Geometric Inference for Measures based on Distance Functions

Frédéric Chazal, David Cohen-Steiner, Quentin Mérigot

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Geometric Inference for Measures based on Distance Functions

Frédéric Chazal — David Cohen-Steiner — Quentin Mérigot

N° 6930 — version Second version

initial version June 22 2010 — revised version Juin 2010

Thème SYM

ISSN 0249-6399 ISRN INRIA/RR--6930--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
SOPHIA ANTIPOLIS - MÉDITERRANÉE

Geometric Inference for Measures based on Distance Functions

Frédéric Chazal , David Cohen-Steiner , Quentin Mérigot

Thème SYM — Systèmes symboliques
Équipe-Projet Geometrica

Rapport de recherche n° 6930 — version Second version — initial version June 22 2010 — revised version Juin 2010 — 25 pages

Abstract: Data often comes in the form of a point cloud sampled from an unknown compact subset of Euclidean space. The general goal of geometric inference is then to recover geometric and topological features (e.g. Betti numbers, normals) of this subset from the approximating point cloud data. In recent years, it appeared that the study of distance functions allows to address many of these questions successfully. However, one of the main limitations of this framework is that it does not cope well with outliers nor with background noise. In this paper, we show how to extend the framework of distance functions to overcome this problem. Replacing compact subsets by measures, we introduce a notion of distance function to a probability distribution in $\mathbb{R}^n$. These functions share many properties with classical distance functions, which makes them suitable for inference purposes. In particular, by considering appropriate level sets of these distance functions, it is possible to associate in a robust way topological and geometric features to a probability measure. Moreover, in settings where empirical measures are considered these functions can be easily evaluated, making them of particular practical interest.

Key-words: density estimation, reconstruction, Wasserstein distance, Mean-Shift

Centre de recherche INRIA Sophia Antipolis – Méditerranée
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Inférence géométrique pour les mesures en utilisant la fonction distance

Résumé : De nombreuses données sont souvent représentées sous forme de nuages de points échantillonnés dans des espaces Euclidiens au voisinage de sous-ensembles compacts. L'objectif général de l'inférence géométrique est de retrouver les caractéristiques topologiques et géométriques (par ex. nombres de Betti, normales) de ces sous-ensembles à partir des données. Ces dernières années, l'études des fonctions distance a permis d'aborder avec succès bon nombre de problèmes d'inférence géométrique. Cependant, une des principales limitation de ce cadre est qu'il ne permet pas de considérer des données qui sont entachées de valeurs aberrantes et/ou d'un bruit de fond. Dans cet article, nous montrons comment étendre le cadre des fonctions distance pour résoudre ce problème. En remplaçant les sous-ensembles compacts par des mesures, nous introduisons une notion de fonction distance à une probabilité dans $\mathbb{R}^n$. Ces fonctions partagent de nombreuses propriétés avec les fonctions distance classiques qui les rendent utiles pour l'inférence géométrique. En particulier, en considérant des niveaux appropriés de ces fonctions, il est possible d'associer de façon robuste des caractéristiques topologiques et géométriques à des mesures de probabilité.

Mots-clés : estimation de densité, reconstruction, distance de Wasserstein, Mean-Shift

1 Introduction

Extracting geometric and topological information from geometric data, such as 3D point clouds obtained from laser scanners, is a requirement for many geometry processing and data analysis algorithms. The need for robust estimation of geometric invariants have been recognized long time ago in geometry processing, and such invariants have found applications in fields as different as shape matching, registration, symmetry detection in 3D models or more generally structure discovery, reconstruction, meshing to name just a few. More recently, it became apparent that such geometric and topological quantities could also be used to analyze more general data sets coming from computational structural biology, large image databases, etc. It turns out that many questions in data analysis can be naturally stated as inferring the geometry of an unknown underlying geometric object. For example, the number of clusters in which a point cloud can be split is related to the number of connected components of this unknown object. Similarly, finding out the number of parameters really needed to faithfully describe a point in the cloud – which is usually much smaller than the dimension of the ambient space – is a matter of estimating the dimension of the underlying set.

1.1 Inference using offsets and distance functions

One approach to geometric inference is to try to build a reconstruction of the unknown set K and to estimate the geometric characteristics of K by the ones of the reconstruction. Perhaps the most obvious way to build such a reconstruction is to consider the r-offset of the point cloud, that is, the union of balls of a suitable radius r whose center lie in the point cloud. It has been recently proven by [19, 12] that this simple idea leads to a correct estimation of the topology of a smooth manifold, under assumptions on the sampling and the choice of r. This result has been extended to a general class of non-smooth compact sets by [3].

An important feature of offsets of point clouds is that their topology can be computed efficiently, at least when the point cloud lies in a low-dimensional ambient space. For instance, [11] has described an algorithm that given a point cloud C builds a simplicial complex, called the α-complex, that has the same topology as the union of balls of radius α centered at points in C. This algorithm requires to compute the Delaunay triangulation of C, and is hence impractical in higher dimensions. However, even in this case, one can resort to Vietoris-Rips complexes and the theory of topological persistence to correctly infer the Betti numbers of offsets of C [8].

A different way to look at offsets, which is equivalent but better suited to the actual proof of inference results, is through the notion of *distance function*. Given a compact subset K of $\mathbb{R}^d$, the distance function d_K maps any point x in $\mathbb{R}^d$ to the minimum distance between x and any point y in

K. The r-offset of K is then nothing but the sublevel set $\mathrm{d}_K^{-1}([0,r])$. The most important property of the distance function for geometric inference is its stability: if a compact set K', e.g. a point cloud, is a good Hausdorff approximation of another compact set K, then the distance functions $\mathrm{d}_{K'}$ and d_K are close to each other. This property, and two other regularity properties that we will describe later, are the only requirements for proving the topological inference result mentioned earlier.

Offset-based topological inference is now mature and has been used in different contexts to estimate the topology and geometry of shapes sampled with a moderate amount of noise [6, 4, 17]. However, these methods obviously fail completely in the presence of outliers. Indeed, adding even a single data point that is far from the original point cloud will increase by one the number of connected components of the offsets of this point cloud, for a large range of parameters. Said otherwise, while the distance function is only slightly perturbed under Hausdorff noise, adding even a single outlier can change it dramatically.

1.2 Contributions

A possible way to solve the problem of outliers for distance-based inference is then to try to replace the usual distance function to a set K by another notion of distance function that is robust to the addition of a certain amount of outliers. To define what is this *certain amount* one can change the way point clouds are interpreted: they are no more purely geometric objects, but also carry a notion of *mass*. Formally, we replace compact subsets of $\mathbb{R}^d$ by *finite (probability) measures* on the space; a k-manifold will be replaced by the uniform k-dimensional measure on it, a point cloud by a finite sum of Dirac masses, etc. The Hausdorff distance is then not meaningful any more; instead, the distance between two probability measures will be measured through Wasserstein distance, which quantifies the minimal cost of transporting one measure onto the other (cf §2.2).

In this article, we introduce a notion of distance function to a probability measure μ, which we denote by d_{μ,m_0} — where m_0 is a "smoothing" parameter in $(0,1)$. We show that this function retains all the required properties for extending offset-based inference results to the case where the data can be corrupted by outliers. Namely, function d_{μ,m_0} shares the same regularity properties as the usual distance function, and it is stable in the Wasserstein sense, meaning that if two measures are Wasserstein-close, then their distance functions are uniformly close. It can also be computed efficiently for point cloud data. This opens the way to the extension of offset-based inference methods to the case where data may be corrupted by outliers. In particular, we show that considering sublevel sets of our distance functions allows for correct inference of the homotopy type of the unknown object under fairly general assumptions. This improves over the main existing pre-

vious work on the subject [18], which assumes a much more restrictive noise model, and is limited to the smooth case.

2 Background: Measures and Wasserstein distances

As explained in the introduction, in order to account for outliers, we consider our objects as mass distributions instead of purely geometric compact sets. Because one of the goals of this article is to give inference results, i.e. comparison between discrete and the continuous representations, we cannot give the definitions and theorems only in the discrete case, but have to deal with the general case of probability measures.

