DISTANCE TO A MEASURE

Contents

It is well known that distance-based methods in TDA may fail completely in the presence of outliers. Indeed, adding even a single outlier to the point cloud can change the distance function d_K dramatically (see Figure 1). To overcome this issue, [CCSM11] introduced an alternative distance function which is robust to noise, the *distance-to-a-measure* (DTM).

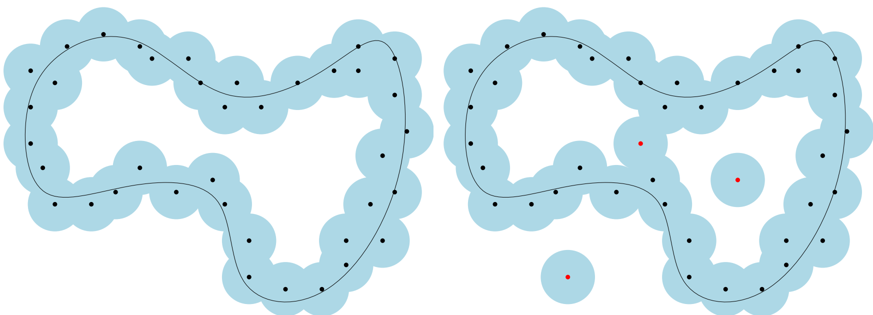


Figure 1. The effect of outliers on the sublevel sets of distance functions. Adding just a few outliers to a point cloud may dramatically change its distance function and the topology of its offsets.

The DTM satisfies properties that are very similar to those of a distance function, yielding nearly identical inference results. In other words, it appears that we can write roughly the same reconstruction result as for offsets of a compact (see previous lesson) for the sublevels sets of the DTM. Such functions are called *distance-like*. Let us first present these functions and their properties in their full generality.

1. Distance-Like Functions

All the inference results of the *Reconstruction of Compact Sets* lesson follow from only three fundamental properties of distance functions:

(i) Stability of the map $K \to \mathrm{d}_K$: for any compact subsets K, K' of $\mathbb{R}^d$ we have

$$\|\mathrm{d}_K - \mathrm{d}_{K'}\|_\infty = \mathrm{d}_\mathrm{H}(K, K').$$

(ii) For any compact set K of $\mathbb{R}^d$, the distance function d_K is 1-Lipschitz: for all $x, x' \in \mathbb{R}^d$, $|\mathrm{d}_K(x) - \mathrm{d}_{K'}(x)| \leqslant \|x - x'\|$.

(iii) For any compact set K of $\mathbb{R}^d$, the distance function d_K^2 is 1-semiconcave: $x \mapsto \|x\|^2 - \mathrm{d}_K^2(x)$ is convex.

The first property is an obvious condition to ensure that the offsets of two close compact sets are close to each other. The second and third properties are the fundamental ingredients to prove the existence and integrability of the gradient of d_K and the isotopy lemma. These results still hold for general proper semiconcave functions, motivating the following definition of functions that are of particular interest for geometric inference.

Definition 1.1 (Distance-Like Function)**.** A non-negative function $\phi : \mathbb{R}^d \to \mathbb{R}_{\geqslant 0}$ is *distance-like* if

(i) ϕ is 1-Lipschitz;
(ii) ϕ^2 is 1-semiconcave, i.e. $x \mapsto \|x\|^2 - \phi^2(x)$ is convex.
(iii) ϕ is proper, i.e. for all compact set $K \subset \mathbb{R}$, $\phi^{-1}(K)$ is compact.

Remark 1.2 (Why *Distance-Like*?)**.** One can show that if ϕ^2 is 1-semiconcave, there exists a closed subset K of $\mathbb{R}^{d+1}$ such that $\phi(x) = \mathrm{d}_K(x)$ for all $x \in \mathbb{R}^d$, where $x \in \mathbb{R}^d$ is identified with $(x, 0)$ in $\mathbb{R}^{d+1}$.

This remark also shows that ϕ^2 being 1-semiconcave and proper yields automatically ϕ distance-like: the Lipschitz property comes from 1-semiconcavity for free.

Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be distance-like. The 1-semiconcavity of ϕ^2 allows to define a notion of gradient vector field $\nabla\phi(x)$ for ϕ, defined everywhere and satisfying $\|\nabla\phi(x)\| \leqslant 1$.

