Advanced Data Visualization

CS 6965

Spring 2018

Prof. Bei Wang Phillips

University of Utah

Lecture 19

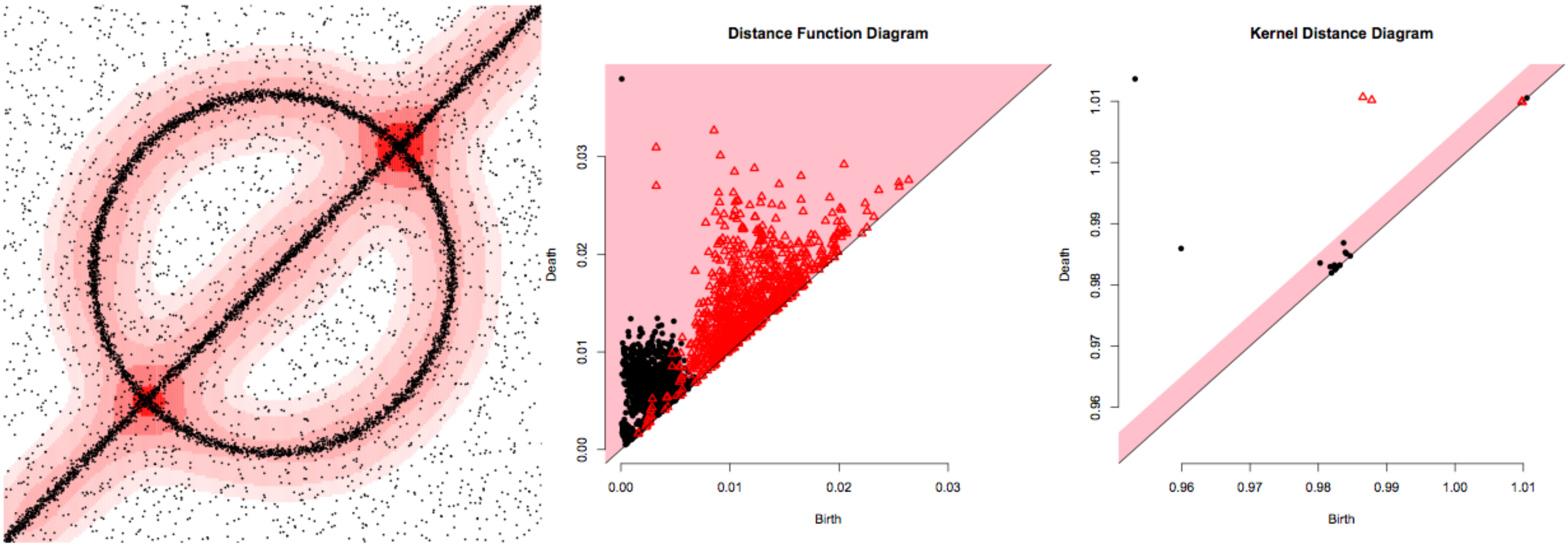
Robust Structural Inference

HD+TOPO

A simple example

- A perfect circle
- A noiseless point cloud sample from the circle
- A point cloud sample with noise
- A point cloud sample with noise and outliers

Another example



Robust Structural Inference

- Kernel distance, kernel density estimate
- Distance to a measure

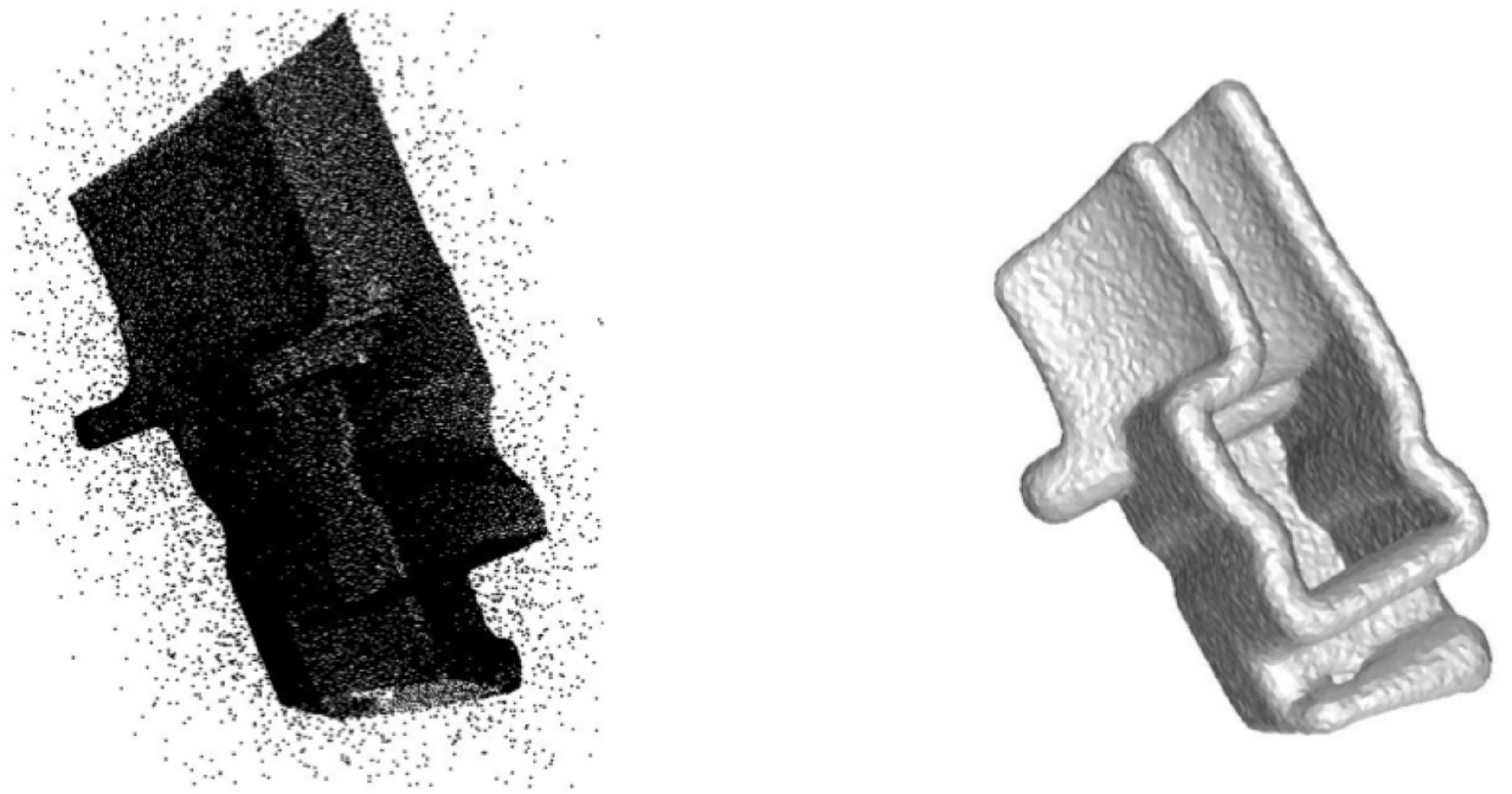
Structural Inference using KDE

Geometric inference

Given:

- An unknown object (e.g. a compact set) $S \subset \mathbb{R}^d$
- A finite point cloud $P \subset \mathbb{R}^d$ that comes from S under some process

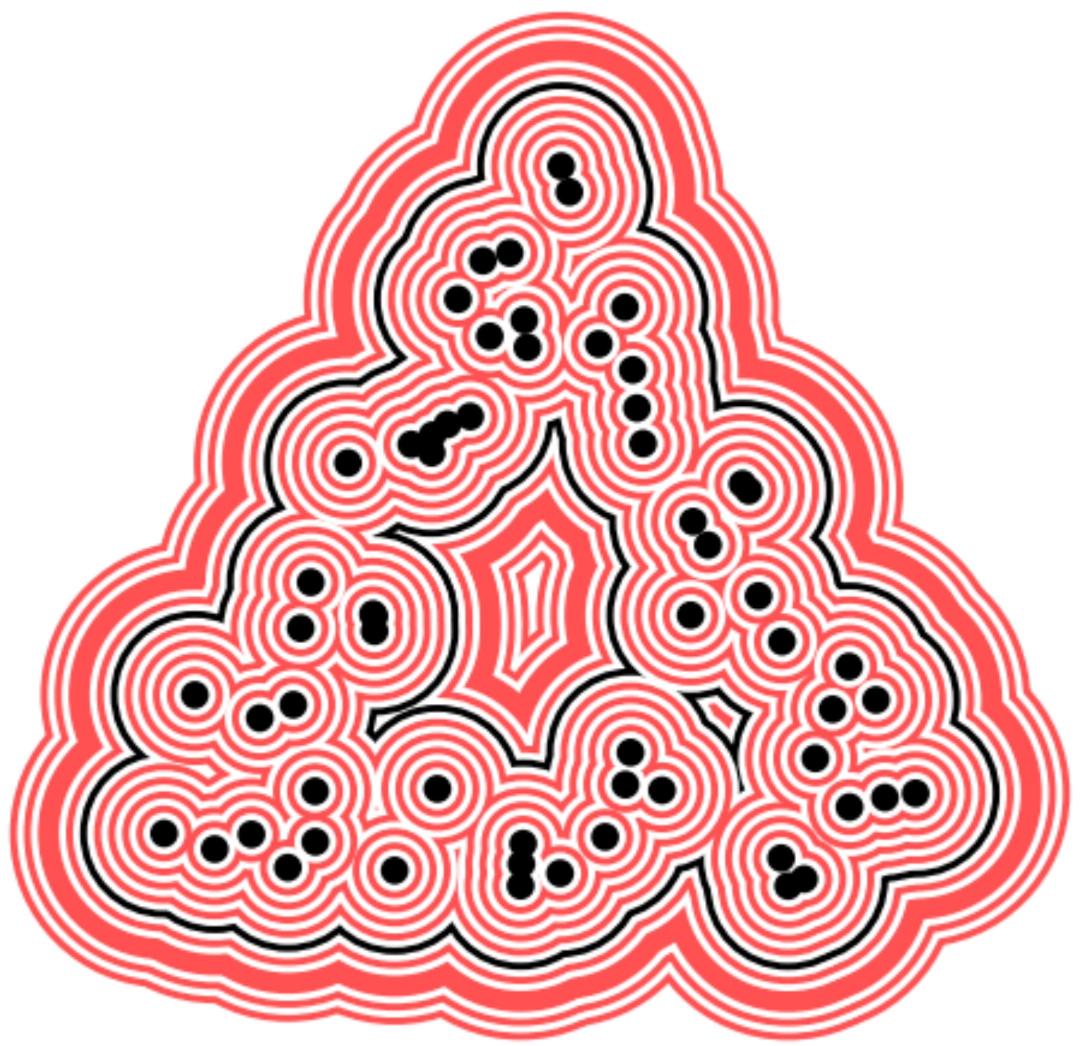
Aim: Recover topological and geometric properties of S from P,
e.g. $\#$ of components, dimension, curvature...
e.g. preserve homeomorphism, homotopy type, or homology of S from P.
e.g. homotopy equivalence: two spaces can be deformed continuously into one another.



[Chazal Cohen-Steiner Merigot 2011]

Distance function based geometric inference

Sample points P from a triangle S with noise; Reconstructs an approximation of S by offsets from P (i.e. union of balls).



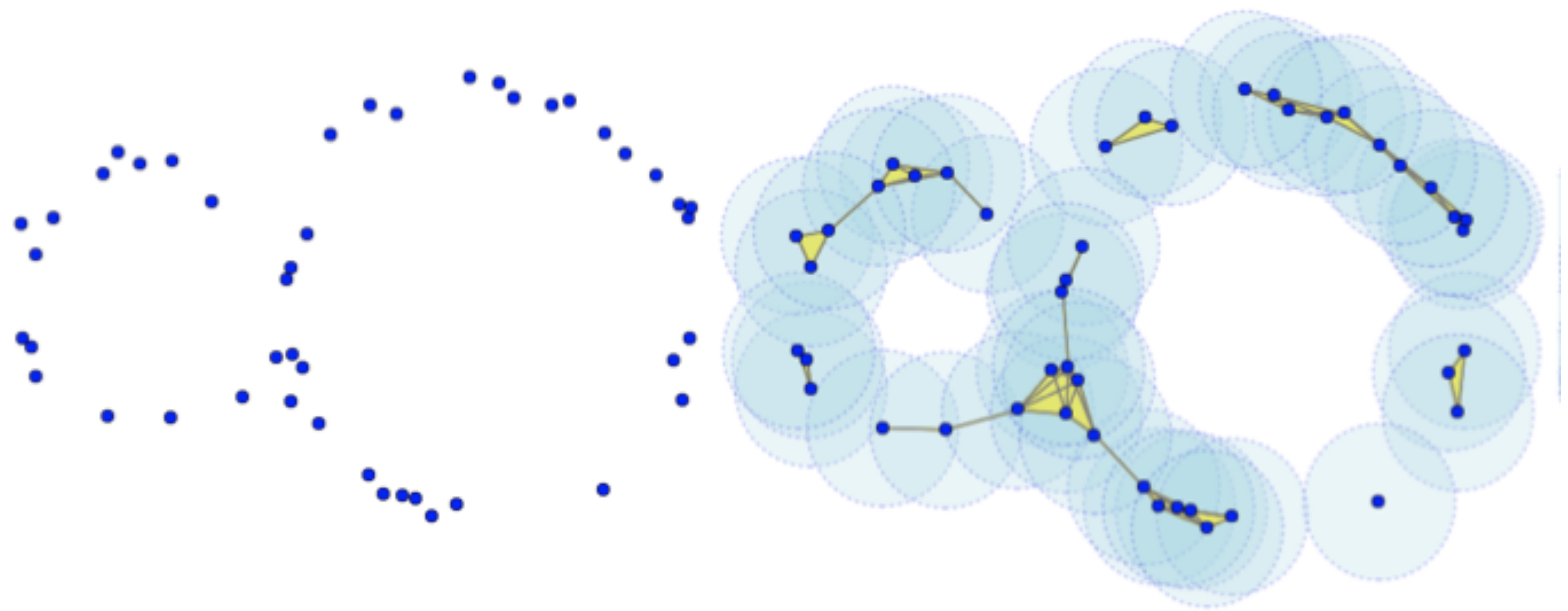
[Chazal, Cohen-Steiner, Lieutier 2009]

Distance function: $f_P(x) = \inf_{y \in P} \|x - y\|$
Offset: $(P)^r = f_P^{-1}([0, r])$

Hausdorff distance (measures sampling quality):
$d_H(S, P) := \|f_S - f_P\|_\infty = \inf_{x \in \mathbb{R}^d} |f_S(x) - f_P(x)| \leq \epsilon$
i.e. smallest $\epsilon \geq 0$ s.t. $S \subseteq (P)^\epsilon$ and $P \subseteq (S)^\epsilon$.

