The distance function to a probability measure

F. Chazal D. Cohen-Steiner Q. Mérigot

2008 January 26

Outline

1. Introduction

2. Distance to a measure
   - Measure, Wasserstein distance
   - Definition of the distance function
   - Properties

3. Results & Applications
   - Distance functions to point clouds with outliers
   - Non-parametric density estimation

Geometric inference

Question

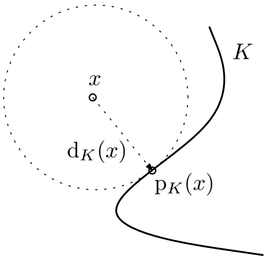
Given an approximation C of a geometric object K, is it possible to estimate the topological and geometric properties of K, knowing only the approximation C ?

- In $\mathbb{R}^3$, without noise, and when K is a smooth surface : ε-sampling condition guarantees a good approximation of the geometry and topology.
- Recent algorithms allow a small amount of *Hausdorff* noise.
- However, dealing with *outliers* is still an open problem.

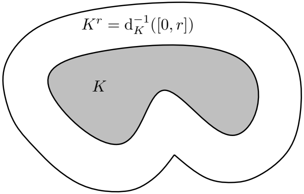
Distance functions for geometric estimations

Distance function

Distance function to a compact $K \subseteq \mathbb{R}^d$:
$\mathrm{d}_K : x \mapsto \inf_{p \in K} \|x - p\|$.



Distance functions for geometric estimations



Offset : $K^r = \mathrm{d}_K^{-1}(K)$
1. If K is smooth, K^r retracts onto K for small r (Grove).

2. The function $r \mapsto \mathrm{vol}_d(K^r)$ is polynomial on $[0, R]$:

- If K is convex : $R = +\infty$ (Steiner, Minkowski);
- If K is a smooth submanifold : $R \ll 1$ (Weyl)

$$\mathrm{vol}_d(K^r) = \sum_{k=1}^{d-1} \left(\int_K \sum_{i_1 < \ldots < i_k} \kappa_{i_1}(x) \ldots \kappa_{i_k}(x) \mathrm{d}x \right) r^i$$

- If K has positive reach : $R = \mathrm{reach}(K)$ (Federer).

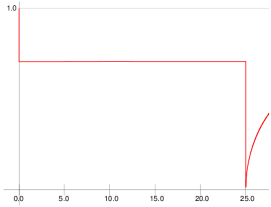
Distance functions for geometric inference

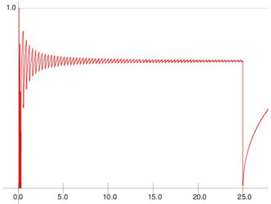
Stability and inference

A stability result for a distance-based geometric estimation method $K \mapsto \mathcal{F}(K)$ bounds the error made in the estimation by replacing the compact set K by a Hausdorff approximation K'.

Exemple of stable properties.

1. Critical function of K : $\chi_K(r) = \inf_{x \in \partial K^r} \|\nabla_x \mathrm{d}_K\|$. Measures the criticality of d_K in offsets of K (CCSL '06).

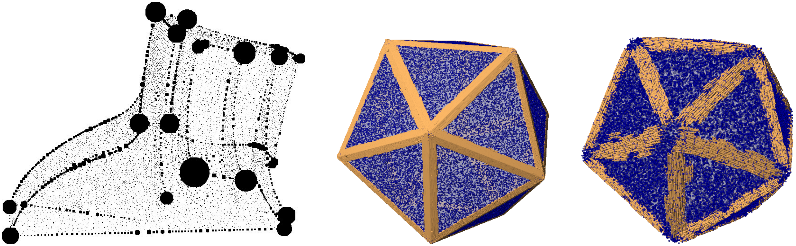




Exemple of stable properties (continued).

2. Approximate normal cones (CCSL '08) ;

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



The proofs of stability relies on three important theoretical facts :

1. the stability of the map $K \mapsto \mathrm{d}_K$:

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

2. the 1-Lipschitz property for d_K ;
3. the 1-concavity on the function d_K^2.

If $K' = K \cup \{x\}$ where $\mathrm{d}_K(x) > R$, then $\|\mathrm{d}_K - \mathrm{d}_{K'}\|_\infty > R$: the properties become unstable both in theory and in practice.

Outline

1. Introduction
2. Distance to a measure
   - Measure, Wasserstein distance
   - Definition of the distance function
   - Properties
3. Results & Applications
   - Distance functions to point clouds with outliers
   - Non-parametric density estimation

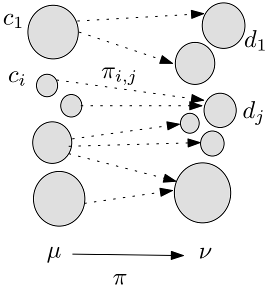
Mathematical background : Measures

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 $C = \{p_1, \ldots, p_n\}$ 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 $\mathrm{vol}_k|_M$.