Although not continuous, the vector field $\nabla\phi$ is sufficiently regular to be integrated in a continuous locally Lipschitz flow $\Phi(\cdot, t) : \mathbb{R}^d \to \mathbb{R}^d$ for $t \geqslant 0$. The flow $\Phi(\cdot, t)$ integrates the gradient $\nabla\phi$ in the sense that for all $x \in \mathbb{R}^d$, the curve $\gamma : t \mapsto \Phi(x, t)$ is right-differentiable, and for all $t > 0$, $\gamma'(t^-) = \nabla\phi(\gamma(t))$. Moreover, for all integral curve $\gamma : [a, b] \to \mathbb{R}^d$ parametrized by arc-length,

$$\phi(\gamma(b)) = \phi(\gamma(a)) + \int_a^b \|\nabla\phi(\gamma(t))\| \,\mathrm{d}t.$$

Remark 1.3. (i) We denote by $\phi^r = \phi^{-1}([0, r])$ the r-sublevel set of ϕ.

(ii) A point $x \in \mathbb{R}^d$ will be called α-*critical*, for $0 \leqslant \alpha \leqslant 1$, if for all $h \in \mathbb{R}^d$,

$$\phi^2(x+h) \leqslant \phi^2(x) + 2\alpha \|h\| \, \phi(x) + \|h\|^2 .$$

A 0-critical point is just called a *critical* point. It follows from the 1-semiconcavity of ϕ^2 that $\|\nabla\phi(x)\|$ is the infimum of the non-negative α's such that x is α-critical.

(iii) The *weak feature size* of ϕ at r, denoted by $\mathrm{wfs}_\phi(r)$, is the maximum $r' > 0$ such that ϕ does not have any critial value between r and $r + r'$. For all $0 < \alpha < 1$, the α-*reach* of ϕ is the maximum r such that $\phi^r = \phi^{-1}([0, r])$ does not contain any α-critical point. Notice that the α-reach is always a lower bound for the weakfeature size, with $r = 0$.

The Isotopy lemma extends to distance-like functions.

Theorem 1.4 (Extended Isotopy Lemma). *Let ϕ be a distance-like function and $r_1 < r_2$ be two positive numbers such that ϕ has no critical points in $\phi^{-1}([r_1, r_2])$. Then all the sublevel sets $\phi^r = \phi^{-1}([0, r])$ are isotopic for $r \in [r_1, r_2]$.*

The proof of the following theorem, showing that the offsets of two uniformly close distance-like functions with large weak feature size have the same homotopy type, relies on Theorem 1.4 and is almost verbatim the same as the one for d_K (see the *Reconstruction of Compact Sets* lesson).

Proposition 1.5. *Let ϕ and ψ be two distance-like functions, such that $\|\phi - \psi\|_\infty \leqslant \varepsilon$. Suppose moreover that $\mathrm{wfs}_\phi(r) > 2\varepsilon$ and $\mathrm{wfs}_\psi(r) > 2\varepsilon$. Then, for every $0 < \eta \leqslant 2\varepsilon$, $\phi^{r+\eta}$ and $\psi^{r+\eta}$ have the same homotopy type.*

The critical point stability theorem also holds for general distance-like functions.

Theorem 1.6. *Let ϕ and ψ be two distance-like functions with $\|\phi - \psi\|_\infty \leqslant \varepsilon$. For any α-critical point x of ϕ, there exists a α'-critical point x' of ψ with $\|x - x'\| \leqslant 2\sqrt{\varepsilon\phi(x)}$ and $\alpha' \leqslant \alpha + 2\sqrt{\varepsilon/\phi(x)}$.*

Proof. Almost verbatim the same as the proof in the *Reconstruction of Compact Sets* lesson. □

Corollary 1.7. *Let ϕ and ψ be two ε-close distance-like functions, and suppose that $\mathrm{reach}_\alpha(\phi) \geqslant R$ for some $\alpha > 0$. Then ψ has no critical value in the interval $(4\varepsilon/\alpha^2, R - 3\varepsilon)$.*

Proof. Almost verbatim the same as the proof in the *Reconstruction of Compact Sets* lesson. □

Theorem 1.8 (Extended Reconstruction Theorem). *Let ϕ and ψ be two ε-close distance-like functions, and suppose that $\mathrm{reach}_\alpha(\phi) \geqslant R$ for some $\alpha > 0$. Then for all $r \in [4\varepsilon/\alpha^2, R - 3\varepsilon]$ and $0 < \eta < R$, the sublevel sets ψ^r and ϕ^η are homotopy equivalent, as soon as*

$$\varepsilon \leqslant \frac{R}{5 + 4/\alpha^2}.$$

Proof. By the extended isotopy Theorem 1.4, all the sublevel sets ψ^r have the same homotopy type for r in the given range. Let us choose $r = 4\varepsilon/\alpha^2$. We have

$$\mathrm{wfs}_\phi(r) \geqslant R - 4\varepsilon/\alpha^2 \text{ and } \mathrm{wfs}_\psi(r) \geqslant R - 3\varepsilon - 4\varepsilon/\alpha^2.$$

By Proposition 1.5, the sublevel sets ϕ^r and ψ^r have the same homotopy type as soon as the uniform distance ε between ϕ and ψ is smaller than $\mathrm{wfs}_\phi(r)/2$ and $\mathrm{wfs}_\psi(r)/2$. This is true provided that $2\varepsilon \leqslant R - \varepsilon(3 + 4\alpha^2)$, which yields the result. □

Remark 1.9. The notion of α-reach could be made dependent on a parameter r, i.e. the (r, α)-reach of ϕ could be defined as the maximum r_0 such that the set $\phi^{-1}([r, r + r_0])$ does not contain any α-critical value. A reconstruction theorem similar to Theorem 1.8 would still hold under the weaker condition that the (r, α)-reach of ϕ is positive.