Distance function based geometric inference

Sample points P from a figure-eight S with noise; Reconstructs an approximation of S by offsets from P (i.e. union of balls).



[Image courtesy: Paul Bruillard]

Distance function: $f_P(x) = \inf_{y \in P} \|x - y\|$
Offset: $(P)^r = f_P^{-1}([0, r])$

Hausdorff distance (measures sampling quality):
$d_H(S, P) := \|f_S - f_P\|_\infty = \inf_{x \in \mathbb{R}^d} |f_S(x) - f_P(x)| \leq \epsilon$
i.e. smallest $\epsilon \geq 0$ s.t. $S \subseteq (P)^\epsilon$ and $P \subseteq (S)^\epsilon$.

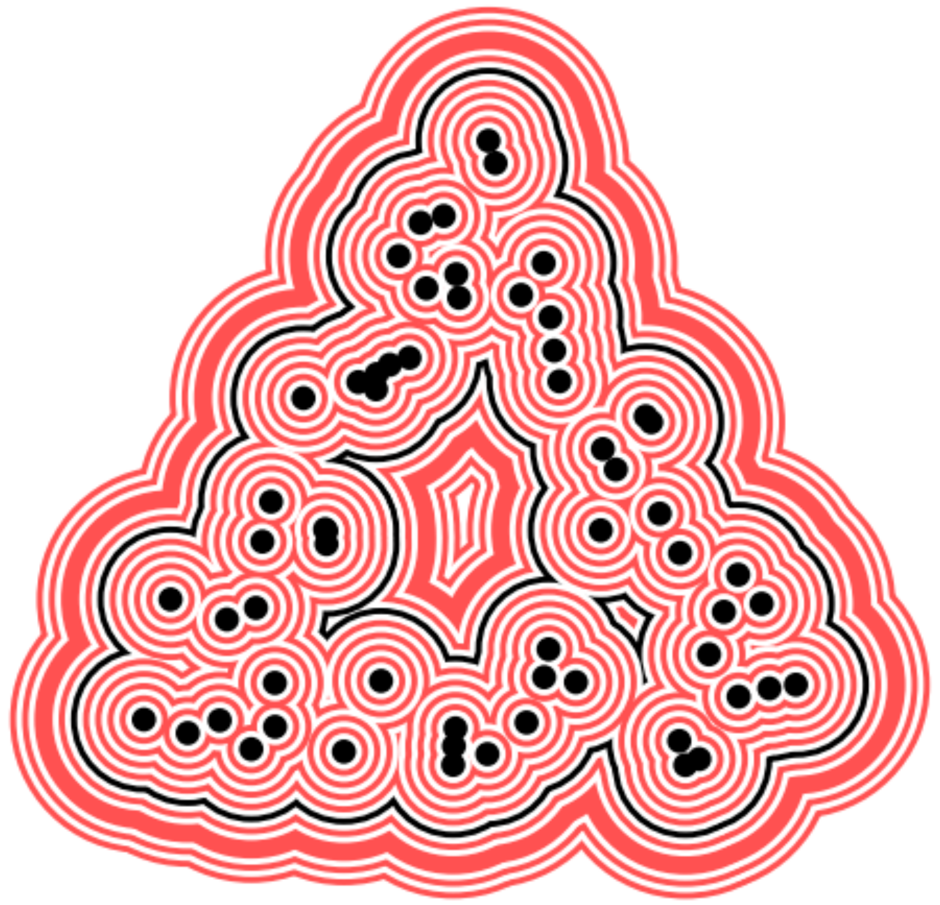
Distance function based geometric inference: the intuition

[Hausdorff stability w.r.t. distance functions]
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.

Theorem (Reconstruction from f_P)

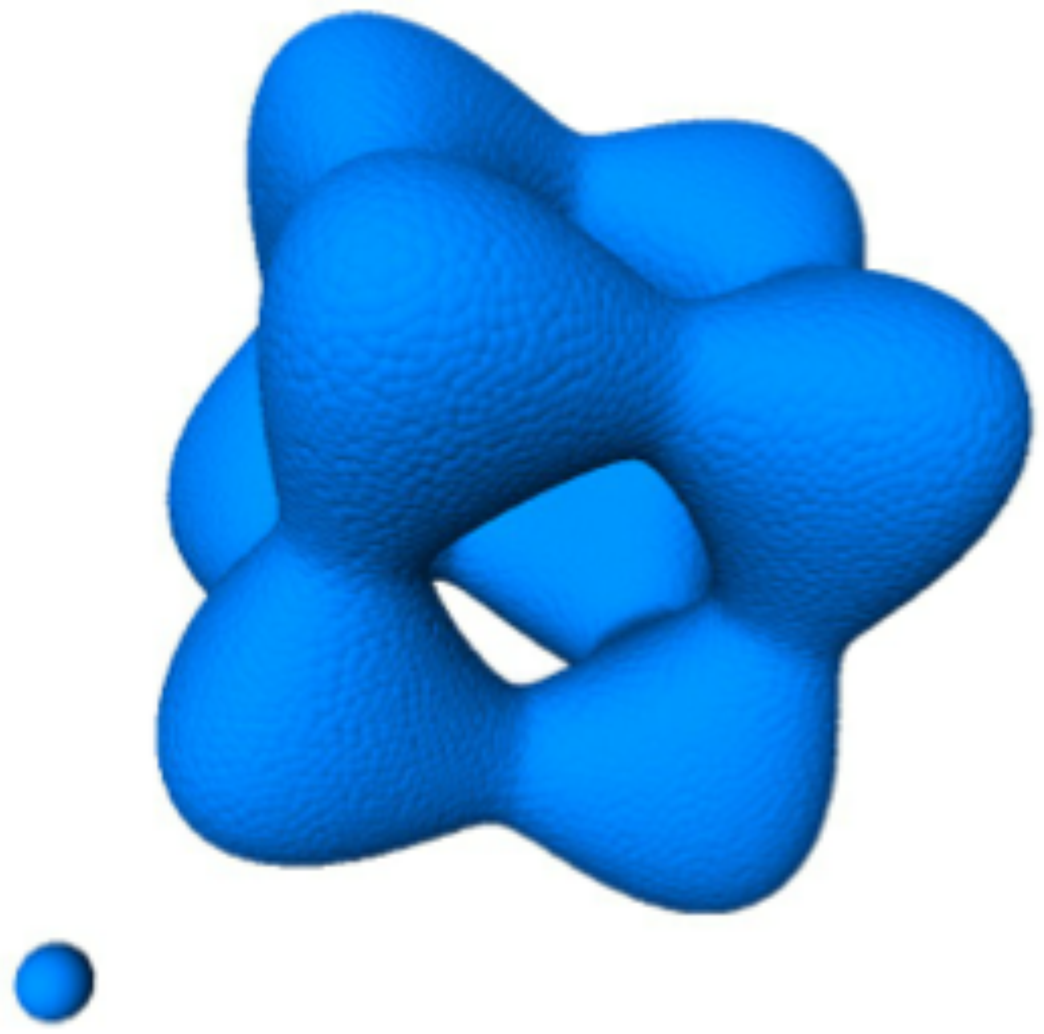
Let $S, P \subset \mathbb{R}^d$ be compact sets such that $\text{reach}(S) > R$ *and* $\varepsilon := d_H(S, P) \leq R/17$. *Then $(S)^\eta$ and $(P)^r$ are homotopy equivalent for sufficiently small η (e.g. $0 < \eta < R$), if* $4\varepsilon \leq r \leq R - 3\varepsilon$. *[Chazal Cohen-Steiner Lieutier 2009] [Chazal Cohen-Steiner Merigot 2011]*

R ensures topological properties of S and $(S)^r$ are the same;
ε ensures $(S)^r$ and $(P)^r$ are close, $\varepsilon \approx$ density of the sample.



Distance function based geometric inference

Not robust to outliers.



[Chazal Cohen-Steiner Merigot 2011]

If $S' = S \cup x$ and $f_S(x) > R$, then $|f_S - f_{S'}|_\infty > R$:
offset-based inference methods fail...

Distance(-like) function that is robust to noise...

Desirable properties for g to be useful in geometric inference:

(D1) g is 1-Lipschitz: for all $x, y \in \mathbb{R}^d$, $|g(x) - g(y)| \leq \|x - y\|$.

(D2) g^2 is 1-semiconcave: $x \in \mathbb{R}^d \mapsto (g(x))^2 - \|x\|^2$ is concave.

(D3) g is proper: $g(x)$ tends the infimum of its domain (e.g., ∞) as x tends to infinity.

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

Kernels

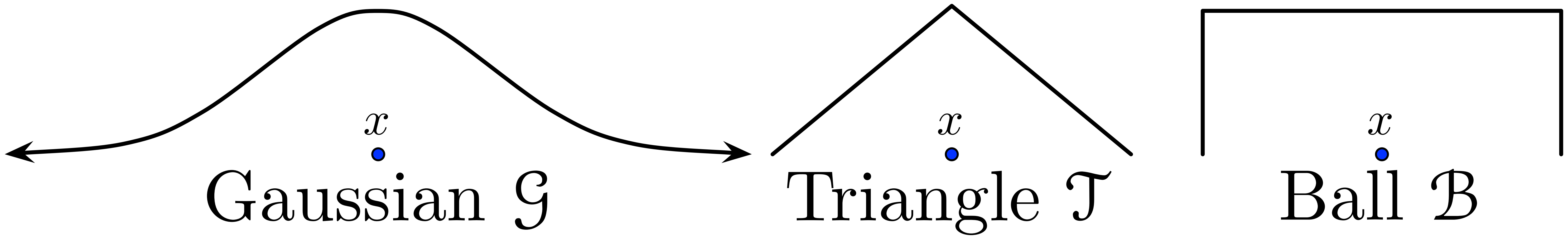
A kernel is a similarity measure, more similar points have higher value,

$$K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^+$$

We focus on the Gaussian kernel (positive definite):

$$K(p,x) = \sigma^2 \exp(-\|p-x\|^2/2\sigma^2)$$

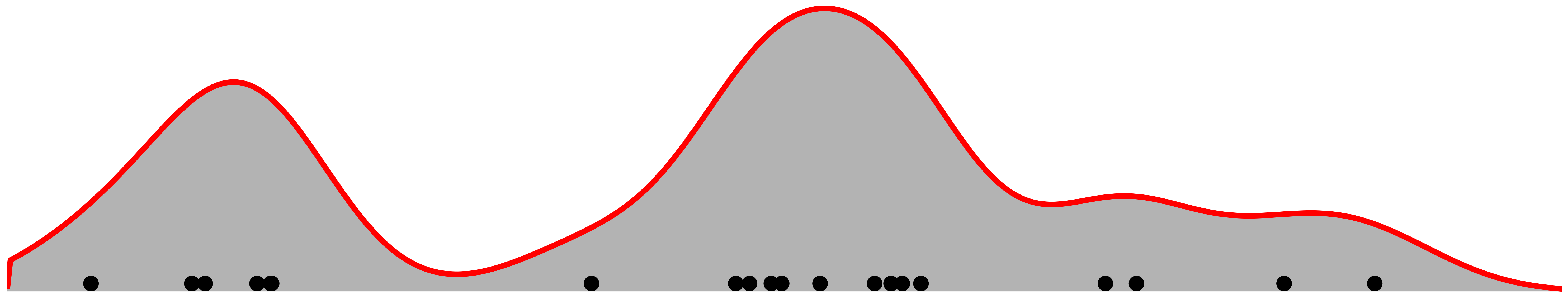
.



Kernel density estimate (KDE)

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

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



More generally, it can be applied to any measure μ (on $\mathbb{R}^d$) as

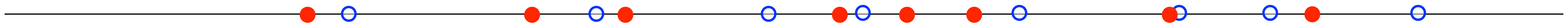
$$\mathrm{KDE}_\mu(x) = \int_{p \in \mathbb{R}^d} K(p, x)\mu(p)\mathrm{d}p$$

Kernel distance

For two point sets P and Q, define similarity

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

If $Q = \{x\}$, $\kappa(P, x) = \mathrm{KDE}_P(x)$.



