HAL open science

Geometric Inference for Measures based on Distance Functions

Frédéric Chazal, David Cohen-Steiner, Quentin Mérigot

▶ **To cite this version:**

Frédéric Chazal, David Cohen-Steiner, Quentin Mérigot. Geometric Inference for Measures based on Distance Functions. [Research Report] RR-6930, 2009. inria-00383685v1

HAL Id: inria-00383685

https://inria.hal.science/inria-00383685v1

Submitted on 13 May 2009 (v1), last revised 23 Jun 2010 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

May 12 2009

Thème SYM

Rapport de recherche

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 — May 12 2009 — 21 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 (Betti numbers, curvatures,...) 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. We also discuss connections between our approach and non parametric density estimation as well as mean-shift clustering.

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 (nombres de Betti, courbures,...) 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. On discute également quelques liens entre l'approche présentée dans ce papier et l'estimation non paramétrique de densité ainsi que l'agorithme de clustering Mean-Shift.

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

I Introduction

The general problem of *geometric inference* can be stated as follows: given a noisy point cloud approximation C of a compact set $K \subseteq \mathbb{R}^d$, how can we recover geometric and topological informations about K, such as its curvature, sharp edges, boundaries, Betti numbers or Euler-Poincaré characteristic, *etc.* knowing only the point cloud C ? Many questions in data analysis can be stated as inferring the geometry of an unknown underlying geometric object: the number of clusters in which a point cloud can be split is related to the number of connected components of this unknown object. Finding out the number of parameters really needed to fully 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 compact set.

Inference using distance functions. By itself, a point cloud carries no geometric or topological information. In order to retrieve information one needs a scale parameter, which allows for instance to connect points which are nearer than some distance threshold. This idea is behind most well-known non-linear dimensionality reduction techniques [20, 18]. A similar idea is to consider a R-offset of the point cloud, instead of the point cloud itself – that is the union of balls of radius R whose center lie in the point cloud. This simple idea recently led to provably good estimation of the topology [4], normal cones [5], and curvature measures [7, 6, 13] of the underlying object. The main tool used in both the geometric and topological results is the notion of *distance function.*

The distance function $\mathrm{d}_K : \mathbb{R}^d \to \mathbb{R}$ to a set K maps a point x in $\mathbb{R}^d$ to its distance to K, defined as the minimum distance between x and any point $y \in K$. It turns out that the inference results mentioned above all depend on three specific properties of distance functions. The most important property of the distance function for geometric inference is its Hausdorff stability:

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

This means that if K' is a good Hausdorff approximation of K, then the distance function $\mathrm{d}_{K'}$ is close to d_K. The two other properties relate to the regularity of these functions:

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

$$|\mathrm{d}_K(x) - \mathrm{d}_K(y)| \leqslant \|x - y\|$$

A consequence of Lipschitz regularity is that the distance function is differentiable almost everywhere (this is Rademacher theorem). This means that the set of non-differentiability points of d_K, which is known to coincide with the *medial axis* of K has zero d-volume.

(ii) d_K^2 **is** 1-***semiconcave***, which is equivalent to the convexity of the map $x \in \mathbb{R}^d \mapsto \|x\|^2 - \mathrm{d}_K^2(x)$. In particular this means, by Alexandrov's theorem [1], that the distance function d_K is not only *almost* $\mathcal{C}^1$, but also twice differentiable almost everywhere. This semiconcavity property is central for the proof of existence of of the flow of the distance function in [12] (cf Lemma 5.1). This flow is one of the main technical tools used in

[4, 5] to prove the stability results. Semiconcavity of the distance function also plays a crucial role in [7, 13]. The main theoretical result proved and used in these papers consists in a L^1 stability property for gradients of uniformly close convex functions.

A non-negative function φ with these two properties, and which is *proper* (meaning that $\lim_{\|x\|\mapsto+\infty} |\varphi(x)| = +\infty$) is called *distance-like*. Most of the Hausdorff-stability results mentioned above can be translated into stability results for uniformly-close distance-like functions.

Contributions. Unfortunately, offset-based methods do not work well at all in the presence of outliers, since *eg.* the number of connected components will be overestimated if one adds just a single data point far from the original point cloud. The problem here is that while the distance function is only slightly perturbed under Hausdorff noise, adding even a single outlier can change it dramatically.

