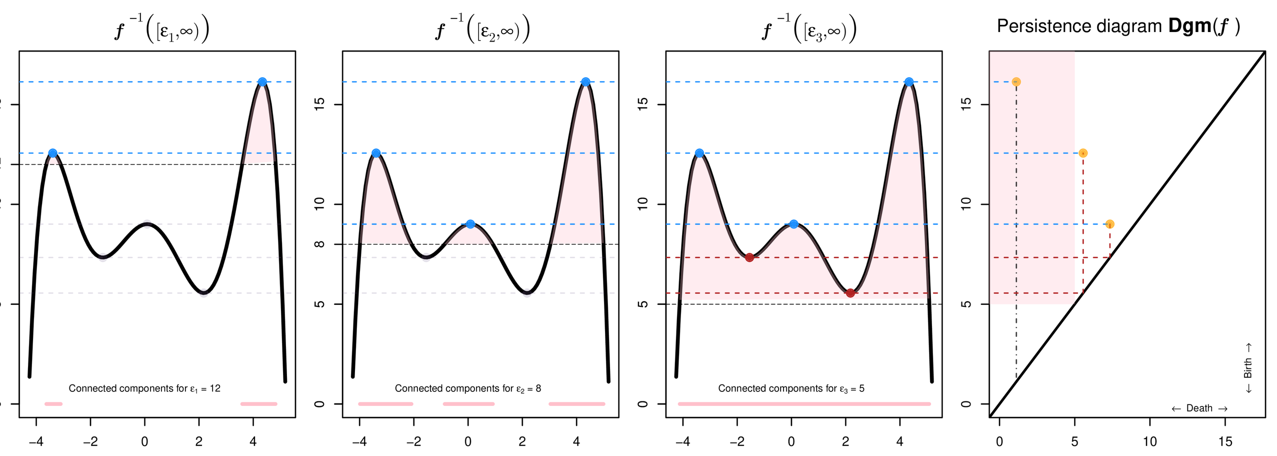
Robust Persistence Diagrams Using Reproducing Kernels

Siddharth Vishwanath, Kenji Fukumizu, Satoshi Kuriki and Bharath Sriperumbudur

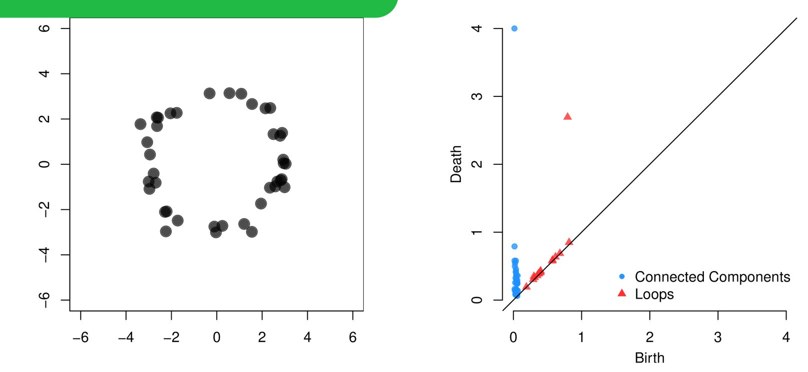
The Pennsylvania State University and The Institute of Statistical Mathematics

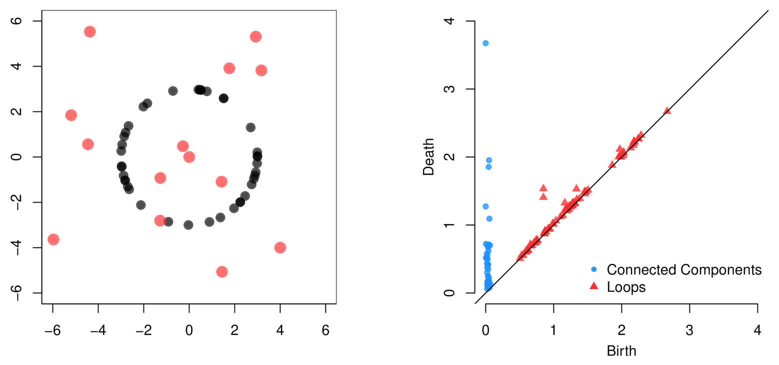
Introduction

- Given $\mathbb{X}_n = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n\}$ sampled i.i.d. from $\mathbb{P}$
- The filter ϕ_n constructed using $\mathbb{X}_n$ is an approximation of $\phi_{\mathbb{P}}$
- The shape of $\mathbb{X}_n$ is encoded in the superlevel sets $\phi_n([\epsilon, \infty))$
- As $\epsilon \downarrow$ new features are born **and/or** existing features die
- This is summarized in $\mathbf{Dgm}(\phi_n) = \mathbf{Dgm}(\mathrm{Sup}(\phi_n))$



