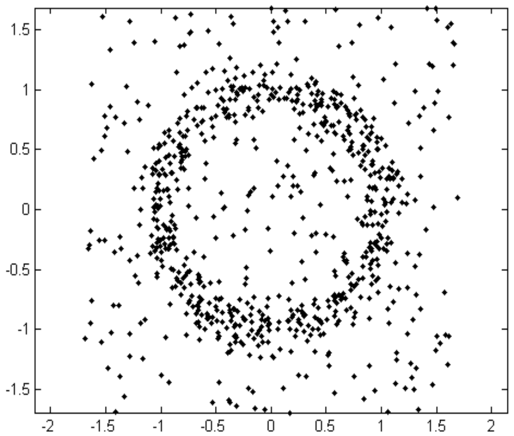
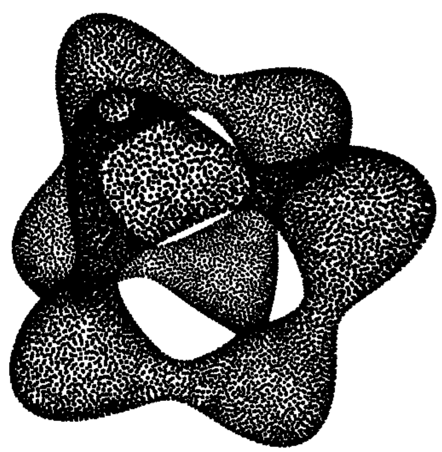
MPRI 2014/15

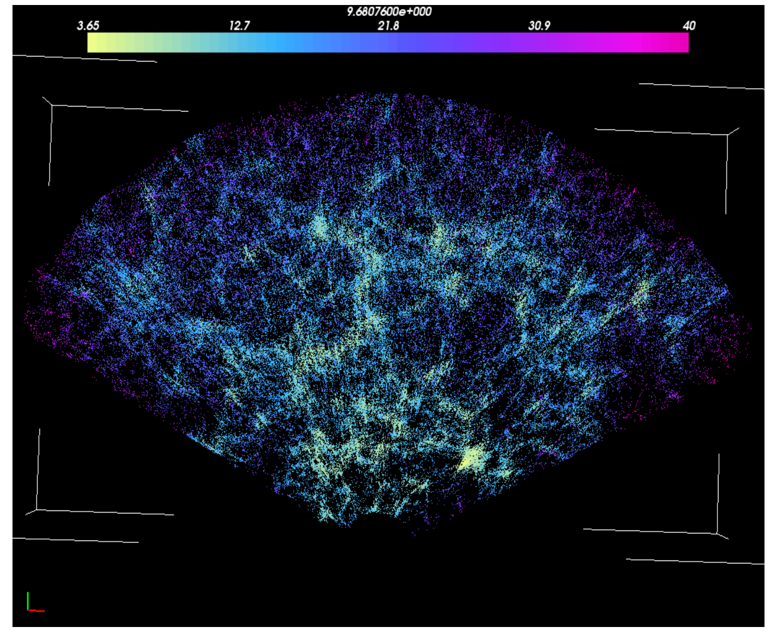
Dealing with noise and outliers: distance functions to measures

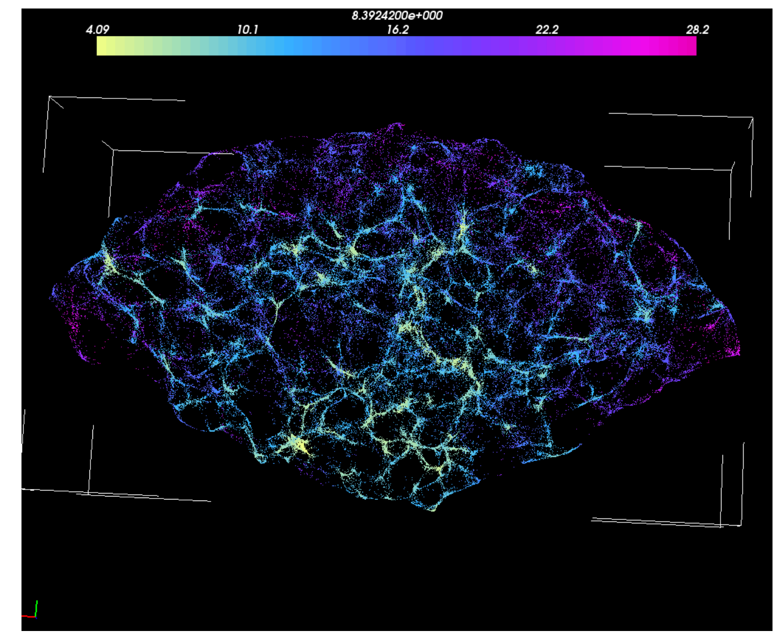
F. Chazal
Geometrica Group
INRIA Saclay

Introduction



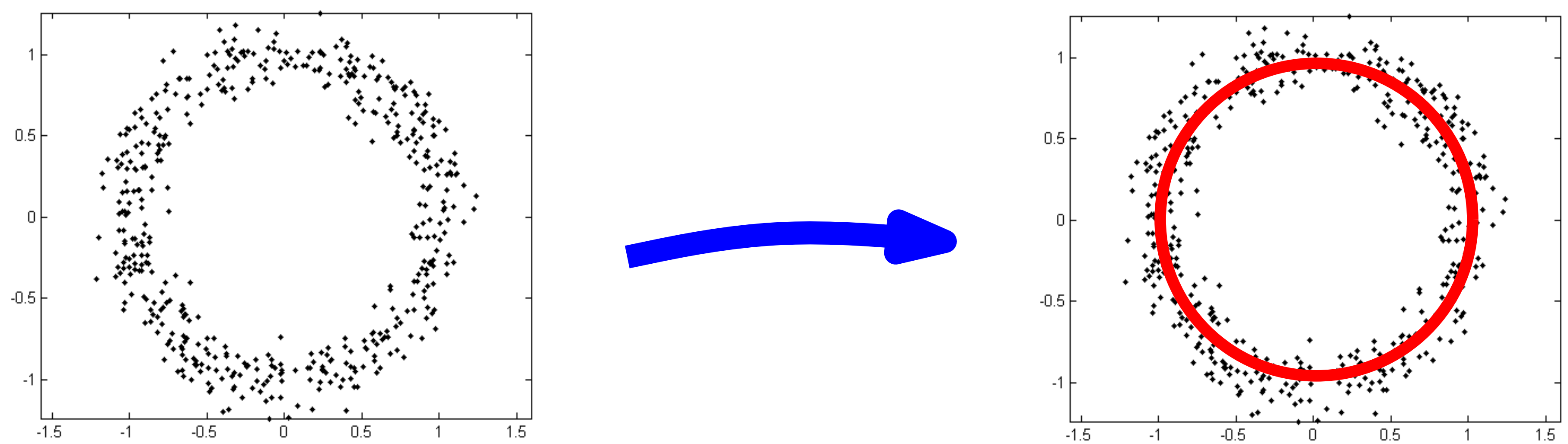






- Data are often corrupted by noise and outliers
- What can we say about the topology/geometry underlying such noisy data?
- Is it possible to generalize the distance based approach?

Geometric Inference



Question: Given an approximation C of a geometric object K, is it possible to reliably estimate the topological and geometric properties of K, knowing only the approximation C?

Question *: Given a point cloud C (or some other more complicated set), is it possible to infer some robust topological or geometric information of C?

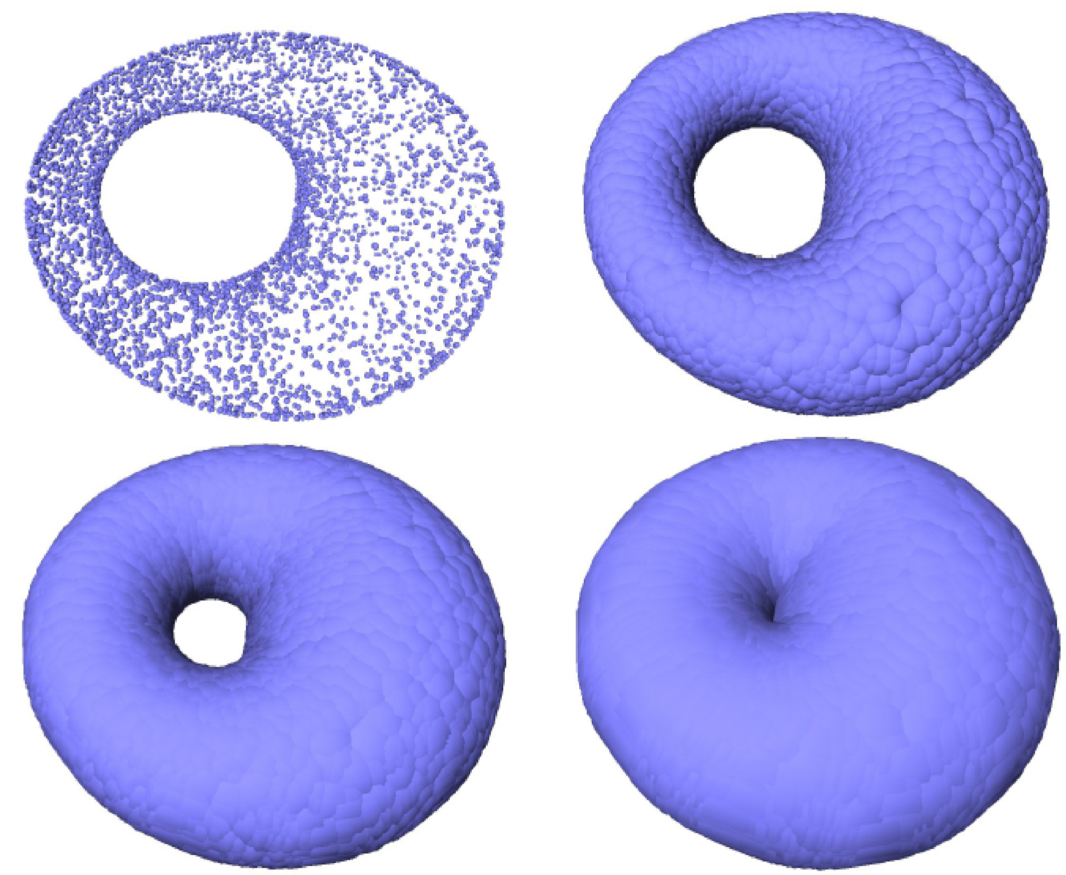
- The answer depends on:
  - the considered class of objects (no hope to get a positive answer in full generality),
  - a notion of distance between the objects (approximation).

Distance functions for geometric inference

Considered objects: compact subsets K of $\mathbb{R}^d$

Distance:
distance function to a compact $K \subset \mathbb{R}^d$: $d_K : x \to \inf_{p \in K} \|x - p\|$
Hausdorf distance between two compact sets:

$$d_H(K, K') = \sup_{x \in \mathbb{R}^d} |d_K(x) - d_{K'}(x)|$$



- Replace K and C by d_K and d_C
- Compare the topology of the offsets $K^r = d_K^{-1}([0, r])$ and $C^r = d_C^{-1}([0, r])$

Distance functions: the three (indeed two) main ingredients of stability

- the stability of the map $K \mapsto d_K$:
  $\|d_K - d_{K'}\|_\infty = d_H(K, K')$

Distance functions: the three (indeed two) main ingredients of stability

- the stability of the map $K \mapsto d_K$:
  $\|d_K - d_{K'}\|_\infty = d_H(K, K')$

- the 1-Lipschitz property for d_K; $\longrightarrow$ d_K is differentiable almost everywhere.

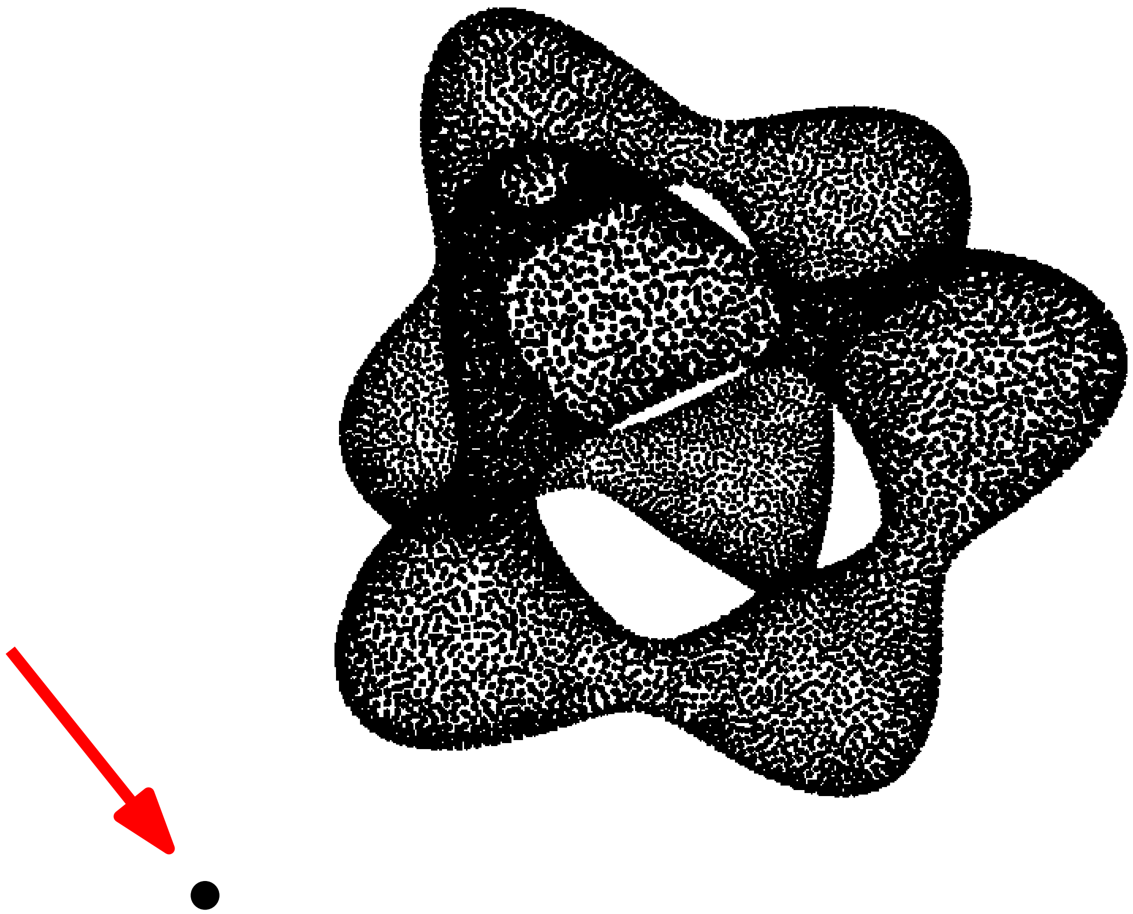
Distance functions: the three (indeed two) main ingredients of stability

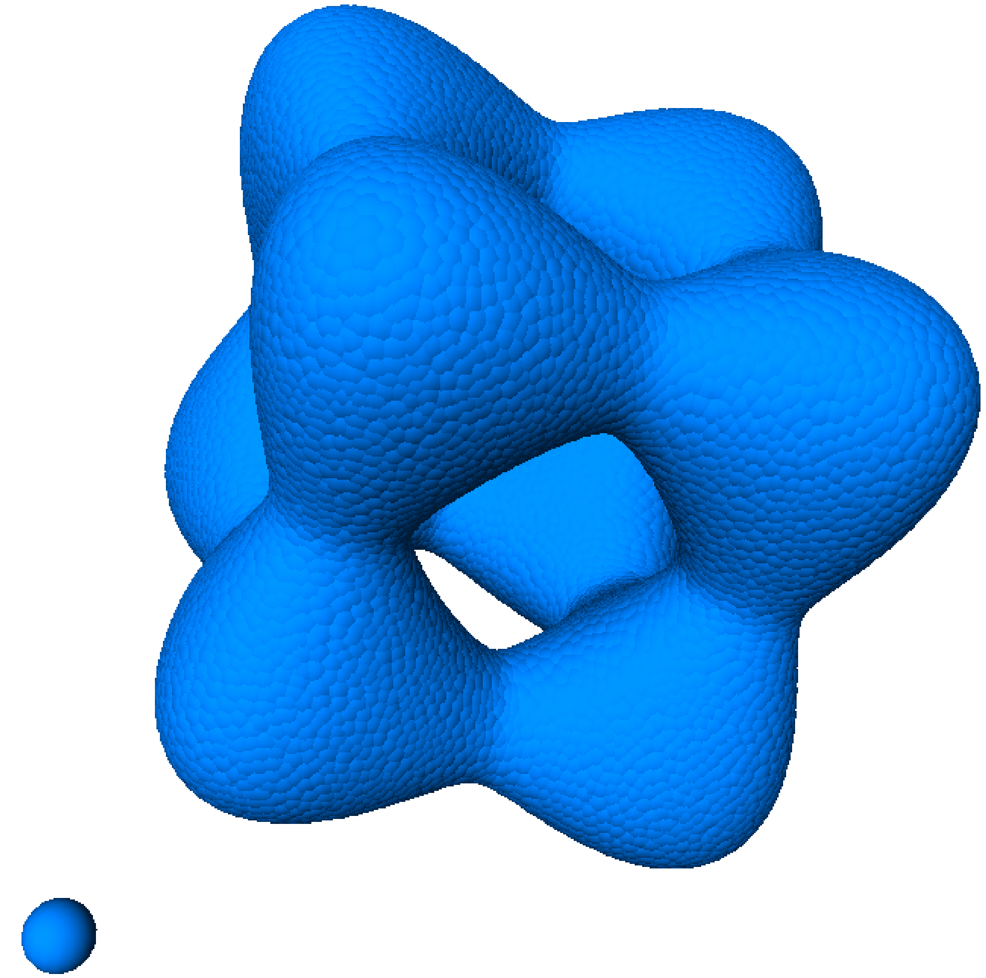
- the stability of the map $K \mapsto d_K$:
  $\|d_K - d_{K'}\|_\infty = d_H(K, K')$

- the 1-Lipschitz property for d_K; $\longrightarrow$ d_K is differentiable almost everywhere.

- the 1-concavity of the function d_K^2: $x \to \|x\|^2 - d_K^2(x)$ is convex. $\longrightarrow$
  - the gradient vector field ∇d_K is well defined and integrable (although not continuous).
  - Isotopy lemma.
  - d_K admits a second derivative almost everywhere.

