

# Geometric Inference on Kernel Density Estimates\*

Jeff M. Phillips<sup>1</sup>, Bei Wang<sup>2</sup>, and Yan Zheng<sup>1</sup>

**1** School of Computing, University of Utah, USA

**2** Scientific Computing and Imaging Institute, University of Utah, USA

---

## Abstract

We show that geometric inference of a point cloud can be calculated by examining its kernel density estimate with a Gaussian kernel. This allows one to consider kernel density estimates, which are robust to spatial noise, subsampling, and approximate computation in comparison to raw point sets. This is achieved by examining the sublevel sets of the *kernel distance*, which isomorphically map to superlevel sets of the kernel density estimate. We prove new properties about the kernel distance, demonstrating stability results and allowing it to inherit reconstruction results from recent advances in distance-based topological reconstruction. Moreover, we provide an algorithm to estimate its topology using weighted Vietoris-Rips complexes.

**1998 ACM Subject Classification** F.2.2: Nonnumerical Algorithms and Problems

**Keywords and phrases** topological data analysis, kernel density estimate, kernel distance

**Digital Object Identifier** 10.4230/LIPIcs.SOCG.2015.857

## 1 Introduction

Geometry and topology have become essential tools in modern data analysis: geometry to handle spatial noise and topology to identify the core structure. Topological data analysis (TDA) has found applications spanning protein structure analysis [24, 40] to heart modeling [32] to leaf science [49], and is the central tool of identifying quantities like connectedness, cyclic structure, and intersections at various scales. Yet it can suffer from spatial noise in data, particularly outliers.

When analyzing point cloud data, classically these approaches consider  $\alpha$ -shapes [23], where each point is replaced with a ball of radius  $\alpha$ , and the union of these balls is analyzed. More recently a distance function interpretation [8] has become more prevalent where the union of  $\alpha$ -radius balls can be replaced by the sublevel set (at value  $\alpha$ ) of the Hausdorff distance to the point set. Moreover, the theory can be extended to other distance functions to the point sets, including the *distance-to-a-measure* [12] which is more robust to noise.

This has more recently led to statistical analysis of TDA. These results show not only robustness in the function reconstruction, but also in the topology it implies about the underlying dataset. This work often operates on persistence diagrams which summarize the persistence (difference in function values between appearance and disappearance) of all homological features in single diagram. A variety of work has developed metrics on these diagrams and probability distributions over them [43, 55], and robustness and confidence intervals on their landscapes [6, 30, 15, 16]). It is now more clear than ever, that these works are most appropriate when the underlying function is robust to noise, e.g., the distance-to-a-measure [12].

---

\* Thanks to supported to JMP by NSF CCF-1350888, IIS-1251019, and ACI-1443046, and for BW by INL 00115847 via DE-AC07ID14517, DOE NETL DEEE0004449, DOE DEFC0206ER25781, DOE DE-SC0007446, and NSF 0904631.



© Jeff M. Phillips, Bei Wang, and Yan Zheng;  
licensed under Creative Commons License CC-BY

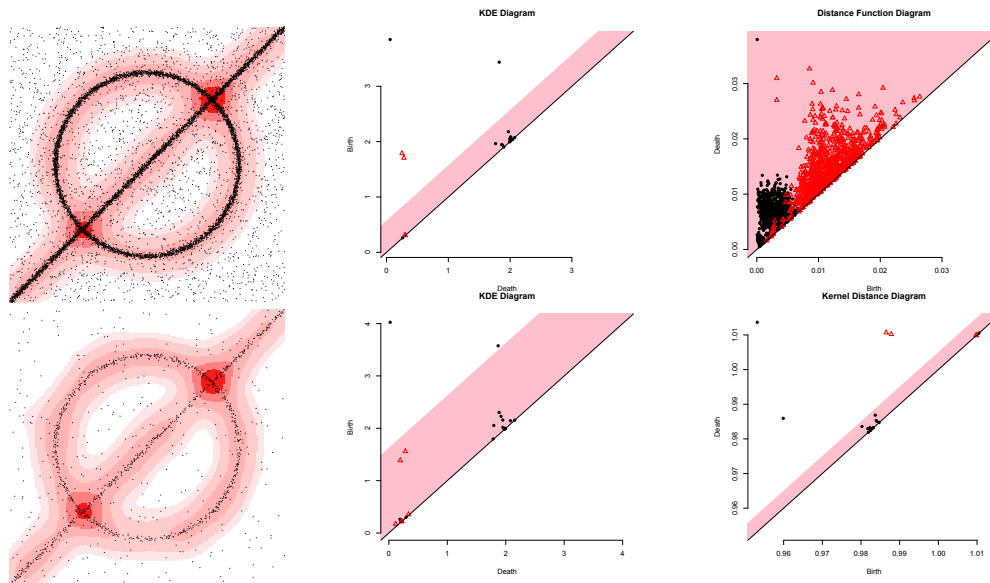
31st International Symposium on Computational Geometry (SoCG'15).

Editors: Lars Arge and János Pach; pp. 857–871



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



■ **Figure 1** Example with 10,000 points in  $[0, 1]^2$  generated on a circle or line with  $N(0, 0.005)$  noise; 25% of points are uniform background noise. The generating function is reconstructed with KDE with  $\sigma = 0.05$  (upper left), and its persistence diagram based on the superlevel set filtration is shown (upper middle). A coreset [58] of the same dataset with only 1,384 points (lower left) and persistence diagram (lower middle) are shown, again using KDE. This associated confidence interval contains the dimension 1 homology features (red triangles) suggesting they are noise; this is because it models data as iid – but the coreset data is not iid, it subsamples more intelligently. We also show persistence diagrams of the original data based on the sublevel set filtration of the standard distance function (upper right, with no useful features due to noise) and the kernel distance (lower right).

A very recent addition to this progression is the new TDA package for R [29]; it includes built in functions to analyze point sets using Hausdorff distance, distance-to-a-measure,  $k$ -nearest neighbor density estimators, kernel density estimates, and kernel distance. The example in Figure 1 used this package to generate persistence diagrams. While, the stability of the Hausdorff distance is classic [8, 23], and the distance-to-a-measure [12] and  $k$ -nearest neighbor distances have been shown robust to various degrees [4], this paper is the first to analyze the stability of kernel density estimates and the kernel distance in the context of geometric inference. Some recent manuscripts show related results. Bobrowski *et al.* [5] consider kernels with finite support, and describe approximate confidence intervals on the superlevel sets, which recover approximate persistence diagrams. Chazal *et al.* [14] explore the robustness of the kernel distance in bootstrapping-based analysis.

In particular, we show that the kernel distance and kernel density estimates, using the Gaussian kernel, inherit some reconstruction properties of distance-to-a-measure, that these functions can also be approximately reconstructed using weighted (Vietoris-)Rips complexes [7], and that under certain regimes can infer homotopy of compact sets. Moreover, we show further robustness advantages of the kernel distance and kernel density estimates, including that they possess small coresets [45, 58] for persistence diagrams and inference.

## 1.1 Kernels, Kernel Density Estimates, and Kernel Distance

A *kernel* is a non-negative similarity measure  $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^+$ ; more similar points have higher value. For any fixed  $p \in \mathbb{R}^d$ , a kernel  $K(p, \cdot)$  can be normalized to be a probability

distribution; that is  $\int_{x \in \mathbb{R}^d} K(p, x) dx = 1$ . For the purposes of this article we focus on the Gaussian kernel defined as  $K(p, x) = \sigma^2 \exp(-\|p - x\|^2 / 2\sigma^2)$ .<sup>1</sup>

A *kernel density estimate* [53, 50, 21, 22] is a way to estimate a continuous distribution function over  $\mathbb{R}^d$  for a finite point set  $P \subset \mathbb{R}^d$ ; they have been studied and applied in a variety of contexts, for instance, under subsampling [45, 58, 2], motion planning [48], multimodality [52, 25], and surveillance [28], road reconstruction [3]. Specifically,

$$\text{KDE}_P(x) = \frac{1}{|P|} \sum_{p \in P} K(p, x).$$

The *kernel distance* [37, 33, 38, 46] (also called *current distance* or *maximum mean discrepancy*) is a metric [44, 54] between two point sets  $P, Q$  (as long as the kernel used is characteristic [54], a slight restriction of being positive definite [1, 57], this includes the Gaussian and Laplace kernels). Define a similarity between the two point sets as

$$\kappa(P, Q) = \frac{1}{|P|} \frac{1}{|Q|} \sum_{p \in P} \sum_{q \in Q} K(p, q).$$

Then the kernel distance between two point sets is defined as

$$D_K(P, Q) = \sqrt{\kappa(P, P) + \kappa(Q, Q) - 2\kappa(P, Q)}.$$

When we let point set  $Q$  be a single point  $x$ , then  $\kappa(P, x) = \text{KDE}_P(x)$ .

