Kernel Distance for Geometric Inference

Jeff M. Phillips
University of Utah

Bei Wang
University of Utah

This abstract considers geometric inference from a noisy point cloud using the kernel distance. Recently Chazal, Cohen-Steiner, and Mérgot [2] introduced *distance to a measure*, which is a distance-like function robust to perturbations and noise on the data. Here we show how to use the kernel distance in place of the distance to a measure; they have very similar properties, but the kernel distance has several advantages.

- The kernel distance has a small coreset, making efficient inference possible on millions of points.
- Its inference works quite naturally using the super-level set of a kernel density estimate.
- The kernel distance is Lipschitz on the outlier parameter σ.

Kernels, Kernel Density Estimates, and Kernel Distance

A *kernel* is a similarity measure $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^+$; more similar points have higher value. For the purposes of this article we will focus on the Gaussian kernel defined $K(p,x) = \sigma^2 \exp(-\|p-x\|^2/2\sigma^2)$.

A *kernel density estimate* represents a continuous distribution function over $\mathbb{R}^d$ for point set $P \subset \mathbb{R}^d$:

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

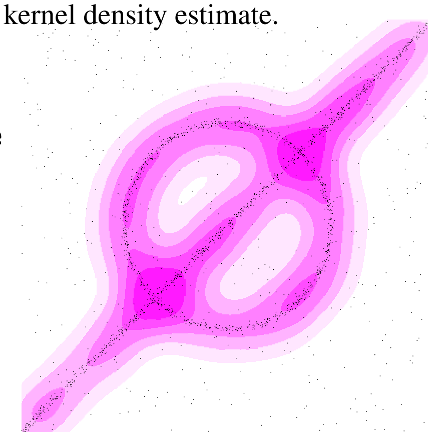

Figure 1: Geometric inference using super-level sets of kernel density estimates on 2000 points.

More generally, it can be applied to any measure μ (on $\mathbb{R}^d$) as $\text{KDE}_\mu(x) = \int_{p \in \mathbb{R}^d} K(p,x)\mu(p)\mathrm{d}p$.

The *kernel distance* [3, 5] is a metric between two point sets P and Q, or more generally two measures μ and ν (as long as K is positive definite, e.g. the Guassian kernel). Define $\kappa(P,Q) = \frac{1}{|P|}\frac{1}{|Q|}\sum_{p\in P}\sum_{q \in Q} K(p,q)$. Then the kernel distance is defined

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

For the kernel distance $D_K(\mu,\nu)$ between two measures μ and ν, we define κ more generally as $\kappa(\mu,\nu) = \int_{p\in\mathbb{R}^d}\int_{q \in \mathbb{R}^d} K(p,q)\mu(p)\mu(q)\mathrm{d}p\mathrm{d}q$. When the points set Q (or measure ν) is a single point x (or unit Dirac mass at x), then the important term in the kernel distance is $\kappa(P,x) = \text{KDE}_P(x)$ (or $\kappa(\mu,x) = \text{KDE}_\mu(x)$).

Distance to a Measure: A Review

Let S be a compact set, and $f_S : \mathbb{R}^d \to \mathbb{R}$ be a distance function to S. As explained in [2], there are a few properties of f_S that are sufficient to make it useful in geometric inference such as [1]:

(F1) f_S is 1-Lipschitz: for all $x, y \in \mathbb{R}^d$, $|f_S(x) - f_S(y)| \leq \|x-y\|$.

(F2) f_S^2 is 1-semiconcave: the map $x \in \mathbb{R}^d \mapsto (f_S(x))^2 - \|x\|^2$ is concave.

Given a probability measure μ on $\mathbb{R}^d$ and let $m_0 > 0$ be a parameter smaller than the total mass of μ, then the distance to a measure $d_{\mu,m_0} : \mathbb{R}^n \to \mathbb{R}^+$ [2] is defined for any point $x \in \mathbb{R}^d$ as

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

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 [2] using d_{μ,m_0} in place of f_S satisfies (F1) and (F2), and furthermore has the following stability property:

(F3) [Stability] If μ and μ' are two probability measures on $\mathbb{R}^d$ and $m_0 > 0$, then $\|d_{\mu,m_0} - d_{\mu',m_0}\|_\infty \leq \frac{1}{\sqrt{m_0}} W_2(\mu,\mu')$, where W_2 is the Wasserstein distance between the two measures.