Motivation





$\mathbf{Dgm}$s are stable w.r.t small perturbations, but even a few **outliers** can dramatically change inference

Robust Persistence Diagrams

- K_σ is a radial kernel, and $\mathcal{H}_\sigma$ its reproducing kernel Hilbert space
- Given $\mathbb{X}_n$ and a robust loss $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ the robust KDE [1] is given by

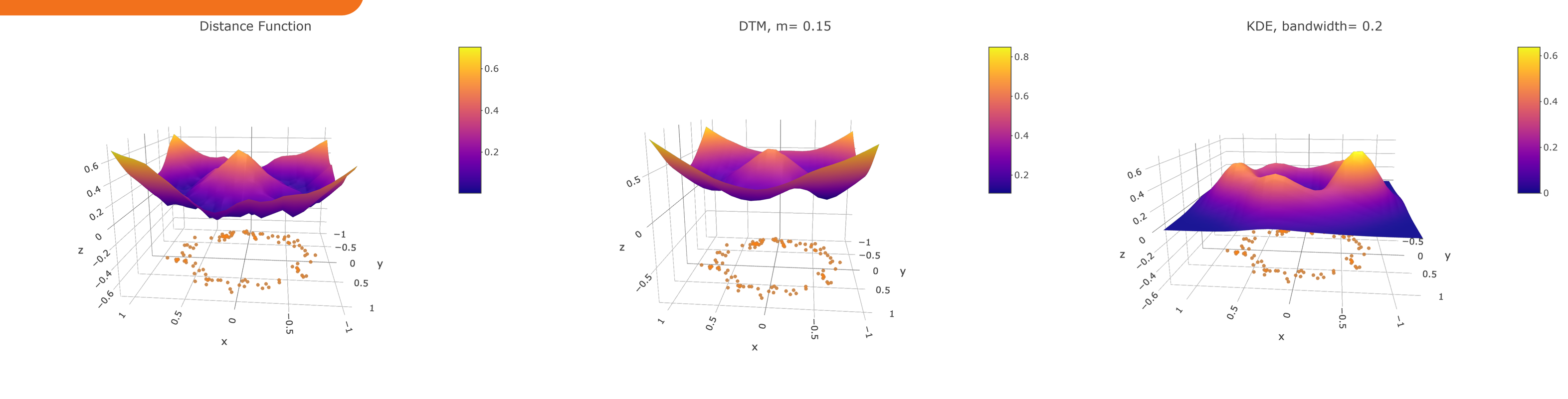
$$f^n_{\rho,\sigma} = \operatorname*{arg\,inf}_{g \in \mathcal{H}_\sigma} \int_{\mathbb{R}^d} \rho\Big(\|g - K_\sigma(\cdot, \boldsymbol{y})\|_{\mathcal{H}_\sigma}\Big)\, d\mathbb{P}_n(\boldsymbol{y}) = \sum_{i=1}^{n} w_\sigma(\boldsymbol{X}_i) K_\sigma(\cdot, \boldsymbol{X}_i)$$

- **w_σ weighs-down extreme outliers**. Solved via reweighted least squares
- **$\mathbf{Dgm}(f^n_{\rho,\sigma}) = \mathbf{Dgm}(\mathrm{Sup}(f^n_{\rho,\sigma}))$ is the robust persistence diagram**

Summary

We can construct robust persistence diagrams for Topological Data Analysis without compromising on statistical efficiency

Examples



Distance Function	Distance-to-measure	Kernel Density Estimator
$d_n(\boldsymbol{x}) = \inf_{\boldsymbol{X}_i \in \mathbb{X}_n} \|\boldsymbol{x} - \boldsymbol{X}_i\|_2$	$d^2_{m,n}(\boldsymbol{x}) = \frac{1}{k} \sum_{\boldsymbol{X}_i \in \mathrm{NN}_k(\boldsymbol{x})} \|\boldsymbol{x} - \boldsymbol{X}_i\|_2^2$	$f^n_\sigma(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} k_\sigma(\boldsymbol{x}, \boldsymbol{X}_i)$