Kernel density estimates applies to any measure  $\mu$  (on  $\mathbb{R}^d$ ) as  $\text{KDE}_\mu(x) = \int_{p \in \mathbb{R}^d} K(p, x) d\mu(p)$ . The similarity between two measures is  $\kappa(\mu, \nu) = \int_{(p, q) \in \mathbb{R}^d \times \mathbb{R}^d} K(p, q) d\mathbf{m}_{\mu, \nu}(p, q)$ , where  $\mathbf{m}_{\mu, \nu}$  is the product measure of  $\mu$  and  $\nu$  ( $\mathbf{m}_{\mu, \nu} := \mu \otimes \nu$ ), and then the kernel distance between two measures  $\mu$  and  $\nu$  is still a metric, defined as  $D_K(\mu, \nu) = \sqrt{\kappa(\mu, \mu) + \kappa(\nu, \nu) - 2\kappa(\mu, \nu)}$ . When the measure  $\nu$  is a Dirac measure at  $x$  ( $\nu(q) = 0$  for  $x \neq q$ , but integrates to 1), then  $\kappa(\mu, x) = \text{KDE}_\mu(x)$ . Given a finite point set  $P \subset \mathbb{R}^d$ , we can work with the empirical measure  $\mu_P$  defined as  $\mu_P = \frac{1}{|P|} \sum_{p \in P} \delta_p$ , where  $\delta_p$  is the Dirac measure on  $p$ , and  $D_K(\mu_P, \mu_Q) = D_K(P, Q)$ .

If  $K$  is positive definite, it is said to have the reproducing property [1, 57]. This implies that  $K(p, x)$  is an inner product in some reproducing kernel Hilbert space (RKHS)  $\mathcal{H}_K$ . Specifically, there is a lifting map  $\phi : \mathbb{R}^d \rightarrow \mathcal{H}_K$  so that  $K(p, x) = \langle \phi(p), \phi(x) \rangle_{\mathcal{H}_K}$ , and moreover the entire set  $P$  can be represented as  $\Phi(P) = \sum_{p \in P} \phi(p)$ , which is a single element of  $\mathcal{H}_K$  and has a norm  $\|\Phi(P)\|_{\mathcal{H}_K} = \sqrt{\kappa(P, P)}$ . A single point  $x \in \mathbb{R}^d$  also has a norm  $\|\phi(x)\|_{\mathcal{H}_K} = \sqrt{K(x, x)}$  in this space.

## 1.2 Geometric Inference and Distance to a Measure: A Review

Given an unknown compact set  $S \subset \mathbb{R}^d$  and a finite point cloud  $P \subset \mathbb{R}^d$  that comes from  $S$  under some process, geometric inference aims to recover topological and geometric properties of  $S$  from  $P$ . The offset-based (and more generally, the distance function-based) approach for geometric inference reconstructs a geometric and topological approximation of  $S$  by offsets from  $P$  (e.g. [10, 11, 12, 17, 18]).

Given a compact set  $S \subset \mathbb{R}^d$ , we can define a *distance function*  $f_S$  to  $S$ ; a common example is  $f_S(x) = \inf_{y \in S} \|x - y\|$ . The offsets of  $S$  are the sublevel sets of  $f_S$ , denoted  $(S)^r = f_S^{-1}([0, r])$ . Now an approximation of  $S$  by another compact set  $P \subset \mathbb{R}^d$  (e.g. a

<sup>1</sup> The choice of coefficient  $\sigma^2$  is not the standard normalization, but it is perfectly valid as it scales everything by a constant. It has the property that  $\sigma^2 - K(p, x) \approx \|p - x\|^2 / 2$  for  $\|p - x\|$  small.

finite point cloud) can be quantified by the Hausdorff distance  $d_H(S, P) := \|f_S - f_P\|_\infty = \inf_{x \in \mathbb{R}^d} |f_S(x) - f_P(x)|$  of their distance functions. The intuition behind the inference of topology is that if  $d_H(S, P)$  is small, thus  $f_S$  and  $f_P$  are close, and subsequently,  $S$ ,  $(S)^r$  and  $(P)^r$  carry the same topology for an appropriate scale  $r$ . In other words, to compare the topology of offsets  $(S)^r$  and  $(P)^r$ , we require Hausdorff stability with respect to their distance functions  $f_S$  and  $f_P$ . An example of an offset-based topological inference result is formally stated as follows (as a particular version of the reconstruction Theorem 4.6 in [11]), where the *reach* of a compact set  $S$ ,  $\text{reach}(S)$ , is defined as the minimum distance between  $S$  and its medial axis [42].

► **Theorem 1** (Reconstruction from  $f_P$  [11]). *Let  $S, P \subset \mathbb{R}^d$  be compact sets such that  $\text{reach}(S) > R$  and  $\varepsilon := d_H(S, P) < R/17$ . Then  $(S)^\eta$  and  $(P)^r$  are homotopy equivalent for sufficiently small  $\eta$  (e.g.,  $0 < \eta < R$ ) if  $4\varepsilon \leq r < R - 3\varepsilon$ .*

Here  $\eta < R$  ensures that the topological properties of  $(S)^\eta$  and  $(S)^r$  are the same, and the  $\varepsilon$  parameter ensures  $(S)^r$  and  $(P)^r$  are close. Typically  $\varepsilon$  is tied to the density with which a point cloud  $P$  is sampled from  $S$ .

For function  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^+$  to be *distance-like* it should satisfy the following properties:

- (D1)  $\phi$  is 1-Lipschitz: For all  $x, y \in \mathbb{R}^d$ ,  $|\phi(x) - \phi(y)| \leq \|x - y\|$ .
- (D2)  $\phi^2$  is 1-semiconcave: The map  $x \in \mathbb{R}^d \mapsto (\phi(x))^2 - \|x\|^2$  is concave.
- (D3)  $\phi$  is proper:  $\phi(x)$  tends to the infimum of its domain (e.g.,  $\infty$ ) as  $x$  tends to infinity.

In addition to the Hausdorff stability property stated above, as explained in [12],  $f_S$  is distance-like. These three properties are paramount for geometric inference (e.g. [11, 41]).

(D1) ensures that  $f_S$  is differentiable almost everywhere and the medial axis of  $S$  has zero  $d$ -volume [12]; and (D2) is a crucial technical tool, e.g., in proving the existence of the flow of the gradient of the distance function for topological inference [11].

**Distance to a measure.** Given a probability measure  $\mu$  on  $\mathbb{R}^d$  and a parameter  $m_0 > 0$  smaller than the total mass of  $\mu$ , the *distance to a measure*  $d_{\mu, m_0}^{\text{CCM}} : \mathbb{R}^n \rightarrow \mathbb{R}^+$  [12] is defined for any point  $x \in \mathbb{R}^d$  as

$$d_{\mu, m_0}^{\text{CCM}}(x) = \left( \frac{1}{m_0} \int_{m=0}^{m_0} (\delta_{\mu, m}(x))^2 dm \right)^{1/2}, \quad \text{where } \delta_{\mu, m}(x) = \inf \{ r > 0 : \mu(\bar{B}_r(x)) \geq m \},$$

and where  $B_r(x)$  is a ball of radius  $r$  centered at  $x$  and  $\bar{B}_r(x)$  is its closure. It has been shown in [12] that  $d_{\mu, m_0}^{\text{CCM}}$  is a distance-like function (satisfying (D1), (D2), and (D3)), and:

- (M4) [Stability] For probability measures  $\mu$  and  $\nu$  on  $\mathbb{R}^d$  and  $m_0 > 0$ , then  $\|d_{\mu, m_0}^{\text{CCM}} - d_{\nu, m_0}^{\text{CCM}}\|_\infty \leq \frac{1}{\sqrt{m_0}} W_2(\mu, \nu)$ , where  $W_2$  is the *Wasserstein distance* [56].

Given a point set  $P$ , the sublevel sets of  $d_{\mu_P, m_0}^{\text{CCM}}$  can be described as the union of balls [35], and then one can algorithmically estimate the topology (e.g., persistence diagram) with weighted alpha-shapes [35] and weighted Rips complexes [7].

### 1.3 Our Results

We show how to estimate the topology (e.g., approximate persistence diagrams, infer homotopy of compact sets) using superlevel sets of the kernel density estimate of a point set  $P$ . We accomplish this by showing that a similar set of properties hold for the kernel distance with respect to a measure  $\mu$ , (in place of distance to a measure  $d_{\mu, m_0}^{\text{CCM}}$ ), defined as

$$d_\mu^K(x) = D_K(\mu, x) = \sqrt{\kappa(\mu, \mu) + \kappa(x, x) - 2\kappa(\mu, x)}.$$

This treats  $x$  as a probability measure represented by a Dirac mass at  $x$ . Specifically, we show  $d_\mu^K$  is distance-like (it satisfies (D1), (D2), and (D3)), so it inherits reconstruction properties of  $d_{\mu, m_0}^{CCM}$ . Moreover, it is stable with respect to the kernel distance:

- (K4) [Stability] If  $\mu$  and  $\nu$  are two measures on  $\mathbb{R}^d$ , then  $\|d_\mu^K - d_\nu^K\|_\infty \leq D_K(\mu, \nu)$ .