The kernel distance (a metric between P and Q):

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

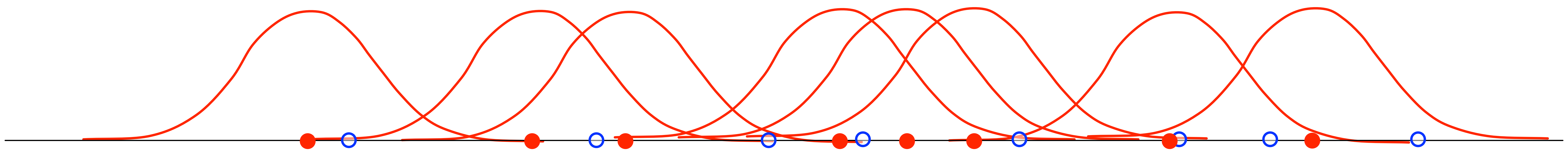
Self similarity minus cross similarity... [Phillips, Venkatasubramanian 2011]

Kernel distance

For two point sets P and Q, define similarity

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

If $Q = \{x\}$, $\kappa(P,x) = \mathrm{KDE}_P(x)$.



The kernel distance (a metric between P and Q):

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

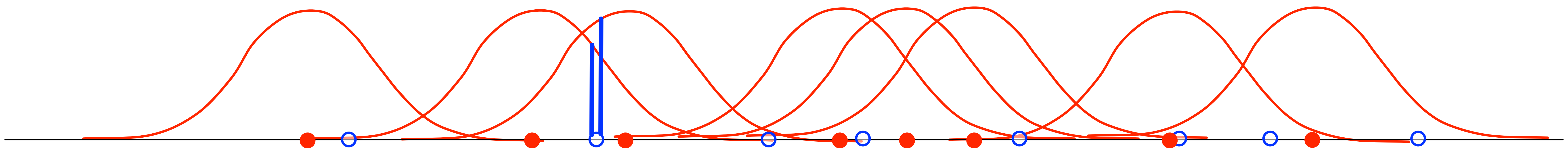
Self similarity minus cross similarity... [Phillips, Venkatasubramanian 2011]

Kernel distance

For two point sets P and Q, define similarity

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

If $Q = \{x\}$, $\kappa(P,x) = \mathrm{KDE}_P(x)$.



The kernel distance (a metric between P and Q):

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

Self similarity minus cross similarity... [Phillips, Venkatasubramanian 2011]

Kernel distance (w.r.t. any measure μ on $\mathbb{R}^d$)

For $D_K(\mu,\nu)$ between two measures μ and ν, define similarity

$$\kappa(\mu,\nu) = \int_{p\in\mathbb{R}^d}\int_{q\in\mathbb{R}^d} K(p,q)\mu(p)\mu(q)\mathrm{d}p\mathrm{d}q$$

The kernel distance (a metric between μ and ν):

$$D_K(\mu,\nu) = \sqrt{\kappa(\mu,\mu)+\kappa(\nu,\nu)-2\kappa(\mu,\nu)}$$

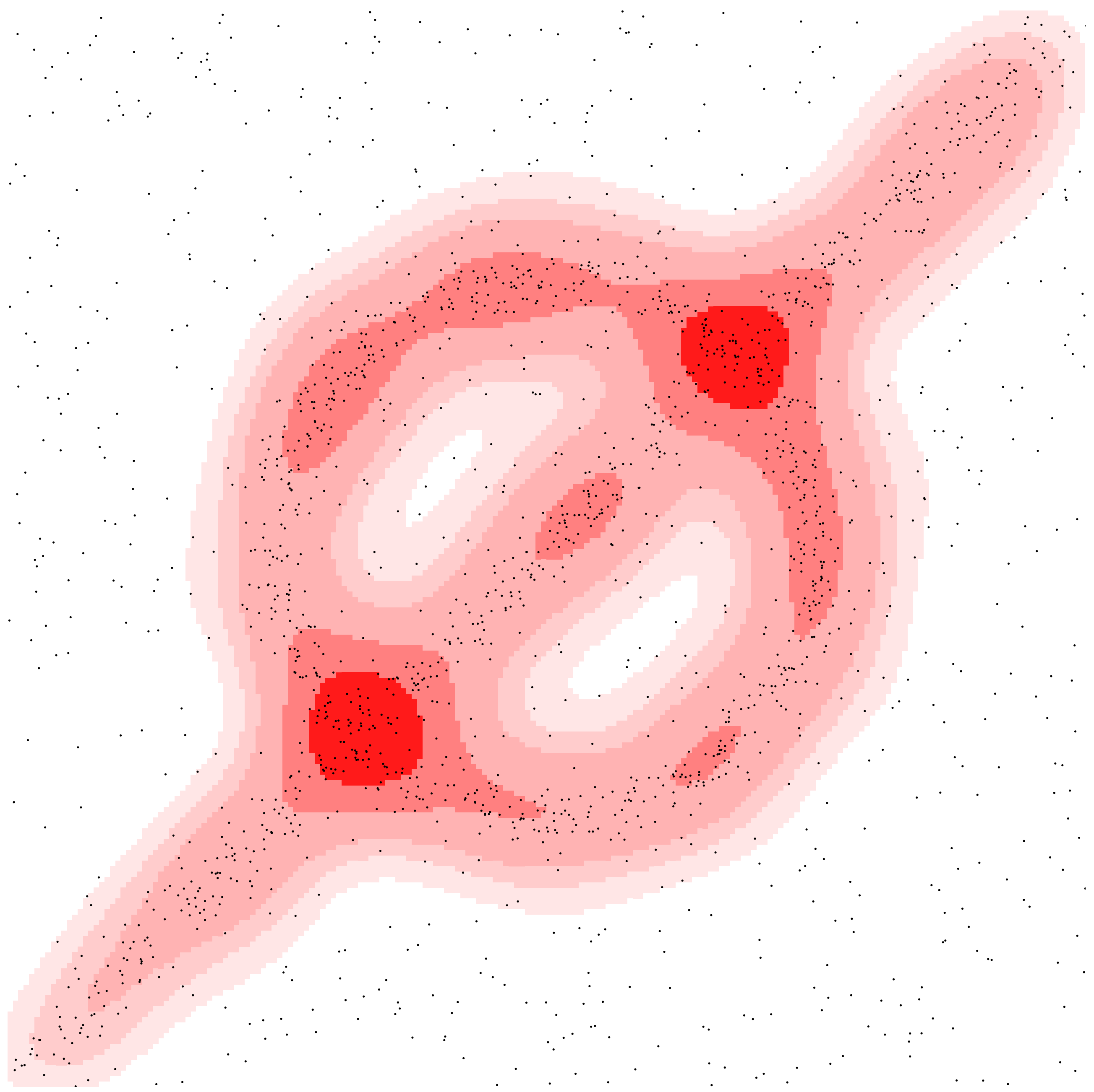
If $\nu =$ unit Dirac mass at x, $\kappa(\mu,x) = \mathrm{KDE}_\mu(x)$,

$$\begin{aligned} D_K(\mu,x) &= \sqrt{\kappa(\mu,\mu)+\kappa(x,x)-2\kappa(\mu,x)} \\ &= \sqrt{c_\mu - 2\mathrm{KDE}_\mu(x)} \end{aligned}$$

Kernel distance (current distance or maximum mean discrepancy) is a metric, if the kernel K is characteristic (a slight restriction of being positive definite, e.g. Gaussian and Laplace kernels).

Take home message

- Geometric inference from a point cloud can be calculated by examining its kernel density estimate (KDE) of Gaussians.
- Such an inference is made possible with provable properties through the vehicle of kernel distance.
- Such an inference is robust to noise and scalable.
- We provide an algorithm to estimate the topology of kernel distance using weighted Vietoris-Rips complexes.



A bit more detail...

Geometric inference using the kernel distance, in place of the distance to a measure [Chazal Cohen-Steiner Merigot 2011].

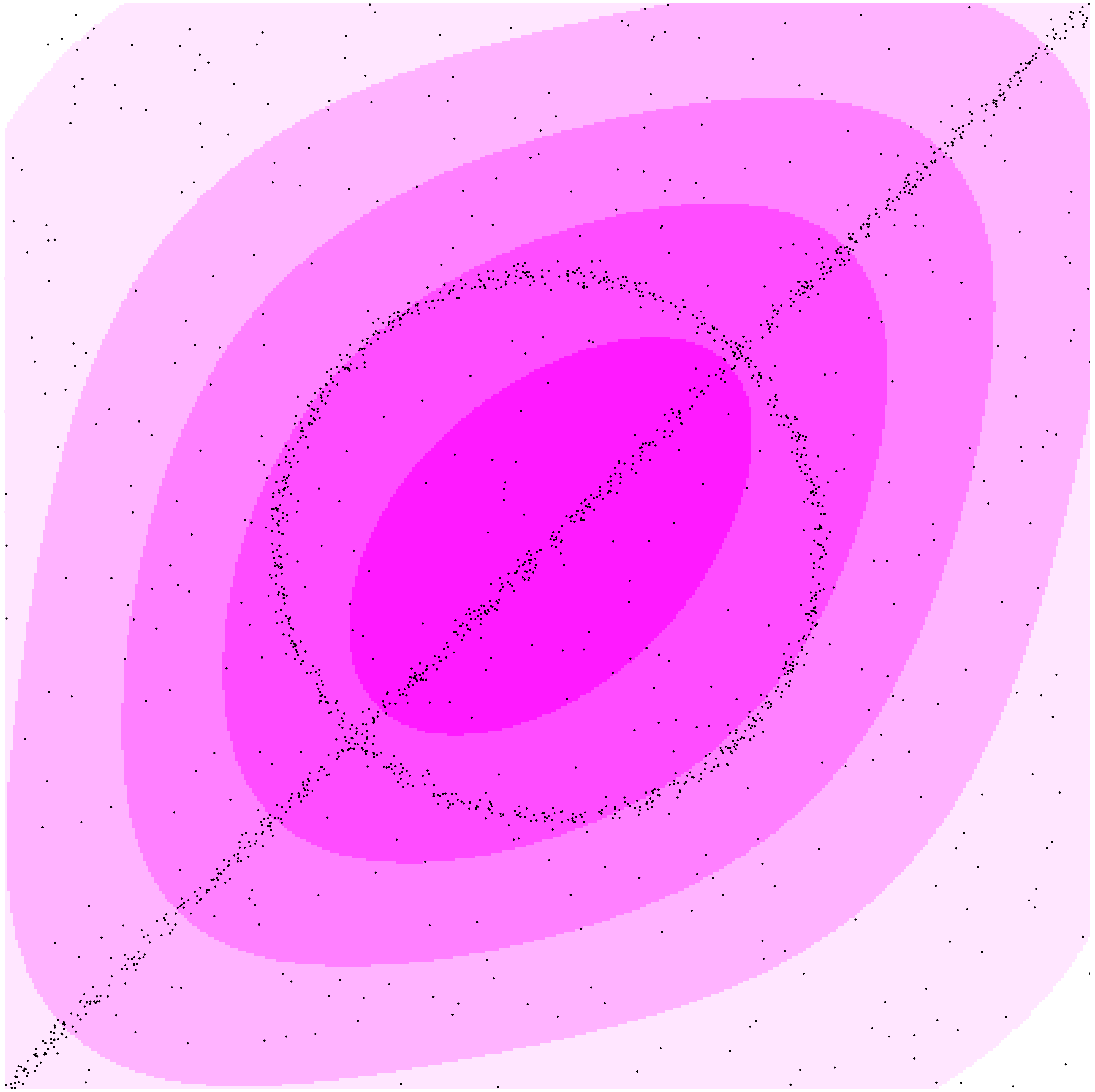
1. [Robustness] Kernel distance is distance-like: 1-Lipschitz, 1-semiconcave, proper and stable.
2. [Scalability] Kernel distance has a small coreset, making efficient inference possible on 100 million points.
3. [Relation to KDE] Geometric inference based on kernel distance works naturally via superlevel sets of KDE: sublevel sets of the kernel distance are superlevel sets of KDE.
4. [Algorithm] to approximate the sublevel set filtration of kernel distance from a point cloud sample.

Why kernel distance?