2.1 Measure theory

A *measure* μ on the space $\mathbb{R}^d$ is a mass distribution. Mathematically, it is defined as a function that maps every (Borel) subset B of $\mathbb{R}^d$ to a non-negative number $\mu(B)$, which is *countably additive* in the sense that whenever (B_i) is a countable family of disjoint Borel subsets of $\mathbb{R}^d$, $\mu\left(\cup_{i\in\mathbb{N}} B_i\right) = \sum_i \mu(B_i)$. The *total mass* of a measure is $\mu(\mathbb{R}^d)$. A measure with finite total mass is called *finite*, while a measure with total mass one is a *probability measure*. The *support* of a measure μ is the smallest closed set K on which the mass of μ is concentrated, i.e. $\mu(\mathbb{R}^d \setminus K) = 0$.

Given a set of N points C, the *uniform measure* on C, which we denote by μ_C, can be defined by $\mu_C(B) = \frac{1}{N}|B \cap C|$. More intuitively, it is the sum of N Dirac masses of weight $1/N$, centered at each point of C. When the points in C are chosen randomly and independently according to an underlying, unknown measure, the measure μ_C is called an *empirical measure*. Formally, we are given a family of independent identically distributed random variables $X_1, \ldots X_N$ who are distributed according to a common measure μ. The uniform probability measure carried by the point cloud $C_N = \{X_1, \ldots, X_n\}$ is known as the *empirical measure*, and simply denoted by μ_N. The uniform law of large numbers asserts that, as N goes to infinity, the empirical measure converges to the underlying measure with probability one — in a sense that will be explained in the next paragraph.

The approach we will describe in this article applies to any measure on Euclidean space. However, to fix ideas, let us describe a family of measures with geometric content that we have in mind when thinking of the underlying measure. One starts from the probability measure μ_M on a compact k-dimensional manifold $M \subseteq \mathbb{R}^d$ given by the rescaled volume form on M, possibly with a non-uniform density. Such measures can be combined, yielding a measure supported on a union of submanifolds of $\mathbb{R}^d$ with various intrinsic dimensions: $\nu = \sum_{i=1}^{\ell} \lambda_i \mu_{M_i}$. Finally, as a simple model of noise, this measure can be convolved with a Gaussian distribution: $\mu = \nu * \mathcal{N}(0, \sigma)$.

This is the same as assuming that each sample that is drawn according to ν is known up to an independant Gaussian error term.

The empirical measure defined by the measure μ we just described could then be obtained by repeatedly *(i)* choosing a random integer $i \in \{0, \ldots, \ell\}$, *(ii)* picking a random sample X_n uniformly distributed in M_i, *(iii)* adding a random Gaussian vector of variance σ^2 to X_n.

2.2 Wasserstein distances

The definition of Wasserstein W_p ($p \geqslant 1$) distance between probability measures rely on the notion of transport plan between measures. It is related to the theory of *optimal transportation* (see e.g. [23]). The Wasserstein distance W_1 is also known as the earth-mover distance, and has been used in vision by [20] and in image retrieval by [22] and others.

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 $A, B \subseteq \mathbb{R}^d$ $\pi(A \times \mathbb{R}^d) = \mu(A)$ and $\pi(\mathbb{R}^d \times B) = \nu(B)$. Intuitively $\pi(A \times B)$ corresponds to the amount of mass of μ contained in A that will be transported to B by the transport plan. Given $p \geqslant 1$, the p-cost of such a transport plan π is given by

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

This cost is finite if both measures μ and ν both have finite *p-moments*, i.e. $\int_{\mathbb{R}^d} \|x\|^p \, \mathrm{d}\mu(x) < +\infty$ and $\int_{\mathbb{R}^d} \|x\|^p \, \mathrm{d}\nu(x) < +\infty$. The set of probability measures on $\mathbb{R}^d$ with finite p-moment includes all probability measures with compact support, such as empirical measures.

Definition 2.1. The *Wasserstein distance* of order p between two probability measures μ and ν on $\mathbb{R}^d$ with finite p-moment is the minimum p-cost $\mathcal{C}_p(\pi)$ of a transport plan π between μ and ν. It is denoted by $W_p(\mu, \nu)$.

As a first example, consider a reference point cloud C with N points, and define a noisy version C' by replacing n points in C by *outliers*, i.e. points o such that $d_C(o) \geqslant R$. The Wasserstein distance between the uniform measures μ_C and μ is at most $\frac{n}{N}(R + \mathrm{diam}(C))$. This can be seen by considering the cost of the transport plan between C' and C that moves the outliers back to their original position, and keeps the other points fixed. On the other hand, the Hausdorff distance between C and C' is at least R. Hence, if the number of outliers is small, i.e. $n \ll N$, the Wasserstein distance is much smaller than the Hausdorff distance.

As mentioned earlier, the question of the convergence of the empirical measure μ_N to the underlying measure μ is fundamendal in the measure-based inference approach we propose. It has been a subject of study in probability and statistics for a long time. If μ is concentrated on a compact

set, then μ_N converges almost surely to μ in the W_p distance. More quantitative convergence statement under different assumptions can be given, as in [2].

If $\chi : \mathbb{R}^d \to \mathbb{R}^+$ defines a probability distribution with finite p-moment $\sigma^p := \int_{\mathbb{R}^d} \|x\|^p \chi(x)\mathrm{d}x$, the Wasserstein distance of order p between any probability measure μ and the convolved measure $\mu * \chi$ can be bounded by: $\mathrm{W}_p(\mu, \mu * \chi) \leqslant \sigma$. If one considers again the example given in the end of §2.1 of an empirical measure μ_N whose samples are drawn according to a "geometric" measure ν convolved with a Gaussian distribution $\mathcal{N}(0, \sigma)$, the combination of the two previous facts gives:

$$\lim_{N \to +\infty} \mathrm{W}_2(\mu_N, \mu) \leqslant \sigma \quad \text{with probability one}$$

Similar bounds are also possible with convolution kernels that are not translation invariant, such as the ones defining the noise model used in [18]. This being said, we would like to stress that the stability results we obtain for the distance functions introduced below do not depend on any noise model; they just depend on the Wasserstein distance between the two probability measures being small.

3 Distance function to a probability measure

In this section we introduce the notion of distance function to a measure that we consider. As explained in the introduction, there are a few constraints for such a definition to be usable in geometric inference, which we now describe in more detail. Let K be a compact set, and d_K be the distance function to K. Then, one can prove the two following properties:

(i) d_K is 1-***Lipschitz.*** For all x, y in $\mathbb{R}^d$, $|\mathrm{d}_K(x) - \mathrm{d}_K(y)| \leqslant \|x - y\|$.

(ii) d_K^2 **is** 1-***semiconcave***. This property is equivalent to the concavity of the map $x \in \mathbb{R}^d \mapsto \mathrm{d}_K^2(x) - \|x\|^2$.

A consequence of Lipschitz regularity is that the distance function is differentiable almost everywhere; in particular, the *medial axis* of K, defined as the set of non-differentiability points of d_K has zero d-volume. Semiconcavity is a stronger regularity property, as thanks to Alexandrov's theorem it implies that the distance function d_K is not only *almost* $\mathcal{C}^1$, but also twice differentiable almost everywhere. The semiconcavity property plays a central role in the proof of existence of the flow of the gradient of the distance function by [16] (Lemma 5.1), which is the main technical tools used in the topological inference results obtained by [3]. The semiconcavity of the squared distance function also plays a crucial role in geometric inference results such as [6] and [17].