2. Wasserstein Distance

There is a whole family of Wasserstein distances W_p, $1 \leqslant p \leqslant \infty$, between probability measures in $\mathbb{R}^d$. Their definition relies on the notion of transport plan between measures. Although some of the results of this chapter can be stated for any distance W_p, for technical reasons that will become clear below, we only consider the W_2 distance.

Definition 2.1 (Transport Plan, Cost)**.** A *transport plan* between two probability measures μ and ν on $\mathbb{R}^d$ is a probability measure π on $\mathbb{R}^d \times \mathbb{R}^d$ such that for every Borel sets $A, B \subset \mathbb{R}^d$,

$$\pi(A \times \mathbb{R}^d) = \mu(A) \text{ and } \pi(\mathbb{R}^d \times B) = \nu(B).$$

The *cost* of such a transport plan π is given by

$$\mathcal{C}(\pi) = \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \|y - x\|^2 \,\mathrm{d}\pi(x, y)\right)^{1/2}.$$

Example 2.2. Consider two probability measures with finite supports

$$\mu = \sum_{i=1}^{n} c_i \delta_{x_i} \text{ and } \nu = \sum_{j=1}^{m} d_j \delta_{y_j}$$

with $\sum_{i=1}^n c_i = \sum_{j=1}^m d_j = 1$. A transport plan between μ and ν can be represented by a $n \times m$ matrix $\pi = (\pi_{i,j})_{i,j}$ with nonnegative entries such that

$$\sum_{j=1}^{m} \pi_{i,j} = c_i \text{ and } \sum_{i=1}^{n} \pi_{i,j} = d_j.$$

The coefficient $\pi_{i,j}$ can be seen as the amount of the mass of μ located at x_i that is transported to y_j. The cost of such a transport is then given by

$$\mathcal{C}(\pi) = \left(\sum_{i=1}^{n} \sum_{j=1}^{m} \|y_j - x_i\|^2\right)^{1/2}.$$

Definition 2.3 (Wasserstein Distance of Order 2)**.** Given two probability measures μ, ν with finite second order moments, $\mathrm{W}_2(\mu, \nu)$ is the minimum cost $\mathcal{C}(\pi)$ of a transport plan π between μ and ν.

Remark 2.4. (i) The space of probability measures with finite moment of order 2 endowed with W_2 is a metric space.
(ii) W_2 provides an interesting notion to quantify the resilience to outliers. To illustrate this, consider a set $\mathcal{P} = \{x_1, x_2, \dots, x_N\}$ of N points in $\mathbb{R}^d$ and a noisy version $\mathcal{P}'$ obtained by replacing the first n points in $\mathcal{P}$ by points y_i such that $\mathrm{d}_{\mathcal{P}}(y) = R > 0$, for $1 \leqslant i \leqslant n$. If we denote by $\mu = \frac{1}{N}\sum_{p\in\mathcal{P}} \delta_p$ and $\nu = \frac{1}{N}\sum_{q\in\mathcal{P}'} \delta_q$ the uniform measures on $\mathcal{P}$ and $\mathcal{P}'$ respectively, then

$$\mathrm{W}_2(\mu, \nu) \leqslant \sqrt{\frac{n}{N}}(R + \mathrm{diam}(\mathcal{P})),$$

while the Hausdorff distance between $\mathcal{P}$ and $\mathcal{P}'$ is at least R. To prove this inequality, consider the transport plan π from ν to μ that moves the outliers back to their original position and leave the other points fixed.

3. Distance to a Measure

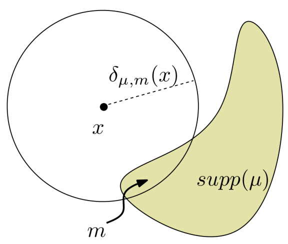
In this section, we associate, to any probability measure in $\mathbb{R}^d$, a family of real-valued functions that are both distance-like and robust with respect to perturbations of the probability measure.