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 $\sum_i p_{ij} = d_j$ and $\sum_j p_{ij} = c_i$.
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 :

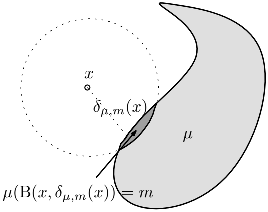
- If C_1 and C_2 are two point clouds, with $\#C_1 = \#C_2$, then $\mathrm{d_W}(\mu_{C_1}, \mu_{C_2})$ is the square root of the cost of a minimal least-square matching between C_1 and C_2.
- If $C = \{p_1, \ldots, \mathrm{p}_n\}$ is a point cloud, and $C' = \{p_1, \ldots, p_{n-k-1}, o_1, \ldots, o_k\}$ with $\mathrm{d}(o_i, C) = R$, then
$$\mathrm{d_H}(C, C') = R \qquad \text{but} \qquad \mathrm{d_W}(\mu_C, \mu_{C'}) \leq \frac{k}{n}(R + \mathrm{diam}(C))$$

Outline

The distance to a measure

Distance to a measure, first attempt

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



- $\delta_{\mu,m}$ 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 $\mathrm{p}_C^k(x)$ denote the kth nearest neighbor to x in C, and set $m_0 = k_0/n$:

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

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

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

Hence, this function is linear inside k_0th-order Voronoï cells, and :

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

1-Concavity of the squared distance function

$$\begin{aligned}
&\mathrm{d}^2_{\mu,m_0}(y) \\
&= \frac{1}{k_0}\sum_{k=1}^{k_0}\left\|y - \mathrm{p}_C^k(y)\right\|^2 \leq \frac{1}{k_0}\sum_{k=1}^{k_0}\left\|y - \mathrm{p}_C^k(x)\right\|^2 \\
&\leq \frac{1}{k_0}\sum_{k=1}^{k_0}\left\|y - x + x - \mathrm{p}_C^k(x)\right\|^2 \\
&= \|y-x\|^2 + \frac{1}{k_0}\sum_{k=1}^{k_0}\left\|x - \mathrm{p}_C^k(x)\right\|^2 + 2\frac{1}{k_0}\sum_{k=1}^{k_0}\langle y - x | x - \mathrm{p}_C^k(x)\rangle \\
&= \|y-x\|^2 + \mathrm{d}^2_{\mu,m_0}(x) + \langle y - x | \nabla \mathrm{d}^2_{\mu,m_0}(x)\rangle
\end{aligned}$$

Proposition

The distance function to a discrete measure is 1-concave, *ie.*

$$(\|y\|^2 - \mathrm{d}^2_{\mu,m_0}(y)) - (\|x\|^2 - \mathrm{d}^2_{\mu,m_0}(x)) \geq \langle \nabla(\|x\|^2 - \mathrm{d}^2_{\mu,m_0})(x) | x - y \rangle$$

This is equivalent to the function $x \mapsto \|x\|^2 - \mathrm{d}^2_{\mu,m_0}(x)$ being convex.

Outline

1. Introduction
2. Distance to a measure
   - Measure, Wasserstein distance
   - Definition of the distance function
   - Properties
3. Results & Applications
   - Distance functions to point clouds with outliers
   - Non-parametric density estimation

Another expression for d_{μ,m_0}

Projection submeasure

For any $x \in \mathbb{R}^d$, and $m > 0$, denote by $\mu_{x,m}$ the restriction of μ on the ball $B = \mathrm{B}(x, \delta_{\mu,m}(x))$, whose trace on the sphere ∂B has been rescaled so that the total mass of $\mu_{x,m}$ is m.

- The measure $\mu_{x,m}$ gives mass to the *multiple* "projections" of x on μ ;
- For the point cloud case, when $m_0 = k_0/n + r$ with $0 < r < 1$, and x is not on a k-Voronoï face,

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

Another expression for d_{μ,m_0}

Proposition

Let μ be a probability measure on $\mathbb{R}^d$ and $m_0 > 0$. Then, for almost every point $x \in \mathbb{R}^d$ d_{μ,m_0} is differentiable at x and :

$$\mathrm{d}^2_{\mu,m_0}(x) = \frac{1}{m_0} \int \|x - h\|^2 \, \mathrm{d}\mu_{x,m_0}(h)$$

$$\nabla(\|x\|^2 - \mathrm{d}^2_{\mu,m_0}(x)) = \frac{2}{m_0} \int h \mathrm{d}\mu_{x,m_0}(h)$$

- compare with : $\nabla(\|x\|^2 - \mathrm{d}^2_{\mu,m_0}(x)) = \frac{2}{k_0} \sum_{k=1}^{k_0} \mathrm{p}^k_C(x)$;
- interpretation : $\mathrm{d}_{\mu,m_0}(x)$ is the (partial) Wasserstein distance between the Dirac mass $m_0\delta_x$ and μ.

Properties

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

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

the proof of **2** and **3** involves the second expression for d_{μ,m_0}.

properties **1** and **2** are related to the regularity of d_{μ,m_0} while **3** is a stability property

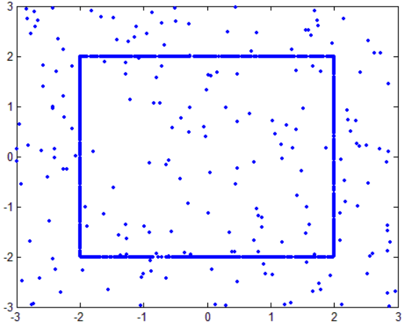
Consequences of the previous properties

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. ...