This motivates the definition of a *distance-like* function as a non-negative function $\varphi : \mathbb{R}^d \to \mathbb{R}^+$ which is 1-Lipschitz, whose square is 1-semiconcave, and which is *proper* in the sense that $\varphi(x)$ tends to infinity as x does. The following proposition gives a characterization of distance-like functions:

Proposition 3.1. *Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a function whose square is 1-semiconcave. There exists a closed subset K of $\mathbb{R}^{d+1}$ such that $\varphi^2(x) = \mathrm{d}_K^2(x)$, where a point x in $\mathbb{R}^d$ is identified with the point $(x, 0)$ in $\mathbb{R}^{d+1}$.*

Proof. Let $x \in \mathbb{R}^d$ and v be a subgradient to φ^2 at x, and $v' = v/2$. Define a function ψ_v by $\psi_v(y) = \varphi^2(x) - \|v'\|^2 + \|x - v' - y\|^2$. The 1-semiconcavity of the function φ^2 yields $\psi_v(y) \geqslant \varphi^2(y)$, with equality at $y = x$. Hence, the function φ^2 is the lower envelope of all the functions ψ_v as defined above. Letting $y = x - v'$, we see that the constant part of ψ_v is positive. Hence, one can define a point z of $\mathbb{R}^{d+1}$, by $(x - v', (\varphi^2(x) - \|v'\|^2)^{1/2})$, such that $\psi_v(x)$ is equal to the squared Euclidean distance between $(x, 0)$ and z in $\mathbb{R}^{d+1}$. Finally, φ^2 is the squared distance to the set $K \subseteq \mathbb{R}^{d+1}$ made of all such points z. □

This proposition proves in particular that a function $\varphi : \mathbb{R}^d \to \mathbb{R}$ whose square is 1-semiconcave and proper is automatically distance-like: the Lipschitz assumption comes with 1-semiconcavity. It also follows from the proof that distance-like functions are simply generalized power distances, with non-positive weights.

3.1 Definition

The distance function to a compact set K at $x \in \mathbb{R}^d$ is by definition the minimum distance between x and a point of K. Said otherwise, the distance $\mathrm{d}_K(x)$ is the minimum radius r such that the ball centered at x of radius r contains at least a point of K. A very natural idea when trying to define the distance function to a probability measure μ on $\mathbb{R}^d$ is to try mimick the definition above. Given a parameter $0 \leqslant m < 1$, define the pseudo-distance $\delta_{\mu,m}$ by

$$\delta_{\mu,m} : x \in \mathbb{R}^d \mapsto \inf\{r > 0\,;\, \mu(\overline{B}(x, r)) > m\}.$$

For instance for $m = 0$, the definition would coincide with the (usual) distance function to the support of the measure μ. For higher values of m, the function $\delta_{\mu,m}$ is 1-Lipschitz, but lacks other features of that a generalization of the usual distance function to a compact should have. For instance, the application that maps a probability measure μ to $\delta_{\mu,m}$ is not continuous in any reasonable sense. Indeed, let δ_x denote the unit Dirac mass at x and $\mu_\varepsilon = (\frac{1}{2} - \varepsilon)\delta_0 + (\frac{1}{2} + \varepsilon)\delta_1$. Then, for $\varepsilon > 0$ one has $\delta_{\mu_\varepsilon,1/2}(t) = |1 - t|$ for $t < 0$ while if $\varepsilon = 0$, one obtains $\delta_{\mu_0,1/2}(t) = |t|$. Said otherwise, the map $\varepsilon \mapsto \delta_{\mu_\varepsilon,1/2}$ is discontinuous at $\varepsilon = 0$.

In order to gain both Wasserstein-stability and regularity, we define the distance function to μ as a L^2 average of the the pseudo-distances $\delta_{\mu,m}$ for a range $[0, m_0]$ of parameters m:

Definition 3.2. Let μ be a (positive) measure on the Euclidean space, and m_0 be a positive mass parameter $m_0 > 0$ smaller than the total mass of μ. We call *distance function to μ with parameter m_0* the function defined by :

$$\mathrm{d}^2_{\mu,m_0} : \mathbb{R}^n \to \mathbb{R}^+,\ x \mapsto \frac{1}{m_0}\int_0^{m_0} \delta_{\mu,m}(x)^2 \mathrm{d}m$$

As an example, let C be a point cloud with N points in $\mathbb{R}^d$, and μ_C be the uniform measure on it. The pseudo-distance function $\delta_{\mu_C,m}$ evaluated at a point $x \in \mathbb{R}^d$ is by definition equal to the distance between x and its kth nearest neighbor in C, where k is the smallest integer larger than $m\,|C|$. Hence, the function $m \mapsto \delta_{\mu_C,m}$ is constant on all ranges $(\frac{k}{N}, \frac{k+1}{N}]$. Using this one obtains the following formula for the squared distance d^2_{μ,m_0}, where $m_0 = k_0/\,|C|$:

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

where $\mathrm{NN}_C^{k_0}(x)$ denote the k_0 nearest neighbors of x in C. In this case, the pointwise evaluation of $\mathrm{d}^2_{\mu_C,k/n}(x)$ reduces to a k-nearest neighbor query in C.

3.2 Equivalent formulation

In this paragraph, we prove that the distance function to a measure d_{μ,m_0} is in fact a real distance to a compact set, but in a infinite-dimensional space. From this fact, we will deduce all of the properties needed for geometric and topological inference.

A measure ν will be called a *submeasure* of another measure μ if for every Borel subset B of $\mathbb{R}^d$, $\nu(B) \leqslant \mu(B)$. This is the same as requiring that $\mu - \nu$ is a measure. The set of all submeasures of a given measure is denoted by $\mathrm{Sub}(\mu)$, while the set of submeasures of μ with a prescribed total mass m_0 is denoted by $\mathrm{Sub}_{m_0}(\mu)$.

Proposition 3.3. *For any measure μ on $\mathbb{R}^d$, the distance function to μ at x is the solution of the following optimal transportation problem:*

$$\mathrm{d}_{\mu,m_0}(x) = \min\left\{m_0^{-1/2}\,\mathrm{W}_2\left(m_0\delta_x, \nu\right); \nu \in \mathrm{Sub}_{m_0}(\mu)\right\} \qquad (1)$$

Then, for any measure μ_{x,m_0} that realizes the above minimum one has:

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

Said otherwise, the distance d_{μ,m_0} evaluated at a point $x \in \mathbb{R}^d$ 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:

$$\mathrm{d}_{\mu,m_0}(x) = \frac{1}{\sqrt{m_0}} \mathrm{dist}_{\mathrm{W}_2}(m_0\delta_x, \mathrm{Sub}_{m_0}(\mu)) \tag{2}$$

The set of minimizers in the above expression corresponds to the "orthogonal" projections, or nearest neighbors, of the Dirac mass $m_0\delta_x$ on the set of submeasures $\mathrm{Sub}_{m_0}(\mu)$. As we will see in the proof of the proposition, these are submeasures μ_{x,m_0} of total mass m_0 whose support is contained in the closed ball $\overline{\mathrm{B}}(x, \delta_{\mu,m}(x))$, and whose restriction to the *open* ball $\mathrm{B}(x, \delta_{\mu,m}(x))$ coincides with μ. Denote these measures by $\mathcal{R}_{\mu,m_0}(x)$.

In order to prove Proposition 3.3, we need a few definitions from probability theory. The *cumulative function* $F_\nu : \mathbb{R}^+ \to \mathbb{R}$ of a measure ν on $\mathbb{R}^+$ is the non-decreasing function defined by $F_\nu(t) = \nu([0,t))$. Its *generalized inverse*, denoted by F_ν^{-1} and defined by $F_\nu^{-1} : m \mapsto \inf\{t \in \mathbb{R} \,;\, F_\nu(t) > m\}$ is left-continuous. Notice that if μ, ν are two measures on $\mathbb{R}^+$, then ν is a submeasure of μ if and only if $F_\nu(t) \leqslant F_\mu(t)$ for all $t > 0$.