3.1. **Definition.** The distance function to a compact set K evaluated at $x \in \mathbb{R}^d$ is the smallest radius r such that $\mathrm{B}(x, r)$ contains at least a point of K. A natural idea to adapt this definition when K is replaced by a measure μ is to consider the smallest radius r such that $\mathrm{B}(x, r)$ contains a given fraction m of the total mass of μ.

Definition 3.1. Let μ be a Borel probability distribution on $\mathbb{R}^d$ and $0 \leqslant m < 1$ a given parameter. We denote by $\delta_{\mu,m} : \mathbb{R}^d \to \mathbb{R}_{\geqslant 0}$ the function

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

where $\mathrm{B}(x, r)$ denotes the closed ball of radius r centered at x.



Remark 3.2. (i) For $m = 0$, the definition coincides with the (usual) distance function to the support of μ.
(ii) For all $0 \leqslant m < 1$, $\delta_{\mu,m}$ is 1-Lipschitz.

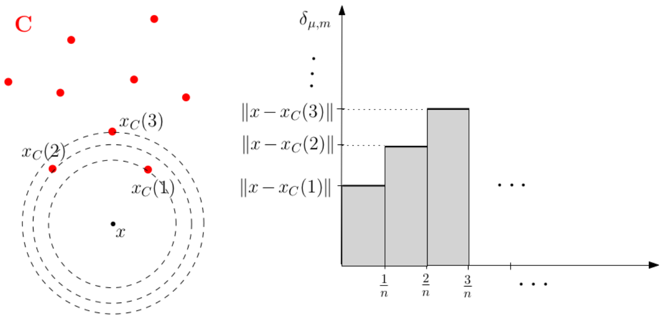


FIGURE 2. Computation of the distance to the empirical measure associated to a point set C. $x_C(k)$ denotes the kth nearest neighbor of x in C.

(iii) Unfortunately, the function $\delta_{\mu,m}$ is not robust to perturbations of the measure μ. More precisely, $\mu \mapsto \delta_{\mu,m}$ is not continuous, as showed by the following example. Let $\mu_\varepsilon : \left(\frac{1}{2} - \varepsilon\right)\delta_0 + \left(\frac{1}{2} + \varepsilon\right)\delta_1$ be the weighted sum of Dirac measures at 0 and 1 in $\mathbb{R}$, and $m = 1/2$. Then, if $t < 0$,
– for $\varepsilon \geqslant 0$, $\delta_{\mu_\varepsilon,1/2}(t) = |1 - t|$;
– for $\varepsilon < 0$, $\delta_{\mu_\varepsilon,1/2}(t) = |t|$.
This means that $\varepsilon \mapsto \delta_{\mu_\varepsilon,1/2}$ is not continuous at $\varepsilon = 0$. To overcome this issue we define the distance function associated to μ as a L^2 average of the pseudo-distances $\delta_{\mu,m}$ for a range $[0, m_0]$ of parameters m.

Definition 3.3 (Distance-to-Measure)**.** Let μ be a Borel probability distribution on $\mathbb{R}^d$ and $0 < m_0 \leqslant 1$ be a mass parameter. We denote by $\mathrm{d}_{\mu,m_0} : \mathbb{R}^d \to \mathbb{R}_{\geqslant 0}$ the function

$$\mathrm{d}^2_{\mu,m_0}(x) = \frac{1}{m_0}\int_0^{m_0} \delta^2_{\mu,m}(x)\mathrm{d}m.$$

3.2. **Distance to Empirical Measures.** An interesting property of the above defined functions is that they have a simple expression in terms of nearest neighbors. More precisely, let C be a point cloud with n points in $\mathbb{R}^d$ and $\mu_C = \frac{1}{n}\sum_{p\in C}\delta_p$ be the uniform distribution on C.

For $0 < m \leqslant 1$, the function $\delta_{\mu_C,m}$ evaluated at a given point $x \in \mathbb{R}^d$ is by definition equal to the distance between x and its k-th nearest neighbor in C, where k is the smallest integer larger than mn. Hence the function $m \mapsto \delta_{\mu_C,m}(x)$ is constant and equal to the distance from x to its kth nearest neighbor in C on each interval $[\frac{k-1}{n}, \frac{k}{n})$. Integrating the square of this piecewise constant functions gives the following expression for $\mathrm{d}^2_{\mu_C,m_0}$,

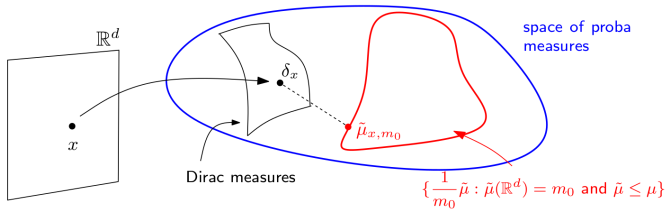


FIGURE 3. The distance function to a measure as a usual distance function in an infinite dimensional space.