In order to solve this problem, we 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. A way to define what is this *certain amount*, is to 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 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. In this setting, the Hausdorff distance is not meaningful any more. The distance between two measures will be measured by a Wasserstein distance, which quantifies the cost of optimally transporting the first one to the second one (see section II.1). In particular, if S a submanifold of the ambient space, and C is a point cloud uniformly distributed on S, with some noise and a few outliers, then the uniform measure on S and the uniform measure on C are close as measured by the Wasserstein distance.

Our distance functions to measures retain the three properties of the usual distance functions to a compact set that were used in the distance-based geometric inference results. Namely, they are stable, 1-Lipschitz, and 1-semiconcave. In this setting, the stability property should be understood in the sense that if two measures are *Wasserstein-close*, then their distance functions are also uniformly close. As a consequence, we are able to extend the offset-based inference results mentioned above to the case where data may be corrupted by outliers. To conclude the paper, we also discuss connections between our approach and non parametric density estimation as well as mean-shift clustering.

II Background

A *non-negative 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 *additive* in the sense that $\mu\left(\cup_{i\in\mathbb{N}} B_i\right) = \sum_i \mu(B_i)$ whenever (B_i) is a countable family of Borel subsets of $\mathbb{R}^d$. A measure μ is *finite* (resp. a *probability measure*) if $\mu(\mathbb{R}^d) < +\infty$ (resp. $\mu(\mathbb{R}^d) = 1$). The set of measures on $\mathbb{R}^d$ is very wide. However, in most applications all of

the measures we will considered are obtained from a few building blocks (Hausdorff measures) and a few operations (restriction, multiplication by a density, convolution and sum) that we describe below.

Hausdorff measures. The simplest example of a probability measure is the *Dirac mass* at p, denoted by δ_p: $\delta_p(B) = 1$ if p is in B and 0 otherwise. Similarly, a point cloud $C = \{p_1, \dots, p_n\}$ with non-negative weights $\mu_{p_1}, \dots, \mu_{p_n}$ defines a measure $\mu = \sum_{i=1}^n \mu_{p_i}\delta_{p_i}$ defined by $\mu(B) = \sum_{p_i \in B} \mu_{p_i}$. The *uniform measure* on C is defined by $\delta_C = \sum_{p \in C} \delta_p$.

Hausdorff measures of dimension k ($k = 0, \dots, d$) on $\mathbb{R}^d$ is used to formalize the notion of k-volume of a subset of $\mathbb{R}^d$. We refer the reader to [14] for more details. Loosely speaking, $\mathcal{H}^k$ maps any k-dimensional subset of $\mathbb{R}^d$ to its volume, any set of dimension more than k to $+\infty$ and any set of dimension less than k to zero. The Hausdorff measure $\mathcal{H}^0$ coincides with the *counting measure*: $\mathcal{H}^0(B) = \#B$, and $\mathcal{H}^d$ coincides with the usual *Lebesgue mesure* vol_d.

Operations on measures. *Restriction of a measure.* Given a subset $K \subseteq \mathbb{R}^d$ and a measure μ, one can define the restriction $\mu|_K$ of μ to K by the formula $\mu|_K(B) = \mu(K \cap B)$. For instance, if C is a point cloud, $\mathcal{H}^0|_C$ is the uniform measure on C. If S is a segment in $\mathbb{R}^d$, $\mathcal{H}^k|_S$ is the uniform lineic mass distribution on S.

Measures with density. Given a (measurable) function $f : \mathbb{R}^d \to \mathbb{R}^+$ and a measure μ on $\mathbb{R}^d$, such that $\int_{\mathbb{R}^d} f \mathrm{d}\mu < +\infty$, one can define the measure $f\mu$ by the formula $f\mu(B) = \int_B f \mathrm{d}\mu$. If μ is the Lebesgue measure, this gives the usual definition of a measure with density.