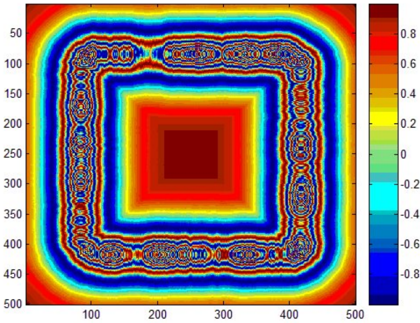
$\Longrightarrow$ the distance functions d_{μ,m_0} share many stability and regularity properties with the usual distance function.

Outline

Square with outliers

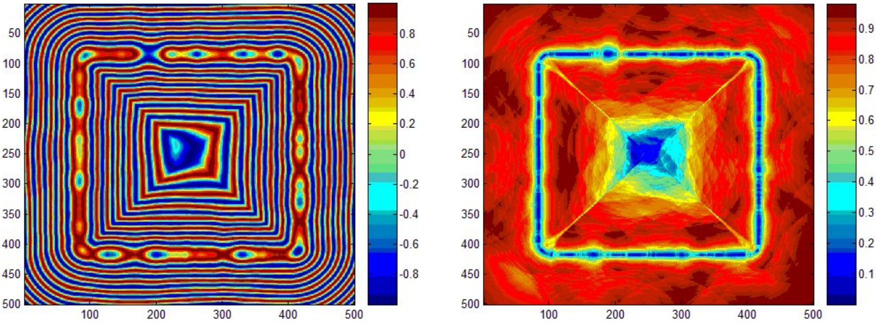


10% outliers, $k = 50$



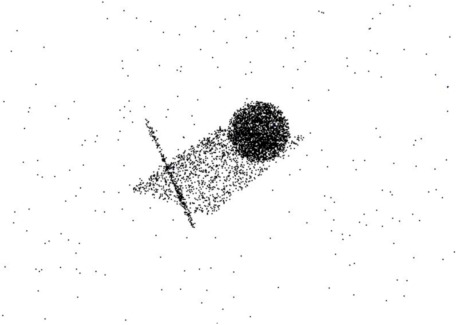
δ_{μ,m_0}

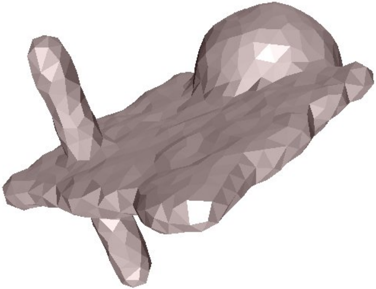
Square with outliers



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

3D square



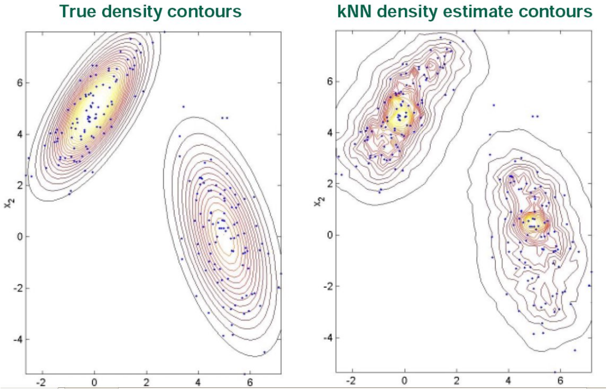


Reconstruction of an offset from a noisy dataset, with 10% outliers

Outline

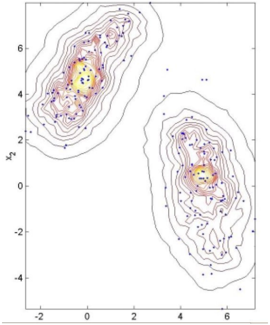
1. Introduction
2. Distance to a measure
   - Measure, Wasserstein distance
   - Definition of the distance function
   - Properties
3. Results & Applications
   - Distance functions to point clouds with outliers
   - Non-parametric density estimation

k-NN density estimation method, from a statistics textbook



Density is estimated using $x \mapsto \dfrac{m_0}{\omega_{d-1}(\delta_{\mu,m_0}(x))}$.

Downsides of kNN density estimation



1. the gradient of the estimated density can behave wildly
2. this density decreases slowly at infinity (it isn't integrable !)
3. exhibits peaks near very dense zone

1. can be fixed using d_{μ,m_0} instead of δ_{μ,m_0} (because of the semiconcavity)
2. and **3.** shows that the *distance function* is a better-behaved geometric object to associate to a measure.

Summary

- We introduced a distance function to a measure $\mu \mapsto \mathrm{d}_{\mu,m_0}$.
- This distance function is robust to Wasserstein perturbations, and thus handles outliers gracefully ;
- d_μ shares regularity properties with the usual distance function to a compact.
- These properties allow to adapt geometric stability results to this measure-theoretic setting.