The problem of “outliers”

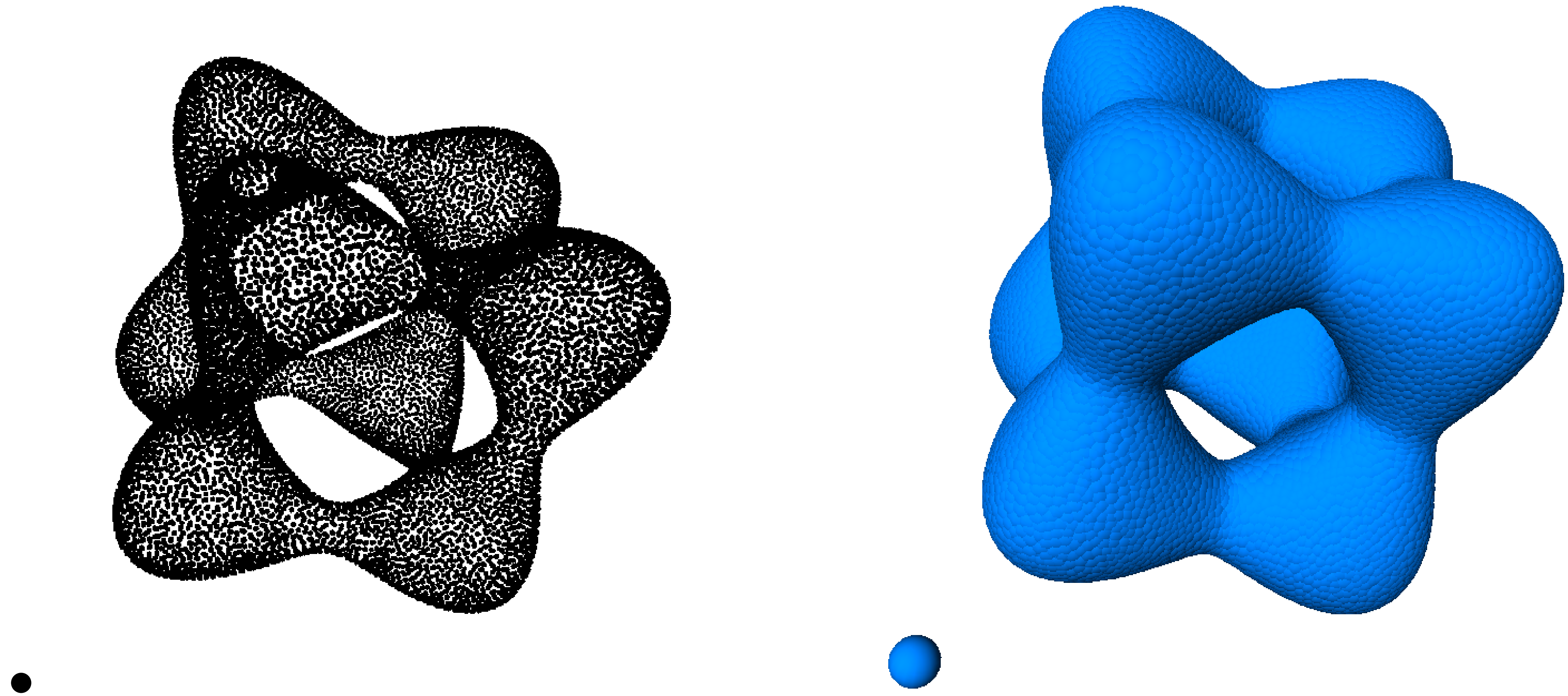




If $K' = K \cup \{x\}$ where $d_K(x) > R$, then $\|d_K - d_{K'}\|_\infty > R$: offset-based inference methods fail!

Question: Can we generalize the previous approach by replacing the distance function by a “distance-like” function having a better behavior with respect to “‘noise” and “outliers”?

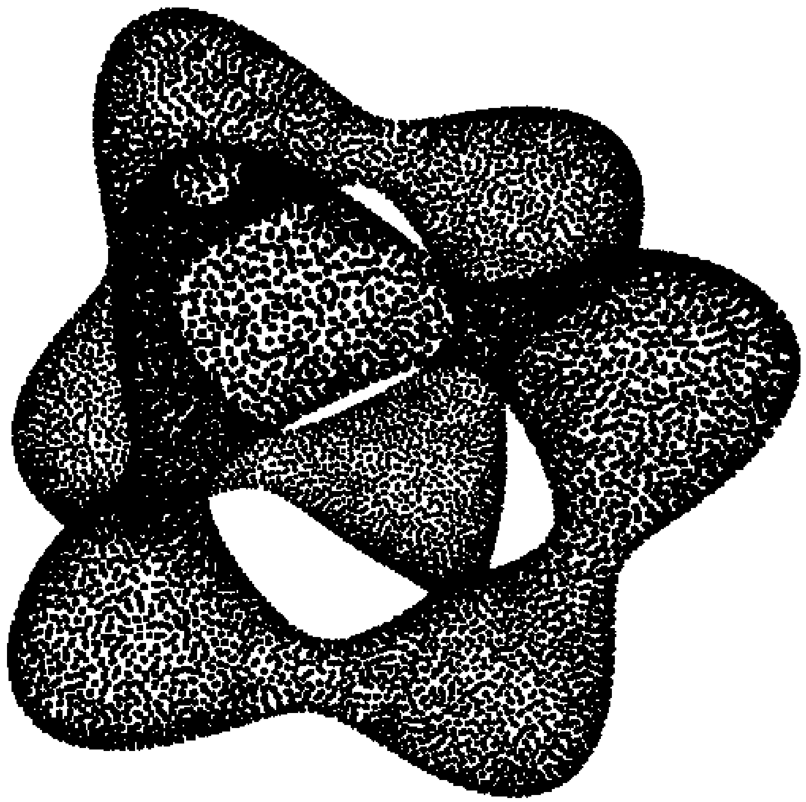
Replacing compact sets by measures

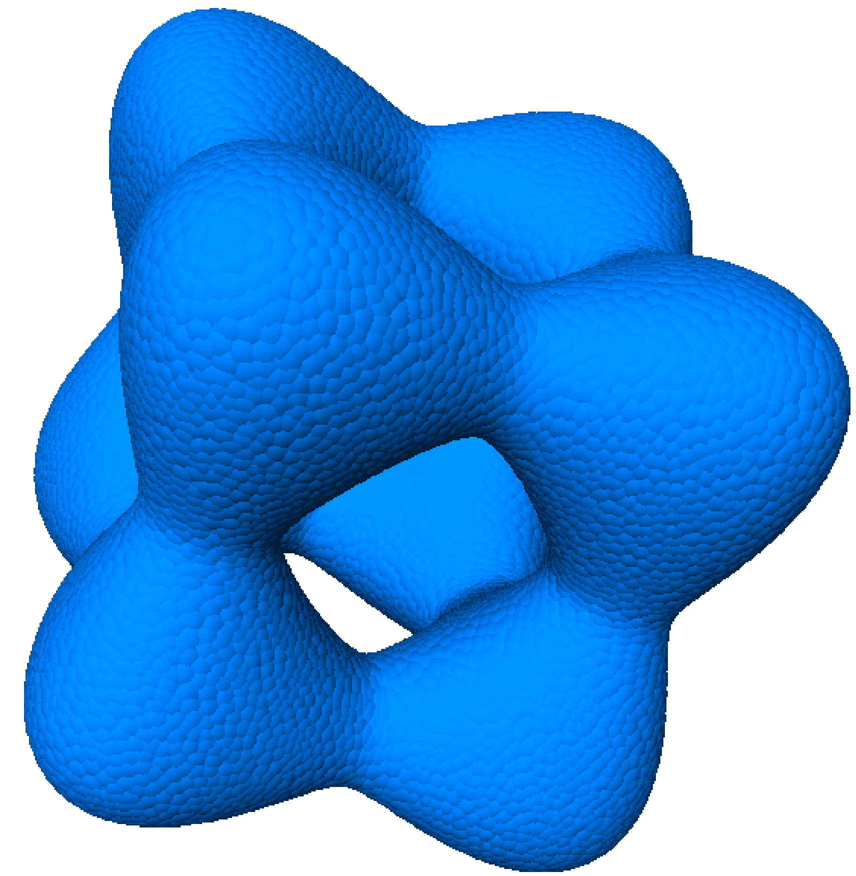


A **measure** μ is a mass distribution on $\mathbb{R}^d$:
mathematically, it is defined as a map μ that takes a (Borel) subset $B \subset \mathbb{R}^d$ and outputs a nonnegative number $\mu(B)$. Moreover we ask that if (B_i) are disjoint subsets, $\mu\left(\bigcup_{i\in\mathbb{N}} B_i\right) = \sum_{i\in\mathbb{N}} \mu(B_i)$.

$\mu(B)$ corresponds to to the mass of μ contained in B

Replacing compact sets by measures

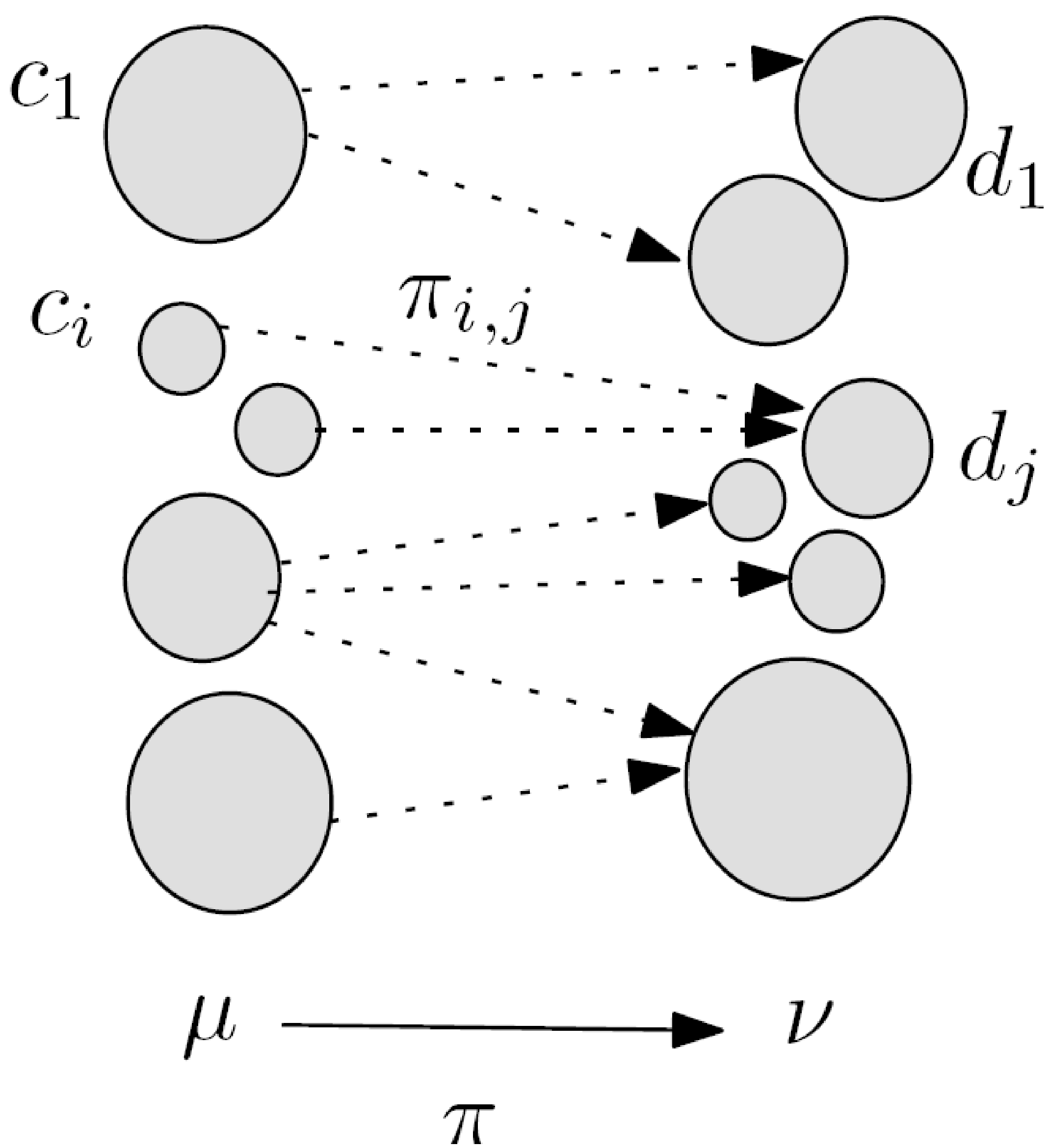




- a point cloud $C = \{p_1, \ldots, p_n\}$ defines a measure $\mu_C = \frac{1}{n}\sum_i \delta_{p_i}$
- the volume form on a k-dimensional submanifold M of $\mathbb{R}^d$ defines a measure $\mathrm{vol}_{k|M}$.
- etc...

Distance between measures

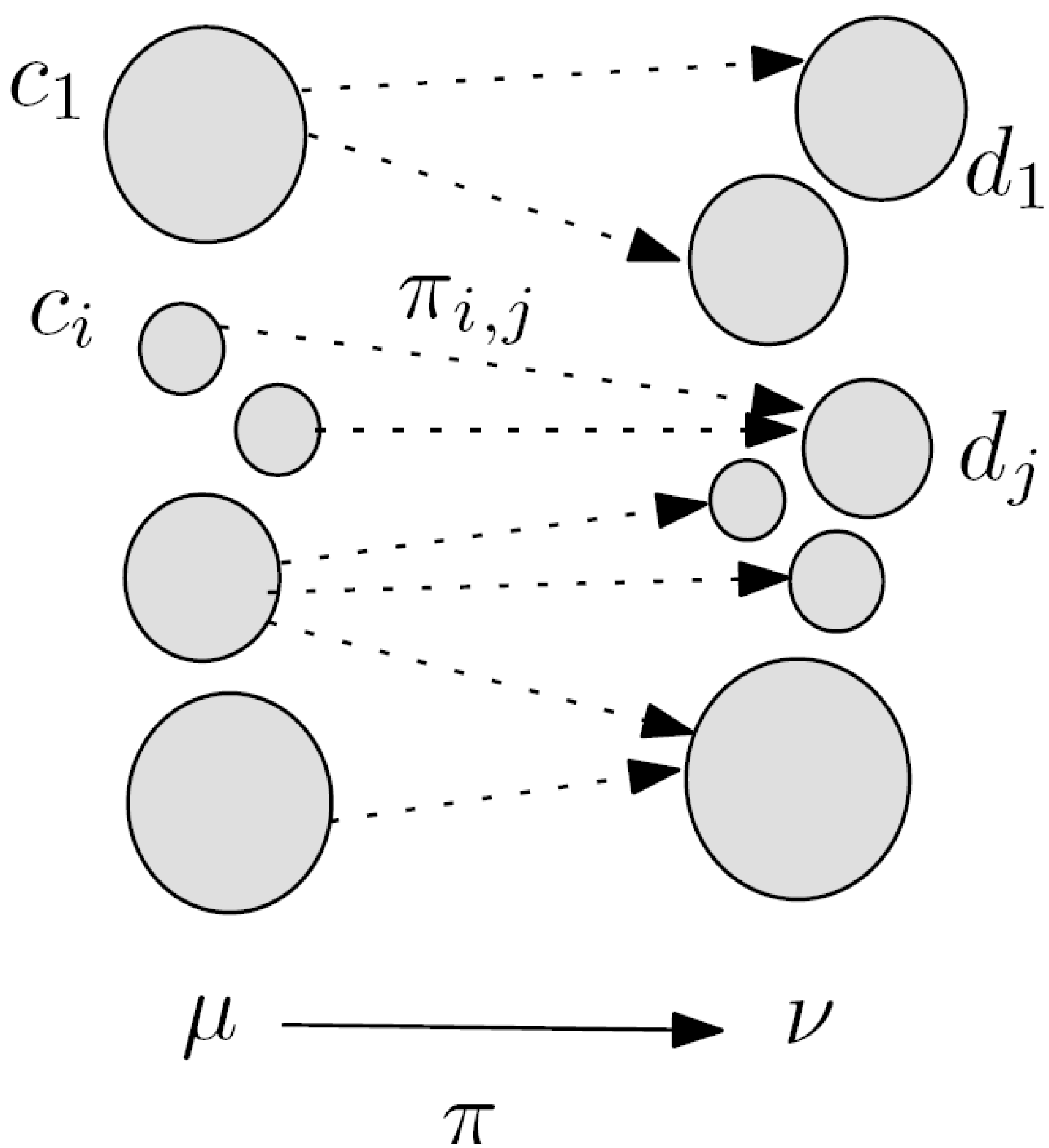
"The" **Wasserstein distance** $d_W(\mu,\nu)$ between two probability measures μ,ν quantifies the optimal cost of pushing μ onto ν, the cost of moving a small mass dx from x to y being $\|x-y\|^2 dx$.



1. μ and ν are discrete measures: $\mu = \sum_i c_i \delta_{x_i}$, $\nu = \sum_j d_j \delta_{y_j}$ with $\sum_j d_j = \sum_i c_i$.
2. *Transport plan*: set of coefficients $\pi_{ij} \geq 0$ with $\sum_i \pi_{ij} = d_j$ and $\sum_j \pi_{ij} = c_i$.
3. Cost of a transport plan
$$C(\pi) = \left(\sum_{ij} \|x_i - y_j\|^2 \pi_{ij}\right)^{1/2}$$
4. $d_W(\mu,\nu) := \inf_\pi C(\pi)$

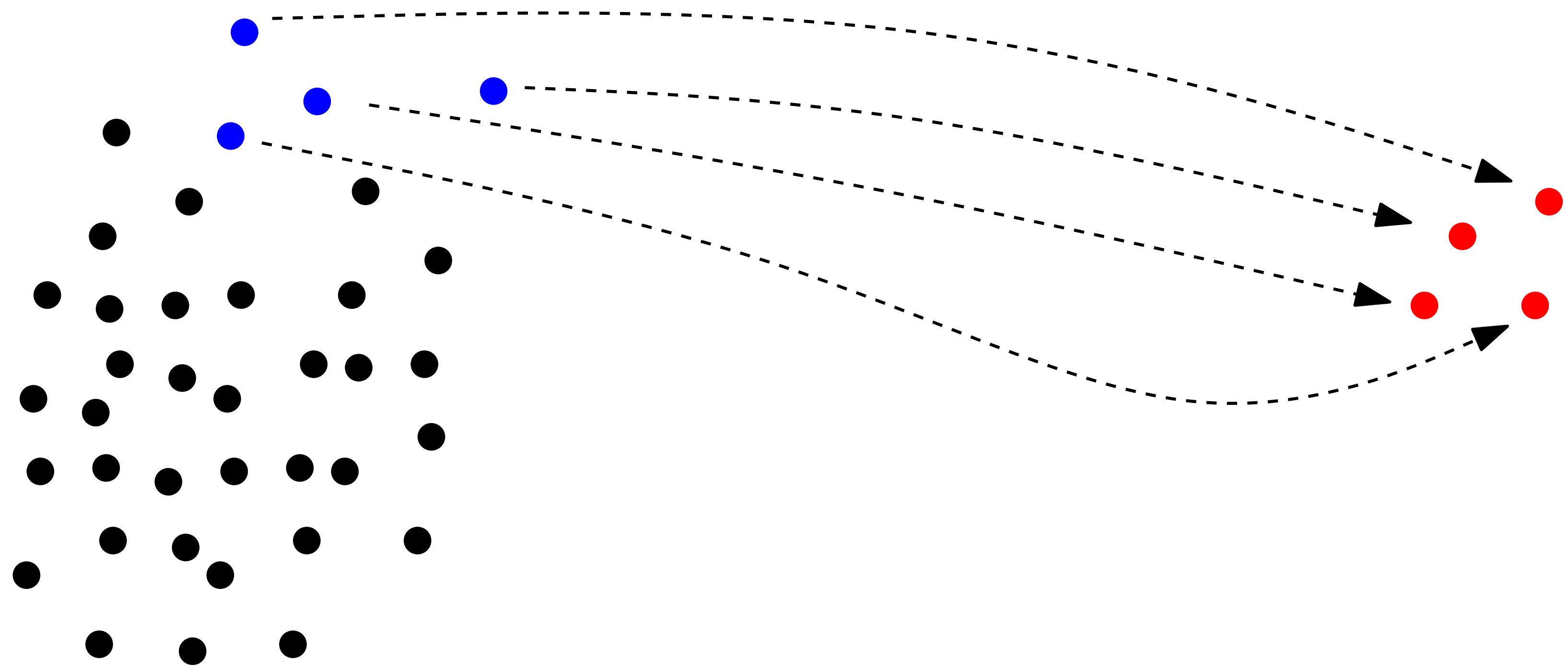
Distance between measures

"The" **Wasserstein distance** $d_W(\mu,\nu)$ between two probability measures μ,ν quantifies the optimal cost of pushing μ onto ν, the cost of moving a small mass dx from x to y being $\|x-y\|^2 dx$.



