Geometric Inference for Probability distributions

F. Chazal [1]   D. Cohen-Steiner [2]   Q. Mérigot [2]

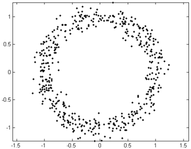
[1]Geometrica, INRIA Saclay,

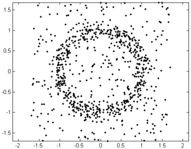
[2]Geometrica, INRIA Sophia-Antipolis

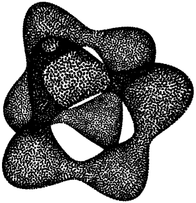
2009 July 8

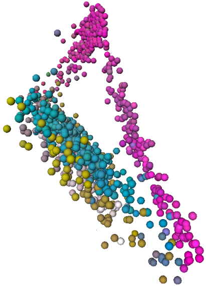
Outline

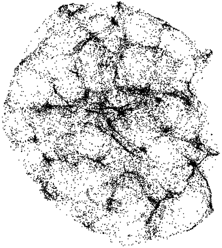
Motivation











What is the (relevant) topology/geometry of a point cloud data set in $\mathbb{R}^d$?

Motivations : Reconstruction, manifold learning and NLDR, clustering and segmentation, etc...

Geometric Inference

Question

Given an approximation C of a geometric object K, what geometric and topological quantities of K is it possible to approximate, knowing only C ?

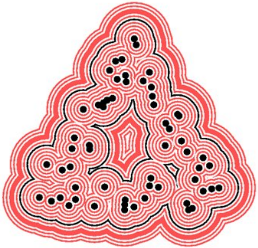
- The answer depends on the considered class of objects and a notion of distance between the objects (approximation) ;
- Some positive answers for a large class of compact sets endowed with the Hausdorff distance.
- In this talk :
  - can the considered objects be probability measures on $\mathbb{R}^d$?
  - motivation : allowing approximations to have outliers or to be corrupted by “non local” noise.

Distance functions for geometric inference

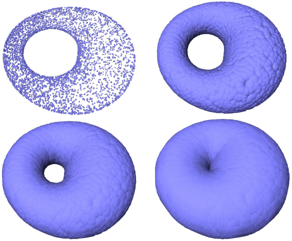
Distance function and Hausforff distance

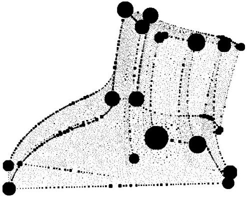
Distance to a compact $K \subseteq \mathbb{R}^d$: $\mathrm{d}_K : x \mapsto \inf_{p \in K} \|x - p\|$ Hausdorff distance between compact sets $K, K' \subseteq \mathbb{R}^d$:
$d_H(K, K') = \inf_{x \in \mathbb{R}^d} |d_K(x) - d_{K'}(x)|$

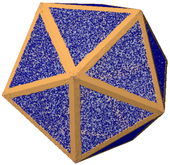
- Replace K and C by d_K and d_C.
- Compare the topology of the *offsets* $K^r = \mathrm{d}_K^{-1}([0, r])$ and $C^r = \mathrm{d}_C^{-1}([0, r])$.



Stability properties of the offsets







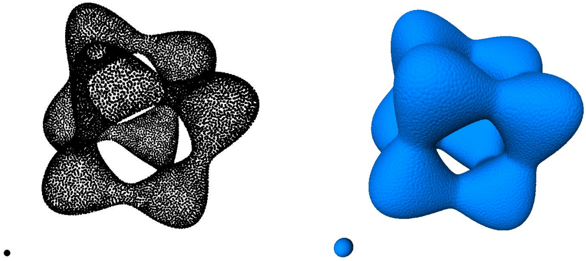
Topological/geometric properties of the offsets of K are stable with respect to Hausdorff approximation :

1. Topological stability of the offsets of K (CCSL'06, NSW'06).

2. Approximate normal cones (CCSL'08).

3. Boundary measures (CCSM'07), curvature measures (CCSLT'09), Voronoi covariance measures (GMO'09).

The problem of outliers



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

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

The three main ingredients for stability

The stability in distance-based geometric inference relies on the three following facts :

1. the 1-Lipschitz property for d_K ;
2. the 1-concavity on the function $\mathrm{d}_K^2 : x \to \|x\|^2 - \mathrm{d}_K^2(x)$ is convex.
3. the stability of the map $K \mapsto \mathrm{d}_K$:

$$\|\mathrm{d}_K - \mathrm{d}_{K'}\|_\infty = \sup_{x \in \mathbb{R}^d} |\mathrm{d}_K(x) - \mathrm{d}_{K'}(x)| = \mathrm{d}_{\mathrm{H}}(K, K')$$

A map $\phi : \mathbb{R}^d \to \mathbb{R}$ which verifies (1) and (2) is called *distance-like*.