In addition, we show how to construct these topological estimates for  $d_\mu^K$  using weighted Rips complexes, following power distance machinery introduced in [7].

We also describe further advantages of the kernel distance. (i) Its sublevel sets conveniently map to the superlevel sets of a kernel density estimate. (ii) It is Lipschitz with respect to the smoothing parameter  $\sigma$  when the input  $x$  is fixed. (iii) As  $\sigma$  tends to  $\infty$  for any two probability measures  $\mu, \nu$ , the kernel distance is bounded by the Wasserstein distance:  $\lim_{\sigma \rightarrow \infty} D_K(\mu, \nu) \leq W_2(\mu, \nu)$ . (iv) It has a small coresets representation, which allows for sparse representation and efficient, scalable computation. In particular, an  $\varepsilon$ -kernel sample [38, 45, 58]  $Q$  of  $\mu$  is a finite point set whose size only depends on  $\varepsilon > 0$  and such that  $\max_{x \in \mathbb{R}^d} |\text{KDE}_\mu(x) - \text{KDE}_{\mu_Q}(x)| = \max_{x \in \mathbb{R}^d} |\kappa(\mu, x) - \kappa(\mu_Q, x)| \leq \varepsilon$ . These coresets preserve inference results and persistence diagrams.

## 2 Kernel Distance is Distance-Like

We prove  $d_\mu^K$  satisfies (D1), (D2), and (D3); hence it is distance-like. Recall we use the  $\sigma^2$ -normalized Gaussian kernel  $K_\sigma(p, x) = \sigma^2 \exp(-\|p - x\|^2/2\sigma^2)$ . For ease of exposition, unless otherwise noted, we will assume  $\sigma$  is fixed and write  $K$  instead of  $K_\sigma$ .

### 2.1 Semiconcave Property for $d_\mu^K$

► **Lemma 2** (D2).  $(d_\mu^K)^2$  is 1-semiconcave: the map  $x \mapsto (d_\mu^K(x))^2 - \|x\|^2$  is concave.

**Proof.** Let  $T(x) = (d_\mu^K(x))^2 - \|x\|^2$ . The proof will show that the second derivative of  $T$  along any direction is nonpositive. We can rewrite

$$\begin{aligned} T(x) &= \kappa(\mu, \mu) + \kappa(x, x) - 2\kappa(\mu, x) - \|x\|^2 \\ &= \kappa(\mu, \mu) + \kappa(x, x) - \int_{p \in \mathbb{R}^d} (2K(p, x) + \|x\|^2) d\mu(p). \end{aligned}$$

Note that both  $\kappa(\mu, \mu)$  and  $\kappa(x, x)$  are absolute constants, so we can ignore them in the second derivative. Furthermore, by setting  $t(p, x) = -2K(p, x) - \|x\|^2$ , the second derivative of  $T(x)$  is nonpositive if the second derivative of  $t(p, x)$  is nonpositive for all  $p, x \in \mathbb{R}^d$ . First note that the second derivative of  $-\|x\|^2$  is a constant  $-2$  in every direction. The second derivative of  $K(p, x)$  is symmetric about  $p$ , so we can consider the second derivative along any vector  $u = x - p$ ,

$$\frac{d^2}{du^2} t(p, x) = 2 \left( \frac{\|u\|^2}{\sigma^2} - 1 \right) \exp \left( -\frac{\|u\|^2}{2\sigma^2} \right) - 2.$$

This reaches its maximum value at  $\|u\| = \|x - p\| = \sqrt{3}\sigma$  where it is  $4 \exp(-3/2) - 2 \approx -1.1$ ; this follows by setting the derivative of  $s(y) = 2(y - 1) \exp(-y/2) - 2$  to 0, ( $\frac{d}{dy} s(y) = (1/2)(3 - y) \exp(-y/2)$ ), substituting  $y = \|u\|^2/\sigma^2$ . ◀

### 2.2 Lipschitz Property for $d_\mu^K$

We generalize a (folklore, see [12]) relation between semiconcave and Lipschitz functions. A function  $f$  is  $\ell$ -semiconcave if the function  $T(x) = (f(x))^2 - \ell\|x\|^2$  is concave.

► **Lemma 3.** *Consider a twice-differentiable function  $g$  and a parameter  $\ell \geq 1$ . If  $(g(x))^2$  is  $\ell$ -semiconcave, then  $g(x)$  is  $\ell$ -Lipschitz.*

We can now state the following lemma as a corollary of Lemma 2 and Lemma 3.

► **Lemma 4 (D1).**  $d_\mu^K$  is 1-Lipschitz on its input.

### 2.3 Properness of $d_\mu^K$

Finally, for  $d_\mu^K$  to be distance-like, we need to show it is proper when its range is restricted to be less than  $c_\mu := \sqrt{\kappa(\mu, \mu) + \kappa(x, x)}$ . This is required for a distance-like version ([12], Proposition 4.2) of the Isotopy Lemma ([34], Proposition 1.8). Here, the value of  $c_\mu$  depends on  $\mu$  not on  $x$  since  $\kappa(x, x) = K(x, x) = \sigma^2$ .

► **Lemma 5 (D3).**  $d_\mu^K$  is proper.

We delay the proof to the full version [47]. The main technical difficulty comes in mapping standard definitions and approaches for distance functions to our function  $d_\mu^K$  with a restricted range. We use two more general, but equivalent definitions of a proper map and the notion of *escape to infinity*. Specifically, a sequence  $\{p_i\}$  in  $\mathbb{X}$  *escapes to infinity* if for every compact set  $G \subset \mathbb{X}$ , there are at most finitely many values of  $i$  for which  $p_i \in G$  ([39], page 46).

By the definition of properness, Lemma 5 implies that it is a closed map and its levelset at any value  $a \in [0, c_\mu)$  is compact. This also means that the sublevel set of  $d_\mu^K$  (for ranges  $[0, a) \subset [0, c_\mu)$ ) is compact. Since the levelset (sublevel set) of  $d_\mu^K$  corresponds to the levelset (superlevel set) of  $\text{KDE}_\mu$ , we have the following corollary.

► **Corollary 6.** *The superlevel sets of  $\text{KDE}_\mu$  for all ranges with threshold  $a > 0$ , are compact.*

The result in [25] shows that given a measure  $\mu_P$  defined by a point set  $P$  of size  $n$ , the  $\text{KDE}_{\mu_P}$  has polynomial in  $n$  modes; hence the superlevel sets of  $\text{KDE}_{\mu_P}$  are compact in this setting. The above corollary is a more general statement as it holds for any measure.

## 3 Power Distance using Kernel Distance

A *power distance* using  $d_\mu^K$  is defined with a point set  $P \subset \mathbb{R}^d$  and a metric  $d(\cdot, \cdot)$  on  $\mathbb{R}^d$ ,

$$f_P(\mu, x) = \sqrt{\min_{p \in P} (d(p, x)^2 + d_\mu^K(p)^2)}.$$

A point  $x \in \mathbb{R}^d$  takes the distance under  $d(p, x)$  to the closest  $p \in P$ , plus a weight from  $d_\mu^K(p)$ ; thus a sublevel set of  $f_P(\mu, \cdot)$  is defined by a union of balls. We consider a particular choice of the distance  $d(p, x) := D_K(p, x)$  which leads to a kernel version of power distance

$$f_P^K(\mu, x) = \sqrt{\min_{p \in P} (D_K(p, x)^2 + d_\mu^K(p)^2)}.$$

In Section 4.2 we use  $f_P^K(\mu, x)$  to adapt the construction introduced in [7] to approximate the persistence diagram of the sublevel sets of  $d_\mu^K$ , using a weighted Rips filtration of  $f_P^K(\mu, x)$ .

Given a measure  $\mu$ , let  $p_+ = \arg \max_{q \in \mathbb{R}^d} \kappa(\mu, q)$ , and let  $P_+ \subset \mathbb{R}^d$  be a point set that contains  $p_+$ . We show below, in Theorem 11 and Theorem 8, that  $\frac{1}{\sqrt{2}} d_\mu^K(x) \leq f_{P_+}^K(\mu, x) \leq \sqrt{14} d_\mu^K(x)$ . However, constructing  $p_+$  exactly seems quite difficult.