Experiments

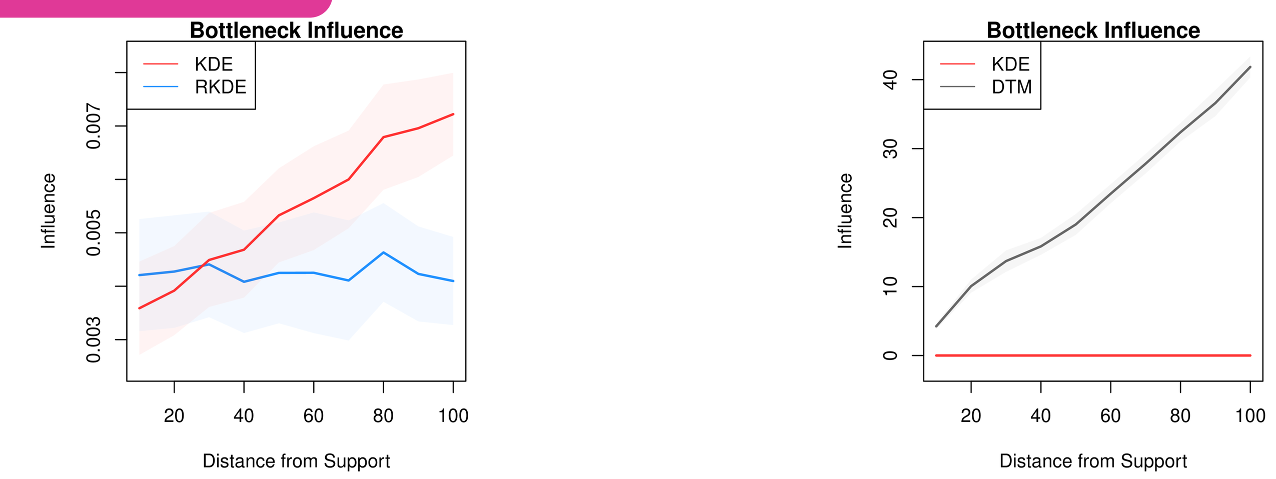


Figure 1. For extreme outliers, there is no noticeable influence on $\mathbf{Dgm}(f^n_{\rho,\sigma})$

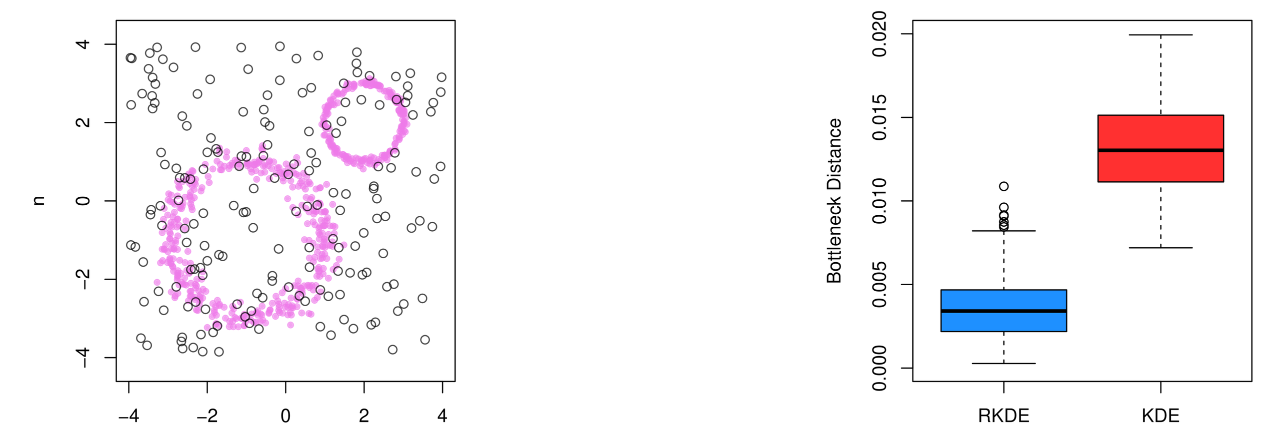


Figure 2. $\mathbf{Dgm}(f^n_{\rho,\sigma})$ recovers the features of signal better than $\mathbf{Dgm}(f^n_\sigma)$

