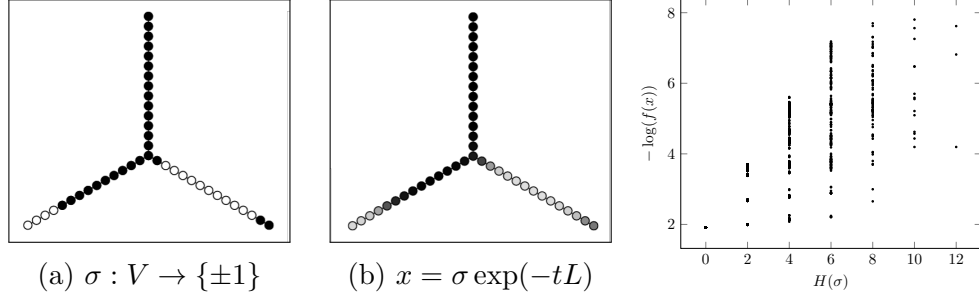


Figure 4.5.1: The result of blending a typical state of the Ising model simulation using the Laplacian operator. On the right, a plot of the energy level versus negative log of kernel density.

discrete spin vector $\sigma : V \rightarrow \{1, -1\}$, we have the Hamiltonian energy

$$H_G(\sigma) = - \sum_{i,j} J_{i,j} \sigma_i \sigma_j = H_{\min} + 2|\{(i,j) \in E : \sigma_i \neq \sigma_j\}|. \quad (4.10)$$

Those pairs $i, j \in E$ 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 $\{1, -1\}^d$ given by

$$P_\beta(\sigma) = \frac{1}{Z_\beta} e^{-\beta H(\sigma)}, \quad Z_\beta = \sum_{\sigma} e^{-\beta H(\sigma)} \quad (4.11)$$

which is usually done using the single-flip Metropolis algorithm.

For the graphs $G \in \{\text{int}(30), \text{circ}(30), \text{flares}(43)\}$, we simulated $N = 20000$ states using a temperature value of $\beta = 3.0$, and interpreted the resulting collections of spin vectors $\{\sigma\}$ as a point clouds $\mathcal{D}_G \subset \{\pm 1\}^d \subset \mathbb{R}^d$ where $d = n$ is the number of vertices. We then took a blended version of $\mathcal{D}_G$ using the left-normalized Laplacian operator $I - D^{-1}A$, where 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 . We then replaced $\mathcal{D}_G$ by sending $\sigma \mapsto \sigma \exp(-tL^t)$ with the value of $t = 10$, so that the vectors are no longer $\{\pm 1\}$ -valued, and considered the corresponding kernel density estimator $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ with $h = 2.0$. An illustration of the result of the convolution, and the distribution of density versus Hamiltonian energy is shown in Figure 4.5.1.

We then computed $\mathcal{X}(f, s, d_0)$ for $s = .5$, and d_0 corresponding to %95 of the data. By taking a random 3D to 2D projection of the values in the

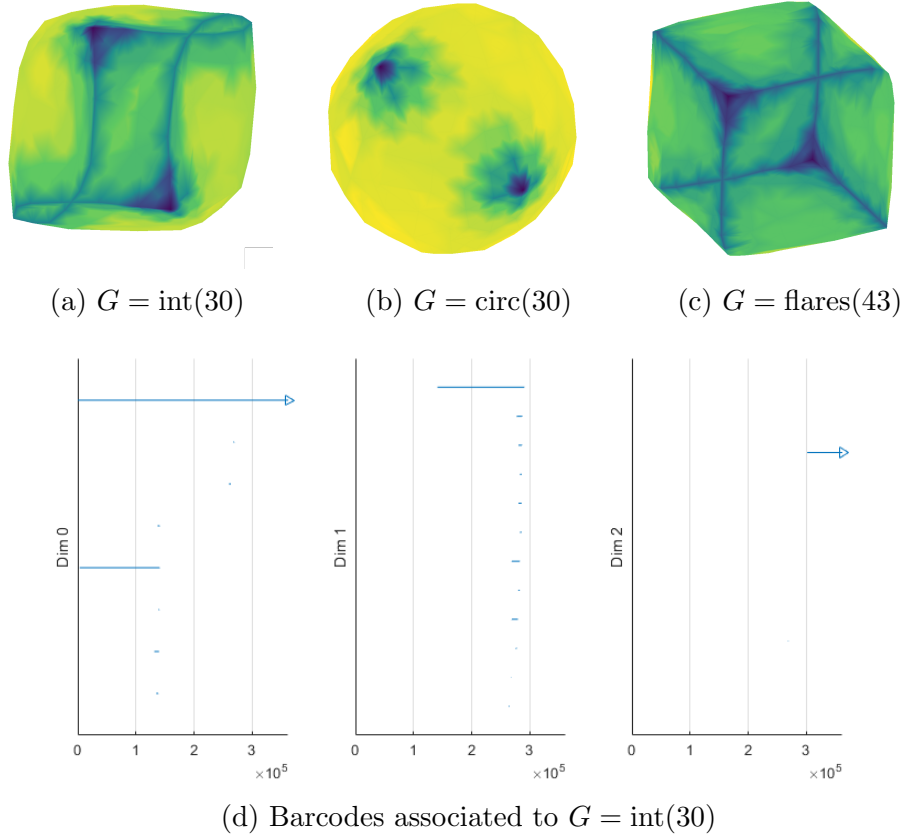


Figure 4.5.2: Low dimensional projections of $\mathcal{X}(f, s, d_0)$ for the graphs $\text{int}(30)$, $\text{circ}(30)$, and $\text{flares}(43)$. In the lower row, the persistence barcodes in the case of the interval.