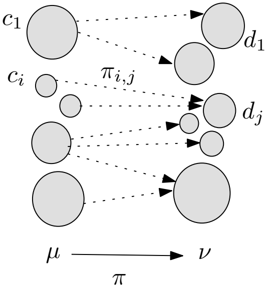
Replacing compact sets by measures

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
- a point cloud $C = \{p_1, \dots, p_n\}$ 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 $\mathrm{vol}_k|_M$.

Distance between measures

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



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. $\mathrm{d_W}(\mu, \nu) := \inf_\pi C(\pi)$

Wasserstein distance

Examples :

1. If $\#C_1 = \#C_2$, then $\mathrm{d}^2_{\mathrm{W}}(\mu_{C_1}, \mu_{C_2})$ is the cost of a minimal least-square matching between C_1 and C_2 ;
2. If $C = \{p_1, \ldots, \mathrm{p}_n\}$ and $C' = \{p_1, \ldots, p_{n-k-1}, o_1, \ldots, o_k\}$ with $\mathrm{d}(o_i, C) = R$, then $\mathrm{d_H}(C, C') \geq R$ while

$$\mathrm{d_W}(\mu_C, \mu_{C'}) \leq m(R + \mathrm{diam}(C)) \text{ with } m = \frac{k}{n};$$

3. If μ is a probability measure, $\mathrm{d_W}(\mu * \mathcal{N}(0,\sigma), \mu) \leq \sigma$;
4. If $X_1, \ldots, X_N$ are iid with law μ, then (in general), $\frac{1}{N}\sum_{i=1}^{N} \delta_{X_i}$ converges to μ whp as N tends to ∞.

Outline

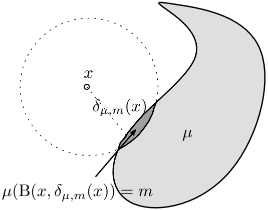
1. Geometric inference for measures.
2. Distance to a probability measures.
3. Applications

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 $\mathbb{R}^d$:

$$\delta_{\mu,m}(x) = \inf\{r > 0;\ \mu(\mathrm{B}(x,r)) > m\}$$



- $\delta_{\mu,m}$ is the smallest distance needed to attain a mass of at least m ;
- Coincides with the distance to the k-th neighbor when $m = k/n$ and $\mu = \frac{1}{n}\sum_{i=1}^{n} \delta_{p_i}$: $\delta_{\mu,k/n}(\mu) = \left\| x - \mathrm{p}_C^k(x) \right\|$.

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

Distance to a measure, first attempt

Let $m \in]0,1[$ be a positive mass, and μ a probability measure on $\mathbb{R}^d$:

$$\delta_{\mu,m}(x) = \inf\{r > 0;\ \mu(\mathrm{B}(x,r)) > m\}$$

Unstability under Wasserstein perturbations :

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

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

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

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

If μ is a measure on $\mathbb{R}^d$ and $m_0 > 0$, one let :

$$\mathrm{d}_{\mu,m_0} : x \in \mathbb{R}^d \mapsto \left(\frac{1}{m_0} \int_0^{m_0} \delta^2_{\mu,m}(x) \mathrm{d}m \right)^{1/2}$$

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

$$\mathrm{d}_{\mu,m_0}(x) = \left(\frac{1}{k_0} \sum_{k=1}^{k_0} \left\| x - \mathrm{p}^k_C(x) \right\|^2 \right)^{1/2}$$

The distance function to a discrete measure.