Proof. Let first remark that if ν is any measure of total mass m_0, there is only one transport plan between ν and the Dirac mass $m_0\delta_x$, which maps any point of $\mathbb{R}^d$ to x. Hence, the Wasserstein distance between ν and δ_x is given by

$$\mathrm{W}_2^2(m_0\delta_x, \nu) = \int_{\mathbb{R}^d} \|h-x\|^2 \, \mathrm{d}\nu(h)$$

Let $\mathrm{d}_x : \mathbb{R}^d \to \mathbb{R}$ denote the distance function to the point x, and let ν_x be the pushforward of ν by the distance function to x, i.e. for any subset I of $\mathbb{R}$, $\nu_x(I) = \nu(\mathrm{d}_x^{-1}(I))$. Using the change-of-variable formula, and the definition of the cumulative function gives us:

$$\int_{\mathbb{R}^d} \|h-x\|^2 \, \mathrm{d}\nu(h) = \int_{\mathbb{R}^+} t^2 \mathrm{d}\nu_x(t) = \int_0^{m_0} F_{\nu_x}^{-1}(m)^2 \mathrm{d}m$$

If ν is a submeasure of μ, then by the remark above, $F_{\nu_x}(t) \leqslant F_{\mu_x}(t)$ for all $t > 0$. From this, one deduces that $F_{\nu_x}^{-1}(m) \geqslant F_{\mu_x}^{-1}(m)$. This gives

$$\begin{aligned} \mathrm{W}_2^2(m_0\delta_x, \nu) = \int_{\mathbb{R}^d} \|h-x\|^2 \, \mathrm{d}\nu(h) &\geqslant \int_0^{m_0} F_{\mu_x}^{-1}(m)^2 \mathrm{d}m \\ &= \int_0^{m_0} \delta_{\mu,m}(x)^2 \mathrm{d}m = m_0 \mathrm{d}_{\mu,m_0}^2(x) \end{aligned} \tag{3}$$

The second inequality is because $F_{\mu_x}(t) = \mu(B(x,t))$, and thus $F^{-1}_{\mu_x}(m) = \delta_{\mu,m}(x)$. This proves that $\mathrm{d}_{\mu,m_0}(x)$ is smaller than the right-hand side of (1).

To conclude the proof, we study the cases of equality in (3). Such a case happens when for almost every $m \leqslant m_0$, $F^{-1}_{\nu_x}(m) = F^{-1}_{\mu_x}(m)$. Since these functions are increasing and left-continuous, equality must in fact hold for every such m. By the definition of the pushforward, this implies that $\nu(\overline{\mathrm{B}}(x, \delta_{\mu,m_0}(x))) = m_0$, i.e. all the mass of ν is contained in the closed ball $\overline{\mathrm{B}}(x, \delta_{\mu,m_0}(x))$ and $\tilde{\mu}(B(x, \delta_{\mu,m_0}(x))) = \mu(\mathrm{B}(x, \delta_{\mu,m_0}(x)))$. Because ν is a submeasure of μ, this can be true if and only iff ν belongs in the set $\mathcal{R}_{\mu,m_0}(x)$ described before the proof.

To finish the proof, we should remark that the set of minimizer $\mathcal{R}_{\mu,m_0}(x)$ always contain a measure μ_{x,m_0}. The only difficulty is when the boundary of the ball carries too much mass. In this case, we uniformly rescale the mass contained in the bounding sphere so that the measure μ_{x,m_0} has total mass m_0. More precisely, we let:

$$\mu_{x,m_0} = \mu|_{B(x,\delta_{\mu,m_0}(x))} + (m_0 - \mu(B(x,\delta_{\mu,m_0}(x)))\frac{\mu|_{\partial B(x,\delta_{\mu,m_0}(x))}}{\mu(\partial B(x,\delta_{\mu,m_0}(x)))}$$

□

3.3 Stability of the distance function to a measure

The goal of this section is to prove that the notion of distance function to a measure that we defined earlier is stable under change of the measure. This follows rather easily from the characterization of d_{μ,m_0} given by Proposition 3.3.

Proposition 3.4. *Let μ and μ' be two probability measures on $\mathbb{R}^d$. Then,*

$$\mathrm{d_H}(\mathrm{Sub}_{m_0}(\mu), \mathrm{Sub}_{m_0}(\mu')) \leqslant \mathrm{W}_2(\mu, \mu')$$

Proof. Let ε be the Wasserstein distance of order 2 between μ and μ', and π be a corresponding optimal transport plan, i.e. a transport plan between μ and μ' such that $\int_{\mathbb{R}^d\times\mathbb{R}^d} \|x-y\|^2 \pi(x,y)\mathrm{d}x\mathrm{d}y = \varepsilon^2$. Then, given a submeasure ν of μ, one can find a submeasure π' of π that transports ν to a submeasure ν' of μ'. Then,

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

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

Theorem 3.5 ((Distance function stability))**.** *If μ and μ' are two probability measures on $\mathbb{R}^d$ and $m_0 > 0$, then $\left\|\mathrm{d}_{\mu,m_0} - \mathrm{d}_{\mu',m_0}\right\|_\infty \leqslant \frac{1}{\sqrt{m_0}}\,\mathrm{W}_2(\mu,\mu')$.*

Proof. The following sequence of equalities and inequalities, that follows from Propositions 3.3 and 3.4, proves the theorem:.

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

□

3.4 The distance to a measure is distance-like.

The subdifferential of a function $f : \Omega \subseteq \mathbb{R}^d \to \mathbb{R}$ at a point x, is the set of vectors v of $\mathbb{R}^d$, denoted by $\partial_x f$, 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 admits a derivative at a point x if and only if the subdifferential $\partial_x f$ is a singleton, in which case the gradient $\nabla_x f$ coincides with its unique element.

Proposition 3.6. *The function $v_{\mu,m_0} : x \in \mathbb{R}^d \mapsto \|x\|^2 - \mathrm{d}^2_{\mu,m_0}$ is convex, and its subdifferential at a point $x \in \mathbb{R}^d$ is given by*

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

Proof. For any two points x and y of $\mathbb{R}^d$, let μ_{x,m_0} and μ_{y,m_0} be in $\mathrm{R}_{\mu,m_0}(x)$ and $\mathrm{R}_{\mu,m_0}(y)$ respectively. Thanks to Proposition 3.3 we have the following sequence of equalities and inequalities:

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

where v is the vector defined by

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

The inequality can be rewritten as:

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

which shows that the vector $(2x - v)$ belongs to the subdifferential of v at x. By the characterization of convex functions by that we recalled above, one deduces that v_{μ,m_0} is convex.

We now turn to the proof of the converse inclusion. This proof is slightly more technical, but not really needed for the remaining of the article. First, let

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

The sets $\mathcal{D}_{\mu,m_0}$ and $\partial_x v_{\mu,m_0}$ are both convex, and we have shown that $\mathcal{D}_{\mu,m_0}$ is contained in $\partial_x v_{\mu,m_0}$. By Theorem 2.5.1 in [9], the subdifferential $\partial_x v_{\mu,m_0}$ can be obtained as the convex hull of the set of limits of gradients $\nabla_{x_n} v_{\mu,m_0}$, where (x_n) is any sequence of points converging to x at which v_{μ,m_0} is differentiable. To sum up, we only need to prove that every such limit also belongs to the set $\mathcal{D}_{\mu,m_0}(x)$. Let (x_n) be a sequence of points at which v_{μ,m_0} is differentiable, and let μ_n be the unique element in $\mathcal{R}_{\mu,m_0}(x_n)$. Necessarily,

$$\nabla_{x_n} v_{\mu,m_0} = 2x_n - 2/m_0 \int_h (x_n - h) \mathrm{d}\mu_n(h)$$

where μ_n is in $\mathcal{R}_{\mu,m_0}(x_n)$. Since every μ_n is a submeasure of μ, by compactness one can extract a subsequence of n such that μ_n weakly converges to a measure μ_∞. This measure belongs to $\mathcal{R}_{\mu,m_0}(x)$, and hence the vector

$$D = 2x - 2/m_0 \int_h (x - h) \mathrm{d}\mu_\infty(h)$$

is in the set $\mathcal{D}_{\mu,m_0}(x)$. Moreover, the weak convergence of μ_n to μ_∞ implies that the sequence $\nabla_{x_n} v_{\mu,m_0}$ converges to D. This concludes the proof of this inclusion. □

Corollary 3.7. *The function* d^2_{μ,m_0} *is* 1*-semiconcave. Moreover,*

(i) d^2_{μ,m_0} *is differentiable at a point* $x \in \mathbb{R}^d$ *if and only if the support of the restriction of* μ *to the sphere* $\partial \mathrm{B}(x, \delta_{\mu,m_0}(x))$ *contains at most one point;*

(ii) d^2_{μ,m_0} *is differentiable almost everywhere in* $\mathbb{R}^d$*, with gradient defined by*

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

where μ_{x,m_0} *is the only measure in* $\mathcal{R}_{\mu,m_0}(x)$*.*

(iii) the function $x \in \mathbb{R}^d \mapsto d_{\mu,m_0}(x)$ *is* 1*-Lipschitz.*