1. μ and ν are proba measures in $\mathbb{R}^d$
2. *Transport plan*: π a proba measure on $\mathbb{R}^d \times \mathbb{R}^d$ s.t. $\pi(A \times \mathbb{R}^d) = \mu(A)$ and $\pi(\mathbb{R}^d \times B) = \nu(B)$.
3. Cost of a transport plan
   $C(\pi) = \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^2 d\pi(x,y)\right)^{\frac{1}{2}}$
4. $d_W(\mu,\nu) := \inf_\pi C(\pi)$

Wasserstein distance



Examples:

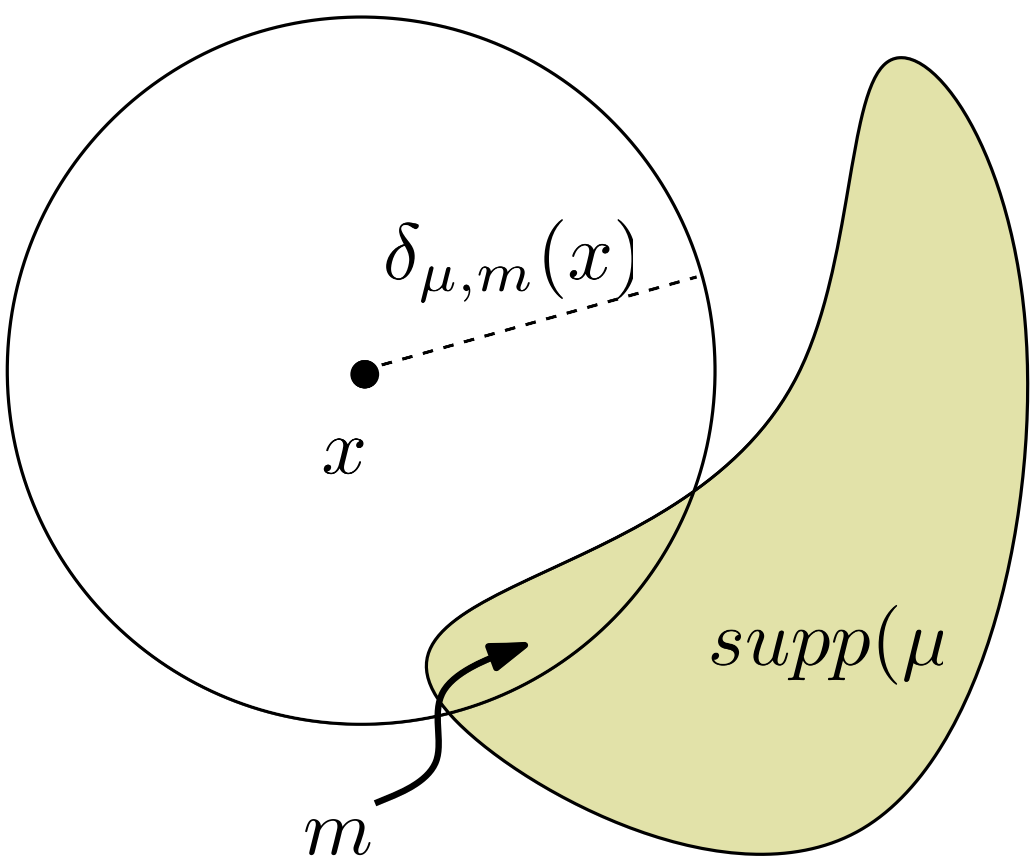
- If C_1 and C_2 are two point clouds, with $\#C_1 = \#C_2$, then $d_W(\mu_{C_1}, \mu_{C_2})$ is the square root of the cost of a minimal least-square matching between C_1 and C_2.

- If $C = \{p_1, \ldots, p_n\}$ is a point cloud, and $C' = \{p_1, \ldots, p_{n-k-1}, o_1, \ldots, o_k\}$ with $d(o_i, C) = R$, then

$$d_H(C, C') \geq R \qquad \text{but} \qquad d_W(\mu_C, \mu_{C'}) \leq \frac{k}{n}(R + \mathrm{diam}(C))$$

The distance to a measure

Distance function to a measure, first attempt:
Let $m \in]0,1[$ be a positive mass, and μ a probability measure on $\mathbb{R}^d$: $\delta_{\mu,m}(x) = \inf \{r > 0 : \mu(\mathbb{B}(x,r)) > m\}$.



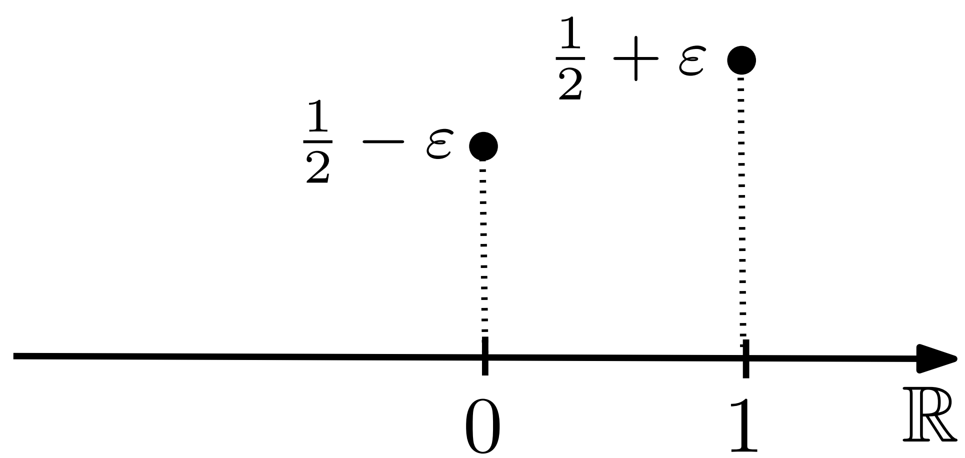
- $\delta_{\mu,m}$ is the smallest distance needed to capture a mass of at least m;
- Coincides with the distance to the k-th neighbor when $m = k/n$ and $\mu = \frac{1}{n}\sum_{i=1}^{n} \delta_{p_i}$:

$$\delta_{\mu,k/n}(\mu) = \|x - p_C^k(x)\|$$

Unstability of $\mu \mapsto \delta_{\mu,m}$

Distance function to a measure, first attempt:
Let $m \in]0,1[$ be a positive mass, and μ a probability measure on $\mathbb{R}^d$:
$\delta_{\mu,m}(x) = \inf\{r > 0 : \mu(\mathbb{B}(x,r)) > m\}$.

Unstability under Wasserstein perturbations:



$$\mu_\varepsilon = (1/2 - \varepsilon)\delta_0 + (1/2 + \varepsilon)\delta_1$$

for $\varepsilon > 0: \ \forall x < 0, \ \delta_{\mu_\varepsilon,1/2}(x) = |x - 1|$

for $\varepsilon = 0: \ \forall x < 0, \ \delta_{\mu_0,1/2}(x) = |x - 0|$

Consequence: the map $\mu \mapsto \delta_{\mu,m} \in \mathcal{C}^0(\mathbb{R}^d)$ is discontinuous whatever the (reasonable) topology on $\mathcal{C}^0(\mathbb{R}^d)$.

The distance function to a measure

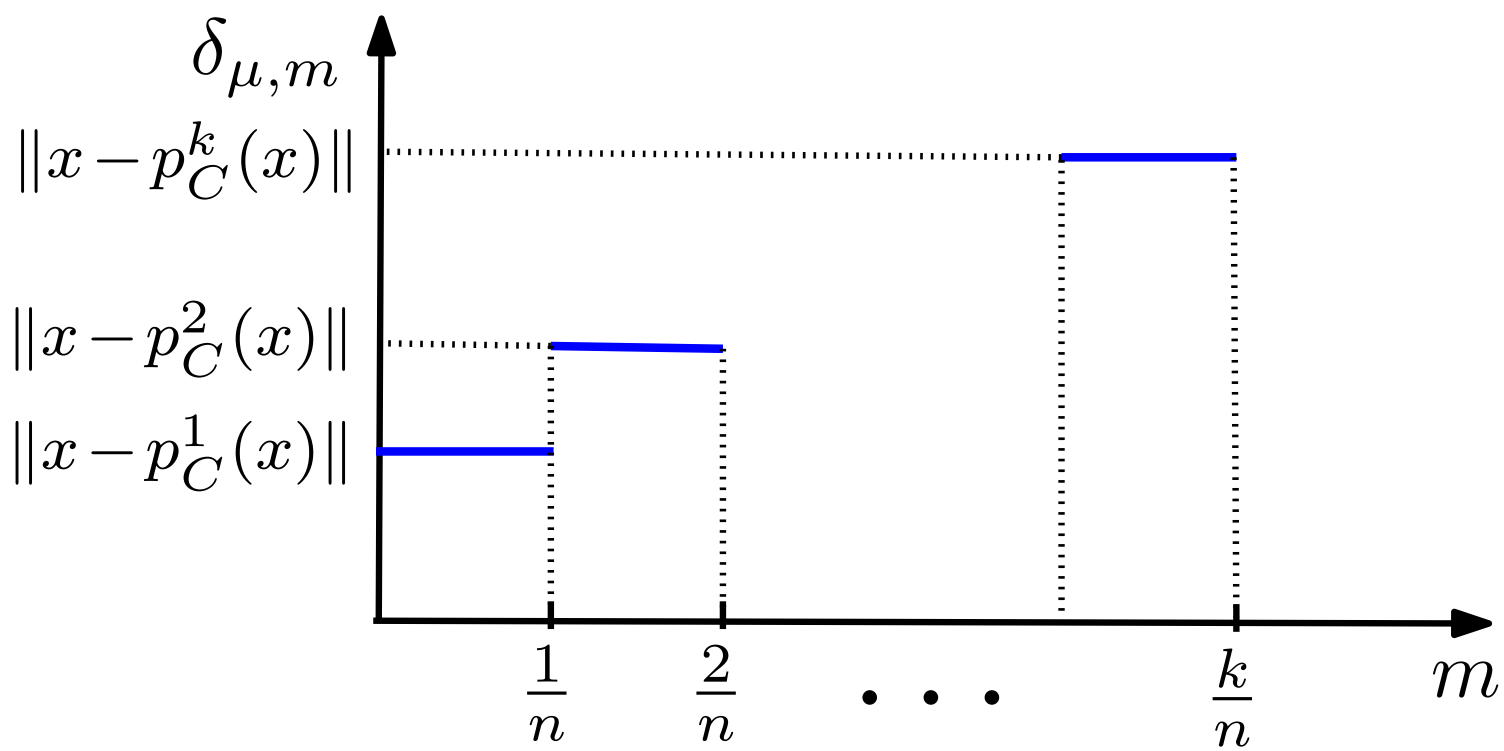
Definition: Given a probability measure μ on $\mathbb{R}^d$ and $m_0 > 0$, one defines:

$$\boxed{d_{\mu,m_0} : x \in \mathbb{R}^d \mapsto \left(\frac{1}{m_0} \int_0^{m_0} \delta_{\mu,m}^2(x) dm \right)^{1/2}}$$

The distance function to a measure

Definition: Given a probability measure μ on $\mathbb{R}^d$ and $m_0 > 0$, one defines:

$$\boxed{d_{\mu,m_0} : x \in \mathbb{R}^d \mapsto \left(\frac{1}{m_0} \int_0^{m_0} \delta_{\mu,m}^2(x) dm \right)^{1/2}}$$

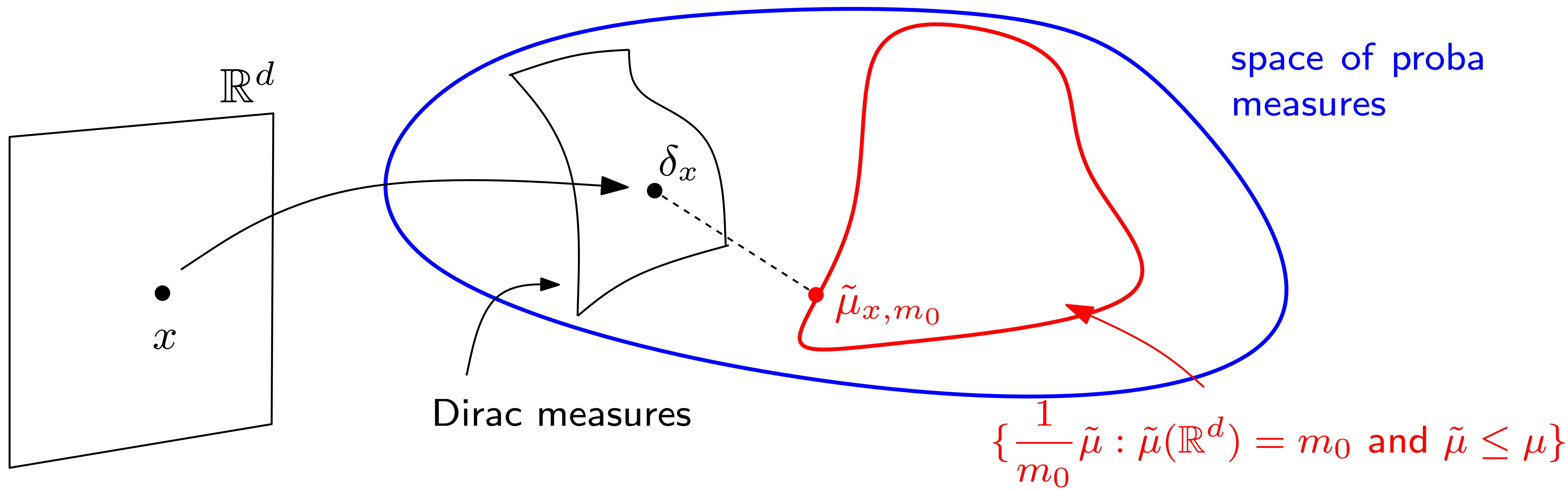


Example. Let $C = \{p_1, \ldots, p_n\}$ and $\mu = \frac{1}{n}\sum_{i=1}^n \delta_{p_i}$. Let $p_C^k(x)$ denote the kth nearest neighbor to x in C, and set $m_0 = k_0/n$:

$$d_{\mu,m_0}(x) = \left(\frac{1}{k_0} \sum_{k=1}^{k_0} \|x - p_C^k(x)\|^2 \right)^{1/2}$$

Another expression for d_{μ,m_0}

$$\boxed{d_{\mu,m_0}(x) = \min_{\tilde\mu} \left\{ d_W\left(\delta_x, \frac{1}{m_0}\tilde\mu\right) : \tilde\mu(\mathbb{R}^d) = m_0 \text{ and } \tilde\mu \le \mu \right\}}$$



"The projection submeasure": $\tilde\mu_{x,m_0}$ = the restriction of μ on the ball $B = \mathbb{B}(x, \delta_{\mu,m_0}(x))$, whose trace on the sphere ∂B has been rescaled so that the total mass of $\tilde\mu_{x,m_0}$ is m_0.

$$d^2_{\mu,m_0}(x) = \frac{1}{m_0}\int_{h\in\mathbb{R}^d} \|h-x\|^2 \, d\tilde\mu_{x,m_0} = d_W^2\left(\delta_x, \frac{1}{m_0}\tilde\mu_{x,m_0}\right)$$

Another expression for d_{μ,m_0}

$$d_{\mu,m_0}(x) = \min_{\tilde{\mu}} \left\{ d_W\left(\delta_x, \frac{1}{m_0}\tilde{\mu}\right) : \tilde{\mu}(\mathbb{R}^d) = m_0 \text{ and } \tilde{\mu} \leq \mu \right\}$$

Proof:

Another expression for d_{μ,m_0}

$$d_{\mu,m_0}(x) = \min_{\tilde{\mu}} \left\{ d_W\left(\delta_x, \frac{1}{m_0}\tilde{\mu}\right) : \tilde{\mu}(\mathbb{R}^d) = m_0 \text{ and } \tilde{\mu} \leq \mu \right\}$$

Proof:

Only one transport plan : $y \in \mathbb{R}^d \to x$

$$\int_{\mathbb{R}^d} \|h - x\|^2 d\tilde{\mu}(h)$$

Another expression for d_{μ,m_0}

$$d_{\mu,m_0}(x) = \min_{\tilde{\mu}} \left\{ d_W\left(\delta_x, \frac{1}{m_0}\tilde{\mu}\right) : \tilde{\mu}(\mathbb{R}^d) = m_0 \text{ and } \tilde{\mu} \le \mu \right\}$$

Proof:

$$\int_{\mathbb{R}^d} \|h - x\|^2 d\tilde{\mu}(h) = \int_{\mathbb{R}_+} t^2 d\tilde{\mu}_x(t) = \int_0^{m_0} F_{\tilde{\mu}_x}^{-1}(m)^2 dm$$

pushforward of $\tilde{\mu}$ by the distance function to x.