Now consider an empirical measure  $\mu_P$  defined by a point set  $P$ . We show (in the full version [47]) how to construct a point  $\hat{p}_+$  (that approximates  $p_+$ ) such that  $D_K(P, \hat{p}_+) \leq (1 + \delta)D_K(P, p_+)$  for any  $\delta > 0$ . For a point set  $P$ , the *median concentration*  $\Lambda_P$  is a radius such that no point  $p \in P$  has more than half of the points of  $P$  within  $\Lambda_P$ , and the *spread*  $\beta_P$  is the ratio between the longest and shortest pairwise distances. The runtime is polynomial in  $n$  and  $1/\delta$  assuming  $\beta_P$  is bounded, and that  $\sigma/\Lambda_P$  and  $d$  are constants.

Then we consider  $\hat{P}_+ = P \cup \{\hat{p}_+\}$ , where  $\hat{p}_+$  is found with  $\delta = 1/2$  in the above construction. Then we can provide the following multiplicative bound, proven in Theorem 12. The lower bound holds independent of the choice of  $P$  as shown in Theorem 8.

► **Theorem 7.** For any point set  $P \subset \mathbb{R}^d$  and point  $x \in \mathbb{R}^d$ , with empirical measure  $\mu_P$  defined by  $P$ , then  $\frac{1}{\sqrt{2}}d_{\mu_P}^K(x) \leq f_{\hat{P}_+}^K(\mu_P, x) \leq \sqrt{71}d_{\mu_P}^K(x)$ .

### 3.1 Kernel Power Distance for a Measure $\mu$

First consider the case for a kernel power distance  $f_P^K(\mu, x)$  where  $\mu$  is an arbitrary measure.

► **Theorem 8.** For measure  $\mu$ , point set  $P \subset \mathbb{R}^d$ , and  $x \in \mathbb{R}^d$ ,  $D_K(\mu, x) \leq \sqrt{2}f_P^K(\mu, x)$ .

**Proof.** Let  $p = \arg \min_{q \in P} (D_K(q, x)^2 + D_K(\mu, q)^2)$ . Then we can use the triangle inequality and  $(D_K(\mu, p) - D_K(p, x))^2 \geq 0$  to show

$$D_K(\mu, x)^2 \leq (D_K(\mu, p) + D_K(p, x))^2 \leq 2(D_K(\mu, p)^2 + D_K(p, x)^2) = 2f_P^K(\mu, x)^2. \quad \blacktriangleleft$$

► **Lemma 9.** For measure  $\mu$ , point set  $P \subset \mathbb{R}^d$ , point  $p \in P$ , and point  $x \in \mathbb{R}^d$  then  $f_P^K(\mu, x)^2 \leq 2D_K(\mu, x)^2 + 3D_K(p, x)^2$ .

**Proof.** Again, we can reach this result with the triangle inequality.

$$\begin{aligned} f_P^K(\mu, x)^2 &\leq D_K(\mu, p)^2 + D_K(p, x)^2 \\ &\leq (D_K(\mu, x) + D_K(p, x))^2 + D_K(p, x)^2 \\ &\leq 2D_K(\mu, x)^2 + 3D_K(p, x)^2. \end{aligned} \quad \blacktriangleleft$$

Recall the definition of a point  $p_+ = \arg \max_{q \in \mathbb{R}^d} \kappa(\mu, q)$ .

► **Lemma 10.** For any measure  $\mu$  and point  $x, p_+ \in \mathbb{R}^d$  we have  $D_K(p_+, x) \leq 2D_K(\mu, x)$ .

**Proof.** Since  $x$  is a point in  $\mathbb{R}^d$ ,  $\kappa(\mu, x) \leq \kappa(\mu, p_+)$  and thus  $D_K(\mu, x) \geq D_K(\mu, p_+)$ . Then by triangle inequality of  $D_K$  to see that  $D_K(p_+, x) \leq D_K(\mu, x) + D_K(\mu, p_+) \leq 2D_K(\mu, x)$ .  $\blacktriangleleft$

► **Theorem 11.** For any measure  $\mu$  in  $\mathbb{R}^d$  and any point  $x \in \mathbb{R}^d$ , using the point  $p_+ = \arg \max_{q \in \mathbb{R}^d} \kappa(\mu, q)$  then  $f_{\{p_+\}}^K(\mu, x) \leq \sqrt{14}D_K(\mu, x)$ .

**Proof.** Combine Lemma 9 and Lemma 10 as

$$f_{\{p_+\}}^K(\mu, x)^2 \leq 2D_K(\mu, x)^2 + 3D_K(p_+, x)^2 \leq 2D_K(\mu, x)^2 + 3(4D_K(\mu, x)^2) = 14D_K(\mu, x)^2. \quad \blacktriangleleft$$

We now need two properties of the point set  $P$  to reach our bound, namely, the spread  $\beta_P$  and the median concentration  $\Lambda_P$ . Typically  $\log(\beta_P)$  is not too large, and it makes sense to choose  $\sigma$  so  $\sigma/\Lambda_P \leq 1$ , or at least  $\sigma/\Lambda_P = O(1)$ .

► **Theorem 12.** Consider any point set  $P \subset \mathbb{R}^d$  of size  $n$ , with measure  $\mu_P$ , spread  $\beta_P$ , and median concentration  $\Lambda_P$ . We can construct a point set  $\hat{P}_+ = P \cup \hat{p}_+$  in  $O(n^2((\sigma/\Lambda_P\delta)^d + \log(\beta)))$  time such that for any point  $x$ ,  $f_{\hat{P}_+}^K(\mu_P, x) \leq \sqrt{71}D_K(\mu_P, x)$ .

**Proof.** We use a result from the full version [47] to find a point  $\hat{p}_+$  such that  $D_K(P, \hat{p}_+) \leq (3/2)D_K(P, p_+)$  in the stated runtime. Thus for any  $x \in \mathbb{R}^d$ , using the triangle inequality

$$\begin{aligned} D_K(\hat{p}_+, x) &\leq D_K(\hat{p}_+, p_+) + D_K(p_+, x) \leq D_K(\mu_P, \hat{p}_+) + D_K(\mu_P, p_+) + D_K(p_+, x) \\ &\leq (5/2)D_K(\mu_P, p_+) + D_K(p_+, x). \end{aligned}$$

Now combine this with Lemma 9 and Lemma 10 as

$$\begin{aligned} f_{\hat{P}_+}^K(\mu_P, x)^2 &\leq 2D_K(\mu_P, x)^2 + 3D_K(\hat{p}_+, x)^2 \\ &\leq 2D_K(\mu_P, x)^2 + 3((5/2)D_K(\mu_P, x) + D_K(p_+, x))^2 \\ &\leq 2D_K(\mu_P, x)^2 + 3((25/4) + (5/2))D_K(\mu_P, x)^2 + (1 + 5/2)D_K(p_+, x)^2 \\ &= (113/4)D_K(\mu_P, x)^2 + (21/2)D_K(p_+, x)^2 \\ &\leq (113/4)D_K(\mu_P, x)^2 + (21/2)(4D_K(\mu_P, x)^2) < 71D_K(\mu_P, x)^2. \quad \blacktriangleleft \end{aligned}$$

## 4 Reconstruction and Topological Estimation using Kernel Distance

Now applying distance-like properties from Section 2 and the power distance properties of Section 3 we can apply known reconstruction results to the kernel distance.

### 4.1 Homotopy Equivalent Reconstruction using $d_\mu^K$

We have shown that the kernel distance function  $d_\mu^K$  is a distance-like function. Therefore the reconstruction theory for a distance-like function [12] holds in the setting of  $d_\mu^K$ . We state the following two corollaries for completeness, whose proofs follow from the proofs of Proposition 4.2 and Theorem 4.6 in [12]. Before their formal statement, we need some notation adapted from [12] to make these statements precise. Let  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^+$  be a distance-like function. A point  $x \in \mathbb{R}^d$  is an  $\alpha$ -critical point if  $\phi^2(x+h) \leq \phi^2(x) + 2\alpha\|h\|\phi(x) + \|h\|^2$  with  $\alpha \in [0, 1]$ ,  $\forall h \in \mathbb{R}^d$ . Let  $(\phi)^r = \{x \in \mathbb{R}^d \mid \phi(x) \leq r\}$  denote the sublevel set of  $\phi$ , and let  $(\phi)^{[r_1, r_2]} = \{x \in \mathbb{R}^d \mid r_1 \leq \phi(x) \leq r_2\}$  denote all points at levels in the range  $[r_1, r_2]$ . For  $\alpha \in [0, 1]$ , the  $\alpha$ -reach of  $\phi$  is the maximum  $r$  such that  $(\phi)^r$  has no  $\alpha$ -critical point, denoted as  $\text{reach}_\alpha(\phi)$ . When  $\alpha = 1$ ,  $\text{reach}_1$  coincides with  $\text{reach}$  introduced in [31].