Proof. For (i), it is enough to remark that $\mathcal{R}_{\mu,m_0}(x)$ is a singleton iff the support of $\mu|_{\partial B(x,\delta_{\mu,m_0}(x))}$ is at most a single point. (ii) This follows from the fact that a convex function is differentiable at almost every point, at which its gradient is the only element of the subdifferential at that point. (iii) The gradient of the distance function d_{μ,m_0} can be written as:

$$\nabla_x \mathrm{d}_{\mu,m_0} = \frac{\nabla_x \mathrm{d}^2_{\mu,m_0}}{2\mathrm{d}_{\mu,m_0}} = \frac{1}{\sqrt{m_0}} \frac{\int_{h\in\mathbb{R}^d}[x-h]\,\mathrm{d}\mu_{x,m_0}(h)}{(\int_{h\in\mathbb{R}^d}\|x-h\|^2\,\mathrm{d}\mu_{x,m_0}(h))^{1/2}}$$

Using the Cauchy-Schwartz inequality we find the bound $\|\nabla_x \mathrm{d}_{\mu,m_0}\| \leqslant 1$ which proves the statement. □

4 Applications to geometric inference

Reconstruction from point clouds with outliers was the main motivation for introducing the distance function to a measure. In this section, we adapt the reconstruction theorem introduced by [3] to our setting. The original version of the theorem states that a regular enough compact set K can be faithfully reconstructed from another close enough compact set C. More precisely, for a suitable choice of r, the offsets C^r and K^η have the same homotopy type for any positive η. The regularity assumption on K is expressed as a lower bound on its so-called μ-reach, which is a generalization of the classical notion of reach [13]. In particular, smooth submanifolds, convex sets and polyhedra always have positive μ-reach for suitable μ, hence the reconstruction theorem may be applied to such sets. In these section, we show that the reconstruction results of [3] can be easily generalized to compare the sub-level sets of two uniformly-close distance-like functions. It is also possible to adapt most of the topological and geometric inference results of [5, 4, 6] in a similar way.

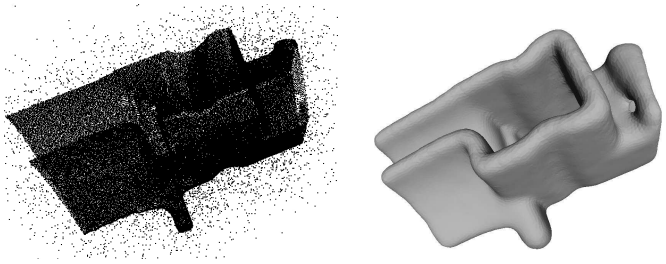


Figure 1: On the left, a point cloud sampled on a mechanical part to which 10% of outliers (uniformly sampled in a box enclosing the model) have been added. On the right, the reconstruction of an isosurface of the distance function d_{μ_C,m_0} to the uniform probability measure on this point cloud.

4.1 Extending the sampling theory for compact sets

In this paragraph we extend the sampling theory of [3] for compact sets to distance-like functions. We don't include all of the results of the paper, but only those that are needed to the reconstruction theorem (Th. 4.6). We refer the interested reader to the original paper for more details.

Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a *distance-like* function. The 1-semiconcavity of φ^2 allows to define a notion of gradient vector field $\nabla_x \varphi$ for φ, defined everywhere and satisfying $\|\nabla_x \varphi\| \leqslant 1$. Although not continuous, the vector field $\nabla\varphi$ is sufficiently regular to be integrated in a continuous locally Lipschitz flow [21] $\Phi^t : \mathbb{R}^d \to \mathbb{R}^d$. The flow Φ^t integrates the gradient $\nabla\varphi$ in the sense that for every $x \in \mathbb{R}^d$, the curve $\gamma : t \mapsto \Phi^t(x)$ is right-differentiable, and for every $t > 0$, $\frac{\mathrm{d}\gamma}{\mathrm{d}\varepsilon}\Big|_{t^-} = \nabla_{\gamma(t)}\varphi$. Moreover, for any integral curve $\gamma : [a, b] \to \mathbb{R}^d$ parametrized by arc-length, one has:

$$\varphi(\gamma(b)) = \varphi(\gamma(a)) + \int_a^b \left\|\nabla_{\gamma(t)}\varphi\right\| \mathrm{d}t.$$

Definition 4.1. Let φ be a distance like function. Following the notation for offset of compact sets, we will denote by $\varphi^r = \varphi^{-1}([0, r])$ the r sublevel set of φ.

(i) A point $x \in \mathbb{R}^d$ will be called α-critical (with $\alpha \in [0, 1]$) if the inequality $\varphi^2(x + h) \leqslant \varphi^2(x) + 2\alpha \|h\| \varphi(x) + \|h\|^2$ is true for all $h \in \mathbb{R}^d$. A 0-critical point is simply called a critical point. It follows from the 1-semiconcavity of φ^2 that $\|\nabla_x \varphi\|$ is the infimum of the $\alpha \geqslant 0$ such that x is α-critical.

(ii) The *weak feature size* of φ at r is the minimum $r' > 0$ such that φ doesn't have any critical value between r and $r + r'$. We denote it by $\mathrm{wfs}_\varphi(r)$. For any $0 < \alpha < 1$, the *α-reach* of φ is the maximum r such that $\varphi^{-1}((0, r])$ does not contain any α-critical point. Obviously, the α-reach is always a lower bound for the weak-feature size, with $r = 0$.

The proof of the Reconstruction Theorem in [3] relies on two important observations. The first one is a consequence of a distance-like version of Grove's isotopy lemma [15, Prop. 1.8], which asserts that the topology of the sublevel sets of φ can only change when one passes critical values. As in [7, Theorem 3], one deduces that the offsets of two uniformly close distance-like functions with large weak feature size have the same homotopy type:

Proposition 4.2 (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 the subset $\varphi^{-1}([r_1, r_2])$. Then all the sublevel sets $\varphi^{-1}([0, r])$ are isotopic for $r \in [r_1, r_2]$.*

Proposition 4.3. *Let φ and ψ be two distance-like functions, such that $\|\varphi - \psi\|_\infty \leqslant \varepsilon$. Suppose moreover that $\mathrm{wfs}_\varphi(r) > 2\varepsilon$ and $\mathrm{wfs}_\psi(r) > 2\varepsilon$. Then, for every $0 < \eta \leqslant 2\varepsilon$, $\varphi^{r+\eta}$ and $\psi^{r+\eta}$ have the same homotopy type.*

Proof. See Appendix. □

The second key observation made in [3] is that the critical points of a distance function are stable in certain sense under small Hausdorff perturbations. This result remains true for uniform approximation by distance-like functions:

Proposition 4.4. *Let φ and ψ be two distance-like functions with $\|\varphi - \psi\|_\infty \leqslant \varepsilon$. For any α-critical point x of φ, there exists a α'-critical point x' of ψ with $\|x - x'\| \leqslant 2\sqrt{\varepsilon\varphi(x)}$ and $\alpha' \leqslant \alpha + 2\sqrt{\varepsilon/\varphi(x)}$.*

Proof. The proof is almost verbatim from [3], and postponed to the Appendix. □

Corollary 4.5. *Let φ and ψ be two ε-close distance-like functions, and suppose that $\mathrm{reach}_\alpha(\varphi) \geqslant R$ for some $\alpha > 0$. Then, ψ has no critical value in the interval $\left]4\varepsilon/\alpha^2, R - 3\varepsilon\right[$.*

Proof. See Appendix. □

Theorem 4.6 (Reconstruction)**.** *Let φ, ψ be two ε-close distance-like functions, with $\mathrm{reach}_\alpha(\varphi) \geqslant R$ for some positive α. Then, for any $r \in [4\varepsilon/\alpha^2, R - 3\varepsilon]$, and for $0 < \eta < R$, 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 isotopy lemma, 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:

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

By Proposition 4.3, 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 $\frac{1}{2}\mathrm{wfs}_\varphi(r)$ and $\frac{1}{2}\mathrm{wfs}_\psi(r)$. This is true, provided that $2\varepsilon \leqslant R - \varepsilon(3 + 4/\alpha^2)$. The theorem follows. □

Remark that in the above definition 4.1 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' such that the set $\varphi^{-1}((r, r + r'])$ does not contain any α-critical value. A reconstruction theorem similar to Theorem 4.6 would still hold under the weaker condition that the (r, α)-reach of φ is positive.