where $m_0 = k_0/n$:

$$\begin{aligned} \mathrm{d}^2_{\mu_C,m_0}(x) &= \frac{1}{m_0}\int_0^{m_0} \delta^2_{\mu_C,m}(x)\mathrm{d}m \\ &= \frac{1}{m_0}\sum_{k=1}^{k_0}\frac{1}{n}\delta^2_{\mu_C,k/n}(x) \\ &= \frac{1}{k_0}\sum_{p\in \mathbf{NN}_C^{k_0}(x)} \|p-x\|^2, \end{aligned}$$

where $\mathbf{NN}_C^{k_0}(x)$ denotes the first k_0th nearest neighbors of x in C. As a consequence the pointwise evaluation of $\mathrm{d}_{\mu_C,k_0/n}(x)$ reduces to a k_0-nearest neighbor query in C.

3.3. **Equivalent Formulation.** In this paragraph, we provide another characterization of the distance function to a measure d_{μ,m_0} showing that it is in fact the distance function to a closed set, but in the non Euclidean space of probability measures endowed with the W_2 metric (see Figure 3). This equivalent formulation will be used to deduce that $\mu \mapsto \mathrm{d}_{\mu,m_0}$ is Lipschitz and $x \mapsto \mathrm{d}_{\mu,m_0}(x)$ is semiconcave.

Definition 3.4 (Submeasure)**.** A measure ν is a *submeasure* of another measure μ if for every Borel set B of $\mathbb{R}^d$, $\nu(B) \leqslant \mu(B)$.

Remark 3.5. (i) The set of all submeasures of a given measure μ is denoted by $\mathrm{Sub}(\mu)$.
(ii) The set of submeasures of μ with a prescribed total mass $m_0 > 0$ is denoted by $\mathrm{Sub}_{m_0}(\mu)$.

PROPOSITION 3.6. *Let μ be a Borel probability distribution on $\mathbb{R}^d$ and $0 < m_0 \leqslant 1$ be a mass parameter. Then for all $x \in \mathbb{R}^d$,*

$$\mathrm{d}_{\mu,m_0}(x) = \min_{\nu\in \mathrm{Sub}_{m_0}(\mu)} \frac{1}{\sqrt{m_0}} \mathrm{W}_2(m_0\delta_x, \nu).$$

Moreover, for any measure μ_{x,m_0} that realizes the above minimum,

$$\mathrm{d}_{\mu,m_0}(x) = \left(\frac{1}{m_0}\|x-h\|^2 \mathrm{d}\mu_{x,m_0}(m)\right)^{1/2}.$$

Remark 3.7. (i) Said otherwise, $\mathrm{d}_{\mu,m_0}(x)$ is the minimal Wasserstein distance between the Dirac mass $m_0\delta_x$ and the set of submeasures of μ with total mass m_0.

(ii) The set $\mathcal{R}_{\mu,m_0}(x)$ of submeasures minimizing the above expression corresponds to the nearest neighbors of the Dirac measure $m_0\delta_x$ on the set of submeasures $\mathrm{Sub}_{m_0}(\mu)$. It is not empty but it might not be reduced to a single element. Indeed, it coincides with the set of submeasures μ_{x,m_0} of total mass m_0 whose support is contained in the closed ball $\mathrm{B}(x, \delta_{\mu,m_0}(x))$, and whose restriction to the open ball $\mathring{\mathrm{B}}(x, \delta_{\mu,m_0}(x))$ coincides with μ.

3.4. **Stability of the Distance to a Measure.** The characterization of d_{μ,m_0} given in Proposition 3.6 provides a pretty simple way to prove the stability of $\mu \mapsto \mathrm{d}_{\mu,m_0}$.

Theorem 3.8 (Stability of the DTM). *Let μ, μ' be two Borel probability distributions on $\mathbb{R}^d$ and $m_0 > 0$. Then,*

$$\left\|\mathrm{d}_{\mu,m_0} - \mathrm{d}_{\mu',m_0}\right\|_\infty \leqslant \frac{1}{\sqrt{m_0}} \mathrm{W}_2(\mu, \mu').$$

The proof of Theorem 3.8 follows from the following proposition.

Proposition 3.9. *Let μ, μ' be two Borel probability distributions on $\mathbb{R}^d$ and $m_0 > 0$. Then,*

$$\mathrm{d}_\mathrm{H}\left(\mathrm{Sub}_{m_0}(\mu), \mathrm{Sub}_{m_0}(\mu')\right) \leqslant \mathrm{W}_2(\mu, \mu'),$$

where d_H stands for the Hausdorff distance in the space of probability measures endowed with the W_2 metric.