Convolution of a measure. A very simple and common noise model is to assume that each sample drawn according to the measure is known up to an independant Gaussian error term. In the measure theoretic setting, this amounts to convolving the measure μ with a Gaussian. Formally, the convolution of a measure μ on $\mathbb{R}^d$ by a (compactly supported, measurable) function $\chi : \mathbb{R}^d \to \mathbb{R}$ is another measure, denoted by $\mu * \chi$ and defined for any set $B \subseteq \mathbb{R}^d$ by the formula:

$$\mu * \chi(B) = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \mathbf{1}_B(x+y) \cdot \chi(x) \mathrm{d}x \right) \mathrm{d}\mu(y)$$

Pushforward of a measure. If $p : X \to Y$ is a measurable function between two measurable spaces X and Y, and μ is a measure on X, $p_\#\mu$ is a measure on Y defined by $(p_\#\mu)(B) = \mu(p^{-1}(B))$.

Empirical measure. We are interested in inferring of geometric properties for a probability measure μ that we know only through finite sampling. Formally this means that we are given a family of independent identically distributed random variables $X_1, \dots X_N$ with common law μ. The uniform probability measure carried by the point cloud $C_N = \{X_1, \dots, X_n\}$ is known as the *empirical measure*, and denoted by $\mu_N = \frac{1}{N}\sum_{i=1}^N \delta_{X_i}$. The theorems concerning the convergence of the empirical measure μ_N to the underlying measure ν are known as uniform law of large numbers.

II.1 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* [21]. The Wasserstein distance W_1 is also known as the earth-mover distance, and has been used in vision [16] and image retrieval [19]. The Wasserstein distance W_2 is related to the problem of optimal least square matching between weighted point clouds [2].

Transport plans between measures. 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 A that will be transported in B by the transport plan. Given a real number $p \geqslant 1$, the p-cost of a transport plan π between μ and ν is $\mathcal{C}_p(\pi) = \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^p \,\mathrm{d}\pi(x,y)\right)^{1/p}$. The p-cost of a transport plan π between μ and ν is defined (and finite) provided that μ and ν both have finite *p-moments*. The p-moment ($p \geqslant 1$) of a measure μ on $\mathbb{R}^d$ is the integral $\int_{\mathbb{R}^d} \|x\|^p \,\mathrm{d}\mu(x)$. We denote by $\mathbb{P}_p(\mathbb{R}^d)$ the set of probability measures on $\mathbb{R}^d$ with finite p-moment. This includes in particular all finite measures with compact support, *e.g.* empirical measures.

Example. We suppose that both μ and ν are supported on point clouds, and can be written as $\mu = \sum_{p \in C} \mu_p \delta_p$ and $\nu = \sum_{q \in D} \nu_q \delta_q$. A transport plan between μ and ν is a measure π in $\mathbb{R}^d \times \mathbb{R}^d$ concentrated on the point cloud $C \times D$ and can be written as:

$$\pi = \sum_{(p,q) \in C \times D} \pi_{p,q} \delta_{(p,q)}$$

with $\sum_{q \in D} \pi_{p,q} = \mu_p$ for all $p \in C$ and $\sum_{p \in C} \pi_{p,q} = \nu_q$ for all $q \in D$. The p-cost of this transport plan is given by:

$$\mathcal{C}_p(\pi) = \left(\sum_{p \in C} \sum_{q \in D} \|p-q\|^p \, \pi_{p,q} \right)^{1/p}$$

Definition. The *Wasserstein distance* between two probability measures μ and ν on $\mathbb{R}^d$ with finite p-moment is the minimum p-cost of a transport plan between these measures:

$$\mathrm{W}_p(\mu, \nu) = \inf\{\mathcal{C}_p(\pi);\ \pi \text{ transport plan between } \mu \text{ and } \nu\}$$

Example.

1. The question of the convergence of the empirical measure μ_N to the underlying measure μ is fundamendal in probability and statistics. 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 more general assumptions can be given, see eg. [3].

2. If $\chi : \mathbb{R}^d \to \mathbb{R}^+$ has finite p-moment $\mathcal{M}_p(\chi) = \int_{\mathbb{R}^d} \|x\|^p \chi(x) \mathrm{d}x$, and $\int_{\mathbb{R}^d} \chi(x) \mathrm{d}x = 1$, then for any probability measure μ, the measure $\mu * \chi$ is Wasserstein-close to μ: $\mathrm{W}_p(\mu, \mu * \chi) \leqslant \mathcal{M}_p(\chi)^{1/p}$.