4.2 Distance to a measure vs. distance to its support

In this paragraph, we compare the distance functions d_{μ,m_0} to a measure μ and the distance function to its support S, and study the convergence properties as the mass parameter m_0 converges to zero. A first obvious remark is that the pseudo-distance δ_{μ,m_0} (and hence the distance d_{μ,m_0}) is always larger than the regular distance function d_S. As a consequence, to obtain a convergence result of d_{μ,m_0} to d_S as m_0 goes to zero, it is necessary to upper bound d_{μ,m_0} by $\mathrm{d}_S + \mathrm{o}(m_0)$. It turns out that the convergence speed of d_{μ,m_0} to d_S depends on the way the mass of μ contained within any ball $\mathrm{B}(p,r)$ centered at a point p of the support decreases with r. Let us define:

(i) We say that a non-decreasing positive function $f : \mathbb{R}^+ \to \mathbb{R}^+$ is a *uniform lower bound on the growth of* μ if for every point p in the support of μ and every $\varepsilon > 0$, $\mu(\mathrm{B}(p,\varepsilon)) \geqslant f(\varepsilon)$;

(ii) The measure μ has *dimension at most* k if there is a constant $C(\mu)$ such that $f(\varepsilon) = C(\mu)\varepsilon^k$ is a uniform lower bound on the growth of μ, for ε small enough.

Lemma 4.7. *Let μ be a probability measure and f be a uniform lower bound on the growth of μ. Then $\|\mathrm{d}_{\mu,m_0} - \mathrm{d}_S\|_\infty < \varepsilon$ as soon as $m_0 < f(\varepsilon)$.*

Proof. Let ε and m_0 be such that $m_0 < f(\varepsilon)$ and let x be a point in $\mathbb{R}^d$, p a projection of x on S, i.e. a point p such that $\|x-p\| = d(p,S)$. By assumption, $\mu(\mathrm{B}(x, \mathrm{d}_S(x)+\varepsilon)) \geqslant \mu(\mathrm{B}(p,\varepsilon)) \geqslant m_0$. Hence, $\delta_{\mu,m_0}(x) \leqslant \mathrm{d}_S(x) + \varepsilon$. The function $m \mapsto \delta_{\mu,m}(x)$ being non-decreasing, we get: $m_0 \mathrm{d}_S^2(x) \leqslant \int_0^{m_0} \delta_{\mu,m}^2(x)\mathrm{d}m \leqslant m_0(\mathrm{d}_S(x)+\varepsilon)^2$. Taking the square root of this expression proves the lemma. □

Corollary 4.8. *(i) If the support S of μ is compact, then d_S is the uniform limit of d_{μ,m_0} as m_0 converges to 0;*

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

$$\|\mathrm{d}_{\mu,m_0} - \mathrm{d}_S\| \leqslant C(\mu)^{-1/k} m_0^{1/k}$$

Proof. (i) If S is compact, there exists a sequence $x_1, x_2, \cdots$ of points in S such that for any $\varepsilon > 0$, $S \subseteq \cup_{i=1}^n B(x_i, \varepsilon/2)$ for some $n = n(\varepsilon)$. By definition of the support of a measure, $\eta(\varepsilon) = \min_{i=1\cdots n} \mu(\mathrm{B}(x_i, \varepsilon/2))$ is positive. Now, for any point $x \in S$, there is a x_i such that $\|x - x_i\| \leqslant \varepsilon/2$. Hence, $\mathrm{B}(x_i, \varepsilon/2) \subseteq \mathrm{B}(x,\varepsilon)$, which means that $\mu(\mathrm{B}(x,\varepsilon)) \geqslant \eta(\varepsilon)$. (ii) Follows straightforwardly from the Lemma. □

For example, the uniform probability measure on a k-dimensional compact submanifold S has dimension at most k. The following proposition gives a more precise convergence speed estimate based on curvature.

Proposition 4.9. *Let S be a smooth k-dimensional submanifold of $\mathbb{R}^d$ whose curvature radii are lower bounded by R, and μ the uniform probability measure on S, then*

$$\|\mathrm{d}_S - \mathrm{d}_{\mu,m_0}\| \leqslant C(S)^{-1/k} m_0^{1/k}$$

for m_0 small enough and $C(S) = (2/\pi)^k \beta_k / \mathcal{H}^k(S)$ where β_k is the volume of the unit ball in $\mathbb{R}^k$.

Notice in particular that the convergence speed of d_{μ,m_0} to d_S depends only on the *intrinsic* dimension k of the submanifold S, and not on the ambient dimension d. In order to prove this result, we make use of the Günther-Bishop theorem (cf [14, §3.101]).

Theorem 4.10 (Günther-Bishop)**.** *If the sectional curvatures of a Riemannian manifold M do not exceed δ, then for every $x \in M$, $\mathcal{H}^k(\mathrm{B}_M(x,r)) \geqslant \beta_{k,\delta}(r)$ where $\beta_{k,\delta}(r)$ is the volume of a ball of radius r in the simply connected k-dimensional manifold with constant sectional curvature δ, provided that r is smaller than the minimum of the injectivity radius of M and $\pi/\sqrt{\delta}$.*

Proof of Proposition 4.9. Since the intrinsic ball $\mathrm{B}_S(x,\varepsilon)$ is always included in the Euclidean ball $\mathrm{B}(x,\varepsilon) \cap S$, the mass $\mu(\mathrm{B}(x,\varepsilon))$ is always larger than $\mathcal{H}^k(\mathrm{B}_S(x,\varepsilon))/\mathcal{H}^k(S)$. Remarking that the sectional curvature of M is upper-bounded by $1/R^2$, Günter-Bishop theorem implies that for any ε smaller than the injectivity radius of S and πR,

$$\mu(\mathrm{B}(x,\varepsilon)) \geqslant \frac{\beta_{k,1/R^2}(\varepsilon)}{\mathcal{H}^k(S)}$$

Hence μ has dimension at most k. Moreover, by comparing the volume of an intrinsic ball of the unit sphere and the volume of its orthogonal projection on the tangent space to its center, one has:

$$\beta_{k,1/R^2}(\varepsilon) = R^k \beta_{k,1}(\varepsilon/R) \geqslant R^k [\sin(\varepsilon/R)]^k \beta_k$$

where β_k is the volume of the k-dimensional unit ball. Using $\sin(\alpha) \geqslant \frac{2}{\pi}\alpha$ gives the announced value for $C(S)$. □

4.3 Shape reconstruction from noisy data

The previous results lead to shape reconstruction theorems from 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 $d_K : \mathbb{R}^d \to \mathbb{R}_+$ be the (Euclidean) distance function to K. If μ' is another probability measure (*eg.* the empirical measure given by a point cloud sampled according to μ),

one has

$$\left\|d_K - d_{\mu',m_0}\right\|_\infty \leqslant \|d_K - d_{\mu,m_0}\|_\infty + \|d_{\mu,m_0} - d_{\mu',m_0}\|_\infty \quad (4)$$

$$\leqslant C(\mu)^{-1/k} m_0^{1/k} + \frac{1}{\sqrt{m_0}} \mathrm{W}_2(\mu,\mu') \quad (5)$$

This inequality insuring the closeness of d_{μ',m_0} to the distance function d_K for the sup-norm follows immediately from the stability theorem 3.5 and the corollary 4.8. As expected, the choice of m_0 is a trade-off: small m_0 lead to better approximation of the distance function to the support, while large m_0 make the distance functions to measures more stable. Eq. 4 leads to the following corollary of Theorem 4.6:

Corollary 4.11. *Let μ be a measure and K its support. Suppose that μ has dimension at most k and that $\mathrm{reach}_\alpha(d_K) \geqslant R$ for some $R > 0$. 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:*

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

Figure 1 illustrates the reconstruction Theorem 4.6 on a sampled mechanical part with 10% of outliers. In this case μ' is the normalized sum of the Dirac measures centered on the data points and the (unknown) measure μ is the uniform measure on the mechanical part.