► **Theorem 13** (Isotopy lemma on  $d_\mu^K$ ). *Let  $r_1 < r_2$  be two positive numbers such that  $d_\mu^K$  has no critical points in  $(d_\mu^K)^{[r_1, r_2]}$ . Then all the sublevel sets  $(d_\mu^K)^r$  are isotopic for  $r \in [r_1, r_2]$ .*

► **Theorem 14** (Reconstruction on  $d_\mu^K$ ). *Let  $d_\mu^K$  and  $d_\nu^K$  be two kernel distance functions such that  $\|d_\mu^K - d_\nu^K\|_\infty \leq \varepsilon$ . Suppose  $\text{reach}_\alpha(d_\mu^K) \geq R$  for some  $\alpha > 0$ . Then  $\forall r \in [4\varepsilon/\alpha^2, R - 3\varepsilon]$ , and  $\forall \eta \in (0, R)$ , the sublevel sets  $(d_\mu^K)^\eta$  and  $(d_\nu^K)^r$  are homotopy equivalent for  $\varepsilon \leq R/(5 + 4/\alpha^2)$ .*

### 4.2 Constructing Topological Estimates using $d_\mu^K$

In order to actually construct a topological estimate using the kernel distance  $d_\mu^K$ , one needs to be able to compute quantities related to its sublevel sets, in particular, to compute the persistence diagram of the sub-level sets filtration of  $d_\mu^K$ . Now we describe such tools needed for the kernel distance based on machinery recently developed by Buchet et al. [7], which shows how to approximate the persistent homology of distance-to-a-measure for any metric space via a power distance construction. Then using similar constructions, we can use the weighted Rips filtration to approximate the persistence diagram of the kernel distance.



To state our results, first we require some technical notions and assume basic knowledge on persistent homology (see [26, 27] for a readable background). Given a metric space  $\mathbb{X}$  with the distance  $d_{\mathbb{X}}(\cdot, \cdot)$ , a set  $P \subseteq \mathbb{X}$  and a function  $w : P \rightarrow \mathbb{R}$ , the (general) *power distance*  $f$  associated with  $(P, w)$  is defined as  $f(x) = \sqrt{\min_{p \in P} (d_{\mathbb{X}}(p, x)^2 + w(p)^2)}$ . Now given the set  $(P, w)$  and its corresponding power distance  $f$ , one could use the weighted Rips filtration to approximate the persistence diagram of  $w$ . Consider the sublevel set of  $f$ ,  $f^{-1}((-\infty, \alpha])$ . It is the union of balls centered at points  $p \in P$  with radius  $r_p(\alpha) = \sqrt{\alpha^2 - w(p)^2}$  for each  $p$ . The weighted Čech complex  $C_{\alpha}(P, w)$  for parameter  $\alpha$  is the union of simplices  $s$  such that  $\bigcap_{p \in s} B(p, r_p(\alpha)) \neq \emptyset$ . The weighted Rips complex  $R_{\alpha}(P, w)$  for parameter  $\alpha$  is the maximal complex whose 1-skeleton is the same as  $C_{\alpha}(P, w)$ . The corresponding weighted Rips filtration is denoted as  $\{R_{\alpha}(P, w)\}$ .

Setting  $w := d_{\mu_P}^K$  and given point set  $\hat{P}_+$  described in Section 3, consider the weighted Rips filtration  $\{R_{\alpha}(\hat{P}_+, d_{\mu_P}^K)\}$  based on the kernel power distance,  $f_{\hat{P}_+}^K$ . We view the persistence diagrams on a logarithmic scale, that is, we change coordinates of points following the mapping  $(x, y) \mapsto (\ln x, \ln y)$ .  $d_B^{\ln}$  denotes the corresponding bottleneck distance between persistence diagrams. We show in the full version [47] that persistence diagrams  $\text{Dgm}(d_{\mu_P}^K)$  and  $\text{Dgm}(\{R_{\alpha}(\hat{P}_+, d_{\mu_P}^K)\})$  follow technical tameness conditions and are well-defined. We now state a corollary of Theorem 7.

► **Corollary 15.** *The weighted Rips filtration  $\{R_{\alpha}(\hat{P}_+, d_{\mu_P}^K)\}$  can be used to approximate the persistence diagram of  $d_{\mu_P}^K$  such that  $d_B^{\ln}(\text{Dgm}(d_{\mu_P}^K), \text{Dgm}(\{R_{\alpha}(\hat{P}_+, d_{\mu_P}^K)\})) \leq \ln(2\sqrt{71})$ .*

**Proof.** To prove that two persistence diagrams are close, one could prove that their filtration are interleaved [9], that is, two filtrations  $\{U_{\alpha}\}$  and  $\{V_{\alpha}\}$  are  $\varepsilon$ -interleaved if for any  $\alpha$ ,  $U_{\alpha} \subseteq V_{\alpha+\varepsilon} \subseteq U_{\alpha+2\varepsilon}$ . The results of Theorem 7 implies an  $\sqrt{71}$  multiplicative interleaving. Therefore for any  $\alpha \in \mathbb{R}$ ,

$$(d_{\mu_P}^K)^{-1}((-\infty, \alpha]) \subset (f_{\hat{P}_+}^K)^{-1}((-\infty, \sqrt{2}\alpha]) \subset (d_{\mu_P}^K)^{-1}((-\infty, \sqrt{71}\sqrt{2}\alpha]).$$

On a logarithmic scale (by taking the natural log of both sides), such interleaving becomes additive,

$$\ln d_{\mu_P}^K - \sqrt{2} \leq \ln f_{\hat{P}_+}^K \leq \ln d_{\mu_P}^K + \sqrt{71}.$$

Theorem 4 of [13] implies

$$d_B^{\ln}(\text{Dgm}(d_{\mu_P}^K), \text{Dgm}(f_{\hat{P}_+}^K)) \leq \sqrt{71}.$$

In addition, by the Persistent Nerve Lemma ([19], Theorem 6 of [51], an extension of the Nerve Theorem [36]), the sublevel sets filtration of  $d_{\mu}^K$ , which correspond to unions of balls of increasing radius, has the same persistent homology as the nerve filtration of these balls (which, by definition, is the Čech filtration). Finally, there exists a multiplicative interleaving between weighted Rips and Čech complexes (Proposition 31 of [13]),  $C_{\alpha} \subseteq R_{\alpha} \subseteq C_{2\alpha}$ . We then obtain the following bounds on persistence diagrams,

$$d_B^{\ln}(\text{Dgm}(f_{\hat{P}_+}^K), \text{Dgm}(\{R_{\alpha}(P_+, d_{\mu_P}^K)\})) \leq \ln(2).$$

We use triangle inequality to obtain the final result:

$$d_B^{\ln}(\text{Dgm}(d_{\mu_P}^K), \text{Dgm}(\{R_{\alpha}(P_+, d_{\mu_P}^K)\})) \leq \ln(2\sqrt{71}). \quad \blacktriangleleft$$

Based on Corollary 15, we have an algorithm that approximates the persistent homology of the sublevel set filtration of  $d_{\mu}^K$  by constructing the weighted Rips filtration corresponding to the kernel-based power distance and computing its persistent homology.

### 4.3 Distance to the Support of a Measure vs. Kernel Distance

Suppose  $\mu$  is a uniform measure on a compact set  $S$  in  $\mathbb{R}^d$ . We now compare the kernel distance  $d_\mu^K$  with the distance function  $f_S$  to the support  $S$  of  $\mu$ . We show how  $d_\mu^K$  approximates  $f_S$ , and thus allows one to infer geometric properties of  $S$  from samples from  $\mu$ .

A generalized gradient and its corresponding flow associated with a distance function are described in [11] and later adapted for distance-like functions in [12]. Let  $f_S : \mathbb{R}^d \rightarrow \mathbb{R}$  be a distance function associated with a compact set  $S$  of  $\mathbb{R}^d$ . It is not differentiable on the medial axis of  $S$ . A *generalized gradient function*  $\nabla_S : \mathbb{R}^d \rightarrow \mathbb{R}^d$  coincides with the usual gradient of  $f_S$  where  $f_S$  is differentiable, and is defined everywhere and can be integrated into a continuous flow  $\Phi^t : \mathbb{R}^d \rightarrow \mathbb{R}^d$  that points away from  $S$ . Let  $\gamma$  be an integral (flow) line. The following technical lemma is proved in the full version [47].