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

$$\mathrm{d}_{\mu,m_0}(x) = \left(\frac{1}{k_0} \sum_{k=1}^{k_0} \left\| x - \mathrm{p}_C^k(x) \right\|^2 \right)^{1/2}$$

$$\nabla \mathrm{d}_{\mu,m_0}(x) = \frac{\frac{1}{k_0} \sum_{k=1}^{k_0} \left[x - \mathrm{p}_C^k(x) \right]}{\mathrm{d}_{\mu,m_0}(x)}$$

Another expression for d_{μ,m_0}

Proposition

The distance $\mathrm{d}_{\mu,m_0}(x)$ coincides with the partial Wasserstein distance between the Dirac mass $m_0\delta_x$ and μ. More precisely :

$$\begin{aligned}\sqrt{m_0}\,\mathrm{d}_{\mu,m_0}(x) &= \min\{\mathrm{d_W}(m_0\delta_x,\nu);\nu\leq\mu \text{ and } \mathrm{mass}(\nu)=m_0\}\\ &= \min\left\{\left(\int_{\mathbb{R}^d}\|y-x\|^2\,\mathrm{d}\nu(y)\right)^{1/2};\nu\leq\mu,\mathrm{mass}(\nu)=m_0\right\}\end{aligned}$$

Let μ_{x,m_0} be a measure realizing this minimum.

- The measure μ_{x,m_0} gives mass to the *multiple* "projections" of x on μ ;
- For the point cloud case, when $m_0 = k_0/n$ and x is not on a k-Voronoï face,

$$\mu_{x,m_0} = \sum_{k=1}^{k_0}\frac{1}{n}\delta_{\mathrm{p}_C^k(x)}$$

1-Concavity of the squared distance function

Regularity

$$
\begin{aligned}
m_0 \mathrm{d}^2_{\mu,m_0}(x+h) &= \int_{\mathbb{R}^d} \|x+h-z\|^d \, \mathrm{d}\mu_{x+h,m_0}(z) \\
&\leq \int_{\mathbb{R}^d} \|x+h-z\|^d \, \mathrm{d}\mu_{x,m_0}(z) \\
&\leq m_0 \mathrm{d}^2_{\mu,m_0}(x) + 2\int_{\mathbb{R}^d} \langle h | x-z \rangle \mathrm{d}\mu_{x,m_0}(z) + m_0 \|h\|^2
\end{aligned}
$$

That is :

$$
\mathrm{d}^2_{\mu,m_0}(x+h) \leq \mathrm{d}^2_{\mu,m_0}(x) + \langle h | \nabla \mathrm{d}^2_{\mu,m_0}(x) \rangle + \|h\|^2
$$

$$
\text{with } \nabla \mathrm{d}^2_{\mu,m_0}(x) := 2m_0^{-1} \int_{\mathbb{R}^d} (x-z) \mathrm{d}\mu_{x,m_0}(z)
$$

Theorem

The distance function d_{μ,m_0} is *distance-like, ie.*

1. the function $x \mapsto \mathrm{d}_{\mu,m_0}(x)$ is 1-Lipschitz ;
2. the function $x \mapsto \|x\|^2 - \mathrm{d}^2_{\mu,m_0}(x)$ is convex ;

Theorem

The map $\mu \mapsto \mathrm{d}_{\mu,m_0}$ from probability measures to continuous functions is $\frac{1}{\sqrt{m_0}}$–Lipschitz, ie

$$\left\|\mathrm{d}_{\mu,m_0} - \mathrm{d}_{\mu',m_0}\right\|_\infty \leq \frac{1}{\sqrt{m_0}} \mathrm{d_W}(\mu,\mu')$$

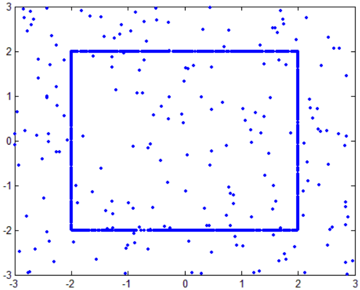
Outline

Consequences of the previous properties

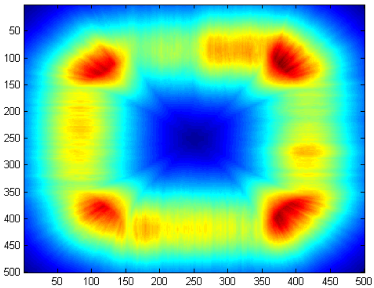
1. existence of an analogous to the medial axis
2. stability of a filtered version of it (as with the μ-medial axis) under Wasserstein perturbation
3. stability of the critical function of a measure
4. the gradient $\nabla \mathrm{d}_{\mu,m_0}$ is L^1-stable
5. ...