$F_{\tilde{\mu}_x}(t) = \tilde{\mu}_x([0,t))$ is the cumulative function of $\tilde{\mu}_x$ and $F_{\tilde{\mu}_x}^{-1}(m) = \inf\{t \in \mathbb{R} : F_{\tilde{\mu}_x}(t) > m\}$ is its generalized inverse

Another expression for d_{μ,m_0}

$$d_{\mu,m_0}(x) = \min_{\tilde{\mu}} \left\{ d_W\left(\delta_x, \frac{1}{m_0}\tilde{\mu}\right) : \tilde{\mu}(\mathbb{R}^d) = m_0 \text{ and } \tilde{\mu} \leq \mu \right\}$$

Proof:

$$\int_{\mathbb{R}^d} \|h - x\|^2 d\tilde{\mu}(h) = \int_{\mathbb{R}_+} t^2 d\tilde{\mu}_x(t) = \int_0^{m_0} F_{\tilde{\mu}_x}^{-1}(m)^2 dm$$

pushforward of $\tilde{\mu}$ by the distance function to x.

$F_{\tilde{\mu}_x}(t) = \tilde{\mu}_x([0,t))$ is the cumulative function of $\tilde{\mu}_x$ and $F_{\tilde{\mu}_x}^{-1}(m) = \inf\{t \in \mathbb{R} : F_{\tilde{\mu}_x}(t) > m\}$ is its generalized inverse

- $\tilde{\mu} \leq \mu \Rightarrow F_{\tilde{\mu}_x}(t) \leq F_{\mu_x}(t) \Rightarrow F_{\tilde{\mu}_x}^{-1}(m) \geq F_{\mu_x}^{-1}(m)$
- $F_{\tilde{\mu}_x}(t) = \mu(\mathbb{B}(x,t))$ and $F_{\tilde{\mu}_x}^{-1}(m) = \delta_{\tilde{\mu},m}(x)$

$$\int_{\mathbb{R}^d} \|h - x\|^2 d\tilde{\mu}(h) \geq \int_0^{m_0} F_{\mu_x}^{-1}(m)^2 dm = \int_0^{m_0} \delta_{\mu,m}(x)^2 dm$$

Another expression for d_{μ,m_0}

$$d_{\mu,m_0}(x) = \min_{\tilde\mu} \left\{ d_W\left(\delta_x, \frac{1}{m_0}\tilde\mu\right) : \tilde\mu(\mathbb{R}^d) = m_0 \text{ and } \tilde\mu \leq \mu \right\}$$

Proof:

$$\int_{\mathbb{R}^d} \|h-x\|^2 d\tilde\mu(h) = \int_{\mathbb{R}_+} t^2 d\tilde\mu_x(t) = \int_0^{m_0} F^{-1}_{\tilde\mu_x}(m)^2 dm$$

pushforward of $\tilde\mu$ by the distance function to x.

$F_{\tilde\mu_x}(t) = \tilde\mu_x([0,t))$ is the cumulative function of $\tilde\mu_x$ and $F^{-1}_{\tilde\mu_x}(m) = \inf\{t \in \mathbb{R} : F_{\tilde\mu_x}(t) > m\}$ is its generalized inverse

Equality iff $F^{-1}_{\tilde\mu_x}(m) = F^{-1}_{\mu_x}(m)$ for almost every m
$\Rightarrow$ equality if $\tilde\mu = \tilde\mu_{x,m_0}$

$$\int_{\mathbb{R}^d} \|h-x\|^2 d\tilde\mu(h) \geq \int_0^{m_0} F^{-1}_{\mu_x}(m)^2 dm = \int_0^{m_0} \delta_{\mu,m}(x)^2 dm$$

Semiconcavity of d^2_{μ,m_0}

Theorem: Let μ be a probability measure in $\mathbb{R}^d$ and let $m_0 \in (0,1)$.

1. d^2_{μ,m_0} is 1-semiconcave, i.e. $x \in \mathbb{R}^d \mapsto \|x\|^2 - d^2_{\mu,m_0}$ is convex.
2. d^2_{μ,m_0} is differentiable almost everywhere in $\mathbb{R}^d$, with gradient defined by
$$\nabla_x d^2_{\mu,m_0} = \frac{2}{m_0} \int_{h \in \mathbb{R}^d} (x-h)\, d\tilde{\mu}_{x,m_0}(h)$$
3. the function $x \in \mathbb{R}^d \mapsto d_{\mu,m_0}(x)$ is 1-Lipschitz.

Example. Let $C = \{p_1, \ldots, p_n\}$ and $\mu = \frac{1}{n}\sum_{i=1}^n \delta_{p_i}$. Let $p^k_C(x)$ denote the kth nearest neighbor to x in C, and set $m_0 = k_0/n$:

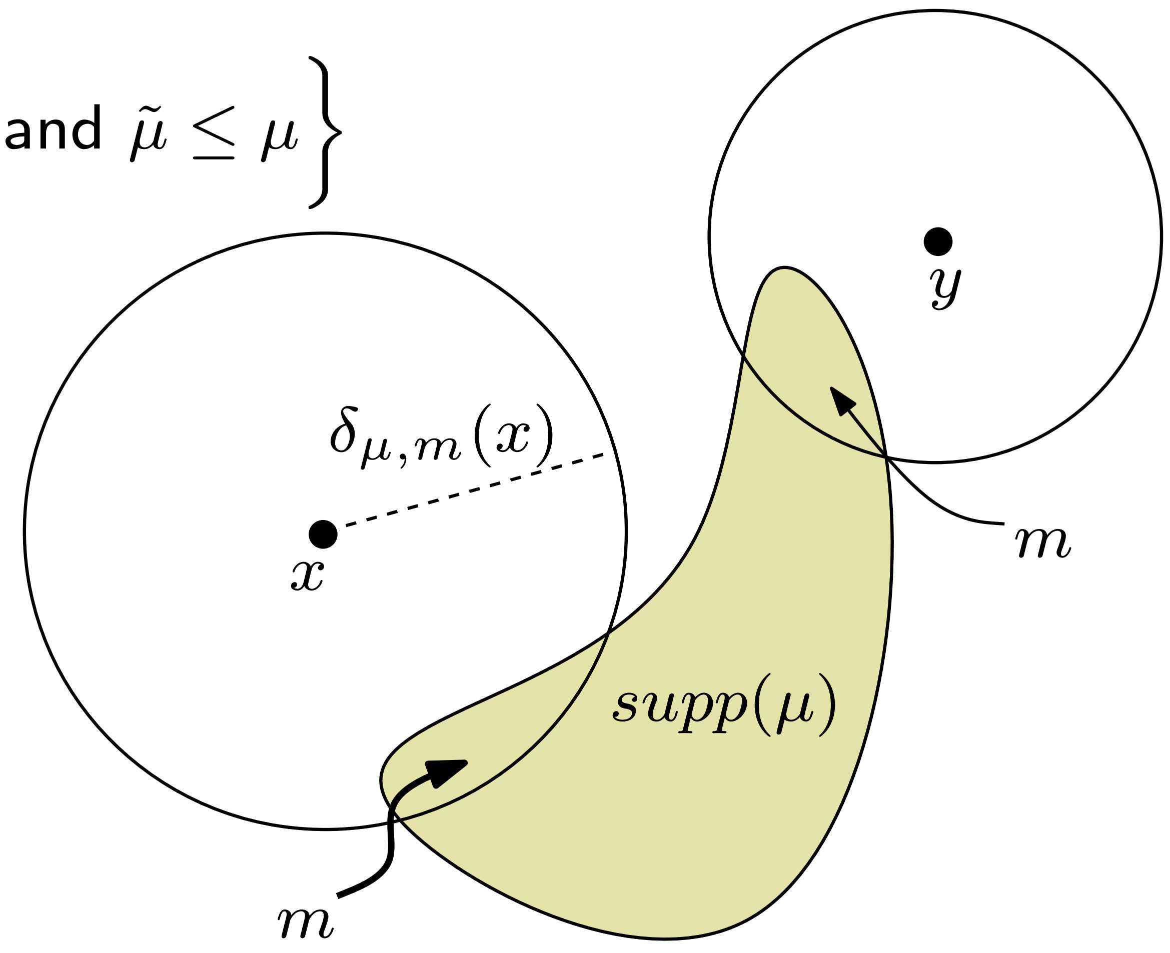
$$\nabla d^2_{\mu,m_0}(x) = 2 d_{\mu,m_0} \nabla d_{\mu,m_0} = \frac{2}{k_0} \sum_{k=1}^{k_0} (x - p^k_C(x))$$

Semiconcavity of d^2_{μ,m_0}

Proof:

$$d^2_{\mu,m_0}(y) = \frac{1}{m_0}\int_{h\in\mathbb{R}^d} \|y-h\|^2 \, d\tilde{\mu}_{y,m_0}(h)$$
$$\leq \frac{1}{m_0}\int_{h\in\mathbb{R}^d} \|y-h\|^2 \, d\tilde{\mu}_{x,m_0}(h)$$

$$d_{\mu,m_0}(x) = \min_{\tilde{\mu}}\left\{ d_W\left(\delta_x, \frac{1}{m_0}\tilde{\mu}\right) : \tilde{\mu}(\mathbb{R}^d) = m_0 \text{ and } \tilde{\mu} \leq \mu \right\}$$



Semiconcavity of d^2_{μ,m_0}

Proof:

$$\begin{aligned} d^2_{\mu,m_0}(y) &= \frac{1}{m_0}\int_{h\in\mathbb{R}^d} \|y-h\|^2 \, d\tilde{\mu}_{y,m_0}(h) \\ &\leq \frac{1}{m_0}\int_{h\in\mathbb{R}^d} \|y-h\|^2 \, d\tilde{\mu}_{x,m_0}(h) \\ &= \frac{1}{m_0}\int_{h\in\mathbb{R}^d} \left(\|x-h\|^2 + 2\langle x-h, y-x\rangle + \|y-x\|^2\ \right) \, d\tilde{\mu}_{x,m_0}(h) \\ &= d^2_{\mu,m_0}(x) + \|y-x\|^2 + \langle V, y-x\rangle \end{aligned}$$

with $V = \frac{2}{m_0}\int_{h\in\mathbb{R}^d}[x-h]\, d\tilde{\mu}_{x,m_0}(h)$.

Semiconcavity of d^2_{μ,m_0}

Proof:

$$\begin{aligned}
d^2_{\mu,m_0}(y) &= \frac{1}{m_0}\int_{h\in\mathbb{R}^d} \|y-h\|^2 \, d\tilde{\mu}_{y,m_0}(h) \\
&\leq \frac{1}{m_0}\int_{h\in\mathbb{R}^d} \|y-h\|^2 \, d\tilde{\mu}_{x,m_0}(h) \\
&= \frac{1}{m_0}\int_{h\in\mathbb{R}^d} \left(\|x-h\|^2 + 2\left\langle x-h, y-x\right\rangle + \|y-x\|^2\right) \, d\tilde{\mu}_{x,m_0}(h) \\
&= d^2_{\mu,m_0}(x) + \|y-x\|^2 + \left\langle V, y-x\right\rangle
\end{aligned}$$

with $V = \frac{2}{m_0}\int_{h\in\mathbb{R}^d} [x-h] \, d\tilde{\mu}_{x,m_0}(h)$.

$$\Rightarrow \left(\|y\|^2 - d^2_{\mu,m_0}(y)\right) - \left(\|x\|^2 - d^2_{\mu,m_0}(x)\right) \geq \left\langle 2x - V, y-x\right\rangle$$

This is the gradient!

Stability of of $\mu \to d_{\mu,m_0}$

Theorem: If μ and ν are two probability measures on $\mathbb{R}^d$ and $m_0 > 0$, then $\|d_{\mu,m_0} - d_{\nu,m_0}\|_\infty \leq \frac{1}{\sqrt{m_0}} d_W(\mu,\nu)$.

Stability of of $\mu \to d_{\mu,m_0}$

Theorem: If μ and ν are two probability measures on $\mathbb{R}^d$ and $m_0 > 0$, then $\|d_{\mu,m_0} - d_{\nu,m_0}\|_\infty \leq \frac{1}{\sqrt{m_0}} d_W(\mu,\nu)$.

Proof:

Set of submeasures of μ of mass m_0.

Proposition: $d_H(\mathrm{Sub}_{m_0}(\mu), \mathrm{Sub}_{m_0}(\nu)) \leq d_W(\mu,\nu)$

$$\begin{aligned}
d_{\mu,m_0}(x) &= \frac{1}{\sqrt{m_0}} d_W(m_0\delta_x, \mathrm{Sub}_{m_0}(\mu)) \\
&\leq \frac{1}{\sqrt{m_0}} (d_H(\mathrm{Sub}_{m_0}(\mu), \mathrm{Sub}_{m_0}(\nu)) + d_W(m_0\delta_x, \mathrm{Sub}_{m_0}(\nu))) \\
&\leq \frac{1}{\sqrt{m_0}} d_W(\mu,\nu) + d_{\nu,m_0}(x)
\end{aligned}$$

To summarize

Theorem

1. the function $x \mapsto d_{\mu,m_0}(x)$ is 1-Lipschitz;
2. the function $x \mapsto \|x\|^2 - d^2_{\mu,m_0}(x)$ is convex;
3. the map $\mu \mapsto d_{\mu,m_0}$ from probability measures to continuous functions is $\frac{1}{\sqrt{m_0}}$–Lipschitz, ie

$$\|d_{\mu,m_0} - d_{\mu',m_0}\|_\infty \leq \frac{1}{\sqrt{m_0}} d_W(\mu, \mu')$$

In practice: d_{μ,m_0} and $\nabla d_{\mu,m_0}$ are very easy to compute for $\mu = \sum_{i=1}^n \delta_{p_i}$, $C = \{p_1, \cdots p_n\} \subset \mathbb{R}^d$, even for pretty large d !

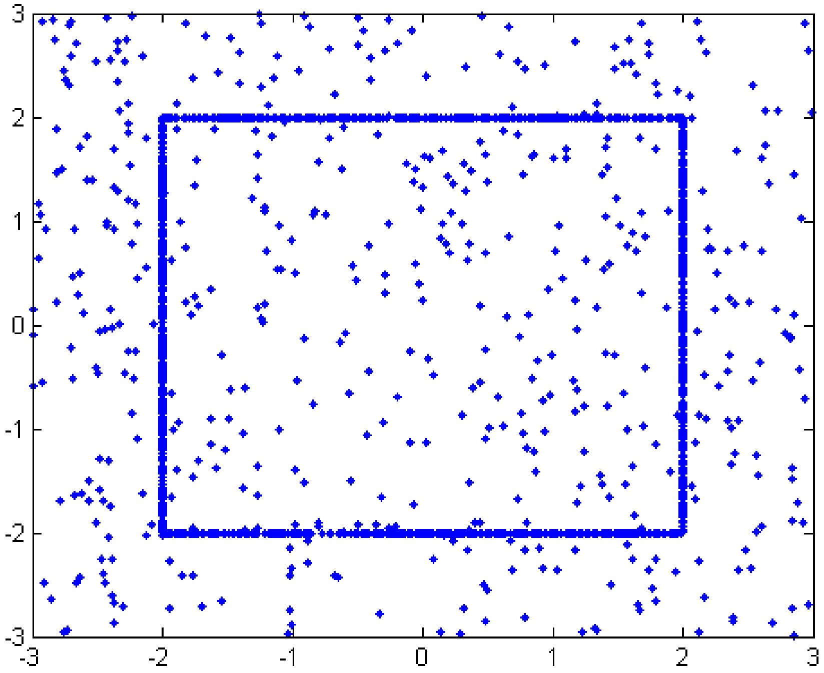
Consequences

Most of the topological and geometric inference for distance functions transpose to distance to a measure functions!

$\Longrightarrow$ This gives a way to associate robust geometric features to any probability measure in an Euclidean space:

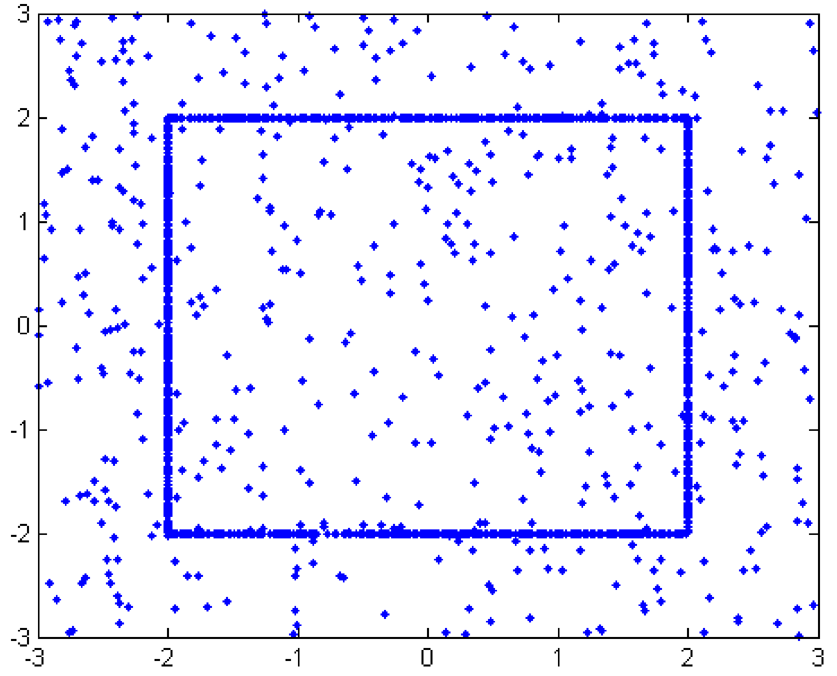
- stable offsets topology and geometry,
- stable persistence diagrams,
- analogous of the notions of medial axes,
- L^1 stability of $\nabla d_{\mu,m_0}$
- $\cdots$