► **Lemma 16.** *Given any flow line  $\gamma$  associated with the generalized gradient function  $\nabla_S$ ,  $d_\mu^K(x)$  is strictly monotonically increasing along  $\gamma$  for  $x$  sufficiently far away from the medial axis of  $S$ , for  $\sigma \leq \frac{R}{6\Delta_G}$  and  $f_S(x) \in (0.014R, 2\sigma)$ . Here  $B(\sigma/2)$  denotes a ball of radius  $\sigma/2$ ,  $G := \frac{\text{Vol}(B(\sigma/2))}{\text{Vol}(S)}$ ,  $\Delta_G := \sqrt{12 + 3 \ln(4/G)}$  and suppose  $R := \min(\text{reach}(S), \text{reach}(\mathbb{R}^d \setminus S)) > 0$ .*

The strict monotonicity of  $d_\mu^K$  along the flow line under the conditions in Lemma 16 makes it possible to define a deformation retract of the sublevel sets of  $d_\mu^K$  onto sublevel sets of  $f_S$ . Such a deformation retract defines a special case of homotopy equivalence between the sublevel sets of  $d_\mu^K$  and sublevel sets of  $f_S$ . Consider a sufficiently large point set  $P \in \mathbb{R}^d$  sampled from  $\mu$ , and its induced measure  $\mu_P$ . We can then also invoke Theorem 14 and a sampling bound (see Section 6) to show homotopy equivalence between the sublevel sets of  $f_S$  and  $d_{\mu_P}^K$ .

## 5 Stability Properties for the Kernel Distance to a Measure

► **Lemma 17 (K4).** *For two measures  $\mu$  and  $\nu$  on  $\mathbb{R}^d$  we have  $\|d_\mu^K - d_\nu^K\|_\infty \leq D_K(\mu, \nu)$ .*

**Proof.** Since  $D_K(\cdot, \cdot)$  is a metric, then by triangle inequality, for any  $x \in \mathbb{R}^d$  we have  $D_K(\mu, x) \leq D_K(\mu, \nu) + D_K(\nu, x)$  and  $D_K(\nu, x) \leq D_K(\nu, \mu) + D_K(\mu, x)$ . Therefore for any  $x \in \mathbb{R}^d$  we have  $|D_K(\mu, x) - D_K(\nu, x)| \leq D_K(\mu, \nu)$ , proving the claim. ◀

Both the Wasserstein and kernel distance are *integral probability metrics* [54], so (M4) and (K4) are both interesting, but not easily comparable. We now attempt to reconcile this.

### 5.1 Comparing $D_K$ to $W_2$

► **Lemma 18.** *There is no Lipschitz constant  $\gamma$  such that for any two probability measures  $\mu$  and  $\nu$  we have  $W_2(\mu, \nu) \leq \gamma D_K(\mu, \nu)$ .*

**Proof.** Consider two measures  $\mu$  and  $\nu$  which are almost identical: the only difference is some mass of measure  $\tau$  is moved from its location in  $\mu$  a distance  $n$  in  $\nu$ . The Wasserstein distance requires a transportation plan that moves this  $\tau$  mass in  $\nu$  back to where it was in  $\mu$  with cost  $\tau \cdot \Omega(n)$  in  $W_2(\mu, \nu)$ . On the other hand,  $D_K(\mu, \nu) = \sqrt{\kappa(\mu, \mu) + \kappa(\nu, \nu) - 2\kappa(\mu, \nu)} \leq \sqrt{\sigma^2 + \sigma^2 - 2 \cdot 0} = \sqrt{2}\sigma$  is bounded. ◀

We conjecture for any two probability measures  $\mu$  and  $\nu$  that  $D_K(\mu, \nu) \leq W_2(\mu, \nu)$ . This would show that  $d_\mu^K$  is at least as stable as  $d_{\mu, m_0}^{\text{CM}}$  since a bound on  $W_2(\mu, \nu)$  would also

bound  $D_K(\mu, \nu)$ , but not vice versa. We leave much of the technical details from this section to the full version [47]. We start with a special case.

► **Lemma 19.** *Consider two probability measures  $\mu$  and  $\nu$  on  $\mathbb{R}^d$  where  $\nu$  is represented by a Dirac mass at a point  $x \in \mathbb{R}^d$ . Then  $d_\mu^K(x) = D_K(\mu, \nu) \leq W_2(\mu, \nu)$  for any  $\sigma > 0$ , where the equality only holds when  $\mu$  is also a Dirac mass at  $x$ .*

Next we show that if  $\nu$  is not a unit Dirac, then this inequality holds in the limit as  $\sigma$  goes to infinity. The technical work is making precise how  $\sigma^2 - K(p, x) \leq \|x - p\|^2/2$  and how this compares to bounds on  $D_K(\mu, \nu)$  and  $W_2(\mu, \nu)$ .

► **Lemma 20.** *For any  $p, q \in \mathbb{R}^d$  we have  $K(p, q) = \sigma^2 - \frac{\|p - q\|^2}{2} + \sum_{i=2}^{\infty} \frac{(-\|p - q\|^2)^i}{2^{i+1}\sigma^{2i-2}i!}$ .*

**Proof.** We use the Taylor expansion of  $e^x = \sum_{i=0}^{\infty} x^i/i! = 1 + x + \sum_{i=2}^{\infty} x^i/i!$ . Then it is easy to see

$$K(p, q) = \sigma^2 \exp\left(-\frac{\|p - q\|^2}{2\sigma^2}\right) = \sigma^2 - \frac{\|p - q\|^2}{2} + \sum_{i=2}^{\infty} \frac{(-\|p - q\|^2)^i}{2^i\sigma^{2i-2}i!}. \quad \blacktriangleleft$$

This lemma illustrates why the choice of coefficient of  $\sigma^2$  is convenient. Since then  $\sigma^2 - K(p, q)$  acts like  $\frac{1}{2}\|p - q\|^2$ , and becomes closer as  $\sigma$  increases. Define  $\bar{\mu} = \int_p p \cdot d\mu(p)$  to represent the mean point of measure  $\mu$ .

► **Theorem 21.** *For any two probability measures  $\mu$  and  $\nu$  defined on  $\mathbb{R}^d$   $\lim_{\sigma \rightarrow \infty} D_K(\mu, \nu) = \|\bar{\mu} - \bar{\nu}\|$  and  $\|\bar{\mu} - \bar{\nu}\| \leq W_2(\mu, \nu)$ . Thus  $\lim_{\sigma \rightarrow \infty} D_K(\mu, \nu) \leq W_2(\mu, \nu)$ .*

### 5.2 Kernel Distance Stability with Respect to $\sigma$

We now explore the Lipschitz properties of  $d_\mu^K$  with respect to the noise parameter  $\sigma$ . We argue any distance function that is robust to noise needs some parameter to address how many outliers to ignore or how far away a point is to be considered as an outlier. Such a parameter in  $d_{\mu, m_0}^{CCM}$  is  $m_0$  which controls the amount of measure  $\mu$  to be used in the distance.

Here we show that  $d_\mu^K$  has a particularly nice property, that it is Lipschitz with respect to the choice of  $\sigma$  for any fixed  $x$ . Many details are deferred to the full version [47].

► **Lemma 22.** *Let  $h(\sigma, z) = \exp(-z^2/2\sigma^2)$ . We can bound  $h(\sigma, z) \leq 1$ ,  $\frac{d}{d\sigma} h(\sigma, z) \leq (2/e)/\sigma$  and  $\frac{d^2}{d\sigma^2} h(\sigma, z) \leq (18/e^3)/\sigma^2$  over any choice of  $z > 0$ .*

► **Theorem 23.** *For any measure  $\mu$  defined on  $\mathbb{R}^d$  and  $x \in \mathbb{R}^d$ ,  $d_\mu^K(x)$  is  $\ell$ -Lipschitz with respect to  $\sigma$ , for  $\ell = 18/e^3 + 8/e + 2 < 6$ .*

**Proof.** (Sketch) Recall that  $m_{\mu, \nu}$  is the product measure of any  $\mu$  and  $\nu$ . Define  $M_{\mu, \nu}$  as  $M_{\mu, \nu}(p, q) = m_{\mu, \mu}(p, q) + m_{\nu, \nu}(p, q) - 2m_{\mu, \nu}(p, q)$ . It is useful to define a function  $f_x(\sigma)$  as

$$f_x(\sigma) = \int_{(p, q)} \exp\left(\frac{-\|p - q\|^2}{2\sigma^2}\right) dM_{\mu, \delta_x}(p, q)$$

$$F(\sigma) = (d_\mu^K(x))^2 - \ell\|\sigma\|^2 = \sigma^2 f_x(\sigma) - \ell\sigma^2.$$

Now  $d_\mu^K(x) = \sigma\sqrt{f_x(\sigma)}$ . Now to prove  $d_\mu^K(x)$  is  $\ell$ -Lipschitz, we can show that  $(d_\mu^K)^2$  is  $\ell$ -semiconcave with respect to  $\sigma$ , and apply Lemma 3. This boils down to showing the second derivative of  $F(\sigma)$  is always non-positive.