- People love and are familiar with KDE, especially with Gaussian kernel
- Kernel distance provides a proper way to relate KDE with properties that are crucial for geometric inference
- We could approximate the topology of kernel distance via point cloud samples

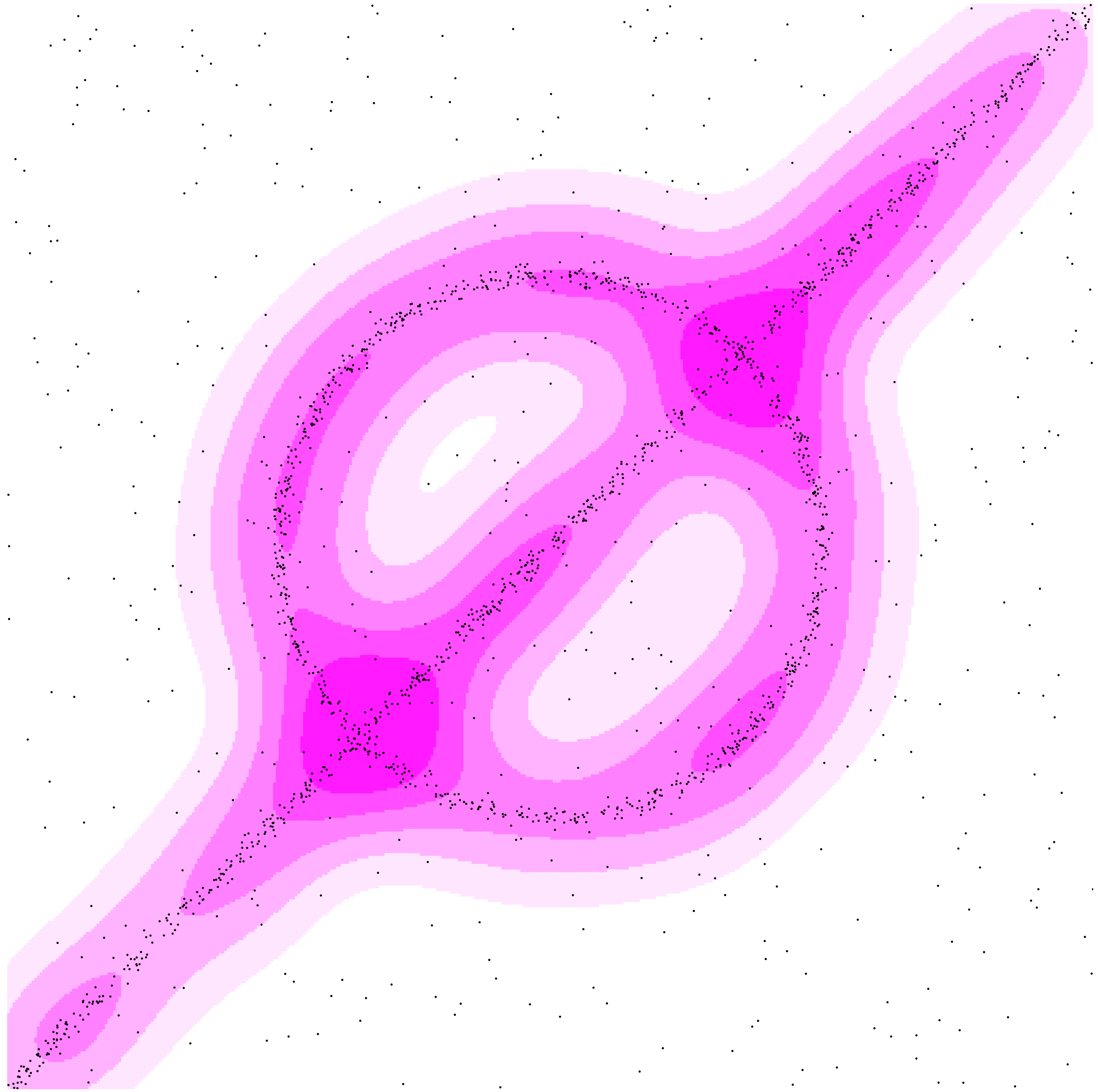
Experiments

An example with 25% of P as noise, $\sigma = 0.05$



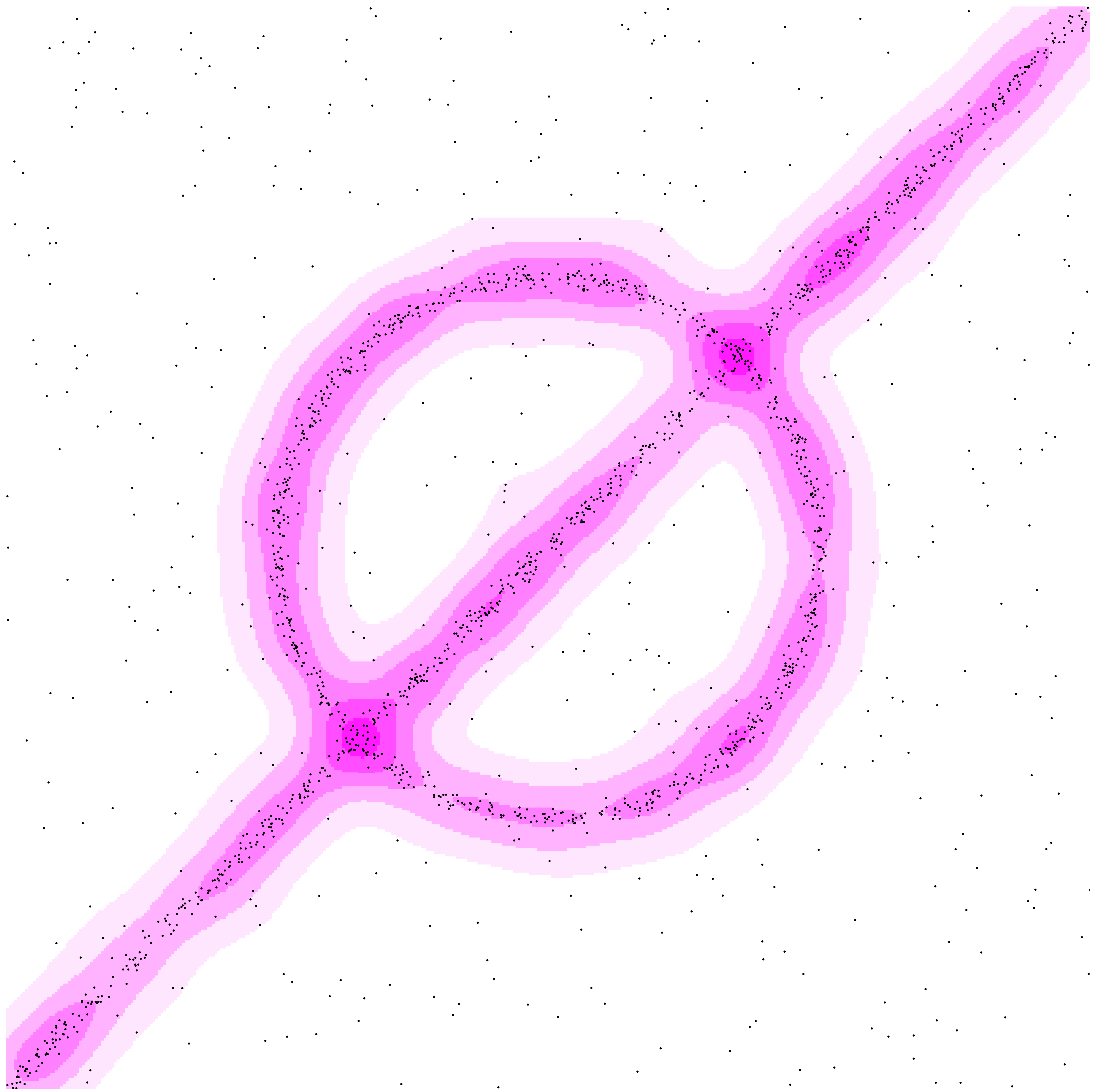
Experiments

An example with 25% of P as noise, $\sigma = 0.003$



Experiments

An example with 25% of P as noise, $\sigma = 0.001$



Kernel Distance is Distance-Like

Similar properties hold for the kernel distance defined as

$$d^K_\mu(x) = D_K(\mu, x) = \sqrt{\kappa(\mu,\mu) + \kappa(x,x) - 2\kappa(\mu,x)}$$
$$= \sqrt{c_\mu^2 - 2\mathrm{KDE}_\mu(x)}$$

For the point cloud setting,

$$d^K_P(x) = D_K(P, x) = \sqrt{\kappa(P,P) + \kappa(x,x) - 2\kappa(P,x)}$$
$$= \sqrt{c_P^2 - 2\mathrm{KDE}_P(x)}$$

Specifically, the following properties of d^K_μ allow it to inherit the reconstruction properties of $d^{\mathrm{CCM}}_{\mu,m_0}$.

(K1) d^K_μ is 1-Lipschitz on its input.
(K2) $(d^K_\mu)^2$ is 1-semiconvave: the map $x \mapsto (d^K_\mu(x))^2 - \|x\|^2$ is concave.
(K3) d^K_μ is proper.
(K4) [Stability] $\|d^K_\mu - d^K_\nu\|_\infty \leq D_K(\mu,\nu)$.

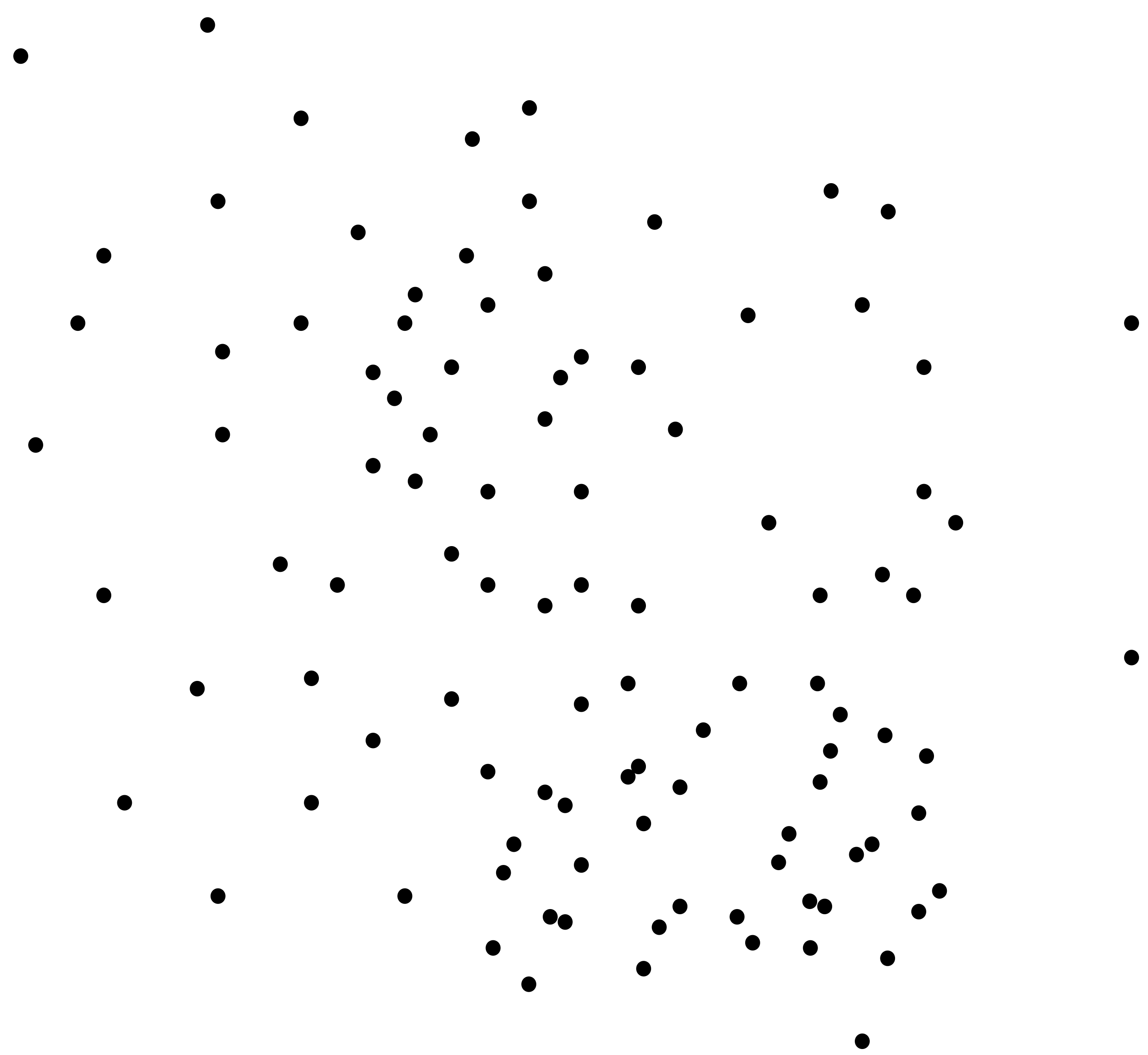
Advantages of the kernel distance summary

(I) Small coreset representation for sparse representation and efficient, scalable computation.

(II) Its inference is easily interpretable and computable through the superlevel sets of a KDE.