Example: a square with outliers

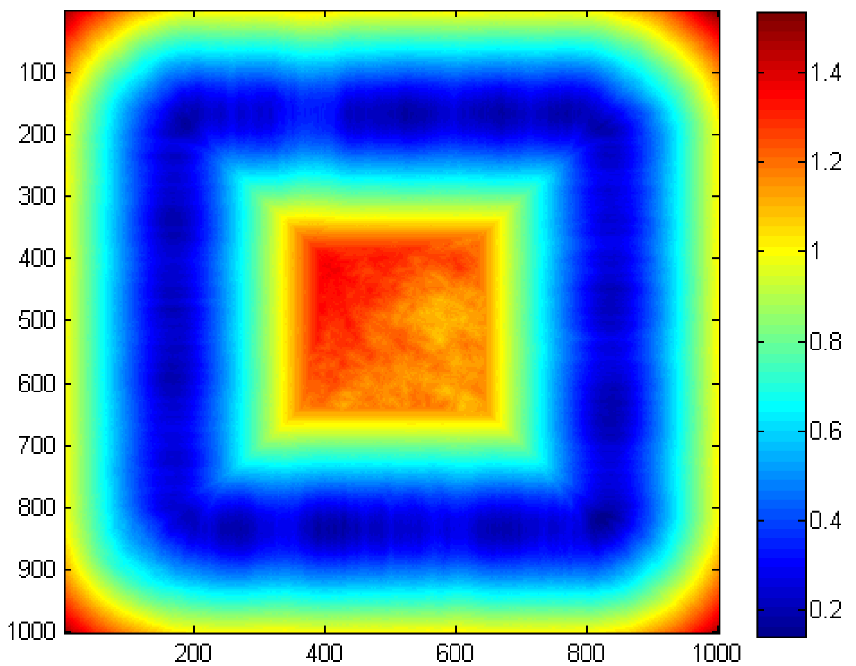


2300 points, 20% outliers

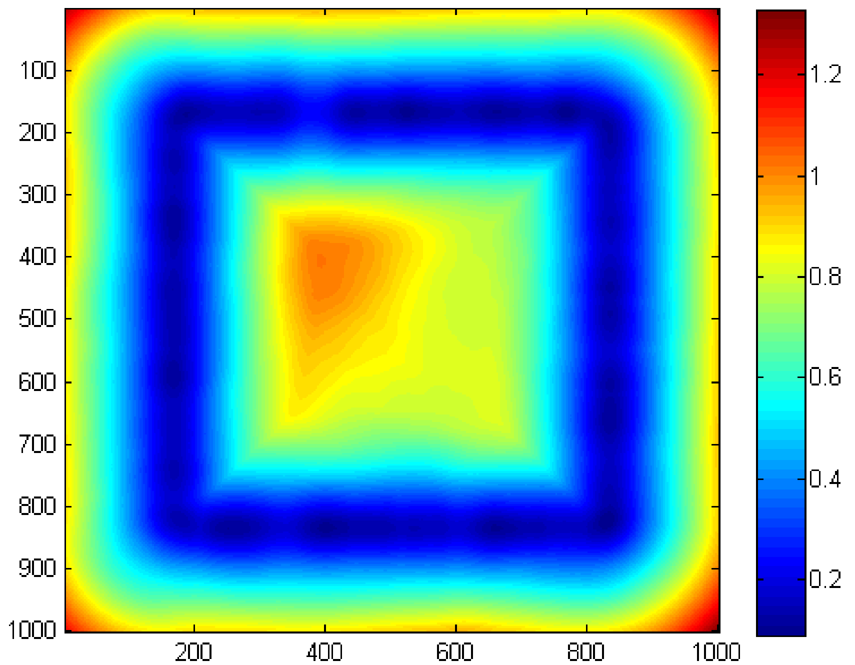
Example: a square with outliers



2300 points, 20% outliers

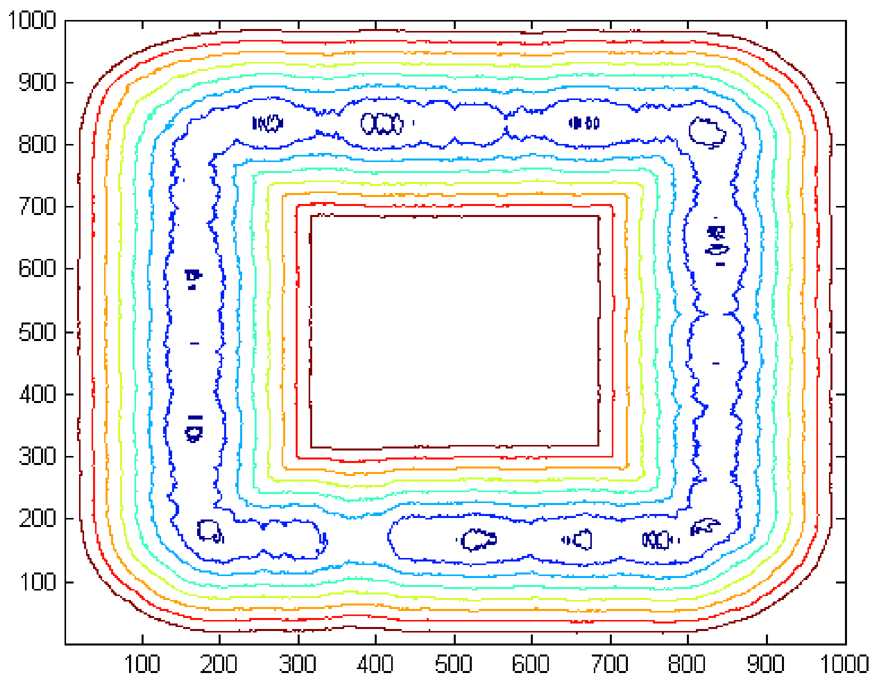


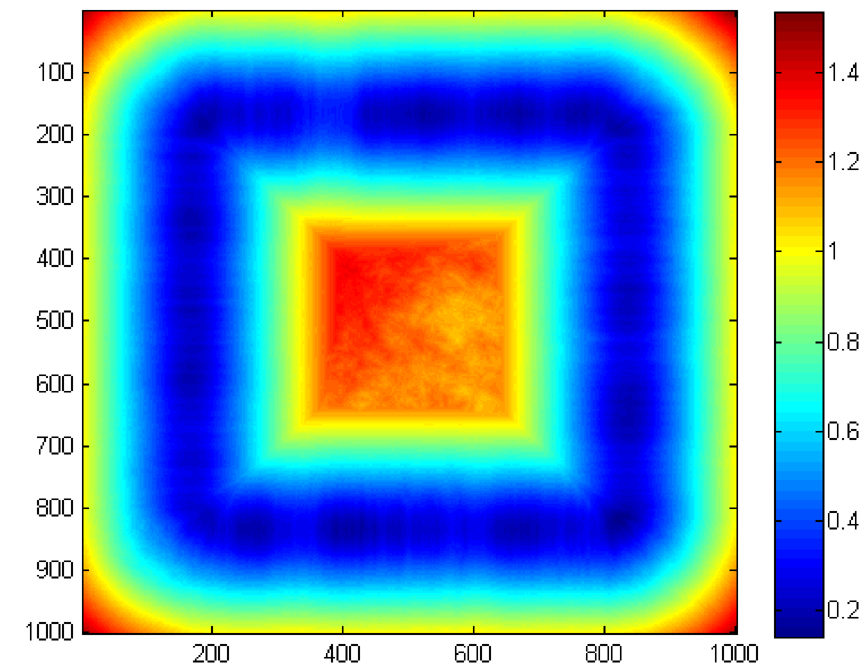
δ_{μ,m_0}, $m_0 = 0.023$ $(k = 50)$



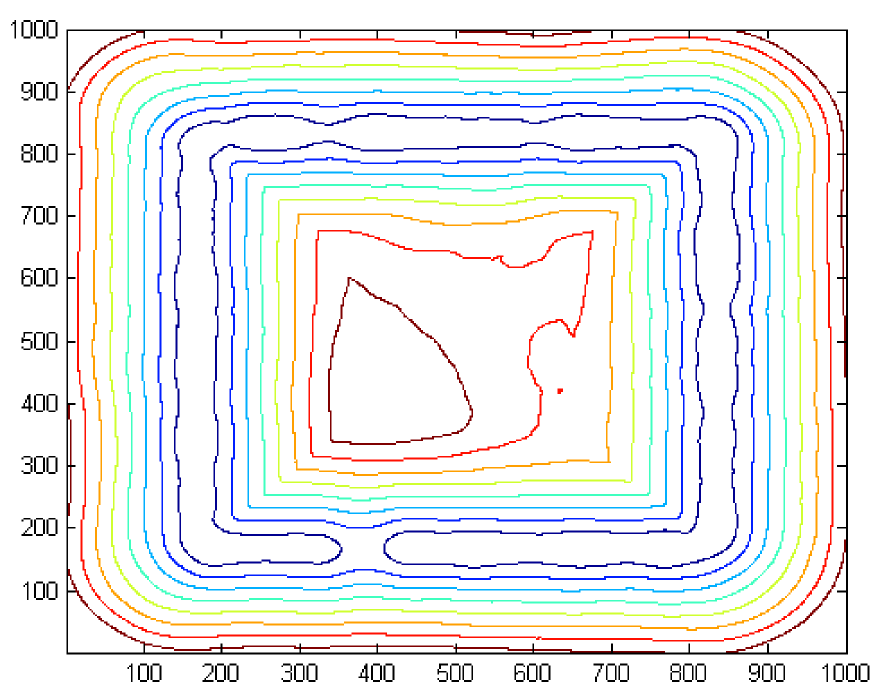
d_{μ,m_0}, $m_0 = 0.023$ $(k = 50)$

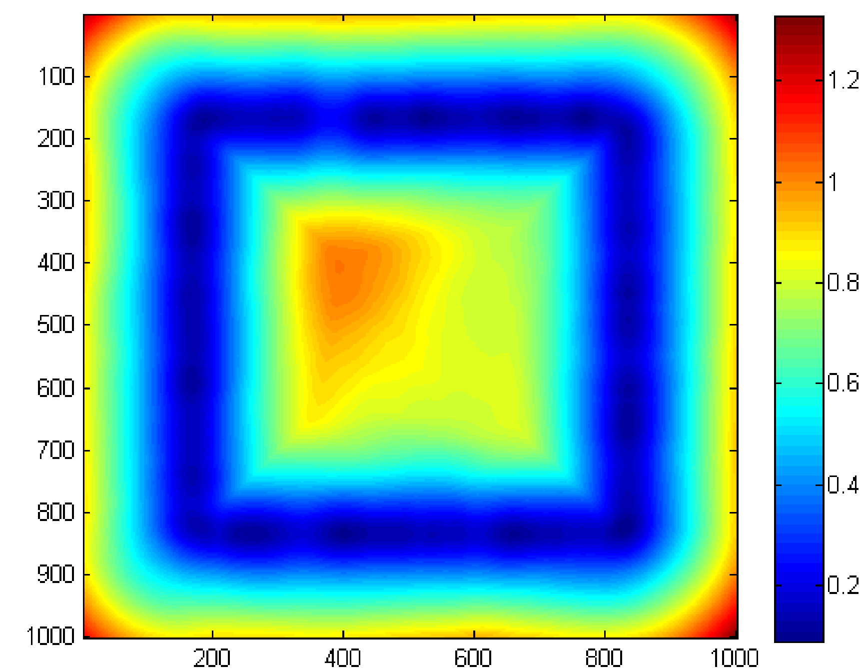
Example: a square with outliers





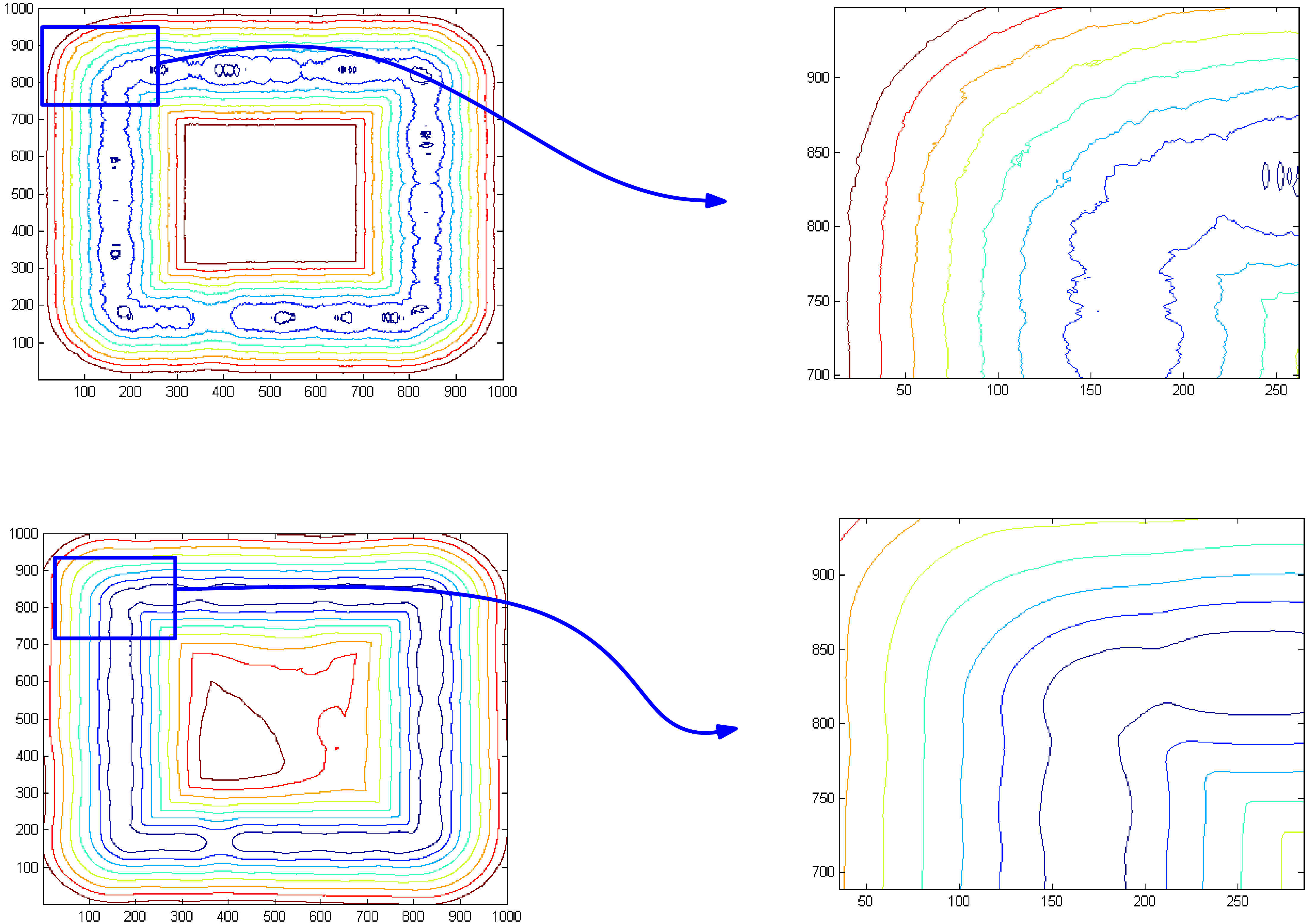
δ_{μ,m_0}, $m_0 = 0.023$ $(k = 50)$



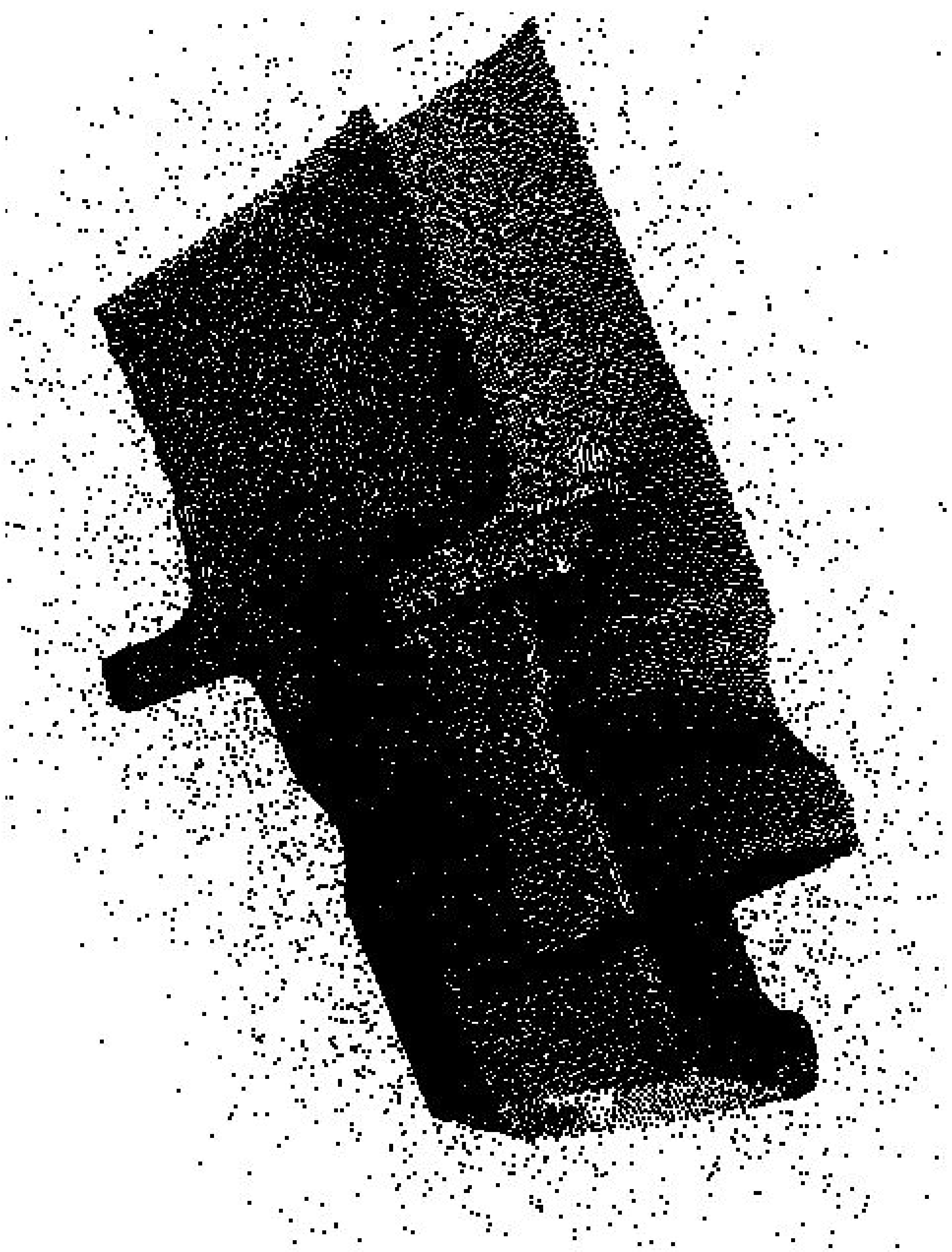


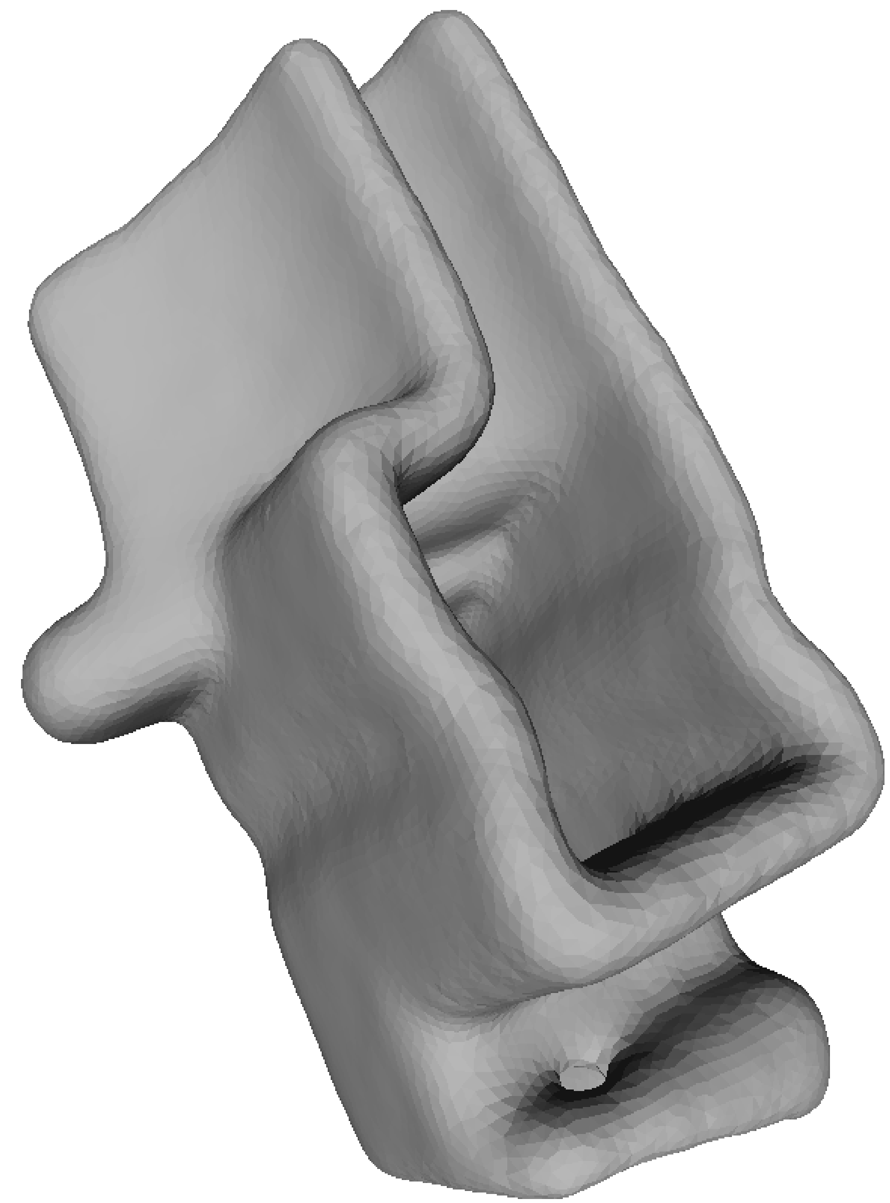
d_{μ,m_0}, $m_0 = 0.023$ $(k = 50)$

Example: a square with outliers



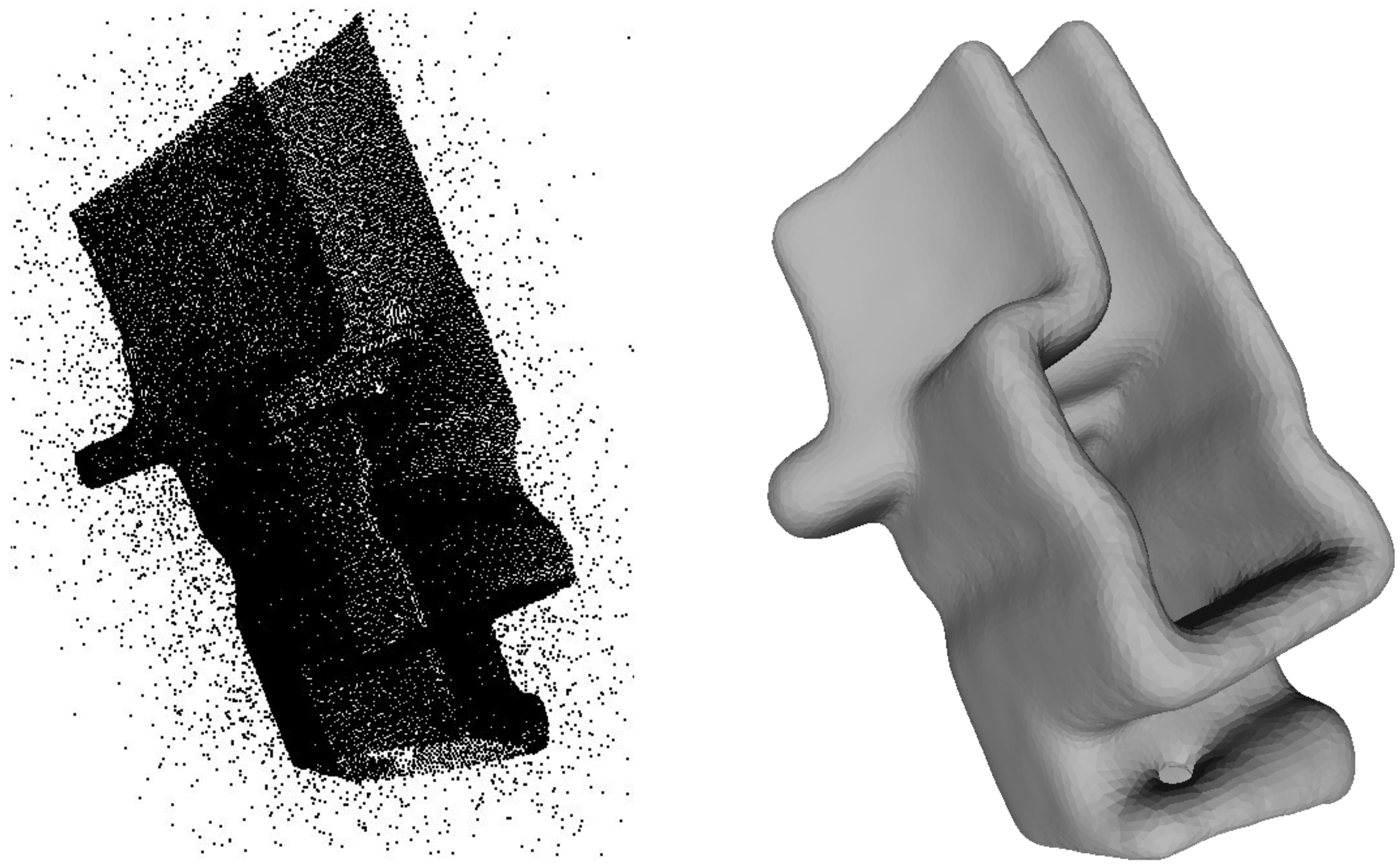
A 3D example





Reconstruction of an offset of a mechanical part from a noisy approximation with 10% outliers

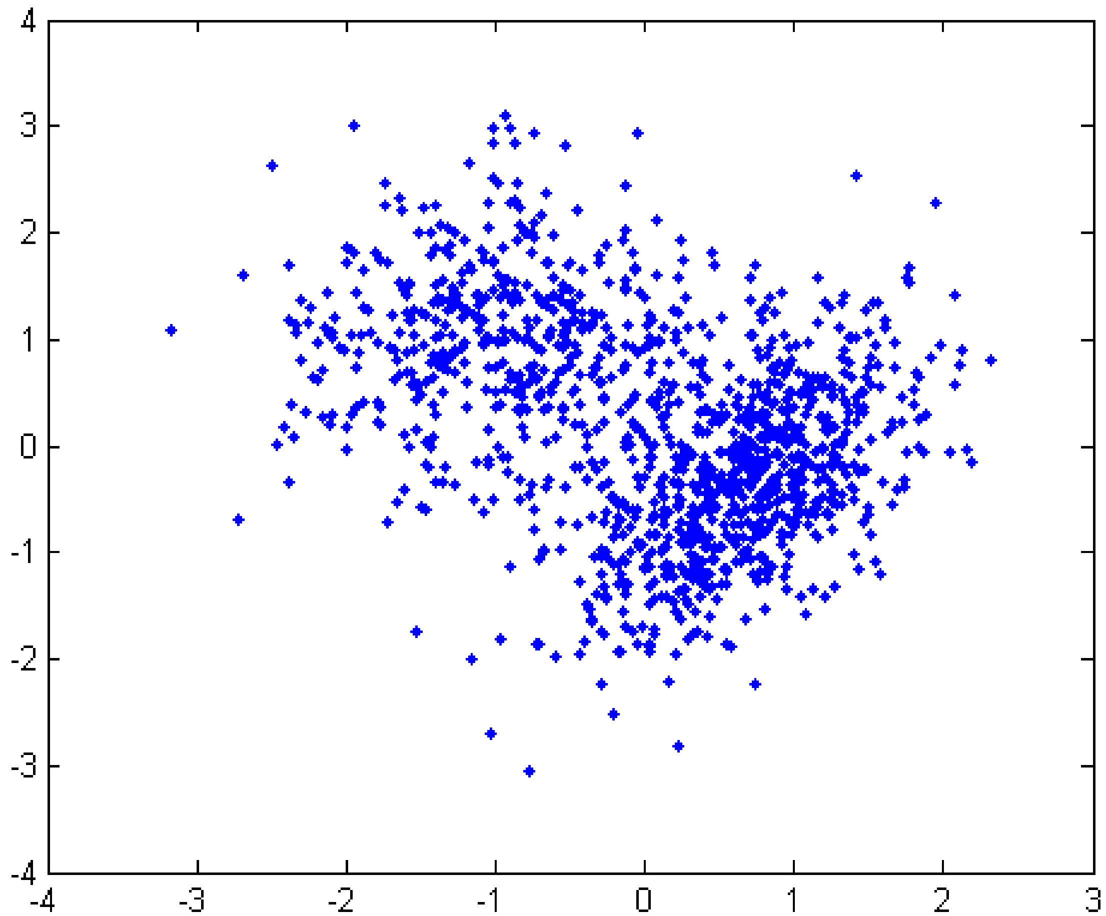
A reconstruction theorem



Theorem: Let μ be a proba measure with compact support $K \subset \mathbb{R}^d$ s. t.