$$\frac{d^2}{d\sigma^2} F(\sigma) = \sigma^2 \frac{d^2}{d\sigma^2} f_x(\sigma) + 4\sigma \frac{d}{d\sigma} f_x(\sigma) + 2f_x(\sigma) - 2\ell.$$

First we note that for any distribution  $\mu$  and Dirac delta that  $\int_{(p,q)} c \cdot dM_{\mu,\delta_x}(p, q) \leq 2c$ . Thus since  $\exp\left(\frac{-\|p-q\|^2}{2\sigma^2}\right)$  is in  $[0, 1]$  for all choices of  $p, q$ , and  $\sigma > 0$ , then  $0 \leq f_x(\sigma) \leq 2$  and  $2f_x(\sigma) \leq 4$ . This bounds the third term in  $\frac{d^2}{d\sigma^2}F(\sigma)$ , we now need to use a similar approach to bound the first and second terms. Using Lemma 22 to obtain

$$\frac{d^2}{d\sigma^2}F(\sigma) \leq 36/e^3 + 16/e + 4 - 2(18/e^3 + 8/e + 2) = 0. \quad \blacktriangleleft$$

**Lipschitz in  $m_0$  for  $d_{\mu, m_0}^{\text{CCM}}$ .** There is no Lipschitz property for  $d_{\mu, m_0}^{\text{CCM}}$ , with respect to  $m_0$ , independent of  $\mu$ . Consider a measure  $\mu_P$  for point set  $P \subset \mathbb{R}$  consisting of two points at  $a = 0$  and at  $b = \Delta$ . When  $m_0 = 1/2 + \alpha$  for  $\alpha > 0$ , then  $d_{\mu_P, m_0}^{\text{CCM}}(a) = \alpha\Delta/(1/2 + \alpha)$  and  $\frac{d}{dm_0}d_{\mu_P, m_0}^{\text{CCM}}(a) = \frac{d}{d\alpha}d_{\mu_P, \frac{1}{2} + \alpha}^{\text{CCM}}(a) = \frac{(1/2 + 2\alpha)\Delta}{(1/2 + \alpha)^2}$ , which is maximized as  $\alpha$  approaches 0 with an infimum of  $2\Delta$ . Hence the Lipschitz constant for  $d_{\mu_P, m_0}^{\text{CCM}}$  with respect to  $m_0$  is  $2\Delta_P$  where  $\Delta_P = \max_{p, p' \in P} \|p - p'\|$ .

## 6 Algorithmic and Approximation Observations

**Kernel coresets.** The kernel distance is robust under random samples [38]. Specifically, if  $Q$  is a point set randomly chosen from  $\mu$  of size  $O((1/\varepsilon^2)(d + \log(1/\delta)))$  then  $\|\text{KDE}_\mu - \text{KDE}_Q\|_\infty \leq \varepsilon$  with probability at least  $1 - \delta$ . We call such a subset  $Q$  and  $\varepsilon$ -kernel sample of  $(\mu, K)$ . Furthermore, it is also possible to construct  $\varepsilon$ -kernel samples  $Q$  with even smaller size of  $|Q| = O(((1/\varepsilon)\sqrt{\log(1/\varepsilon\delta)})^{2d/(d+2)})$  [45]; in particular in  $\mathbb{R}^2$  the required size is  $|Q| = O((1/\varepsilon)\sqrt{\log(1/\varepsilon\delta)})$ . Exploiting the above constructions, recent work [58] builds a data structure to allow for efficient approximate evaluations of  $\text{KDE}_P$  where  $|P| = 100,000,000$ .

These constructions of  $Q$  also immediately imply that  $\|(d_\mu^K)^2 - (d_Q^K)^2\|_\infty \leq 4\varepsilon$  since  $(d_\mu^K(x))^2 = \kappa(\mu, \mu) + \kappa(x, x) - 2\text{KDE}_\mu(x)$ , and both the first and third term incur at most  $2\varepsilon$  error in converting to  $\kappa(Q, Q)$  and  $2\text{KDE}_Q(x)$ , respectively. Thus, an  $(\varepsilon^2/4)$ -kernel sample  $Q$  of  $(\mu, K)$  implies that  $\|d_\mu^K - d_Q^K\|_\infty \leq \varepsilon$ .

This implies algorithms for geometric inference on enormous noisy data sets, or when input  $Q$  is assumed to be drawn iid from an unknown distribution  $\mu$ .

► **Corollary 24.** Consider a measure  $\mu$  defined on  $\mathbb{R}^d$ , a kernel  $K$ , and a parameter  $\varepsilon \leq R(5 + 4/\alpha^2)$ . We can create a coreset  $Q$  of size  $|Q| = O(((1/\varepsilon^2)\sqrt{\log(1/\varepsilon\delta)})^{2d/(d+2)})$  or randomly sample  $|Q| = O((1/\varepsilon^4)(d + \log(1/\delta)))$  points so, with probability at least  $1 - \delta$ , any sublevel set  $(d_\mu^K)^\eta$  is homotopy equivalent to  $(d_Q^K)^r$  for  $r \in [4\varepsilon/\alpha^2, R - 3\varepsilon]$  and  $\eta \in (0, R)$ .

**Stability of persistence diagrams.** Furthermore, the stability results on persistence diagrams [20] hold for kernel density estimates and kernel distance of  $\mu$  and  $Q$  (where  $Q$  is a coreset of  $\mu$  with the same size bounds as above). If  $\|f - g\|_\infty \leq \varepsilon$ , then  $d_B(\text{Dgm}(f), \text{Dgm}(g)) \leq \varepsilon$ , where  $d_B$  is the bottleneck distance between persistence diagrams.

► **Corollary 25.** Consider a measure  $\mu$  defined on  $\mathbb{R}^d$  and a kernel  $K$ . We can create a coreset  $Q$  of size  $|Q| = O(((1/\varepsilon)\sqrt{\log(1/\varepsilon\delta)})^{2d/(d+2)})$  or randomly sample  $|Q| = O((1/\varepsilon^2)(d + \log(1/\delta)))$  points which will have the following properties with probability at least  $1 - \delta$ .

- $d_B(\text{Dgm}(\text{KDE}_\mu), \text{Dgm}(\text{KDE}_Q)) \leq \varepsilon$ .
- $d_B(\text{Dgm}((d_\mu^K)^2), \text{Dgm}((d_Q^K)^2)) \leq \varepsilon$ .

► **Corollary 26.** Consider a measure  $\mu$  defined on  $\mathbb{R}^d$  and a kernel  $K$ . We can create a coreset  $Q$  of size  $|Q| = O(((1/\varepsilon^2)\sqrt{\log(1/\varepsilon\delta)})^{2d/(d+2)})$  or randomly sample  $|Q| = O((1/\varepsilon^4)(d + \log(1/\delta)))$  points which will have the following property with probability at least  $1 - \delta$ .

- $d_B(\text{Dgm}(d_\mu^K), \text{Dgm}(d_Q^K)) \leq \varepsilon$ .

Another bound was independently derived to show an upper bound on the size of a random sample  $Q$  such that  $d_B(\text{Dgm}(\text{KDE}_{\mu_P}), \text{Dgm}(\text{KDE}_Q)) \leq \varepsilon$  in [2]; this can, as above, also be translated into bounds for  $\text{Dgm}((d_Q^K)^2)$  and  $\text{Dgm}(d_Q^K)$ . This result assumes  $P \subset [-C, C]^d$  and is parametrized by a bandwidth parameter  $h$  that retains that  $\int_{x \in \mathbb{R}^d} K_h(x, p) dx = 1$  for all  $p$  using that  $K_1(\|x - p\|) = K(x, p)$  and  $K_h(\|x - p\|) = \frac{1}{h^d} K_1(\|x - p\|^2/h)$ . This ensures that  $K(\cdot, p)$  is  $(1/h^d)$ -Lipschitz and that  $K(x, x) = \Theta(1/h^d)$  for any  $x$ . Then their bound requires  $|Q| = O(\frac{d}{\varepsilon^2 h^d} \log(\frac{Cd}{\varepsilon \delta h}))$  random samples.

To compare directly against the random sampling result we derive from Joshi *et al.* [38], for kernel  $K_h(x, p)$  then  $\|\text{KDE}_{\mu_P} - \text{KDE}_Q\|_\infty \leq \varepsilon K_h(x, x) = \varepsilon/h^d$ . Hence, our analysis requires  $|Q| = O((1/\varepsilon^2 h^{2d})(d + \log(1/\delta)))$ , and is an improvement when  $h = \Omega(1)$  or  $C$  is not known or bounded, as well as in some other cases as a function of  $\varepsilon$ ,  $h$ ,  $\delta$ , and  $d$ .