Sketch of proof. Let $\varepsilon = \mathrm{W}_2(\mu, \mu')$ and π be a corresponding optimal transport plan, i.e. a transport plan between μ and μ' that achieves $\mathcal{C}(\pi) = \varepsilon$. Given a submeasure ν of μ, one can find a submeasure π' of π that transports ν to a submeasure ν' of μ' (notice that this latter claim is not completely obvious and its formal proof is beyond the scope of this lesson. It can be proven using the Radon-Nykodim's theorem). Then,

$$\mathrm{W}_2(\nu, \nu')^2 \leqslant \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \,\mathrm{d}\pi'(x, y) \leqslant \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \,\mathrm{d}\pi(x, y) = \varepsilon^2.$$

This shows that $\mathrm{d}_{\mathrm{Sub}_{m_0}(\mu')}(\nu) \leqslant \varepsilon$ for every submeasure $\nu \in \mathrm{Sub}_{m_0}(\mu)$. The same holds exchanging the roles of μ and μ', thus proving the bound on the Hausdorff distance. □

Proof of Theorem 3.8. The following sequence of equalities and inequalities, that follows from Propositions 3.6 and 3.9, proves the result:

$$\begin{aligned}
\mathrm{d}_{\mu,m_0}(x) &= \mathrm{d}_{\mathrm{Sub}_{m_0}(\mu)}(m_0\delta_x) \\
&\leqslant \frac{1}{\sqrt{m_0}}\Big(\mathrm{d}_\mathrm{H}\left(\mathrm{Sub}_{m_0}(\mu), \mathrm{Sub}_{m_0}(\mu')\right) + \mathrm{d}_{\mathrm{Sub}_{m_0}(\mu')}(m_0\delta_x)\Big) \\
&\leqslant \frac{1}{\sqrt{m_0}} \mathrm{W}_2(\mu, \mu') + \mathrm{d}_{\mu',m_0}(x).
\end{aligned}$$

□

3.5. **The Distance to a Measure is a Distance-Like Function.** Recall that the subdifferential of a function $f : \Omega \subset \mathbb{R}^d \to \mathbb{R}$ at a point x, denoted by $\partial_x f$, is the set of vectors $v \in \mathbb{R}^d$ such that for all small enough vector h, $f(x+h) \geqslant f(x) + \langle h, v\rangle$. This gives a characterization of convexity: a function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if and only if its subdifferential $\partial_x f$ is non-empty for every point x. If this is the case, then f is differentiable at x if and only if its subdifferential $\partial_x f$ is a singleton, in which case the gradient $\nabla f(x)$ coincides with its unique element.

PROPOSITION 3.10. *The function* $v_{\mu,m_0} : \mathbb{R}^d \ni x \mapsto \|x\|^2 - \mathrm{d}^2_{\mu,m_0}(x)$ *is convex, and its subdifferential at* $x \in \mathbb{R}^d$ *is given by*

$$\partial_x v_{\mu,m_0} = \left\{ 2x - \frac{2}{m_0}\int_{\mathbb{R}^d}(x-h)\mathrm{d}\mu_{x,m_0}(h) \,\middle|\, \mu_{x,m_0} \in \mathcal{R}_{\mu,m_0}(x) \right\}.$$

Proof. For any two points x and y of $\mathbb{R}^d$, let μ_{x,m_0} and μ_{y,m_0} be measures belonging to $\mathcal{R}_{\mu,m_0}(x)$ and $\mathcal{R}_{\mu,m_0}(y)$ respectively. Thanks to Proposition 3.6, we have the following sequence of equalities and inequalities:

$$\begin{aligned}
\mathrm{d}^2_{\mu,m_0}(y) &= \frac{1}{m_0}\int_{\mathbb{R}^d} \|y-h\|^2 \,\mathrm{d}\mu_{y,m_0}(h) \\
&\leqslant \frac{1}{m_0}\int_{\mathbb{R}^d} \|y-h\|^2 \,\mathrm{d}\mu_{x,m_0}(h) \\
&\leqslant \frac{1}{m_0}\int_{\mathbb{R}^d} \Big(\|x-h\|^2 + 2\langle x-h, y-x\rangle + \|y-x\|^2\Big)\,\mathrm{d}\mu_{x,m_0}(h) \\
&\leqslant \mathrm{d}^2_{\mu,m_0}(x) + \langle v, y-x\rangle + \|y-x\|^2,
\end{aligned}$$

where v is the vector defined by

$$v = \frac{2}{m_0}\int_{\mathbb{R}^d}(x-h)\mathrm{d}\mu_{x,m_0}(h).$$

The inequality can be rewritten as

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

which shows that the vector $(2x-v)$ belongs to the subdifferential of v at x. By the characterization of convexity that we recalled above, we get that v_{μ,m_0} is convex. The proof of the reverse inclusion is slightly more technical and beyond the scope of the lesson. □