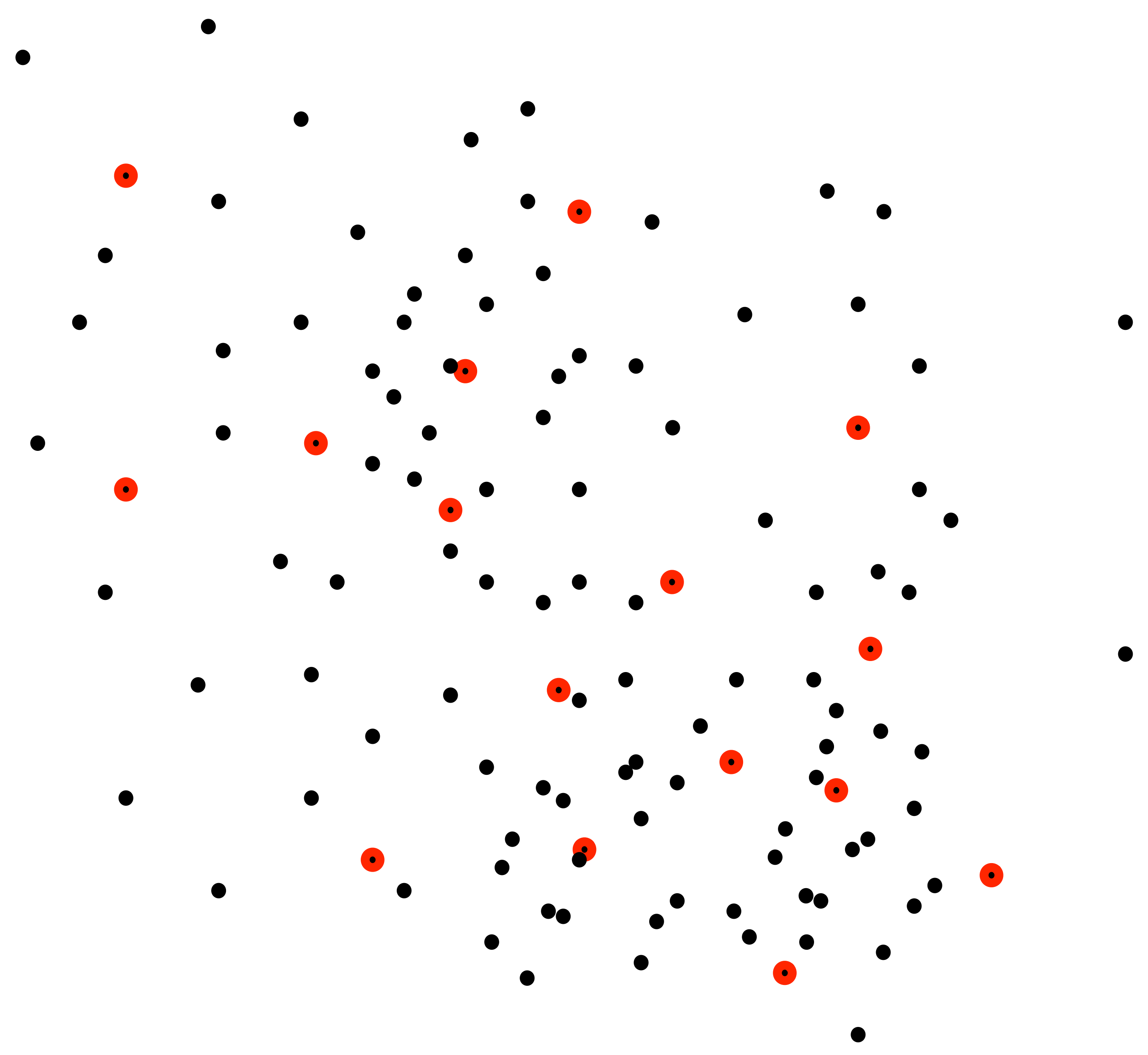
Small coreset

- There exists a small ϵ-coreset $Q \subset P$ s.t. $\|d_P^K - d_Q^K\|_\infty \leq \varepsilon$ and $\|\text{KDE}_P - \text{KDE}_Q\|_\infty \leq \varepsilon$ with probability at least $1 - \delta$.
- Size $O(((1/\varepsilon)\sqrt{\log(1/\varepsilon\delta)})^{2d/(d+2)})$ [Phillips 2013].
- The same holds under a random sample of size $O((1/\varepsilon^2)(d + \log(1/\delta)))$ [Joshi Kommaraju Phillips 2011].
- Operate with $|P| = 100{,}000{,}000$ [Zheng Jestes Phillips Li 2013].
- Stability of persistence diagram is preserved: $d_B(\text{Dgm}(\text{KDE}_P), \text{Dgm}(\text{KDE}_Q)) \leq \varepsilon$.



Small coreset

- There exists a small ϵ-coreset $Q \subset P$ s.t. $\|d_P^K - d_Q^K\|_\infty \leq \varepsilon$ and $\|\text{KDE}_P - \text{KDE}_Q\|_\infty \leq \varepsilon$ with probability at least $1 - \delta$.
- Size $O(((1/\varepsilon)\sqrt{\log(1/\varepsilon\delta)})^{2d/(d+2)})$ [Phillips 2013].
- The same holds under a random sample of size $O((1/\varepsilon^2)(d + \log(1/\delta)))$ [Joshi Kommaraju Phillips 2011].
- Operate with $|P| = 100{,}000{,}000$ [Zheng Jestes Phillips Li 2013].
- Stability of persistence diagram is preserved: $d_B(\text{Dgm}(\text{KDE}_P), \text{Dgm}(\text{KDE}_Q)) \leq \varepsilon$.



Geometric inference with KDE

Recall $d_P^K(x) = \sqrt{c_P^2 - 2\text{KDE}_P(x)}$ where c_P^2 is a constant that depends only on P. Perform geometric inference on noisy P by considering the super-level sets of KDE_P,

$$\{x \in \mathbb{R}^d \mid \text{KDE}_P(x) \geq \tau\}$$

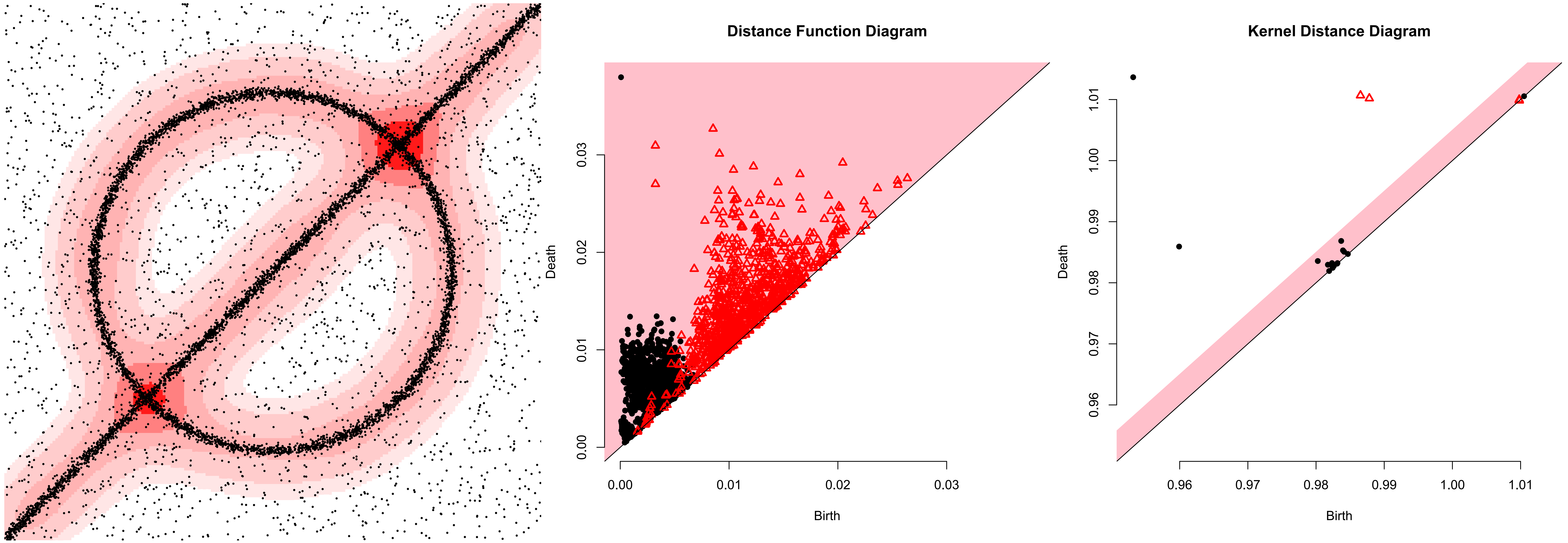
Key:

- $d_P^K(\cdot)$ is *monotonic* with $\text{KDE}_P(\cdot)$; as $d_P^K(x)$ gets smaller, $\text{KDE}_P(x)$ gets larger.
- A clean and natural interpretation of the reconstruction problem through the well-studied lens of KDE. Geometric inference with sublevel sets of d_P^K (superlevel sets of KDE_P).

Experiments

Experiments: Power of Kernel Distance

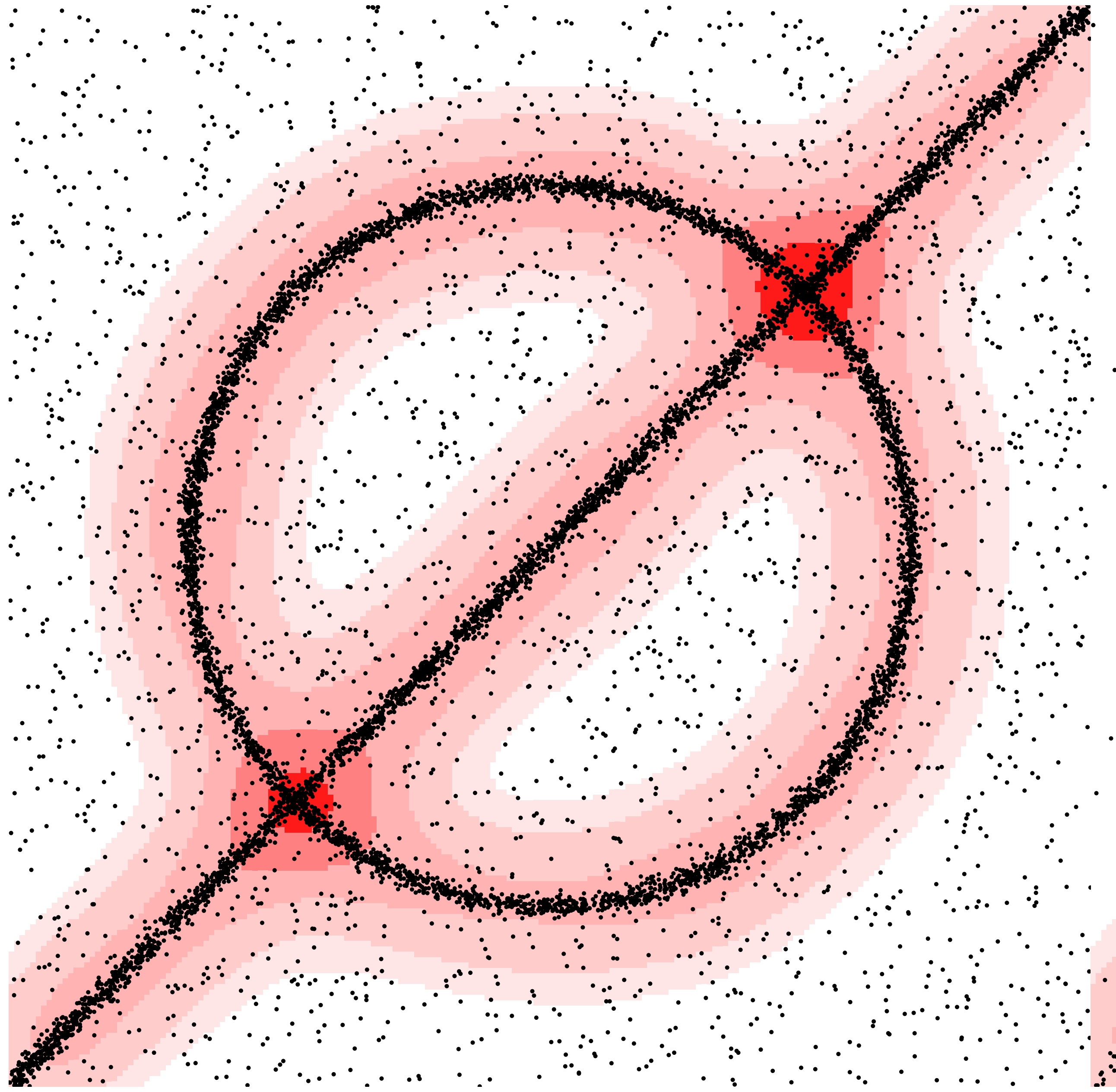
$10K$ points in $[0,1]^2$, noise $N(0, 0.005)$, 25% of P as noise