3 most dominant eigenvalues of L , we obtain a visualization of the energy landscape as in 4.1, shown in Figure 4.5.2. In the first frame, associated to the interval, we have two densest types of points with energy level zero, corresponding to all spins equal to plus or minus 1, realized at the corners. We then have two curved line segments consisting of energy states with exactly one transition joining those points, starting from either side. The points shown in green/yellow correspond to states with two transitions, filling in the resulting circle to form a sphere. This is reflected in the persistence barcodes, shown in the figure.

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] Aaron Adcock, Erik Carlsson, and Gunnar Carlsson. The ring of algebraic functions on persistence bar codes. *Homology, Homotopy and Applications*, 18, 04 2013.
- [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] Ulrich Bauer, Michael Kerber, Fabian Roll, and Alexander Rolle. A unified view on the functorial nerve theorem and its variations. *Expositiones Mathematicae*, 2023.
- [5] Andrew Blumberg, Itamar Gal, Michael Mandell, and Matthew Pancia. Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces. *Foundations of Computational Mathematics*, 14, 06 2012.
- [6] Andrew Blumberg and Michael Lesnick. Stability of 2-parameter persistent homology. *arXiv:2010.09628*, 10 2020.
- [7] Omer Bobrowski, Sayan Mukherjee, and Jonathan E. Taylor. Topological consistency via kernel estimation. *Bernoulli*, 23(1):288 – 328, 2017.
- [8] Karol Borsuk. On the imbedding of systems of compacta in simplicial complexes. *Fundamenta Mathematicae*, 35(1):217–234, 1948.
- [9] Peter Bubenik. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*, 16:77–102, 2012.
- [10] Erik Carlsson and John Carlsson. Computing the alpha complex using dual active set quadratic programming, 2023.
- [11] Gunnar Carlsson, Gurjeet Singh, and Afra Zomorodian. Computing multidimensional persistence. In Yingfei Dong, Ding-Zhu Du, and Oscar

- Ibarra, editors, *Algorithms and Computation*, pages 730–739, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.
- [12] Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete and Computational Geometry*, 42:71–93, 06 2007.
 - [13] Gunnar E. Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46:255–308, 2009.
 - [14] N. L. Carr. Kinetics of catalytic isomerization of n -pentane. *Indust. Eng. Chem.*, 52:391–396, 1960.
 - [15] Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11:733–751, 2011.
 - [16] Frédéric Chazal, Brittany Fasy, Fabrizio Lecci, Alessandro Rinaldo, Aarti Singh, and Larry Wasserman. On the bootstrap for persistence diagrams and landscapes. *Modelirovanie i Analiz Informacionnyh Sistem*, 20, 11 2013.
 - [17] Frédéric Chazal, Leonidas Guibas, Steve Oudot, and Primoz Skraba. Persistence-based clustering in riemannian manifolds. *Journal of the ACM*, 60, 06 2011.
 - [18] F.R. Cohen. On configuration spaces, their homology, and lie algebras. *Journal of Pure and Applied Algebra*, 100(1):19–42, 1995.
 - [19] Li Deng. The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.
 - [20] H. Edelsbrunner. The union of balls and its dual shape. *Discrete and Computational Geometry*, pages 415–440, 1995.
 - [21] 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.
 - [22] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, 29(4):551–559, 1983.

- [23] Herbert Edelsbrunner, Michael Facello, and Jie Liang. On the definition and the construction of pockets in macromolecules. *Discrete Applied Mathematics*, 88(1):83–102, 1998. Computational Molecular Biology DAM - CMB Series.
- [24] Herbert Edelsbrunner and John Harer. *Computational Topology - an Introduction*. American Mathematical Society, 2010.
- [25] Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. Confidence sets for persistence diagrams. *The Annals of Statistics*, 42(6):2301 – 2339, 2014.
- [26] Robin Forman. Morse theory for cell complexes. *Advances in Mathematics*, 134(1):90–145, 1998.
- [27] Eric M. Friedlander. *Etale Homotopy of Simplicial Schemes. (AM-104), Volume 104*. Princeton University Press, Princeton, 1983.
- [28] 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.
- [29] 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.
- [30] E. Ising. Beitrag zur theorie des ferromagnetismus. *Z. Physik*, 31:253–258, 1925.
- [31] 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.
- [32] Jean-L. Leray. L’anneau spectral et l’anneau filtré d’homologie d’un espace localement compact et d’une application continue. *J. Math. Pures Appl*, 9, 1950.
- [33] Michael Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Foundations of Computational Mathematics*, 15, 06 2011.

- [34] Nina Otter, Mason Porter, Ulrike Tillmann, Peter Grindrod, and Heather Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6, 06 2015.
- [35] Jeff M. Phillips, Bei Wang, and Yan Zheng. Geometric Inference on Kernel Density Estimates. In Lars Arge and János Pach, editors, *31st International Symposium on Computational Geometry (SoCG 2015)*, volume 34 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 857–871, Dagstuhl, Germany, 2015. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- [36] Steven J. Phillips, Robert P. Anderson, and Robert E. Schapire. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3):231–259, 2006.
- [37] Vanessa Robins, Peter John Wood, and Adrian P. Sheppard. Theory and algorithms for constructing discrete morse complexes from grayscale digital images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1646–1658, 2011.
- [38] G. Seber and C. Wild. *Nonlinear Regression (1st ed.)*. Wiley, 2005.
- [39] Vin de Silva and Gunnar Carlsson. Topological estimation using witness complexes. In Markus Gross, Hanspeter Pfister, Marc Alexa, and Szymon Rusinkiewicz, editors, *SPBG’04 Symposium on Point - Based Graphics 2004*. The Eurographics Association, 2004.
- [40] Sumio Watanabe. *Algebraic Geometry and Statistical Learning Theory*. Cambridge University Press, USA, 2009.