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

1. Geometric inference for measures.

2. Distance to a probability measures.

3. Applications
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...
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 Hausdorff distance

- **Distance to a compact \( K \subseteq \mathbb{R}^d \):** \( d_K : x \mapsto \inf_{p \in K} \| x - p \| \) 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
  \[ K' = d_K^{-1}([0, r]) \] and
  \[ C' = d_C^{-1}([0, r]). \]
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).
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 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 $d^2_K : x \rightarrow \|x\|^2 - d^2_K(x)$ is convex.
3. the stability of the map $K \mapsto d_K$:

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

A map $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ which verifies (1) and (2) is called distance-like.
A **measure** $\mu$ is a mass distribution on $\mathbb{R}^d$.

Mathematically, it is defined as a map $\mu$ 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 the mass of $\mu$ 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 $\text{vol}_k|_M$.  

F. Chazal, D. Cohen-Steiner, Q. Mérigot (Geometrica)

Geometric Inference for Probability distributions

2009/07/08 9 / 32
The **Wasserstein distance** $d_W(\mu, \nu)$ between two probability measures $\mu, \nu$ quantifies the optimal cost of pushing $\mu$ onto $\nu$, the cost of moving a small mass $dx$ from $x$ to $y$ being $\|x - y\|^2 \, dx$.

1. $\mu$ and $\nu$ 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. $d_W(\mu, \nu) := \inf_\pi C(\pi)$
Wasserstein distance

Examples:

1. If \( \#C_1 = \#C_2 \), then \( d_W^2(\mu_{C_1}, \mu_{C_2}) \) is the cost of a minimal least-square matching between \( C_1 \) and \( C_2 \);

2. If \( C = \{p_1, \ldots, p_n\} \) and \( C' = \{p_1, \ldots, p_{n-k-1}, o_1, \ldots, o_k\} \) with \( d(o_i, C) = R \), then \( d_H(C, C') \geq R \) while

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

3. If \( \mu \) is a probability measure, \( d_W(\mu \ast \mathcal{N}(0, \sigma), \mu) \leq \sigma \);

4. If \( X_1, \ldots, X_N \) are iid with law \( \mu \), then (in general), \( \frac{1}{N} \sum_{i=1}^{N} \delta x_i \) converges to \( \mu \) whp as \( N \) tends to \( \infty \).
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 $\mu$ a probability measure on $\mathbb{R}^d$:

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

- $\delta_{\mu,m}$ 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 $\mu = \frac{1}{n} \sum_{i=1}^{n} \delta_{p_i}$:
  $$\delta_{\mu,k/n}(\mu) = \| x - p^k_C(x) \|.$$
Unstability of $\mu \mapsto \delta_{\mu,m}$

**Distance to a measure, first attempt**

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

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

**Unstability under Wasserstein perturbations**:

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

for $\varepsilon > 0$:

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

for $\varepsilon = 0$:

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

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

**Definition**

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

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

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

$$d_{\mu,m_0}(x) = \left( \frac{1}{k_0} \sum_{k=1}^{k_0} \|x - p_k^C(x)\|^2 \right)^{1/2}$$
The distance function to a discrete measure.

Example (continued) 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 $k$th nearest neighbor to $x$ in $C$, and set $m_0 = k_0/n$:

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

$$\nabla d_{\mu,m_0}(x) = \frac{1}{k_0} \sum_{k=1}^{k_0} \frac{[x - p_C^k(x)]}{d_{\mu,m_0}(x)}$$
Another expression for $d_{\mu,m_0}$

**Proposition**

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

$$\sqrt{m_0}d_{\mu,m_0}(x) = \min\{d_W(m_0\delta_x, \nu); \nu \leq \mu \text{ and } \text{mass}(\nu) = m_0\}$$

$$= \min \left\{ \left(\int_{\mathbb{R}^d} \|y - x\|^2 \, d\nu(y)\right)^{1/2}; \nu \leq \mu, \text{mass}(\nu) = m_0 \right\}$$

Let $\mu_{x,m_0}$ be a measure realizing this minimum.

- The measure $\mu_{x,m_0}$ gives mass to the *multiple* “projections” of $x$ on $\mu$;
- For the point cloud case, when $m_0 = k_0/n$ and $x$ is not on a $k$-Voronoi face,

$$\mu_{x,m_0} = \sum_{k=1}^{k_0} \frac{1}{n} \delta_{p_k^C}(x)$$
1-Concavity of the squared distance function

Regularity

\[ m_0 d_{\mu,m_0}^2(x+h) = \int_{\mathbb{R}^d} \|x+h-z\|^d d\mu_{x+h,m_0}(z) \]
\[ \leq \int_{\mathbb{R}^d} \|x+h-z\|^d d\mu_{x,m_0}(z) \]
\[ \leq m_0 d_{\mu,m_0}^2(x) + 2 \int_{\mathbb{R}^d} \langle h|x-z \rangle d\mu_{x,m_0}(z) + m_0 \|h\|^2 \]

That is:

\[ d_{\mu,m_0}^2(x+h) \leq d_{\mu,m_0}^2(x) + \langle h|\nabla d_{\mu,m_0}^2(x) \rangle + \|h\|^2 \]

with \( \nabla d_{\mu,m_0}^2(x) := 2m_0^{-1} \int_{\mathbb{R}^d} (x-z) d\mu_{x,m_0}(z) \)
Theorem

The distance function $d_{\mu,m_0}$ is distance-like, ie.

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

Theorem

The map $\mu \mapsto d_{\mu,m_0}$ from probability measures to continuous functions is $\frac{1}{\sqrt{m_0}}$-Lipschitz, ie

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

1. Geometric inference for measures.

2. Distance to a probability measures.

3. Applications
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 $\mu$-medial axis) under Wasserstein perturbation
3. stability of the critical function of a measure
4. the gradient $\nabla d_{\mu,m_0}$ is $L^1$-stable
5. ...

$\implies$ the distance functions $d_{\mu,m_0}$ share many stability and regularity properties with the usual distance function.
Example: square with outliers

10% outliers, $k = 150$

$\delta_{\mu, m_0}, m_0 = 1/10$
Example: square with outliers

\[ d_{\mu, m_0} \]

\[ \| \nabla d_{\mu, m_0} \| \]
A 3D example

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

**Theorem**

Let $\mu$ 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 $\mu'$ such that $W_2(\mu, \mu') < C \sqrt{m_0}$, the sublevel set $d_{\mu', m_0}^{-1}((-\infty, \eta])$ 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

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 is estimated using $x \mapsto \frac{m_0}{\text{vol}_d(B(x, \delta_{\mu}, m_0(x)))}$, $m_0 = 150/1200$ (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 \(d_{\mu,m_0}\) (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 $d_{\mu,m_0}$ (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 $d_{\mu, m_0}$

- 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 $d_{\mu,m_0}$
Summary

- $\mu \mapsto d_{\mu,m_0}$ provide a way to associate geometry to a measure in Euclidean space.
- $d_{\mu,m_0}$ is robust to Wasserstein perturbations: outliers and noise are easily handled (no assumption on the nature of the noise).
- $d_{\mu,m_0}$ 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 $d_{\mu,m_0}$, stable notion of persistence diagram for $\mu$.
- Algorithm: for finite point clouds $d_{\mu,m_0}$ and $\nabla (d_{\mu,m_0})$ can be easily and efficiently computed pointwise in any dimension.