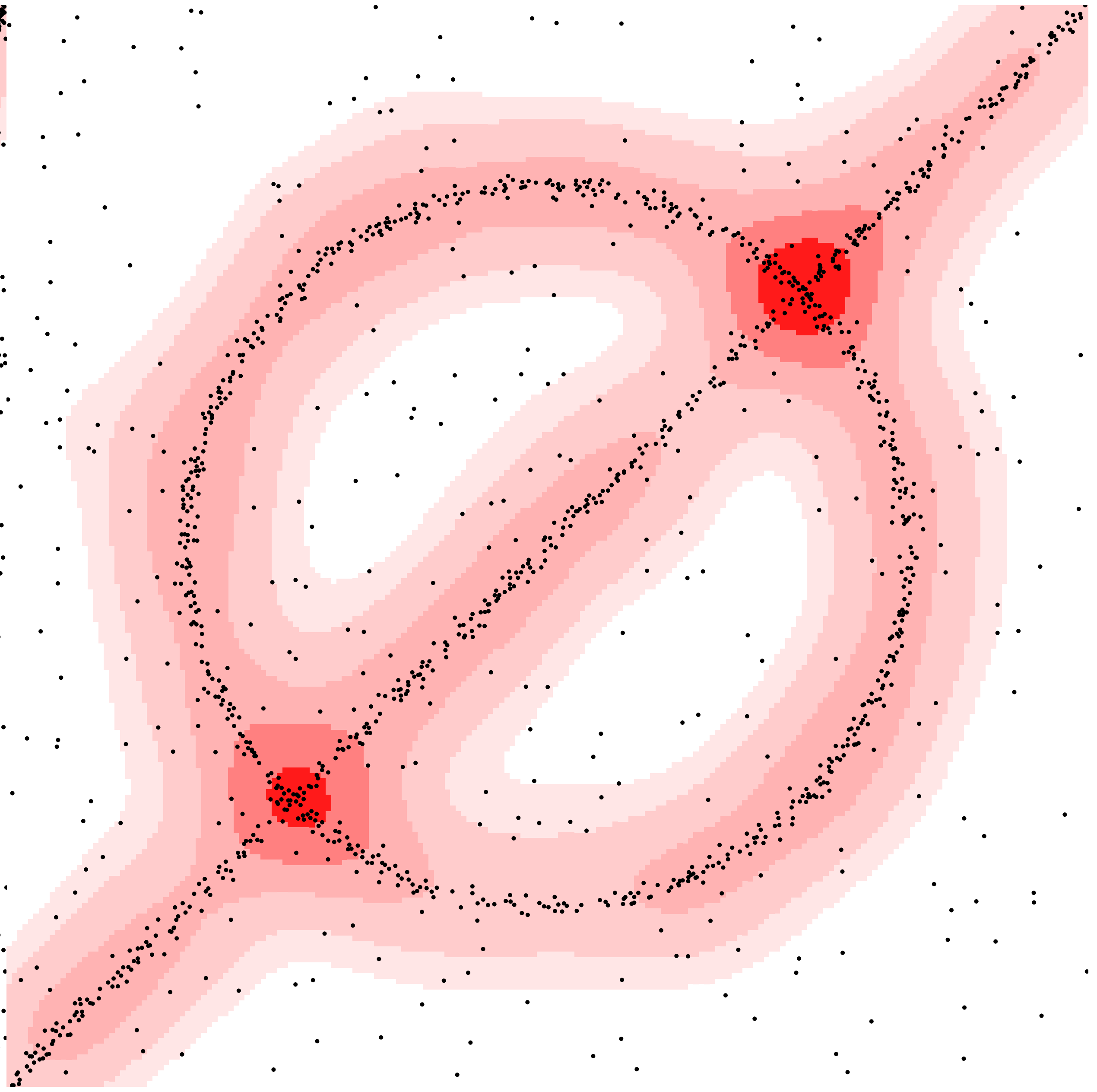


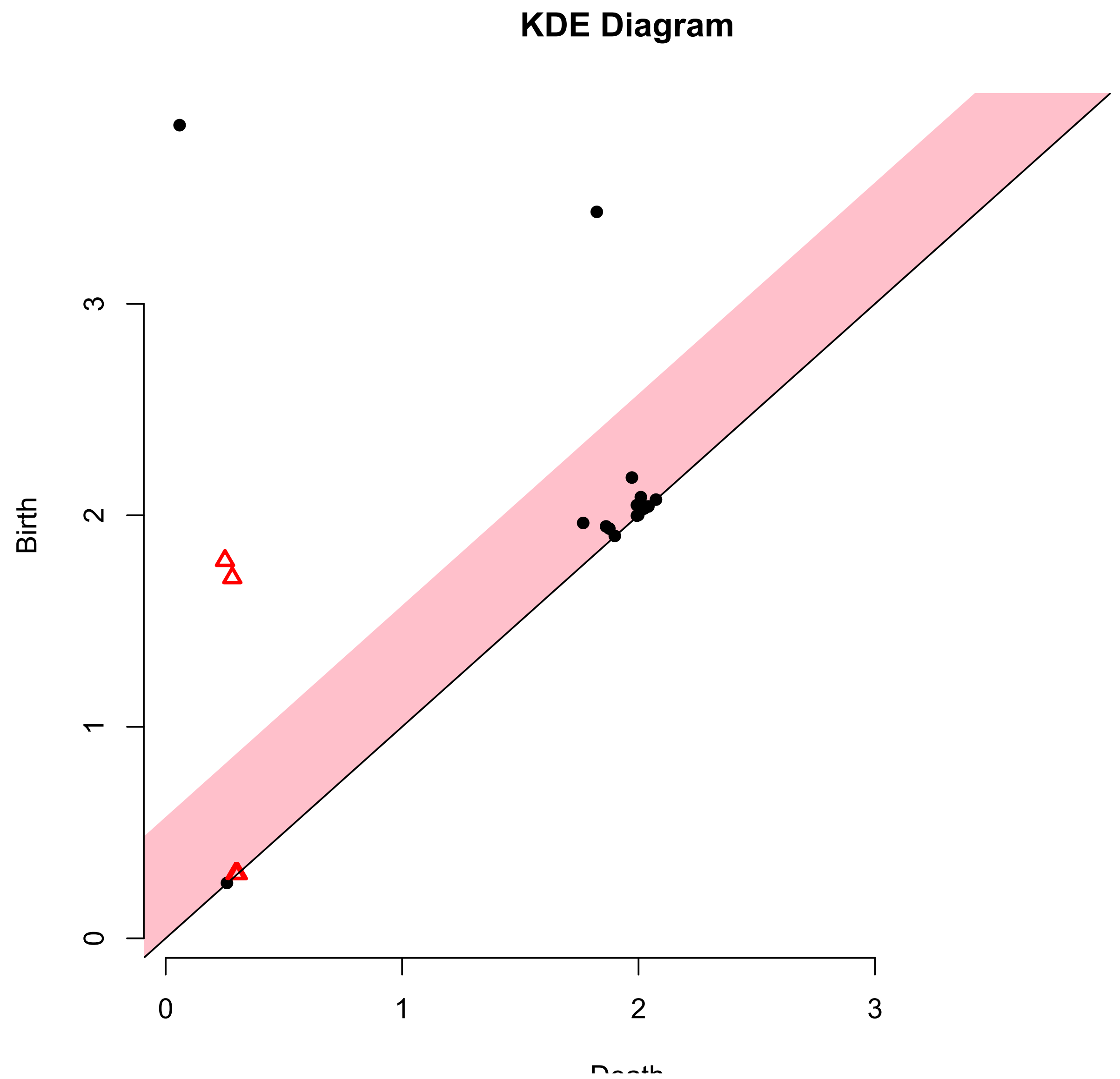
Persistence diagram using standard distance function (no useful features due to noise) and kernel distance.

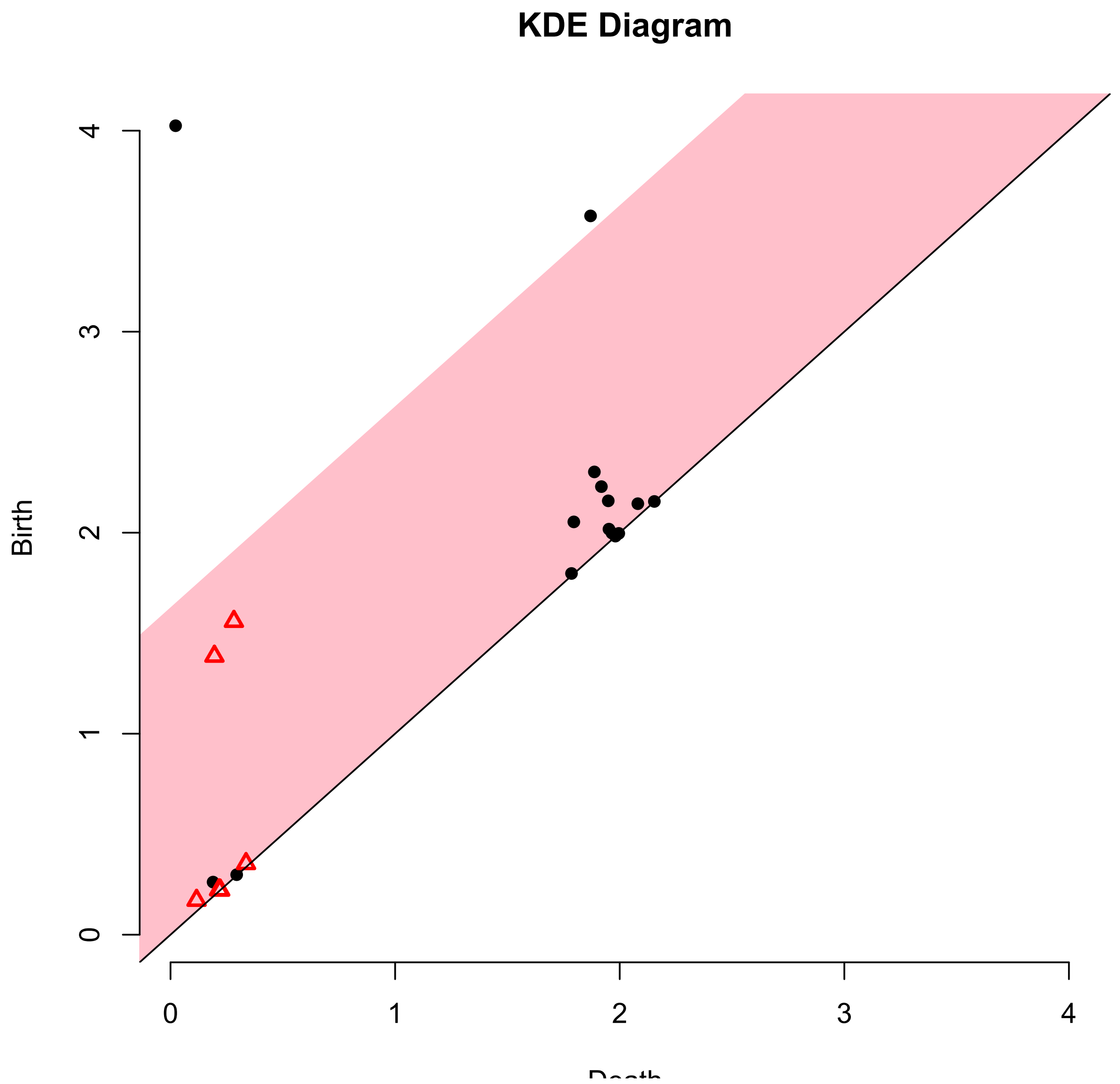
Experiments: Coreset

Original data v.s. Coreset, $10K$ vs. 1384 points









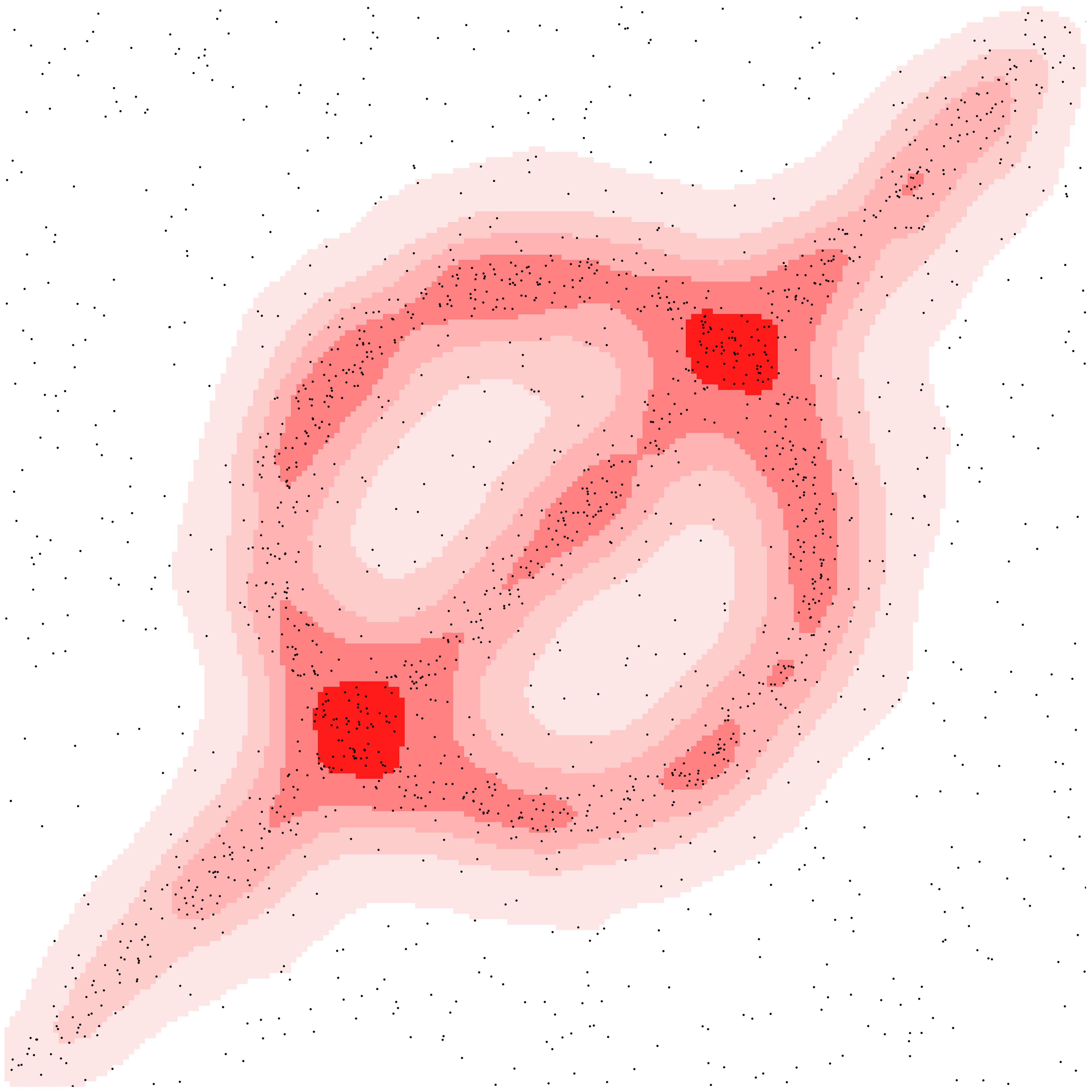
Other Kernels

Beyond Gaussian kernels

- More general theory for KDE with systematic understanding of family of kernels: distance to a measure (KNN kernel), kernel distance (a larger class of kernels, e.g. Gaussian, *Laplace*; triangle kernel may work OK in practice with less perfect properties).