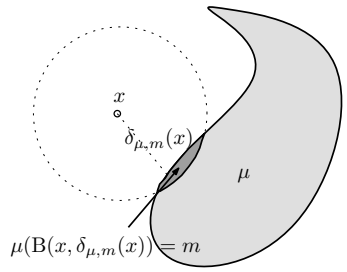

Figure III.1: The pseudo-distance $\delta_{\mu,m}$ to a measure μ

3. Let μ be the uniform probability measure on an hypersurface S embedded in $\mathbb{R}^d$ and denote by $C_{N,\sigma}$ a point cloud of N points drawn independently on the surface, each of them being perturbed by a Gaussian noise of variance σ. By the previous two results, one can prove that the uniform probability measure carried by the point cloud $C_{N,\sigma}$ converges to μ as N converges to $+\infty$ and σ to zero, with high probability, in the Wasserstein sense (for all value p).

   However we should stress that the stability results we obtain for the distance function introduced below do not depend on such a noise model, they just depend on the 2-Wasserstein distance between the two probability measures.

III The distance function to a measure

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: $\mathrm{d}_K(x) = \min_{y \in K} \|x - y\|$. Or, said otherwise, $\mathrm{d}_K(x)$ is the minimum radius of a ball centered at x which contains a point in K: $\mathrm{d}_K(x) = \min\{r > 0; \mathrm{B}(x, r) \cap K \neq \emptyset\}$. A possible idea when trying to define the distance function to a given probability measure μ on $\mathbb{R}^d$ is to mimick the definition above:

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

given a parameter $0 < m < 1$. 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 $\delta_{\mu,m}$ retains some of the features of a distance function. In particular, it is 1-Lipschitz, which means that $|\delta_{\mu,m}(x) - \delta_{\mu,m}(y)| \leqslant \|x - y\|$.

However it is a poor generalization of the usual distance function to a compact. The first reason is that $\delta^2_{\mu,m}(x)$ is not semi-concave. The second, and more important reason is that the map $\mu \mapsto \delta_{\mu,m}$ is not continuous in any reasonable sense. 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, as soon as $\varepsilon > 0$, $\delta_{\mu_\varepsilon,1/2}(t) = |1 - t|$ for $t < 0$, while $\delta_{\mu_0,1/2}(t) = |t|$ for $t < 0$. This proves that $\delta_{\mu_\varepsilon,1/2}$ does not converge for any reasonable topology to $\delta_{\mu_0,1/2}$ even though μ_ε converges to μ_0 for the weak topology on measures.

Definition. For any measure μ with finite second moment, and a positive mass parameter $m_0 > 0$, the distance function to μ with parameter m_0 is defined by

the formula:

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

Example. Let $C = \{p_1, \dots, p_n\}$ be a point cloud, and μ_C be the sum of Dirac masses $\frac{1}{n}\sum_i \delta_{p_i}$. Then, function δ_{μ,m_0} with $m_0 = k/n$ evaluated at $x \in \mathbb{R}^d$ is simply equal to the distance between x and its kth nearest neighbor in C. A *kth order Voronoï cell* of C is a subset $\mathrm{Vor}_C(S)$ of $\mathbb{R}^d$ where the k first nearest neighbors in C are exactly the points in S, where S is a cardinal k subset of C: $x \in \mathrm{Vor}_C(S)$ iff $\forall p_i \notin S, \mathrm{d}(x,p_i) > \mathrm{d}(x,S)$. On such a cell, one obtains the following expression for the distance function to μ_C:

$$\forall x \in \mathrm{Vor}_C(S),\ \mathrm{d}^2_{\mu_C,\frac{k}{n}}(x) = \frac{n}{k}\sum_{p\in S} \|x-p\|^2$$

Note that in this case, the computation of $\mathrm{d}^2_{\mu_C,\frac{k}{n}}(x)$ just reduces to finding the k nearest neigbhors of x.