$\Longrightarrow$ the distance functions d_{μ,m_0} share many stability and regularity properties with the usual distance function.

Example : square with outliers

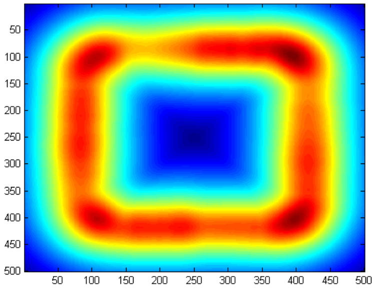


10% outliers, $k = 150$

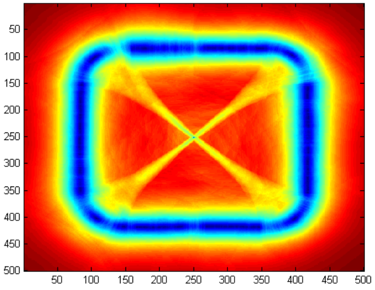


δ_{μ,m_0}, $m_0 = 1/10$

Example : square with outliers

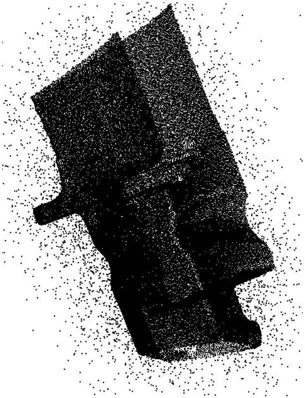


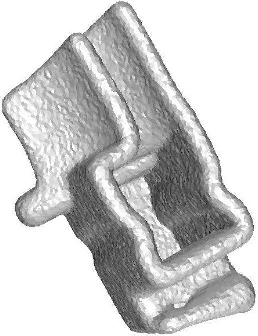
d_{μ,m_0}



$\|\nabla \mathrm{d}_{\mu,m_0}\|$

A 3D example





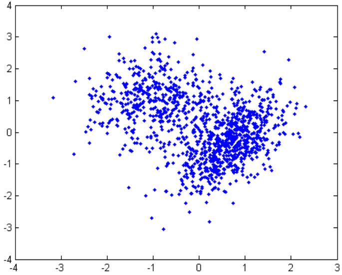
Reconstruction of an offset from a noisy dataset, with 10% outliers

A reconstruction theorem

Theorem

Let μ be a probability measure of dimension at most $k > 0$ with compact support $K \subset \mathbb{R}^d$ such that $r_\alpha(K) > 0$ for some $\alpha \in (0, 1]$. For any $0 < \eta < r_\alpha(K)$, there exists positive constants $m_1 = m_1(\mu, \alpha, \eta) > 0$ and $C = C(m_1) > 0$ such that :
for any $m_0 < m_1$ and any probability measure μ' such that $W_2(\mu, \mu') < C\sqrt{m_0}$, the sublevel set $d^{-1}_{\mu',m_0}((-\infty, \eta])$ is homotopy equivalent (and even isotopic) to the offsets $d_K^{-1}([0, r])$ of K for $0 < r < r_\alpha(K)$.

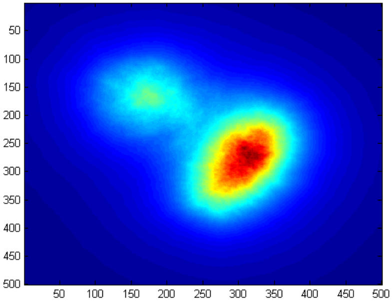
k-NN density estimation vs distance to a measure



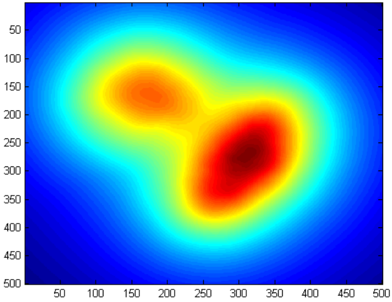
data

Density is estimated using $x \mapsto \frac{m_0}{\omega_{d-1}(\delta_{\mu,m_0}(x))}$, $m_0 = 150/1200$.