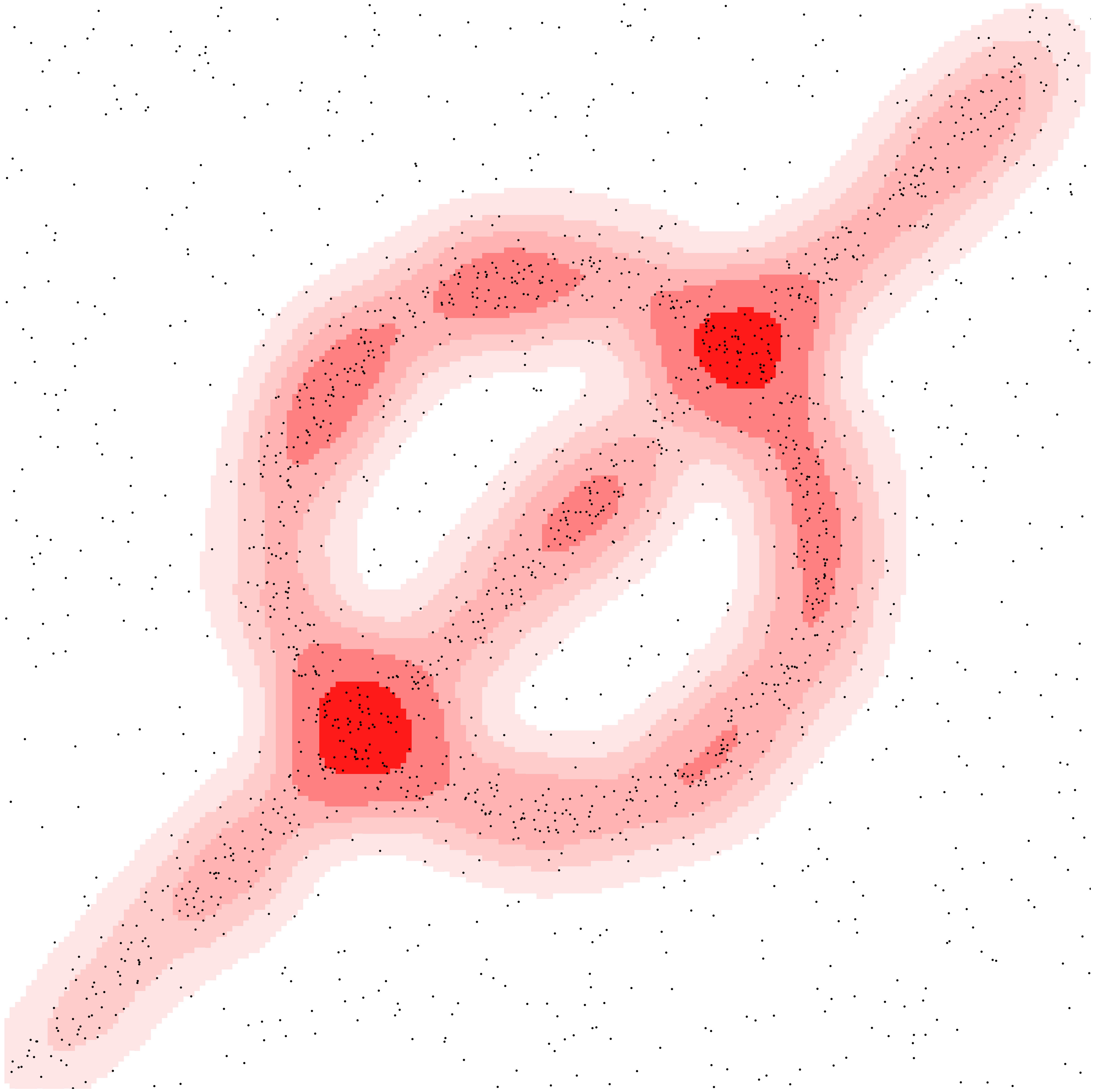
Alternative KDE

Laplace kernel $K(p, x) = \exp(-2\|x - y\|/\sigma)$



Alternative KDE

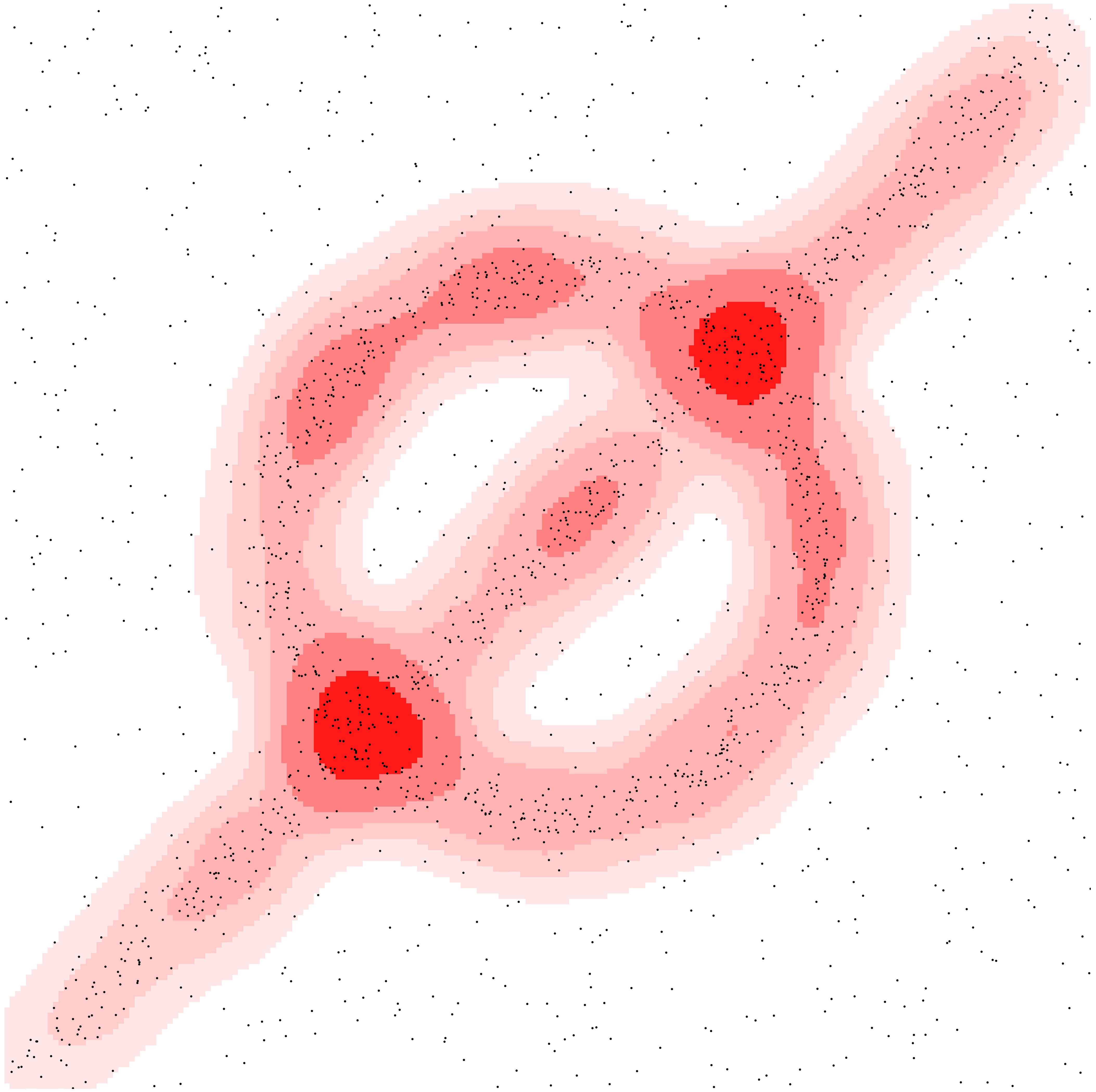
Triangle kernel: $K(x,p) = \max\left\{0, 1 - \frac{\|p-x\|}{\sigma=0.05}\right\}$



Alternative KDE

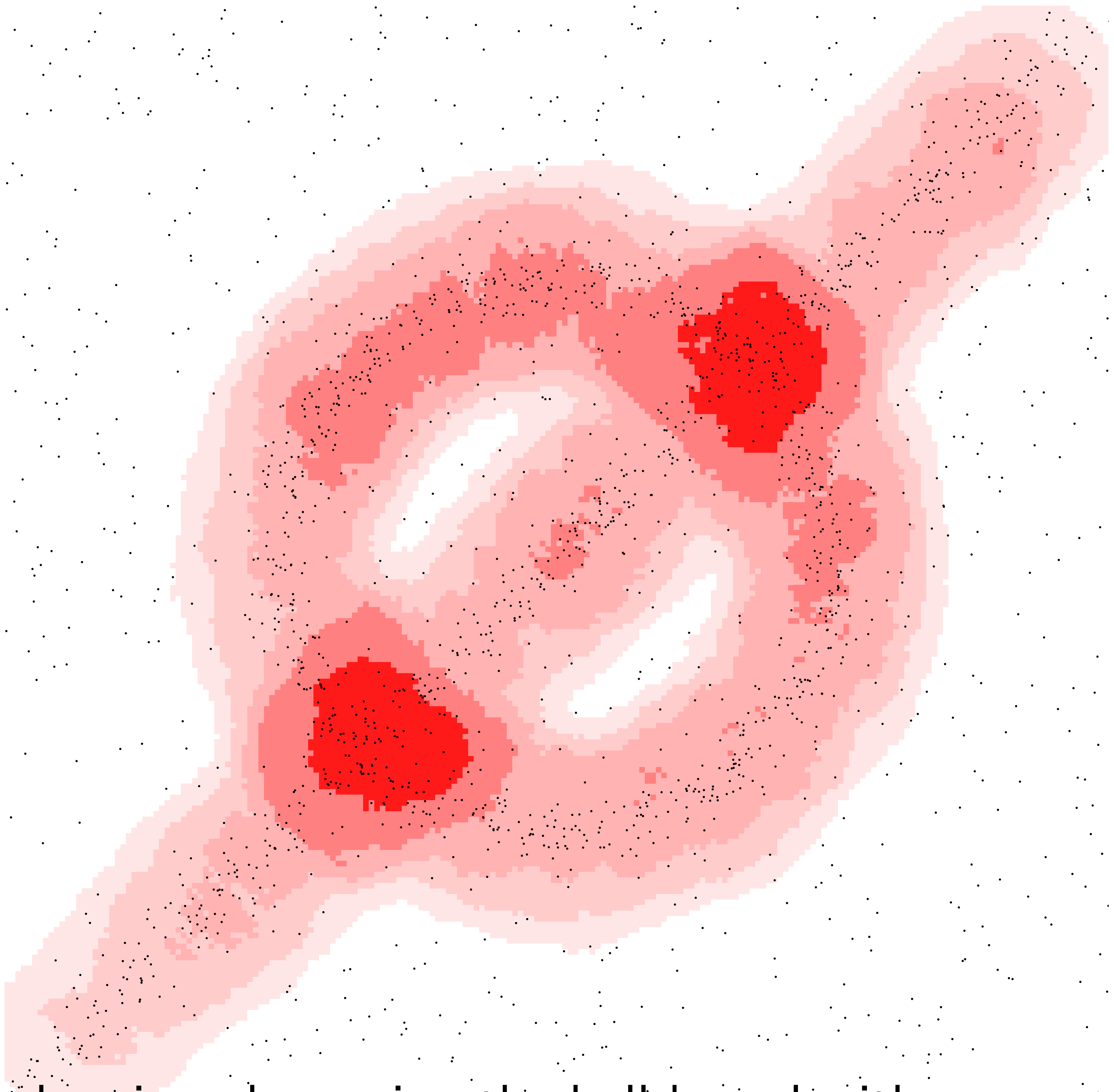
Epanechnikov kernel: (reconstruction)

$$K(x,p) = \max\left\{0, 1 - \frac{\|p-x\|^2}{(\sigma{=}0.05)^2}\right\}$$



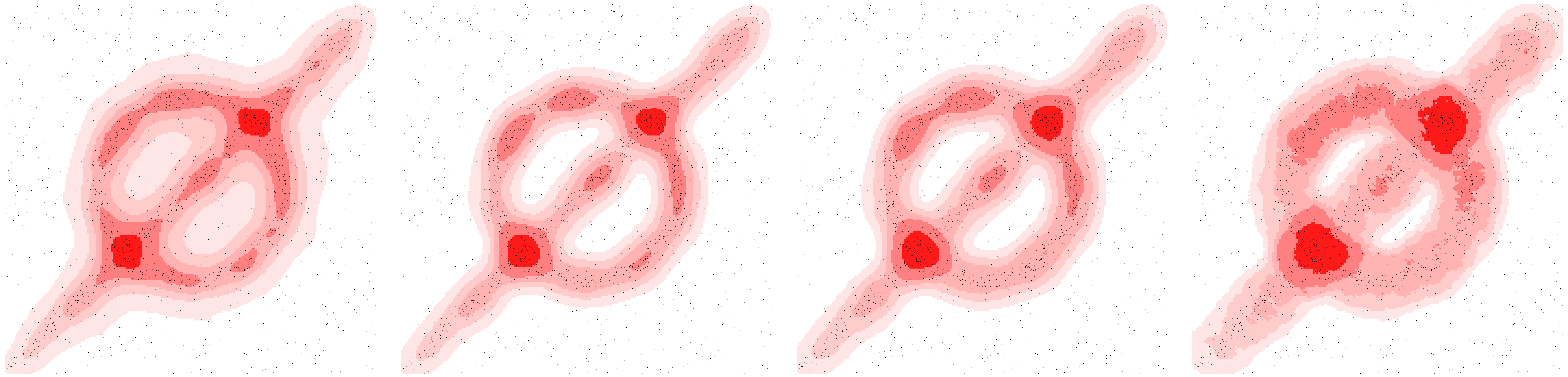
Alternative KDE

Ball kernel: $K(x,p) = \begin{cases} 1 & \textbf{if } \|p - x\| \leq \sigma = 0.05 \\ 0 & \textbf{otherwise}. \end{cases}$



$\alpha-$shape can be viewed as using the ball kernel with $\sigma = \alpha$ and $r = 1/n$.

Alternative KDEs



Multi-dim / scale space persistence? Parameter selection?

Two parameters r (isolevel) and σ (outlier/bandwidth) that control the scale.