(i) $r_\alpha(K) > 0$ for some $\alpha \in (0,1]$,
(ii) $\exists C > 0$ s.t. $\forall x \in K$, $\mu(\mathbb{B}(x,r)) \geq Cr^k$
Let μ' be another measure, and ε be an upper bound on the uniform distance between d_K and d_{μ',m_0}. Then, for any $r \in [4\varepsilon/\alpha^2, R - 3\varepsilon]$, the r-sublevel sets of d_{μ,m_0} and the offsets K^η, for $0 < \eta < R$ are homotopy equivalent, as soon as:

$$W_2(\mu,\mu') \leq \frac{R\sqrt{m_0}}{5 + 4/\alpha^2} - C^{-1/k} m_0^{1/k+1/2}$$

Comparison to kNN density estimation



$$\hat{\mu} = \frac{1}{1200} \sum_{i=1}^{1200} \delta_{p_i}$$

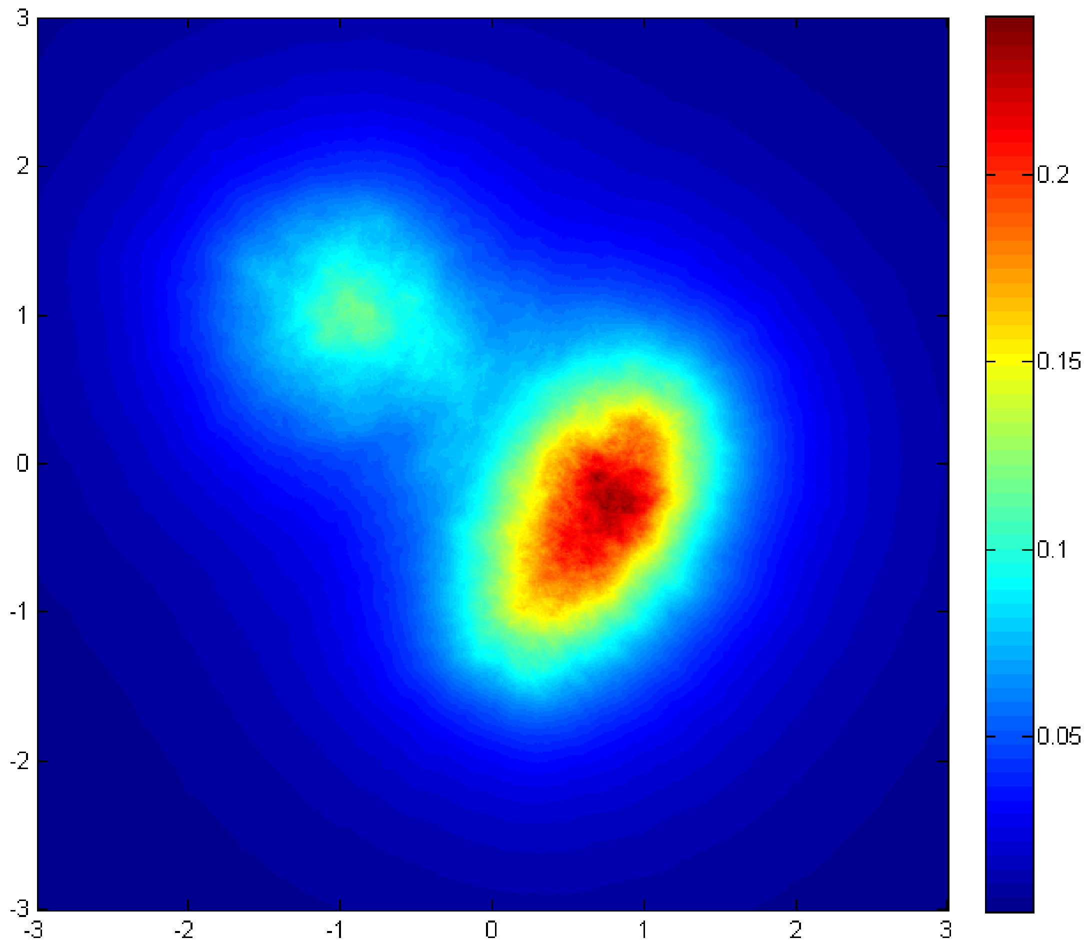
Data: 1200 points $p_1, \cdots, p_{1200}$

Density is estimated using

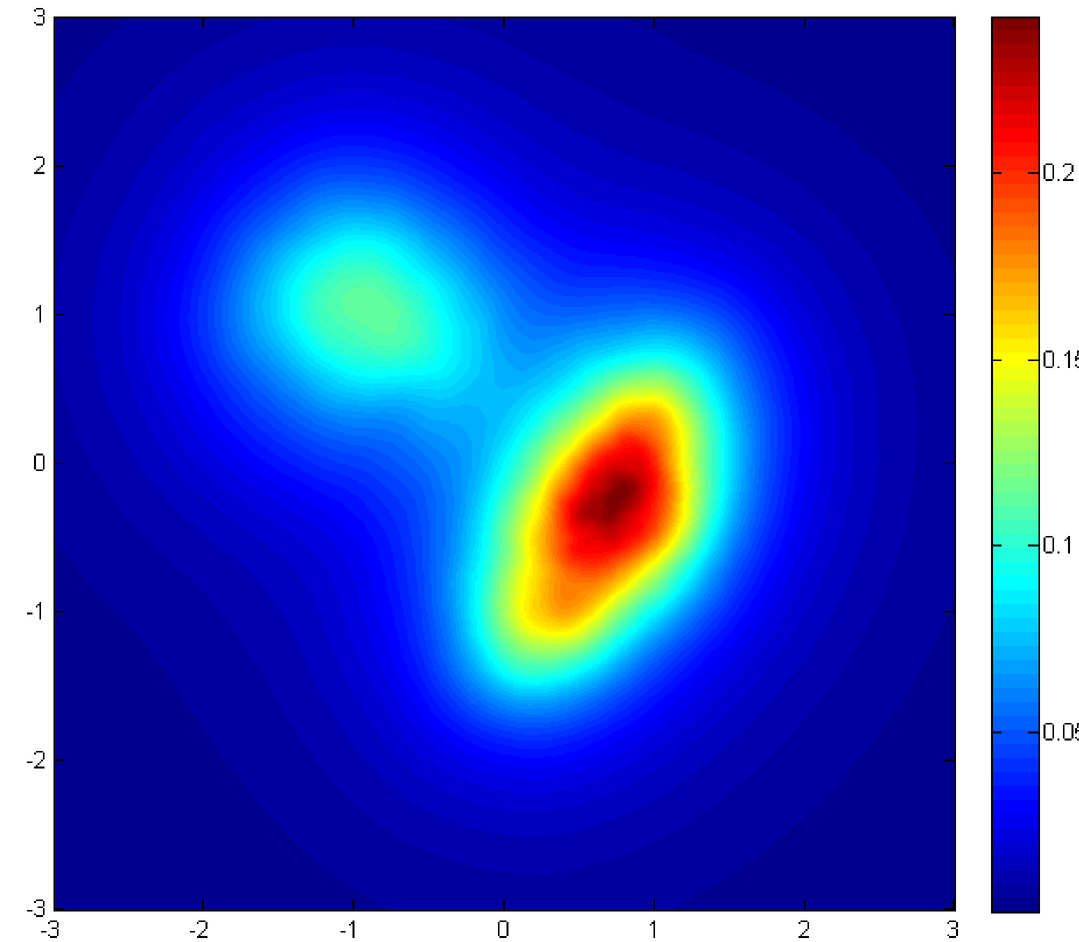
1. $x \mapsto \frac{m_0}{\omega_{d-1}(\delta_{\hat{\mu},m_0}(x))}$, $m_0 = 150/1200$ $(k = 150)$ (Devroye-Wagner'77).

2. $\frac{m_0}{2\pi d_{\hat{\mu},m_0}(x)^2}$, $m_0 = 150/1200$ $(k = 150)$.

Comparison to kNN density estimation



1.