Our Results

We demonstrate (with proof sketches) that similar properties hold for the kernel distance defined as $d_P(x) = D_K(P,x)$. These properties also hold on $d_\mu(\cdot) = D_K(\mu,\cdot)$ for a measure μ in place of P.

(K1) d_P is 1-Lipschitz.

This is implied by d_P^2 being 1-semiconcave.

(K2) d_P^2 is 1-semiconvave: The map $x \mapsto (d_P(x))^2 - \|x\|^2$ is concave.

In any direction, the second derivative of $(d_P(x))^2$ is at most that of a single kernel $K(p,x)$ for any p, and this is maximized at $x = p$. The second derivative of $\|x\|^2$ is 2 everywhere, thus the second derivative of $(d_P(x))^2 - \|x\|^2$ is non-positive, and hence is concave.

(K3) [Stability] If P and Q are two point sets in $\mathbb{R}^d$, then $\|d_P - d_Q\|_\infty \leq D_K(P,Q)$.

Using that $D_K(\cdot,\cdot)$ is a metric, we compare $D_K(P,Q)$, $D_K(P,x)$ and $D_K(Q,x)$. Note: Wasserstein and kernel distance are different *integral probability metrics* [5], so (F3) and (K3) are not comparable.

Advantages of the kernel distance.

- There exists a *coreset* $Q \subset P$ of size $O(((1/\varepsilon)\sqrt{\log(1/\varepsilon\delta)})^{2d/(d+2)})$ [4] such that $\|d_P - d_Q\|_\infty \leq \varepsilon$ and $\|\mathrm{KDE}_P - \mathrm{KDE}_Q\|_\infty \leq \varepsilon$ with probability at least $1-\delta$. The same holds under a random sample of size $O((1/\varepsilon^2)(d + \log(1/\delta)))$ [3]. In ongoing work, this allows us to operate with $|P| = 100{,}000{,}000$.

  Bottleneck distance between persistence diagrams $d_B(\mathsf{Dgm}(\mathrm{KDE}_P), \mathsf{Dgm}(\mathrm{KDE}_Q)) \leq \varepsilon$ is preserved.
- We can perform geometric inference on noisy P by considering the superlevel sets of KDE_P; the τ-superlevel set of KDE_P is $\{x \in \mathbb{R}^d \mid \mathrm{KDE}_P(x) \geq \tau\}$. This follows since $d_P(\cdot)$ is *monotonic* with $\mathrm{KDE}_P(\cdot)$; as $d_P(x)$ gets smaller, $\mathrm{KDE}_P(x)$ gets larger. This arguably is a more natural interpretation than using the sublevel sets of some f_S. Figure 1 shows an example with 25% of P as noise.
- Both the distance to a measure and the kernel distance have parameters that control the amount of outliers allowed (m_0 for d_{μ,m_0} and σ for d_P). For d_P the smoothing effect of σ has been well-studied, and in fact $d_P(x)$ is Lipschitz continuous with respect to σ (for σ greater than a fixed constant). Alternatively, $d_{P,m_0}(x)$, for fixed x, is not known to be Lipschitz (for arbitrary P) with respect to m_0 and fixed x; we suspect that the Lipschitz constant for m_0 is a function of $\Delta(P) = \max_{p,p' \in P} \|p - p'\|$.

References

[1] Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Boundary measures for geometric inference. *Foundations of Computational Mathematics*, 10(2):221–240, 2010.

[2] Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11(6):733–751, 2011.

[3] Sarang Joshi, Raj Varma Kommaraju, Jeff M. Phillips, and Suresh Venkatasubramanian. Comparing distributions and shapes using the kernel distance. *Proceedings 27th Annual Symposium on Computational Geometry*, 2011.

[4] Jeff M. Phillips. eps-samples for kernels. *Proceedings 24th Annual ACM-SIAM Symposium on Discrete Algorithms (to appear)*, 2013.

[5] Jeff M. Phillips and Suresh Venkatasubramanian. A gentle introduction to the kernel distance. arXiv:1103.1625, March 2011.