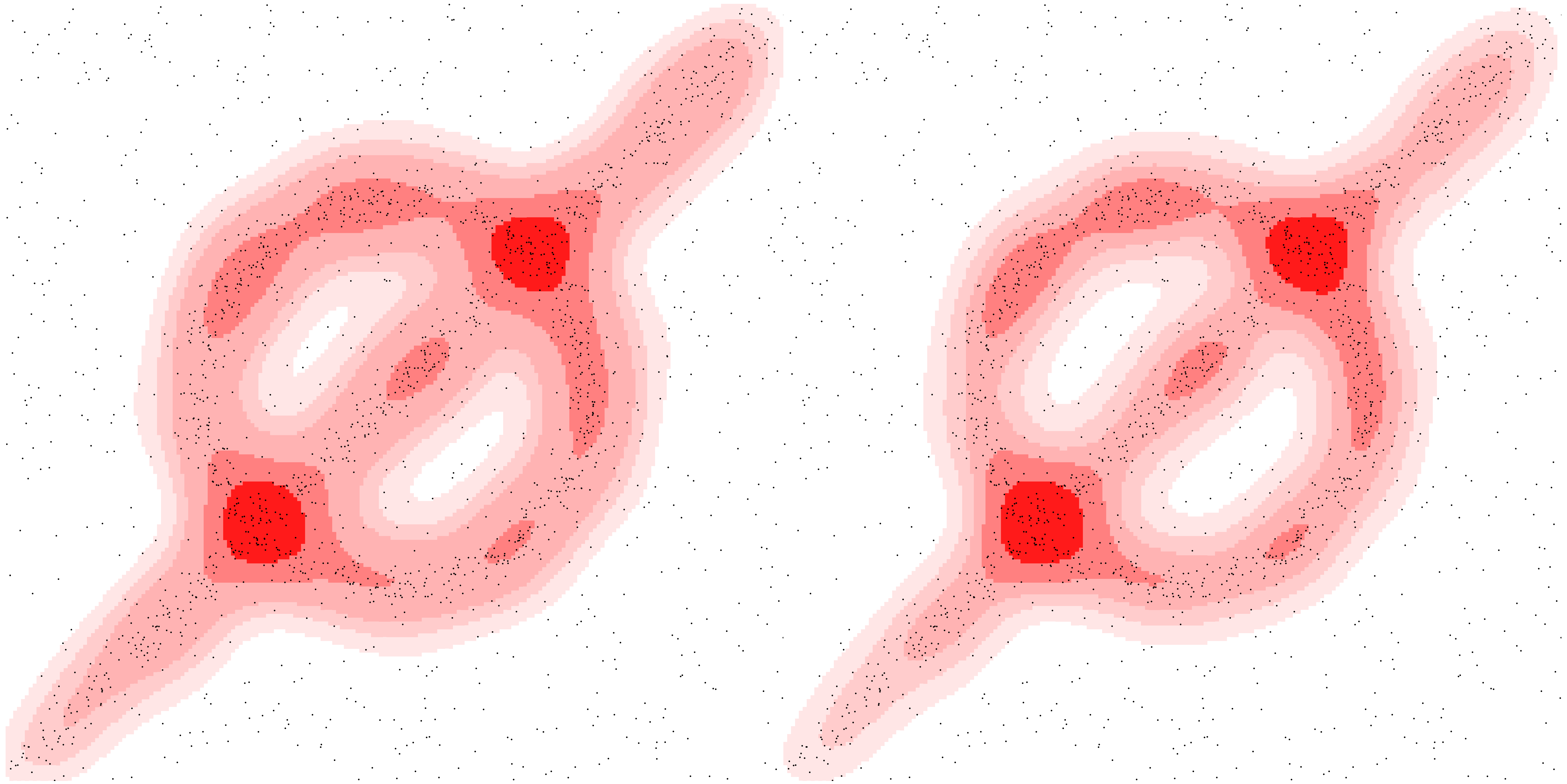


Figure: Sublevel sets for the kernel distance. Left: fix σ, vary r. Right: fix r, vary σ. The values of σ and r are chosen to make the plots similar.

Structural Inference using distance to measure

Distance to a measure [Chazal Cohen-Steiner Merigot 2011]

Intuition: W_2 distance to m_0 fraction of the space.

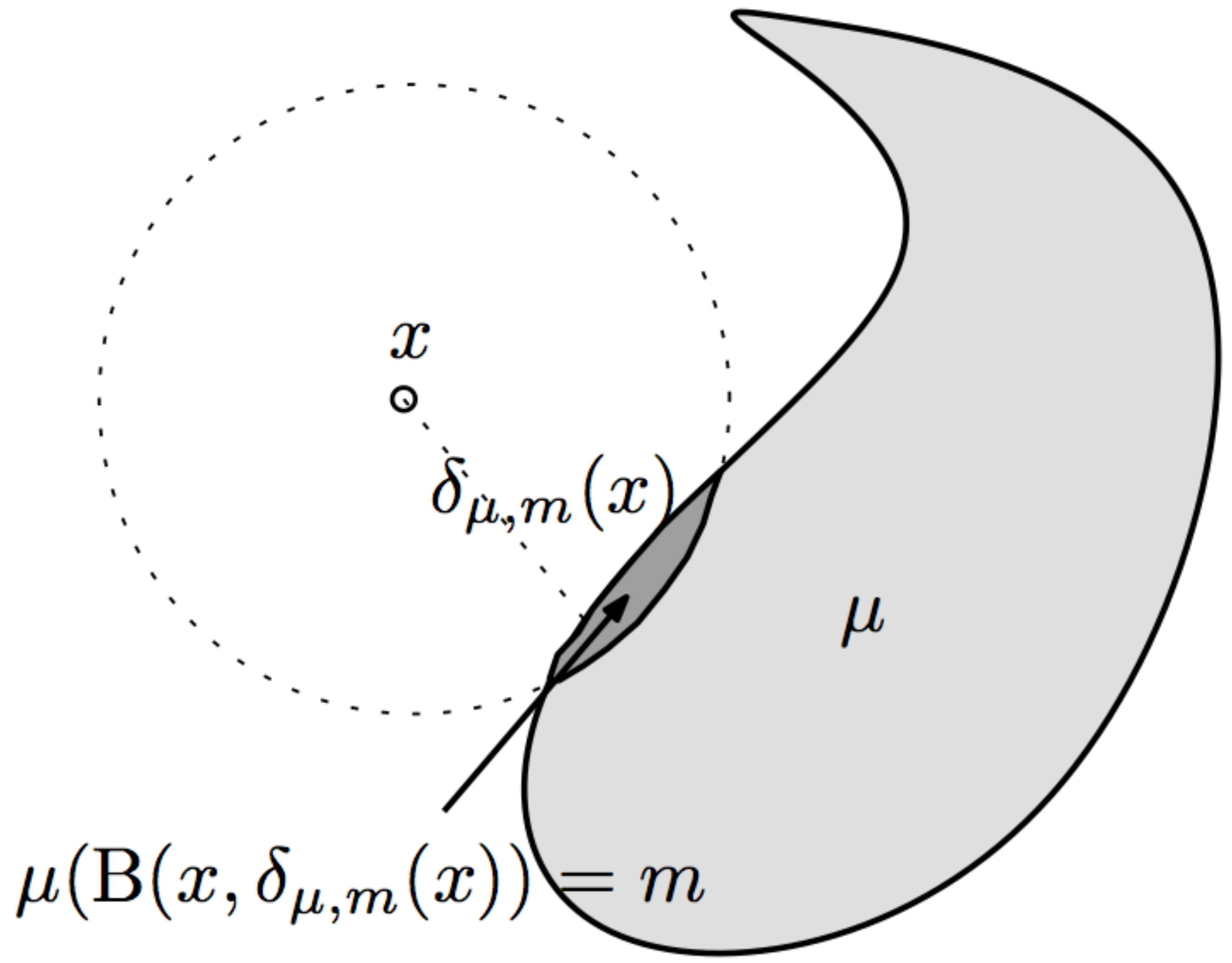
μ: probability measure on $\mathbb{R}^d$

$m_0 > 0$: a parameter smaller than the total mass of μ

The distance to a measure $d^{\mathrm{CCM}}_{\mu,m_0} : \mathbb{R}^n \to \mathbb{R}^+$, $\forall x \in \mathbb{R}^d$,

$$d^{\mathrm{CCM}}_{\mu,m_0}(x) = \left(\frac{1}{m_0} \int_{m=0}^{m_0} (\delta_{\mu,m}(x))^2 \mathrm{d}m \right)^{1/2}$$

where $\delta_{\mu,m}(x) = \inf \left\{ r > 0 : \mu(\bar{B}_r(x)) \leq m \right\}$.



Wasserstein-2 distance $W_2(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} ||x-y||^2 \mathrm{d}\pi(x,y) \right)^{1/2}$

Distance to a measure $d^{\mathrm{CCM}}_{\mu,m_0}$ is distance-like

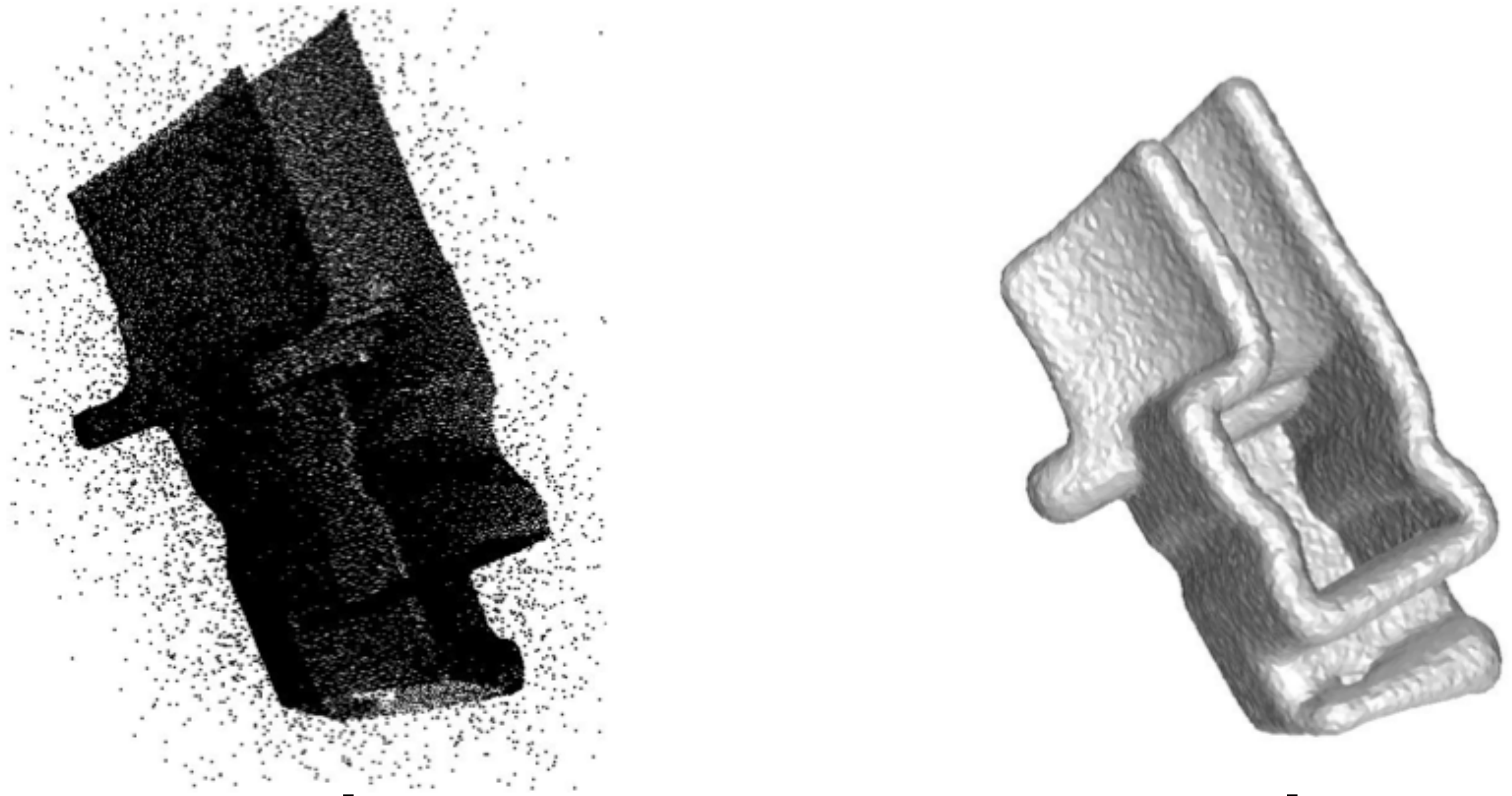
(D1) 1-Lipschitz

(D2) 1-semiconcave

(D3) Proper (for Groves Isotopy Lemma).

(D4) [Stability] For probability measures μ and ν on $\mathbb{R}^d$ and $m_0 > 0$, then $\|d^{\mathrm{CCM}}_{\mu,m_0} - d^{\mathrm{CCM}}_{\nu,m_0}\|_\infty \leq \frac{1}{\sqrt{m_0}} W_2(\mu,\nu)$.

Stability: two distance to a measure are close if their corresponding measures are close.



[Chazal Cohen-Steiner Merigot 2011]

Tools for persistent homology computation

Computing PH

- Ripser:
  - https://github.com/Ripser/ripser
  - http://live.ripser.org/
- TDA-R:
  - https://cran.r-project.org/web/packages/TDA/index.html
- DIPHA
  - https://github.com/DIPHA/dipha
- PHAT
  - https://github.com/blazs/phat
- GUDHI
  - https://project.inria.fr/gudhi/software/

https://people.maths.ox.ac.uk/otter/PH_programs

Thanks!

Any questions?

You can find me at: **beiwang@sci.utah.edu**

CREDITS

Special thanks to all people who made and share these awesome resources for free:

- Presentation template designed by Slidesmash
- Photographs by unsplash.com and pexels.com
- Vector Icons by Matthew Skiles

Presentation Design

This presentation uses the following typographies and colors:

Free Fonts used:

http://www.1001fonts.com/oswald-font.html

https://www.fontsquirrel.com/fonts/open-sans

Colors used