COROLLARY 3.11. *(i) The function* d^2_{μ,m_0} *is* 1*-semiconcave.*
(ii) d^2_{μ,m_0} *is differentiable almost everywhere in* $\mathbb{R}^d$*, with gradient*

$$\nabla_x \mathrm{d}^2_{\mu,m_0} = \frac{2}{m_0}\int_{\mathbb{R}^d}(x-h)\mathrm{d}\mu_{x,m_0}(h),$$

where μ_{x,m_0} *is the only measure in* $\mathcal{R}_{\mu,m_0}(x)$*.*
(iii) The function $\mathbb{R}^d \ni x \mapsto \mathrm{d}_{\mu,m_0}(x)$ *is* 1*-Lipschitz.*

Proof. (i) Already proved.
(ii) Follows from the fact that a convex function is differentiable almost everywhere, with gradient given by the only element of the sub-differential at the considered points.

(iii) The gradient of d_{μ,m_0} can be written as

$$\nabla_x d_{\mu,m_0} = \frac{\nabla_x d^2_{\mu,m_0}}{2 d_{\mu,m_0}} = \frac{1}{\sqrt{m_0}} \frac{\int_{\mathbb{R}^d} (x-h) d\mu_{x,m_0}(h)}{\left(\int_{\mathbb{R}^d} \|x-h\|^2 d\mu_{x,m_0}(h)\right)^{1/2}}.$$

Using the Cauchy-Schwarz inequality we find $\|\nabla_x d_{\mu,m_0}\| \leqslant 1$, which proves the statement.

□

4. Application to Geometric Inference

Reconstruction from point clouds in presence of outliers was the main motivation for introducing the distance to a measure. In this section, we show how the extended reconstruction Theorem 1.8 can be applied to distance to measure functions. It is also possible to adapt most of the topological and geometric inference results of the *Reconstruction of Compact Sets* lesson in a similar way.

4.1. **Distance to a Measure vs. Distance to its Support.** In this section, we compare the DTM d_{μ,m_0} of a measure μ and the distance function d_S to its support $S = \text{supp}(\mu)$, and study the convergence properties of d_{μ,m_0} to d_S as the mass parameter m_0 goes to zero.

Note that the function δ_{μ,m_0} (and hence the DTM d_{μ,m_0}) is always larger than the distance function d_S, i.e. for all $x \in \mathbb{R}^d, d_S(x) \leqslant d_{\mu,m_0}(x)$. As a consequence, to obtain a convergence result of d_{μ,m_0} towards d_S as m_0 goes to zero, we just need to upper bound $d_{\mu,m_0} - d_S$ by a function converging to 0 as m_0 goes to 0. It turns out that the convergence rate of d_{μ,m_0} towards d_S depends on the way that μ charges the balls $B(p,r)$ centered at points $p \in S$, as r decreases. For this, we need to define a few notions:

(i) We say that a non-decreasing positive function $f : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ is a *uniform lower bound on the growth of* μ if for every point $p \in S$ and every $\varepsilon > 0$, $\mu(B(p,\varepsilon)) \geqslant f(\varepsilon)$.
(ii) The measure μ is said to have *dimension at most* k if there is a constant $C_\mu > 0$ such that $f(\varepsilon) = C_\mu \varepsilon^k$ is a uniform lower bound on the growth of μ, for ε small enough.

Lemma 4.1. *Let μ be a Borel probability distribution and f be a uniform lower bound on the growth of μ. Then $\|d_{\mu,m_0} - d_S\|_\infty < \varepsilon$ as soon as $m_0 < f(\varepsilon)$.*

Proof. Assume that $m_0 < f(\varepsilon)$ and let $x \in \mathbb{R}^d$ be fixed. Write $p \in S$ for any point such that $\|x-p\| = d_S(x)$. By assumption, $\mu(B(x, d_S + \varepsilon)) \geqslant \mu(B(p,\varepsilon)) \geqslant m_0$. Hence, $\delta_{\mu,m_0}(x) \leqslant d_S(x) + \varepsilon$. The map $m \mapsto \delta_{\mu,m}(x)$ being non-decreasing, we get

$$m_0 d_S^2(x) \leqslant \int_0^{m_0} \delta^2_{\mu,m}(x) dm \leqslant m_0 (d_S(x) + \varepsilon)^2.$$

Taking the square root of this expression proves the lemma. □

Corollary 4.2. *(i) If the support S of μ is compact, then d_S is the uniform limit of d_{μ,m_0} as m_0 goes to 0:*

$$\|\mathrm{d}_{\mu,m_0} - \mathrm{d}_S\|_\infty \xrightarrow[m_0\to 0]{} 0.$$

(ii) If μ has dimension at most $k > 0$, then

$$\|\mathrm{d}_{\mu,m_0} - \mathrm{d}_S\|_\infty \leqslant C_\mu^{-1/k} m_0^{1/k}.$$