**Acknowledgements.** The authors thank Don Sheehy, Frédéric Chazal and the rest of the Geometrica group at INRIA-Saclay for enlightening discussions on geometric and topological reconstruction. We also thank Don Sheehy for personal communications regarding the power distance constructions, and Yusu Wang for ideas towards Lemma 16. Finally, we are also indebted to the anonymous reviewers for many detailed suggestions leading to improvements in results and presentation.

---

## References

- 1 N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.
- 2 Sivaraman Balakrishnan, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Aarti Singh, and Larry Wasserman. Statistical inference for persistent homology. Technical report, ArXiv:1303.7117, March 2013.
- 3 James Biagioni and Jakob Eriksson. Map inference in the face of noise and disparity. In *ACM SIGSPATIAL GIS*, 2012.
- 4 Gérard Biau, Frédéric Chazal, David Cohen-Steiner, Luc Devroye, and Carlos Rodriguez. A weighted  $k$ -nearest neighbor density estimate for geometric inference. *Electronic Journal of Statistics*, 5:204–237, 2011.
- 5 Omer Bobrowski, Sayan Mukherjee, and Jonathan E. Taylor. Topological consistency via kernel estimation. Technical report, arXiv:1407.5272, 2014.
- 6 Peter Bubenik. Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 2014.
- 7 Mickael Buchet, Frederic Chazal, Steve Y. Oudot, and Donald R. Sheehy. Efficient and robust persistent homology for measures. In *SODA*, 2015.
- 8 Frédéric Chazal and David Cohen-Steiner. Geometric inference. *Tessellations in the Sciences*, 2012.
- 9 Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Y. Oudot. Proximity of persistence modules and their diagrams. In *SOCG*, 2009.
- 10 Frédéric Chazal, David Cohen-Steiner, and André Lieutier. Normal cone approximation and offset shape isotopy. *CGTA*, 42:566–581, 2009.
- 11 Frédéric Chazal, David Cohen-Steiner, and André Lieutier. A sampling theory for compact sets in Euclidean space. *DCG*, 41(3):461–479, 2009.
- 12 Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Geometric inference for probability measures. *FOCM*, 11(6):733–751, 2011.
- 13 Frederic Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. The structure and stability of persistence modules. arXiv:1207.3674, 2013.

- 14 Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Robust topological inference: Distance-to-a-measure and kernel distance. Technical report, arXiv:1412.7197, 2014.
- 15 Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Aarti Singh, and Larry Wasserman. On the bootstrap for persistence diagrams and landscapes. *Modeling and Analysis of Information Systems*, 20:96–105, 2013.
- 16 Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. Stochastic convergence of persistence landscapes. In *SOCG*, 2014.
- 17 Frédéric Chazal and André Lieutier. Weak feature size and persistent homology: computing homology of solids in  $R^n$  from noisy data samples. In *SOCG*, pages 255–262, 2005.
- 18 Frédéric Chazal and André Lieutier. Topology guaranteeing manifold reconstruction using distance function to noisy data. In *SOCG*, 2006.
- 19 Frédéric Chazal and Steve Oudot. Towards persistence-based reconstruction in euclidean spaces. In *SOCG*, 2008.
- 20 David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *DCG*, 37:103–120, 2007.
- 21 Luc Devroye and László Györfi. *Nonparametric Density Estimation: The  $L_1$  View*. Wiley, 1984.
- 22 Luc Devroye and Gábor Lugosi. *Combinatorial Methods in Density Estimation*. Springer-Verlag, 2001.
- 23 Herbert Edelsbrunner. The union of balls and its dual shape. In *SOCG*, 1993.
- 24 Herbert Edelsbrunner, Michael Facello, Ping Fu, and Jie Liang. Measuring proteins and voids in proteins. In *Proceedings 28th Annual Hawaii International Conference on Systems Science*, 1995.
- 25 Herbert Edelsbrunner, Brittany Terese Fasy, and Günter Rote. Add isotropic Gaussian kernels at own risk: More and more resilient modes in higher dimensions. In *SOCG*, 2012.
- 26 Herbert Edelsbrunner and John Harer. Persistent homology. *Contemporary Mathematics*, 453:257–282, 2008.
- 27 Herbert Edelsbrunner and John Harer. *Computational Topology: An Introduction*. American Mathematical Society, Providence, RI, USA, 2010.
- 28 Ahmed Elgammal, Ramani Duraiswami, David Harwood, and Larry S. Davis. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proc. IEEE*, 90:1151–1163, 2002.
- 29 Brittany Terese Fasy, Jisu Kim, Fabrizio Lecci, and Clément Maria. Introduction to the R package TDA. Technical report, arXiv:1411.1830, 2014.
- 30 Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. Statistical inference for persistent homology: Confidence sets for persistence diagrams. In *The Annals of Statistics*, volume 42, pages 2301–2339, 2014.
- 31 H. Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93:418–491, 1959.
- 32 Mingchen Gao, Chao Chen, Shaoting Zhang, Zhen Qian, Dimitris Metaxas, and Leon Axel. Segmenting the papillary muscles and the trabeculae from high resolution cardiac CT through restoration of topological handles. In *Proceedings International Conference on Information Processing in Medical Imaging*, 2013.
- 33 Joan Glaunès. *Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l’anatomie numérique*. PhD thesis, Université Paris 13, 2005.
- 34 Karsten Grove. Critical point theory for distance functions. *Proceedings of Symposia in Pure Mathematics*, 54:357–385, 1993.

- 35 Leonidas Guibas, Quentin Mérigot, and Dmitriy Morozov. Witnessed  $k$ -distance. In *SOCG*, 2011.
- 36 Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.
- 37 Matrial Hein and Olivier Bousquet. Hilbertian metrics and positive definite kernels on probability measures. In *Proceedings 10th International Workshop on Artificial Intelligence and Statistics*, 2005.
- 38 Sarang Joshi, Raj Varma Kommaraju, Jeff M. Phillips, and Suresh Venkatasubramanian. Comparing distributions and shapes using the kernel distance. In *SOCG*, 2011.
- 39 John M. Lee. *Introduction to smooth manifolds*. Springer, 2003.
- 40 Jie Liang, Herbert Edelsbrunner, Ping Fu, Pamidighantam V. Sudharkar, and Shankar Subramanian. Analytic shape computation of macromolecules: I. molecular area and volume through alpha shape. *Proteins: Structure, Function, and Genetics*, 33:1–17, 1998.
- 41 André Lieutier. Any open bounded subset of  $\mathbb{R}^n$  has the same homotopy type as its medial axis. *Computer-Aided Design*, 36:1029–1046, 2004.
- 42 Quentin Mérigot. *Geometric structure detection in point clouds*. PhD thesis, Université de Nice Sophia-Antipolis, 2010.
- 43 Yuriy Mileyko, Sayan Mukherjee, and John Harer. Probability measures on the space of persistence diagrams. *Inverse Problems*, 27(12), 2011.
- 44 A. Müller. Integral probability metrics and their generating classes of functions. *Advances in Applied Probability*, 29(2):429–443, 1997.
- 45 Jeff M. Phillips.  $\varepsilon$ -samples for kernels. *SODA*, 2013.
- 46 Jeff M. Phillips and Suresh Venkatasubramanian. A gentle introduction to the kernel distance. arXiv:1103.1625, March 2011.
- 47 Jeff M. Phillips, Bei Wang, and Yan Zheng. Geometric inference on kernel density estimates. In *arXiv:1307.7760*, 2015.
- 48 Florian T. Pokorny, Carl Henrik, Hedvig Kjellström, and Danica Kragic. Persistent homology for learning densities with bounded support. In *Neural Informations Processing Systems*, 2012.
- 49 Charles A. Price, Olga Symonova, Yuriy Mileyko, Troy Hilley, and Joshua W. Weitz. Leaf gui: Segmenting and analyzing the structure of leaf veins and areoles. *Plant Physiology*, 155:236–245, 2011.
- 50 David W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, 1992.
- 51 Donald R. Sheehy. A multicover nerve for geometric inference. *CCCG*, 2012.
- 52 Bernard W. Silverman. Using kernel density estimates to investigate multimodality. *J. R. Statistical Society B*, 43:97–99, 1981.
- 53 Bernard W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, 1986.
- 54 Bharath K. Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert R.G. Lanckriet. Hilbert space embeddings and metrics on probability measures. *JMLR*, 11:1517–1561, 2010.
- 55 Kathryn Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. Fréchet means for distributions of persistence diagrams. *DCG*, 2014.
- 56 Cédric Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.
- 57 Grace Wahba. Support vector machines, reproducing kernel Hilbert spaces, and randomization. In *Advances in Kernel Methods – Support Vector Learning*, pages 69–88. The MIT Press, 1999.
- 58 Yan Zheng, Jeffrey Jests, Jeff M. Phillips, and Feifei Li. Quality and efficiency in kernel density estimates for large data. In *SIGMOD*, 2012.