Sensitivity Analysis in the space $(\mathfrak{D}, W_\infty)$

- Given a statistical functional $T : \mathbb{P} \mapsto (V, \|\cdot\|_V)$ and $\boldsymbol{x} \in \mathbb{R}^d$, the **influence function, $\mathrm{IF}(T; \mathbb{P}, \boldsymbol{x})$**, measures the sensitivity of T to outliers
- $(\mathfrak{D}, W_\infty)$ is not a normed space 🙁 $\Rightarrow$ generalize via metric derivative 🙂

Given a **filter** $\phi_{\mathbb{P}}$ and $\mathbb{P}^\epsilon_{\boldsymbol{x}} = (1-\epsilon)\mathbb{P} + \epsilon\delta_{\boldsymbol{x}}$, the **persistence influence** is given by

$$\Psi\Big(\phi_{\mathbb{P}}; \mathbb{P}, \boldsymbol{x}\Big) \doteq \lim_{\epsilon \to 0} \frac{1}{\epsilon} W_\infty\Big(\mathbf{Dgm}(\phi_{\mathbb{P}^\epsilon_{\boldsymbol{x}}}), \mathbf{Dgm}(\phi_{\mathbb{P}})\Big)$$

Theorem. For a robust loss ρ and $\sigma, m > 0$

$$\begin{aligned}
\text{Robust KDE:} \quad & \Psi\Big(f_{\rho,\sigma}; \mathbb{P}, \boldsymbol{x}\Big) \leq \sigma^{-d/2} w_\sigma(\boldsymbol{x}) \|K(\cdot, \boldsymbol{x}) - f_{\rho,\sigma}\|_{\mathcal{H}_\sigma} \\
\text{KDE:} \quad & \Psi\Big(f_\sigma; \mathbb{P}, \boldsymbol{x}\Big) \leq \sigma^{-d/2} \|K(\cdot, \boldsymbol{x}) - f_\sigma\|_{\mathcal{H}_\sigma} \\
\text{DTM:} \quad & \Psi\Big(d_m; \mathbb{P}, \boldsymbol{x}\Big) \leq \frac{1}{\sqrt{m}} \sup_{\|\nabla f\|_{L_2(\mathbb{P})} \leq 1} \left\{ \left| f(\boldsymbol{x}) - \int f(\boldsymbol{y}) d\mathbb{P}(\boldsymbol{y}) \right| \right\}
\end{aligned}$$

- The presence of w_σ makes $\mathbf{Dgm}(f^n_{\rho,\sigma})$ more resilient to noise

Statistical Analysis: Gain with no pain

Consistency. Given $\mathbb{X}_n$ from $\mathbb{P}$ with density f and $\sigma > 0$

$$W_\infty\Big(\mathbf{Dgm}(f^n_{\rho,\sigma}), \mathbf{Dgm}(f_{\rho,\sigma})\Big) \xrightarrow{p} 0 \quad \text{as } n \to \infty$$

furthermore, $W_\infty\Big(\mathbf{Dgm}(f_{\rho,\sigma}), \mathbf{Dgm}(f)\Big) \longrightarrow 0$ as $\sigma \to 0$

Theorem. For $a_\sigma > 1$ and $p \in (0, 1)$, if the entropy numbers of the kernel K_σ satisfy $e_n(\mathcal{H}_\sigma) \leq a_\sigma n^{-1/2p}$, then with probability $\geq 1 - \alpha$ and uniformly over $\mathbb{P}$

$$W_\infty\Big(\mathbf{Dgm}(f^n_{\rho,\sigma}), \mathbf{Dgm}(f_{\rho,\sigma})\Big) \leq 2C\sigma^{-d/2}\left(\xi(n,p) + \sqrt{\frac{2\log(1/\alpha)}{n}}\right)$$

where

$$\xi(n,p) = \begin{cases} \mathcal{O}(n^{-1/2}) & \text{if } p \in (0, \frac{1}{2}) \\ \mathcal{O}(n^{-1/2}\log n) & \text{if } p = \frac{1}{2} \\ \mathcal{O}(n^{-1/4p}) & \text{if } p \in (\frac{1}{2}, 1) \end{cases}$$

References

[1] J. Kim and C. D. Scott. "Robust kernel density estimation". In: *Journal of Machine Learning Research* 13.Sep (2012), pp. 2529-2565.

[2] F. Chazal, V. De Silva, M. Glisse, and S. Oudot. *The Structure and Stability of Persistence Modules*. AG, CH: Springer, 2016.