Proof. (i) If S is compact, there exists a sequence $(x_i)_{i\geqslant 0}$ of points in S such that for all $\varepsilon > 0$, $S \subset \cup_{i=1}^n \mathrm{B}(x_i, \varepsilon/2)$ for some $n = n(\varepsilon)$. By definition of the support of a measure, $\eta(\varepsilon) = \ min_{1\leqslant i\leqslant n}\mu(\mathrm{B}(x_i, \varepsilon/2)) > 0$. Now, for all $x \in S$, there is a x_i such that $\|x - x_i\| \leqslant \varepsilon/2$. Hence, $\mathrm{B}(x_i, \varepsilon/2) \subset \mathrm{B}(x, \varepsilon)$, which means that $\mu(\mathrm{B}(x, \varepsilon)) \geqslant \eta(\varepsilon)$.
(ii) Follows straightforwardly from Lemma 4.1.

□

4.2. **Shape Reconstruction from Noisy Data.** The previous results lead to shape reconstruction theorems that work for noisy data with outliers. To fit in our framework we consider shapes that are defined as supports of probability measures. Let μ be a probability measure of dimension at most $k > 0$ with compact support $K \subset \mathbb{R}^d$ and let $\mathrm{d}_K : \mathbb{R}^d \to \mathbb{R}$ be the (Euclidean) distance function to K. If μ' is another probability measure (e.g. the empirical measure given by a point cloud sampled according to μ), one has

$$\begin{aligned}\left\|\mathrm{d}_K - \mathrm{d}_{\mu',m_0}\right\|_\infty &\leqslant \|\mathrm{d}_K - \mathrm{d}_{\mu,m_0}\|_\infty + \left\|\mathrm{d}_{\mu,m_0} - \mathrm{d}_{\mu',m_0}\right\|_\infty \\ &\leqslant \|\mathrm{d}_K - \mathrm{d}_{\mu,m_0}\|_\infty + \frac{1}{\sqrt{m_0}}\,\mathrm{W}_2(\mu, \mu').\end{aligned}$$

As expected, the choice of m_0 is a trade-off:
– small m_0 leads to better approximation of the distance function to the support, while
– large values of m_0 make the distance functions to measures more stable.

The previous bound together with Theorem 1.8 yield the following result.

Corollary 4.3. *Let μ be a Borel probability measure and $K = \mathrm{supp}(\mu)$ its support. Assume that μ has dimension at most k and that $\mathrm{reach}_\alpha(K) \geqslant R$, for some $R > 0$. Let μ' be another measure, and $\varepsilon \geqslant \left\|\mathrm{d}_K - \mathrm{d}_{\mu',m_0}\right\|_\infty$.*

Then for all $r \in [4\varepsilon/\alpha^2, R - 3\varepsilon]$, the r-sublevel set of d_{μ,m_0} and the offsets K^η, for $0 < \eta < R$, are homotopy equivalent as soon as

$$\mathrm{W}_2(\mu, \mu') \leqslant \frac{R\sqrt{m_0}}{5 + 4/\alpha^2} - C_\mu^{-1/k} m_0^{1/k+1/2}.$$

Figure 4 illustrates Corollary 4.3 on a sampled mechanical part with 10% outliers. In this case, μ' is the normalized sum of the Dirac measures centered at the data points and the (unknown) measure μ is the uniform measure on the mechanical part.

5. Further Sources

These notes mainly follow [BCY18] and [CCSM11] .

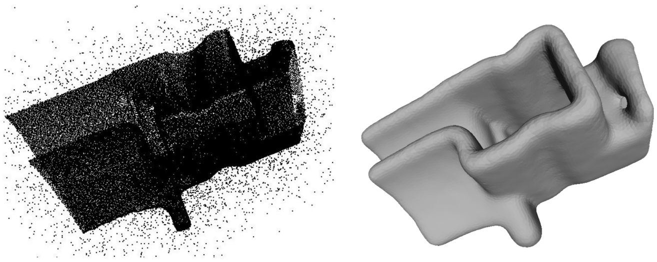


FIGURE 4. On the left, a point cloud C sampled on a mechanical part to which 10% outliers uniformly sampled in a box enclosing the model have been added. On the right, the reconstruction of an isosurface of the distance function d_{μ_C,m_0} to the uniform probability measure on this point cloud.

References

[BCY18] Jean-Daniel Boissonnat, Frédéric Chazal, and Mariette Yvinec. *Geometric and Topological Inference*. Cambridge University Press, 2018. Cambridge Texts in Applied Mathematics.

[CCSM11] Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Geometric inference for probability measures. *Found. Comput. Math.*, 11(6):733–751, 2011.