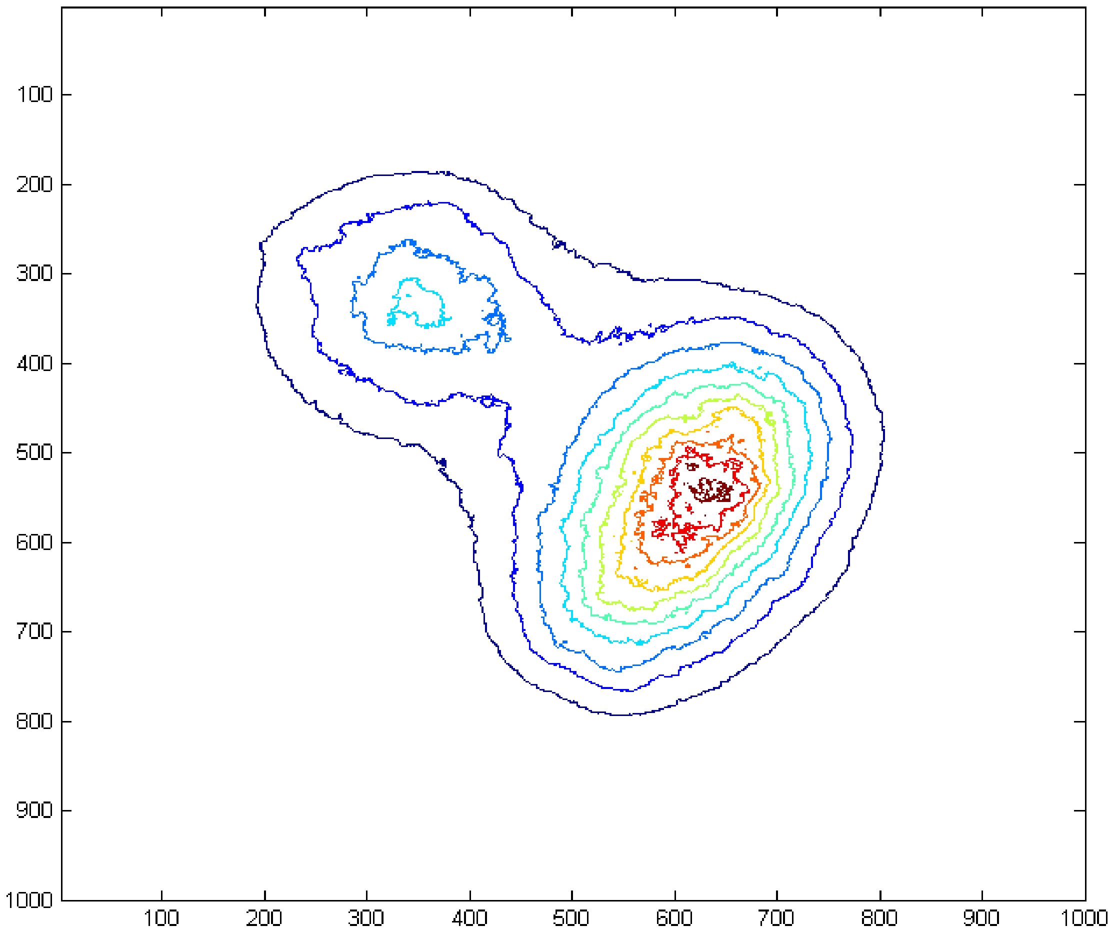


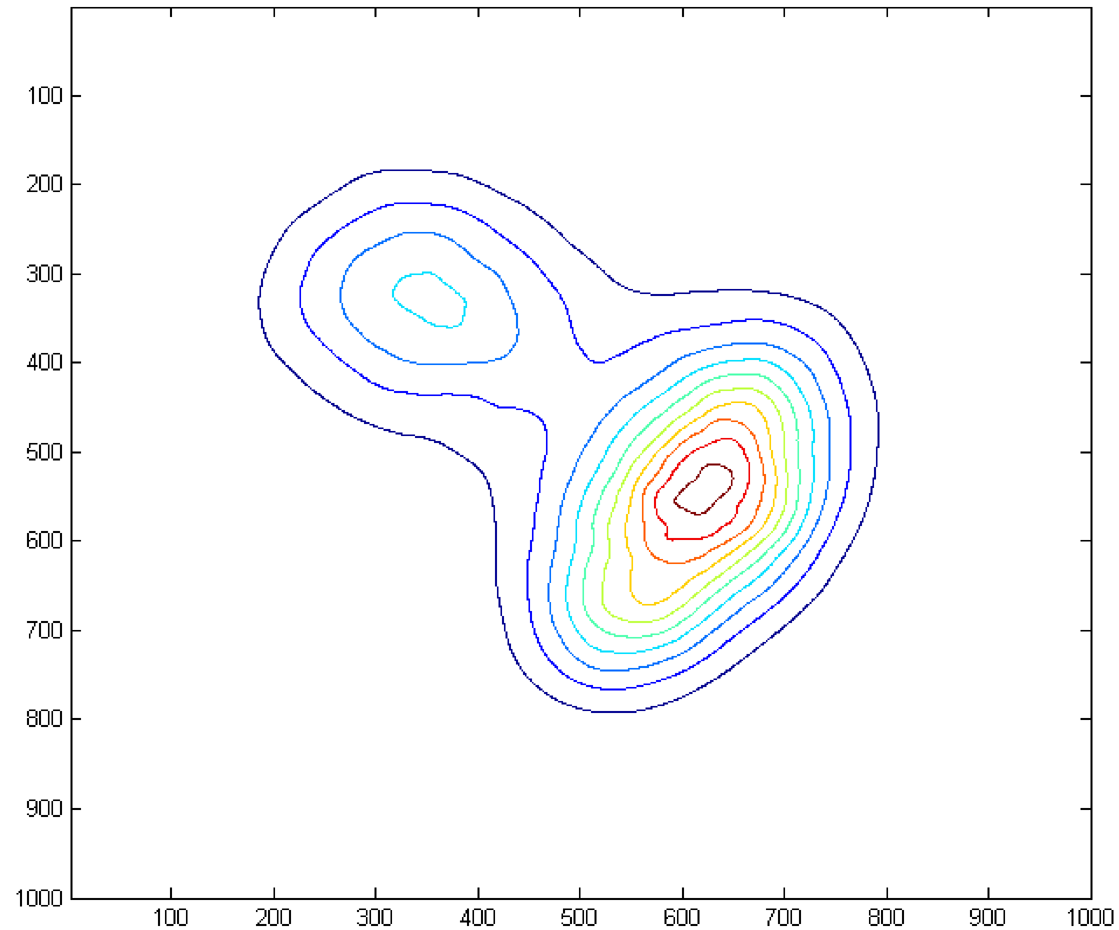
2.

Density is estimated using

1. $x \mapsto \frac{m_0}{\omega_{d-1}(\delta_{\hat{\mu},m_0}(x))}$, $m_0 = 150/1200$ $(k = 150)$ (Devroye-Wagner'77).
2. $\frac{m_0}{2\pi d_{\hat{\mu},m_0}(x)^2}$, $m_0 = 150/1200$ $(k = 150)$.

Comparison to kNN density estimation



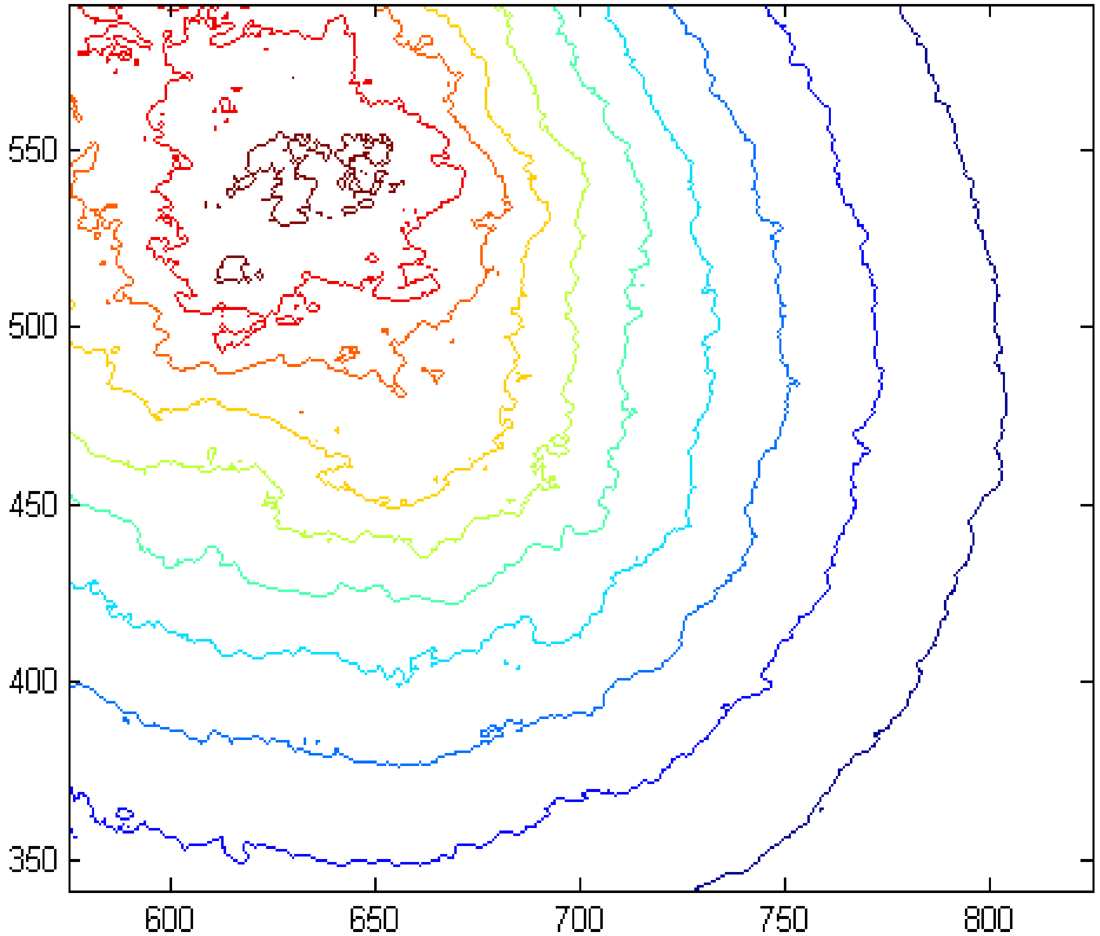


Density is estimated using

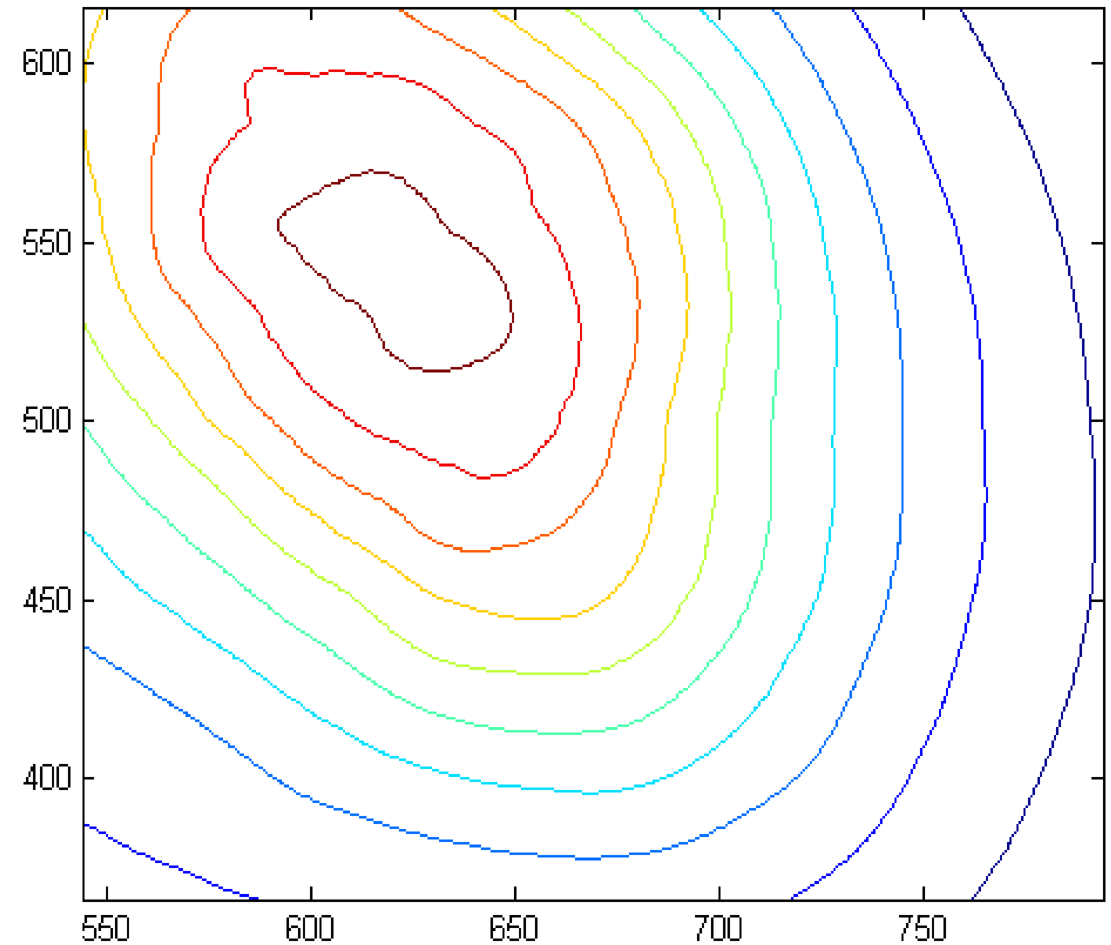
1. $x \mapsto \frac{m_0}{\omega_{d-1}(\delta_{\hat{\mu},m_0}(x))}$, $m_0 = 150/1200$ $(k = 150)$ (Devroye-Wagner'77).

2. $\frac{m_0}{2\pi d_{\hat{\mu},m_0}(x)^2}$, $m_0 = 150/1200$ $(k = 150)$.

Comparison to kNN density estimation



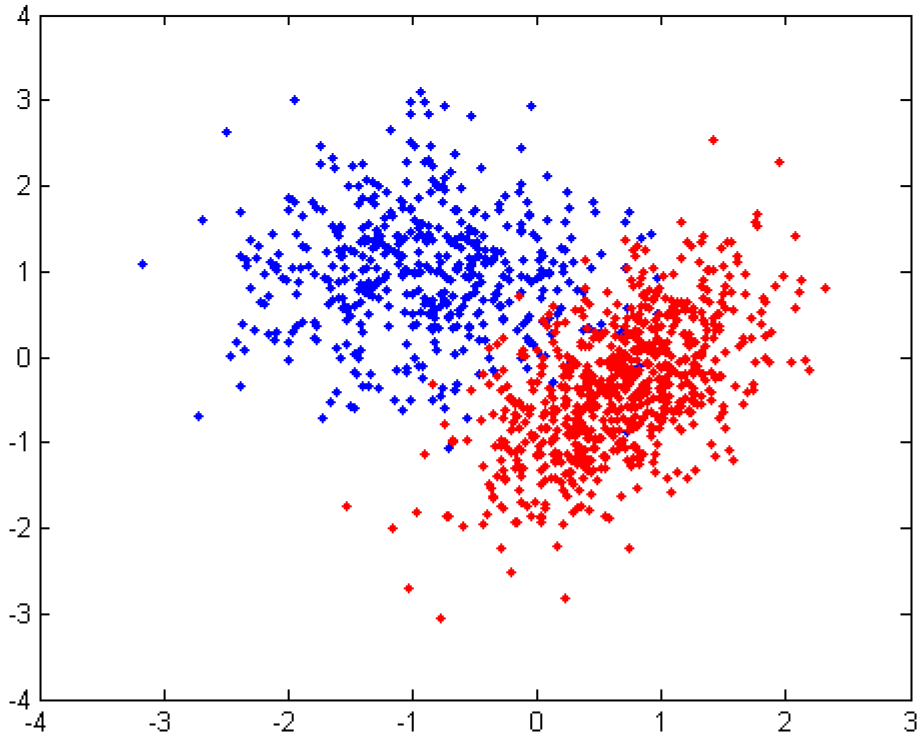
1.

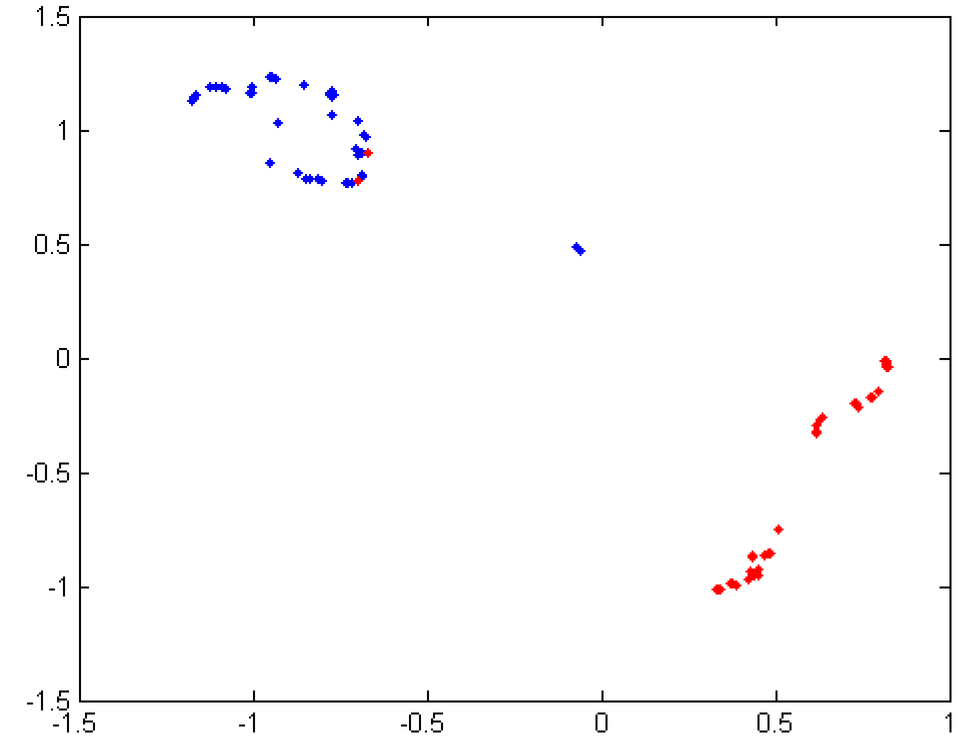


2.

[Biau, C., Cohen-Steiner, Devroye, Rodrìguez 2011]: d_{μ,m_0} can be turned into a density estimator whose level sets foliation is the same as the one of d_{μ,m_0}.