5 Relation with non-parametric density estimation

Nearest-neighbor estimators are a family of non-parametric density estimators, that has been extensively used in nonparametric discrimination, pattern recognition and spatial analysis problems. Suppose that C is a point cloud whose points are independently drawn with respect to a probability measure μ on $\mathbb{R}^d$ which admits a densiy f with respect to the Lebesgue measure. Nearest neighbors estimators estimate the density function by:

$$\tilde{f}(x) = \frac{k}{\#C\omega_d(\delta_{C,k}(x))} \quad (6)$$

where $\omega_d(r)$ is the volume of the d-sphere of radius r, and $\delta_{C,k}$ denotes the distance to the kth nearest neighbor in C.

Our claim in this section is that or some of the applications where density estimators are used (such as mean shift, see below), distance to measure functions could also be used as well. The advantages of the distance function

d_{μ,m_0} over nearest-neighbor density estimators are multiple. First of all, they are well defined even when the underlying probability measure does not have a density with respect to the Lebesgue measure, e.g. if it is concentrated on a lower-dimensional subset. Second, the distance to measure d_{μ,m_0} is always uniformly stable with respect to Wasserstein perturbations of the data. Third, because of its 1-semiconcavity, the distance function d_{μ,m_0} is much more regular than the distance δ_{μ,m_0}, as illustrated by Figure 5. This has consequences for gradient descent algorithms such as mean shift.

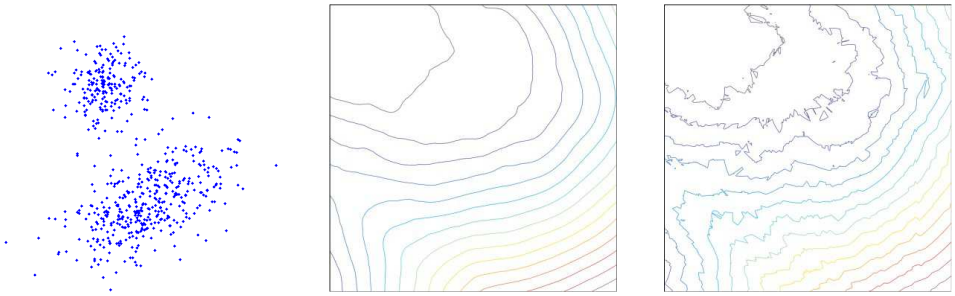


Figure 2: The distance functions to an empirical measure μ associated to a 600 points 2D data set P sampled independently according two gaussians (first figure). The second and third figures represent close-ups on the level sets of the functions $\mathrm{d}_{\mu,1/30}$ and $\delta_{\mu,1/30}$ respectively, illustrating the difference of regularity between the two distance functions.

As a consequence of this regularity, it is possible to prove higher order convergence properties of d_{μ_n,m_0} to d_{μ,m_0} as μ_n converges to μ. For example, it can be shown that $\nabla d_{\mu_n,m_0}$ converges to $\nabla d_{\mu,m_0}$ as locally integrable vector fields. Pointwise convergence results can also be obtained at points where $\nabla d_{\mu,m_0}$ is bounded away from 0.

5.1 Mean-Shift Methods using Distance Functions

Kernel-based mean-shift clustering. Mean-shift clustering [10] is a non-parametric clustering method that works on point cloud drawn from an unknown probability measure with density. Specifically, one is given a point cloud $C \subseteq \mathbb{R}^d$ and a radial kernel K. The underlying probability density is estimated by:

$$f(x) = \frac{1}{h^d \#C} \sum_{p \in C} K\left(\frac{p-x}{h}\right)$$

where h is a given bandwidth parameter. Starting from a point x in the space, one iteratively constructs a sequence of points (x_i), descending the

Figure 3: 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.

gradient of the estimated density:

$$\begin{cases} x_0 = x \\ x_{i+1} = \dfrac{\sum_{p\in C} K\left(\frac{p-x}{h}\right) p}{\sum_{p\in C} K\left(\frac{p-x}{h}\right)} \end{cases} \tag{7}$$

The clustering method works as follows: for each point x_0 in the point cloud, one iterates the sequence x_i until convergence. This defines a mapping from C to the set of critical points of the kernel-based density estimate. A *cluster* of C is simply the set of points of C which correspond to the same critical point under this mapping.

Distance-based mean-shift. We propose a method similar to mean shift, but where the distance function replaces the estimated density. Our iterative scheme is a simple gradient descent for the squared distance function:

$$\begin{cases} x_0 = x \\ x_{i+1} = x_i - \dfrac{1}{2}\nabla_{x_i} \mathrm{d}^2_{\mu,m_0} \end{cases} \tag{8}$$

In practice, μ is the uniform probability measure on a point cloud C and $m_0 = k_0/\#C$. In this context, x_{i+1} is simply the isobarycenter of the k_0 nearest neighbor of x_i in C:

$$x_{i+1} = x_i - \frac{1}{k}\sum_{k=1}^{k_0}(x_i - \mathrm{p}^k_C(x_i)) = \frac{1}{k}\sum_{k=1}^{k_0}\mathrm{p}^k_C(x_i)$$

Proposition 5.1. *Let x be a point in $\mathbb{R}^d$ and $x_t = x - \frac{t}{2}\nabla_{x_i}\mathrm{d}^2_{\mu,m_0}$. Then,*

1. $\mathrm{d}_{\mu,m_0}(x_t) \leqslant \mathrm{d}_{\mu,m_0}(x)$
2. $\langle \nabla_{x_t}\mathrm{d}^2_{\mu,m_0} | \nabla_{x_0}\mathrm{d}^2_{\mu,m_0}\rangle \geqslant 0$

Proof. 1. This is a simple application of Prop. 3.3.(1).
2. Since d^2_{μ,m_0} is 1-concave,

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

Now, if we set $y = x_t$,

$$\begin{aligned}\langle x-y|\nabla_x \mathrm{d}^2_{\mu,m_0} - \nabla_y \mathrm{d}^2_{\mu,m_0}\rangle &= \langle x-y|\nabla_x \mathrm{d}^2_{\mu,m_0}\rangle - \langle x-y|\nabla_y \mathrm{d}^2_{\mu,m_0}\rangle \\ &= \frac{2}{t}\|x-y\|^2 - \frac{t}{2}\langle \nabla_x \mathrm{d}^2_{\mu,m_0}|\nabla_y \mathrm{d}^2_{\mu,m_0}\rangle\end{aligned}$$

This proves that $\langle \nabla_x \mathrm{d}^2_{\mu,m_0}|\nabla_{x_t} \mathrm{d}^2_{\mu,m_0}\rangle \geqslant \frac{4}{t}\left(\frac{1}{t}-1\right)\|x-y\|^2$. □

Both properties indicate good convergence properties for our iterative scheme: the first one prevents any infinite loop, while the second shows that trajectories are not too wiggly (more precisely, consecutive edges never make an acute angle). The first property of this proposition had been proved for classical mean-shift, when the kernel K has convex and monotonically decreasing profile (Theorem 1 in [10]). On the other hand, the second property is shown to hold for mean-shift when K is Gaussian (Theorem 2 in [10]), in which case it is not convex. We are not aware of any choice of kernel such that the resulting mean-shift scheme satisfies these two properties simultaneously.

6 Discussion

We have extended the notion of distance function to a compact subset of $\mathbb{R}^d$ to the case of measures, and showed that this permits to reconstruct sampled shapes with the correct homotopy type even in the presence of outliers. It also seems very likely that a similar statement showing that the sublevel sets of d_{μ,m_0} are isotopic to the offsets of K can be proved, using the same sketch of proof as in [4]. Moreover, in the case of point clouds/empirical measures (finite sums of Dirac measures), the computation of the distance function to a measure (and its gradient) at a given point boils down to a computation of nearest neighbors making it easy to use in practice. However, we note that in the important case where the unknown shape is a submanifold, our reconstructions are clearly not homeomorphic since they do not have the correct dimension. Is there a way to combine our framework with the classical techniques developped for homeomorphic surface reconstruction (see *e.g.* [1]) to make them robust to outliers while retaining their guarantees?