k-NN density estimation vs distance to a measure



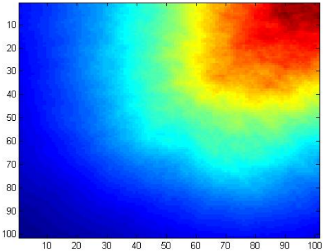
density

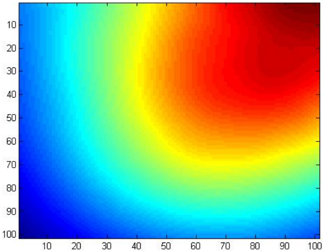


$\log(1 + 1/d_{\mu,m_0})$

Density is estimated using $x \mapsto \frac{m_0}{vol_d(B(x,\delta_{\mu,m_0}(x)))}$, $m_0 = 150/1200$ (Devroye-Wagner '77).

k-NN density estimation vs distance to a measure

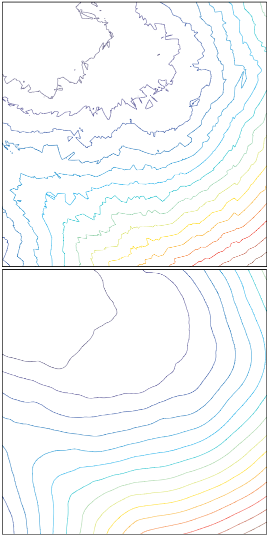




1. the gradient of the estimated density can behave wildly
2. exhibits peaks near very dense zone

1. can be fixed using d_{μ,m_0} (because of the semiconcavity)
2. shows that the *distance function* is a better-behaved geometric object to associate to a measure.

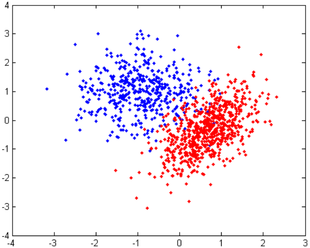
k-NN density estimation vs distance to a measure

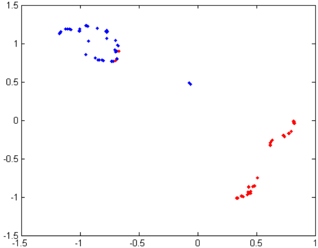


1. the gradient of the estimated density can behave wildly
2. exhibits peaks near very dense zone

1. can be fixed using d_{μ,m_0} (because of the semiconcavity)
2. shows that the *distance function* is a better-behaved geometric object to associate to a measure.

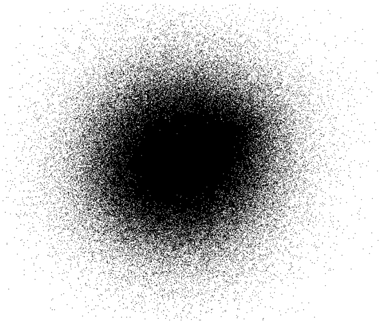
Pushing data along the gradient of d_{μ,m_0}

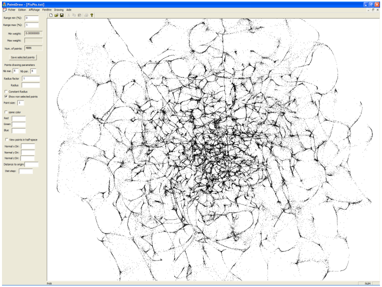


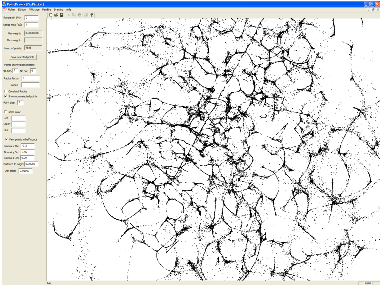


- Mean-Shift like algorithm (Comaniciu-Meer '02)
- Theoretical guarantees on the convergence of the algorithm and "smoothness" of trajectories.

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







Summary

- $\mu \mapsto \mathrm{d}_{\mu,m_0}$ 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,.
- Algorithm : for finite point clouds d_{μ,m_0} and $\nabla(\mathrm{d}_{\mu,m_0})$ can be easily and efficiently computed pointwise in any dimension.