Pushing data along the gradient of d_{μ,m_0}



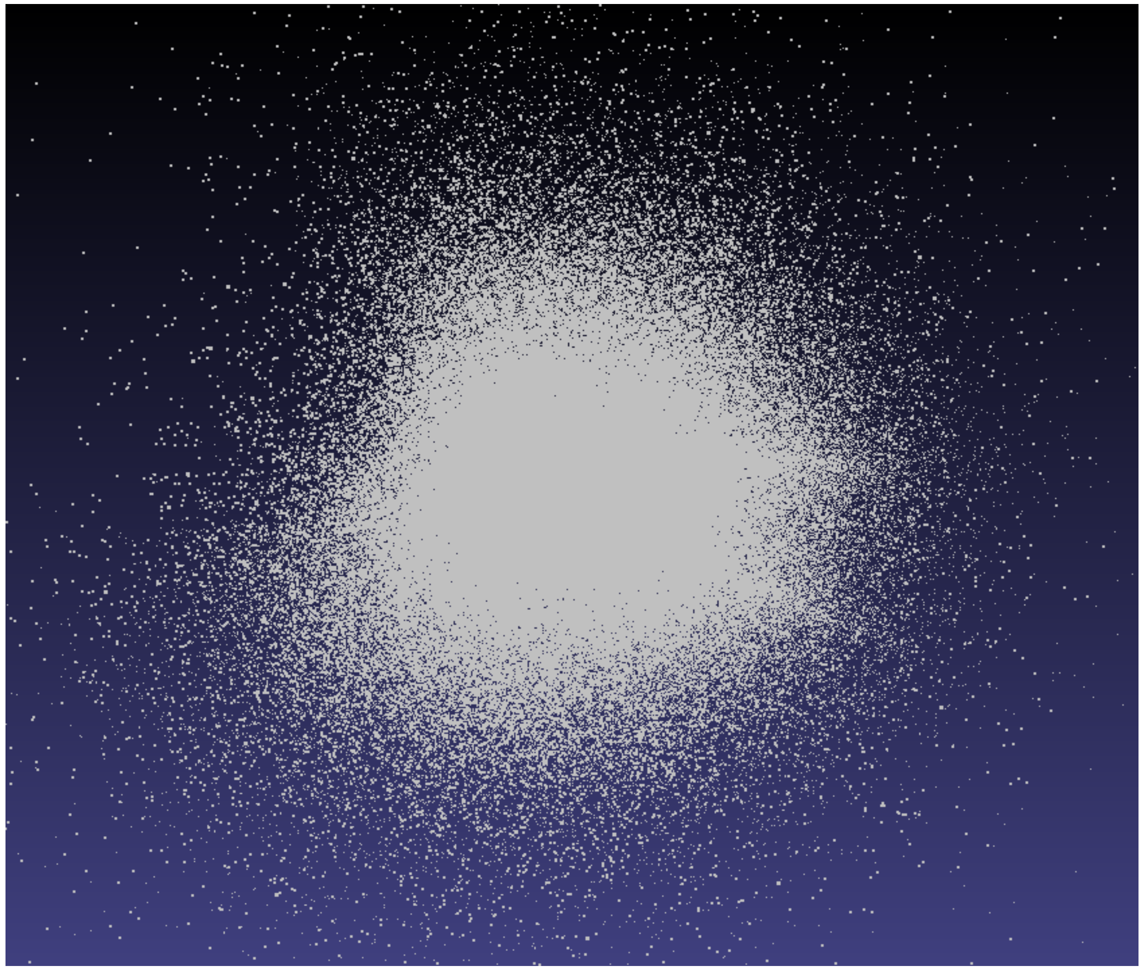


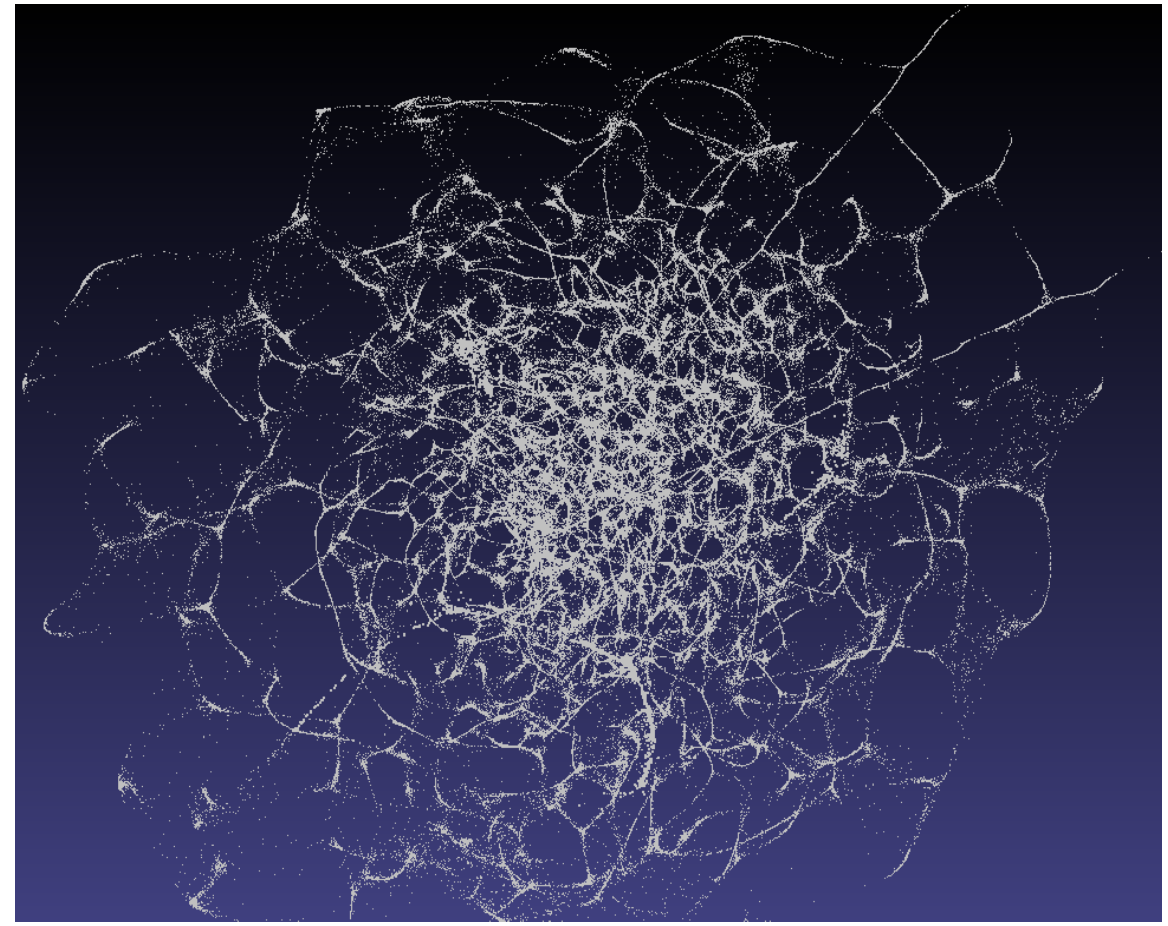
- Mean-Shift like algorithm (Fukunaga-Hostetler'75, Comaniciu-Meer '02)
- Theoretical guarantees on the convergence of the algorithm and "smoothness" of trajectories.
- "Fast concentration of mass" around underlying geometric structures?

Pushing data along the gradient of d_{μ,m_0}

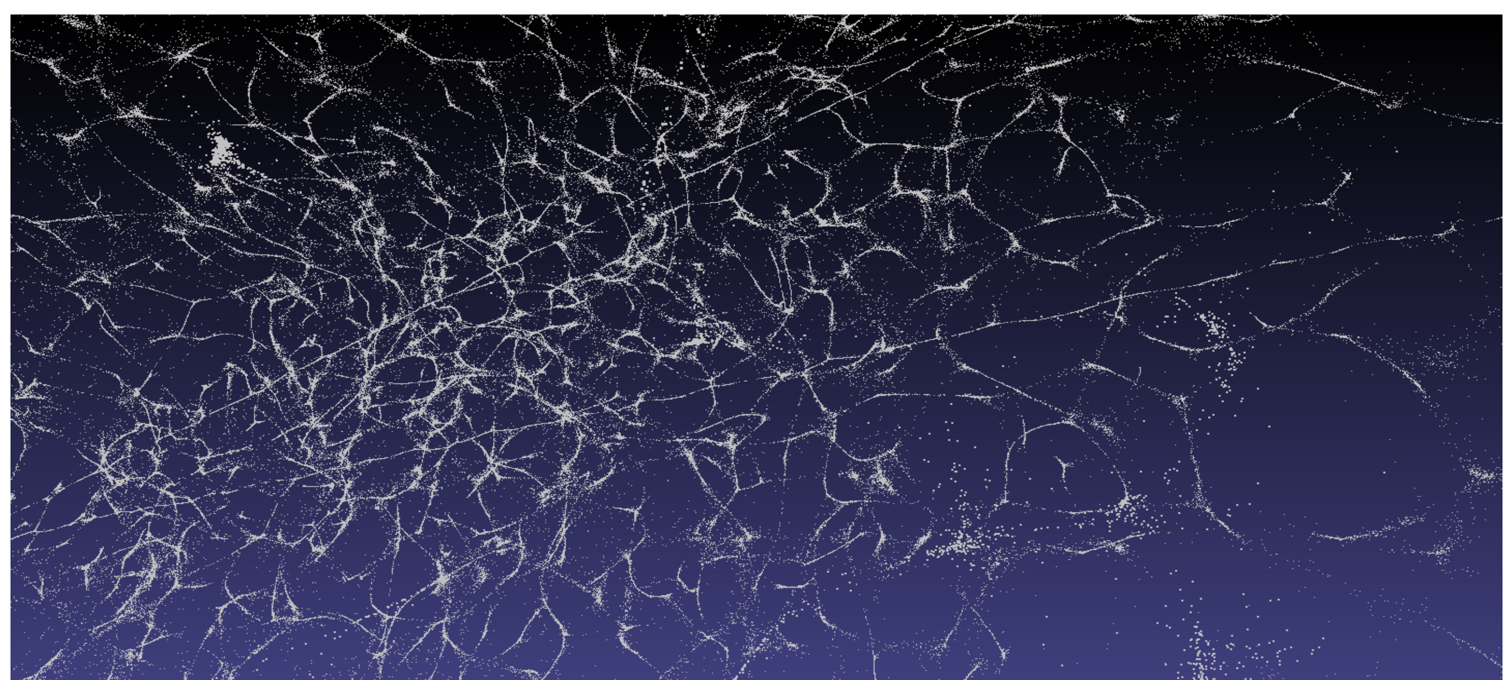
Distance-based mean-shift followed by k-Means clustering on the point cloud made of LUV colors of the pixels of the picture on the right (10 clusters).

Pushing data along the gradient of d_{μ,m_0}

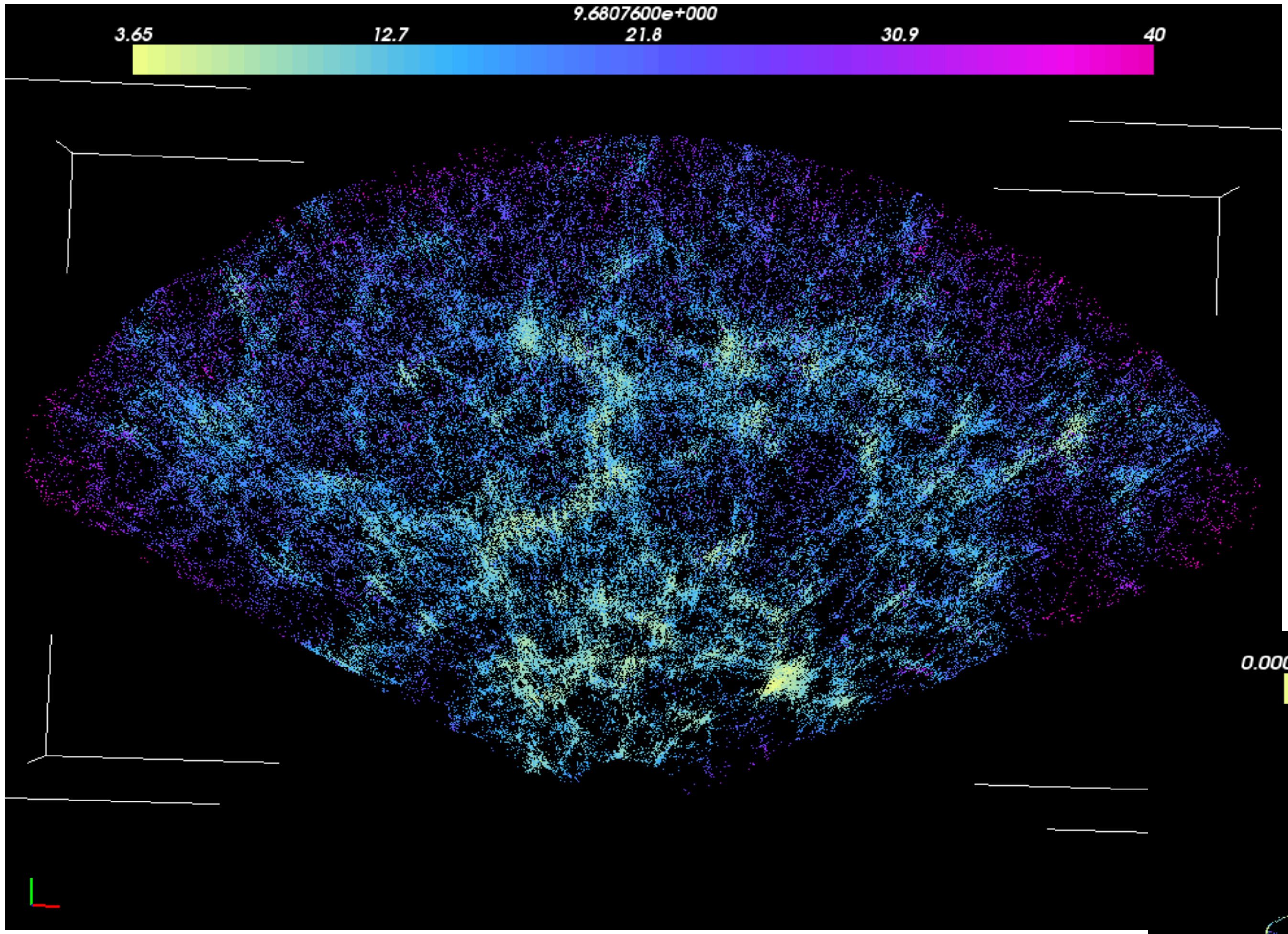


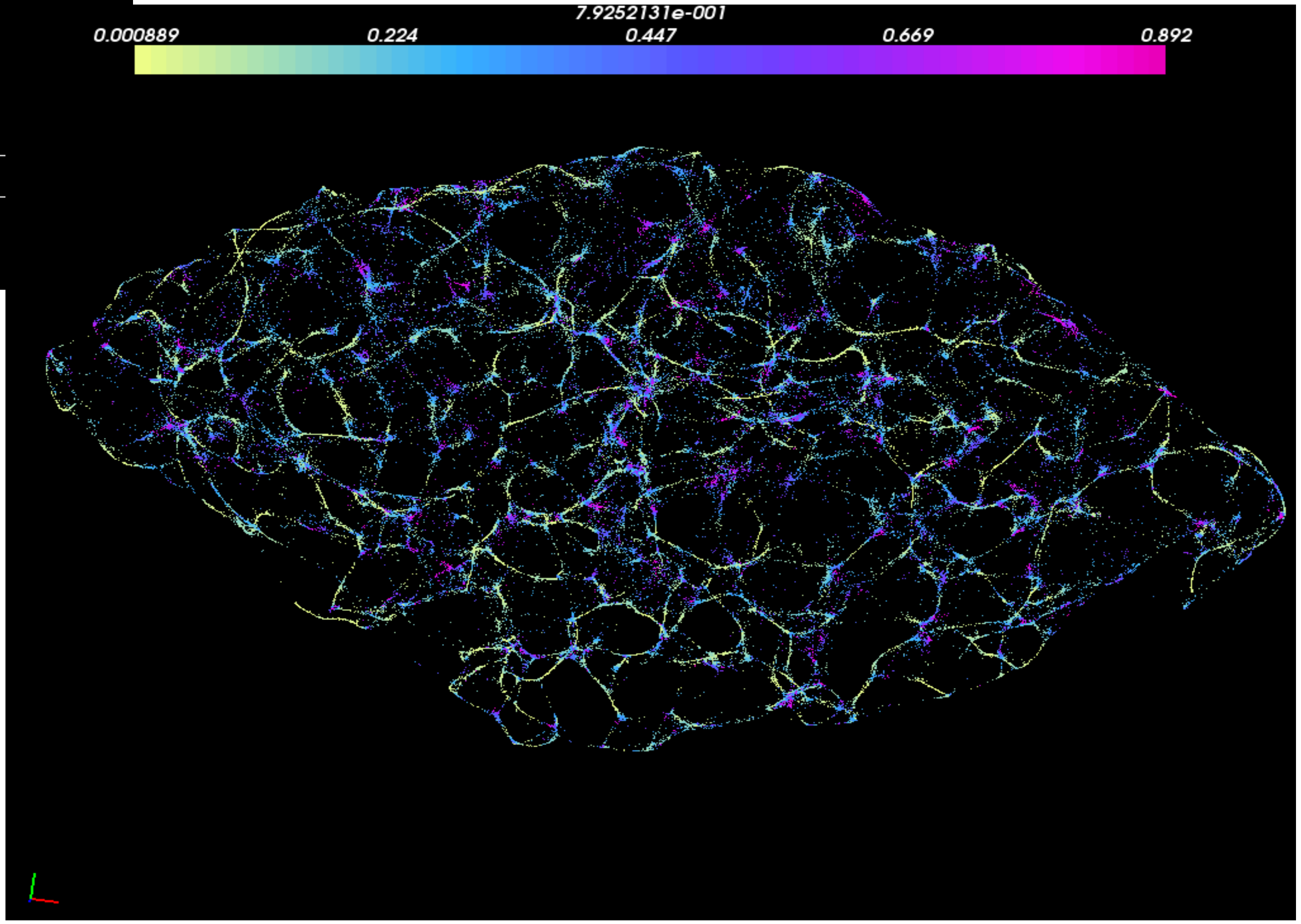


Galaxies data set



Pushing data along the gradient of d_{μ,m_0}





Take-home messages

- $\mu \mapsto d_{\mu,m_0}$ provide a way to associate geometry to a measure in Euclidean space.
- d_{μ,m_0} is robust to Wasserstein perturbations : outliers and noise are easily handled (no assumption on the nature of the noise).
- d_{μ,m_0} shares regularity properties with the usual distance function to a compact.
- Geometric stability results in this measure-theoretic setting : topology/geometry of the sublevel sets of d_{μ,m_0}, stable notion of persistence diagram for μ,...
- No need of statistical models.
- Algorithm: for finite point clouds d_{μ,m_0} and $\nabla(d_{\mu,m_0})$ can be easily and efficiently computed in any dimension.

To get more details: C., Cohen-Steiner, Mérigot, Geometric Inference for Probability Measures, J. Foundations of computational Mathematics, vol. 11, 6, 2011