References

[1] N. Amenta, S. Choi, T. K. Dey, and N. Leekha. A simple algorithm for homeomorphic surface reconstruction. *Int. J. Comput. Geom. Ap.*, 12(1-2):125–141, 2002.

[2] F. Bolley, A. Guillin, and C. Villani. Quantitative Concentration Inequalities for Empirical Measures on Non-compact Spaces. *Probab. Theory Rel.*, 137(3):541–593, 2007.

[3] F. Chazal, D. Cohen-Steiner, and A. Lieutier. A sampling theory for compact sets in Euclidean space. *Discrete Comput. Geom.*, 41(3):461–479, 2009.

[4] F. Chazal, D. Cohen-Steiner, and A. Lieutier. Normal cone approximation and offset shape isotopy. *Comp. Geom. Theor. Appl.*, 42(6-7):566–581, 2009.

[5] F. Chazal, D. Cohen-Steiner, A. Lieutier, and B. Thibert. Stability of Curvature Measures. *Computer Graphics Forum (proc. SGP 2009)*, pages 1485–1496, 2008.

[6] F. Chazal, D. Cohen-Steiner, and Q. Mérigot. Boundary measures for geometric inference. *J. Found. Comput. Math*, 10:221–240, 2010.

[7] F. Chazal and A. Lieutier. Stability and Computation of Topological Invariants of Solids in $\mathbb{R}^n$. *Discrete Comput. Geom.*, 37(4):601–617, 2007.

[8] F. Chazal and S. Y. Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proc. 24th ACM Sympos. Comput. Geom.*, pages 232–241, 2008.

[9] F.H. Clarke. *Optimization and nonsmooth analysis*. Wiley New York, 1983.

[10] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[11] H. Edelsbrunner. The union of balls and its dual shape. *Discrete Comput. Geom.*, 13:415–440, 1995.

[12] Chazal F. and Lieutier A. Smooth manifold reconstruction from noisy and non-uniform approximation with guarantees. *Comp. Geom. Theor. Appl.*, 40(2):156–170, 2008.

[13] H. Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 93, 1959.

[14] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Springer-Verlag, 1990.

[15] K. Grove. Critical point theory for distance functions. In *Proc. of Symposia in Pure Mathematics*, volume 54, 1993.

[16] A. Lieutier. Any open bounded subset of $\mathbb{R}^n$ has the same homotopy type as its medial axis. *Comp. Aid. Geom. Des.*, 36(11):1029–1046, 2004.

[17] Q. Mérigot, M. Ovsjanikov, and L. Guibas. Robust Voronoi-based Curvature and Feature Estimation. In *Proc. SIAM/ACM Joint Conference on Geom. and Phys. Modeling*, pages 1–12, 2009.

[18] P. Niyogi, S. Smale, and S. Weinberger. A Topological View of Unsupervised Learning from Noisy Data. Preprint, 2008.

[19] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39(1):419–441, 2008.

[20] S. Peleg, M. Werman, and H. Rom. A unified approach to the change of resolution: space and gray-level. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(7):739–742, 1989.

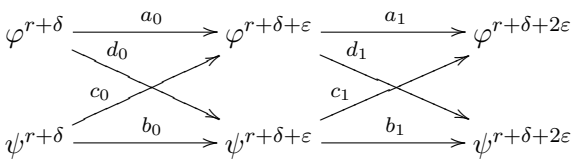
[21] A. Petrunin. Semiconcave functions in Alexandrov's geometry. In *Surveys in differential geometry. Vol. XI*, pages 137–201. Int. Press, Somerville, MA, 2007.

[22] Y. Rubner, C. Tomasi, and L.J. Guibas. The Earth Mover's Distance as a Metric for Image Retrieval. *Int. J. Comput. Vision*, 40(2):99–121, 2000.

[23] C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.

Appendix

Proof of Proposition 4.3 Let $\delta > 0$ be such that $\mathrm{wfs}_\varphi(r) > 2\varepsilon + \delta$ and $\mathrm{wfs}_\psi(r) > 2\varepsilon + \delta$. Since $\|\varphi - \psi\|_\infty \leqslant \varepsilon$, we have the following commutative diagram where each map is an inclusion.



It follows from the isotopy lemma 4.2 that the inclusions a_0, a_1, b_0 and b_1 are homotopy equivalences. Let s_0, s_1, r_0 and r_1 be homotopic inverses of a_0, a_1, b_0 and b_1 respectively. Now a straightforward computation shows that c_1 is an homotopy equivalence with homotopic inverse $r_1 \circ d_1 \circ s_1$:

$$\begin{aligned} c_1 \circ r_1 \circ d_1 \circ s_1 &\cong c_1 \circ (r_1 \circ b_1) \circ d_0 \circ s_0 \circ s_1 \\ &\cong (c_1 \circ d_0) \circ s_0 \circ s_1 \\ &\cong a_1 \circ a_0 \circ s_0 \circ s_1 \cong id_{\varphi^{r+\delta+2\varepsilon}} \end{aligned}$$

Similarly, we get $r_1 \circ d_1 \circ s_1 \circ c_1 \cong id_{\psi^{r+\delta+\varepsilon}}$ proving the proposition 4.3.

Proof of Proposition 4.4 Let $\rho > 0$ and let γ be an integral curve of the flow defined by $\nabla\psi$, starting at x and parametrized by arclength. If γ reaches a critical point of ψ before length ρ, we are done. Assume this is not the case. Then, with $y = \gamma(\rho)$, one has $\psi(y) - \psi(x) = \int_0^\rho \left\|\nabla_{\gamma(t)}\psi\right\| \mathrm{d}t$. As a consequence, there exists a point $p(\rho)$ on the integral curve such that $\left\|\nabla_{p(\rho)}\varphi\right\| \leqslant \frac{1}{\rho}(\varphi(y) - \varphi(x))$.

Now, by the assumption on the uniform distance between φ and ψ, $\psi(y) \leqslant \varphi(y) + \varepsilon$ and $\psi(x) \geqslant \varphi(x) - \varepsilon$. Using the fact that x is α-critical, one obtains:

$$\begin{aligned} &\varphi(y)^2 \leqslant \varphi(x)^2 + 2\alpha \left\|x-y\right\| \varphi(x) + \left\|x-y\right\|^2 \\ \textit{ie. } &\varphi(y) \leqslant \varphi(x) \left(1 + 2\alpha \frac{\left\|x-y\right\|}{\varphi(x)} + \frac{\left\|x-y\right\|^2}{\varphi(x)^2}\right)^{1/2} \leqslant \varphi(x) + \alpha \left\|x-y\right\| + \frac{1}{2}\frac{\left\|x-y\right\|^2}{\varphi(x)} \end{aligned}$$

Putting things together, we get $\left\|\nabla_{p(\rho)}\varphi\right\| \leqslant \alpha + \frac{2\varepsilon}{\rho} + \frac{1}{2}\frac{\rho}{\varphi(x)}$. The minimum of this upper bound is $\alpha + 2\sqrt{\varepsilon/\varphi(x)}$ and is attained for $\rho = 2\sqrt{\varepsilon\varphi(x)}$. This concludes the proof.

Proof of Corollary 4.5 Assume that there exists a critical point x of ψ such that $\psi(x)$ belongs to the range $[4\varepsilon/\alpha^2, R']$. Then, there would exist an α'-critical point y of φ at distance at most D of x. By the previous proposition,

$$\alpha' \leqslant 2\sqrt{\varepsilon/\psi(x)} \leqslant 2\sqrt{\varepsilon/(4\varepsilon/\alpha^2)} = \alpha \text{ and } D \leqslant 2\sqrt{\varepsilon R'}$$

Hence, using the fact that x is a critical point for ψ,

$$\varphi(y) \leqslant \psi(y) + \varepsilon \leqslant \left(\psi^2(x) + \left\|x-y\right\|^2\right)^{1/2} + \varepsilon \leqslant R'\left(1 + D^2/R'^2\right)^{1/2} + \varepsilon \leqslant R' + 3\varepsilon$$

This last term is less than R if $R' < R - 3\varepsilon$. With these values, one gets the desired contradiction.

Centre de recherche INRIA Sophia Antipolis – Méditerranée
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399