III.1 Regularity properties for the distance to a measure

In this section we will need a few definitions from measure theory. If μ, ν are two measures on $\mathbb{R}^d$, we will write $\mu \leqslant \nu$ iff for all Borel subset $B \subseteq \mathbb{R}^d$, $\mu(B) \leqslant \nu(B)$. If ν is a measure on $\mathbb{R}^+$ its *cumulative function* is defined by $F_\nu(t) = \nu([0,t[)$. The generalized inverse F_ν^{-1} of F_ν is defined by $F_\nu^{-1} : m \mapsto \inf\{t \in \mathbb{R}\,;\, F_\nu(t) > m\}$. Notice that if μ, ν are two measures on $\mathbb{R}^+$, then $\nu \leqslant \mu$ iff $F_\nu(t) \leqslant F_\mu(t)$ for all $t > 0$.

Equivalent formulation. We start by giving an integral formulation of our distance function.

Proposition III.1. *1. The distance d_{μ,m_0} evaluated at a point $x \in \mathbb{R}^d$ is the minimal cost of the following transportation problem:*

$$\mathrm{d}_{\mu,m_0}(x) = \min_{\tilde\mu}\left\{\mathrm{W}_2\left(\delta_x, \frac{1}{m_0}\tilde\mu\right)\ ;\ \tilde\mu(\mathbb{R}^d) = m_0 \text{ and } \tilde\mu \leqslant \mu\right\}$$

where $\tilde\mu$ is any measure of total mass m_0 such that $\tilde\mu \leqslant \mu$.

2. The set of minimizers coincides with the set $\mathcal{R}_{\mu,m_0}(x)$ of measures $\tilde\mu_{x,m_0}$ whose total mass is m_0 and is contained in the closed ball $\overline{\mathrm{B}}(x, \delta_{\mu,m}(x))$, whose restriction to the open *ball $\mathrm{B}(x,\delta_{\mu,m}(x))$ coincides with μ, and such that $\tilde\mu_{x,m_0} \leqslant \mu$ as defined above.*

3. In particular, for any measure $\tilde\mu_{x,m_0}$ in $\mathcal{R}_{\mu,m_0}(x)$,

$$\mathrm{d}^2_{\mu,m_0}(x) = \frac{1}{m_0}\int_{h\in\mathbb{R}^d} \|h-x\|^2\ \mathrm{d}\tilde\mu_{x,m_0} = \mathrm{W}_2^2\left(\delta_x, \frac{1}{m_0}\tilde\mu_{x,m_0}\right)$$

Proof. 1. Let first remark that if $\tilde\mu$ is any measure of total mass m_0, there is only one transport plan between $\tilde\mu$ and $m_0\delta_x$: namely, the one which maps any point of $\mathbb{R}^d$ to x. Hence,

$$\mathrm{W}_2^2\left(\delta_x, \frac{1}{m_0}\tilde\mu\right) = \int_{\mathbb{R}^d} \|h-x\|^2\,\mathrm{d}\tilde\mu(h)$$

Let $\tilde{\mu}_x$ denote the pushforward of $\tilde{\mu}$ by the distance function to x. Using the property of the cumulative function gives us:

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

The assumption that $\tilde{\mu} \leqslant \mu$ can be rewritten as $F_{\tilde{\mu}_x}(t) \leqslant F_{\mu_x}(t)$ for all $t > 0$, from which one deduces that $F_{\tilde{\mu}_x}^{-1}(m) \geqslant F_{\mu_x}^{-1}(m)$. This gives us

$$\int_{\mathbb{R}^d} \|h-x\|^2 \,\mathrm{d}\tilde{\mu}(h) \geqslant \int_0^{m_0} F_{\mu_x}^{-1}(m)^2 \mathrm{d}m$$

Since by definition $F_{\mu_x}(t) = \mu(B(x,t))$, it follows that $F_{\mu_x}^{-1}(m) = \delta_{\mu,m}(x)$, thus proving

$$\int_{\mathbb{R}^d} \|h-x\|^2 \,\mathrm{d}\tilde{\mu}(x) \geqslant \int_0^{m_0} \delta_{\mu,m}(x)^2 \mathrm{d}m = m_0 \mathrm{d}^2_{\mu,m_0}(x)$$

2. The above inequality is an equality iff $F_{\tilde{\mu}_x}^{-1}(m) = F_{\mu_x}^{-1}(m)$ for almost every $m \leqslant m_0$. Since these function are increasing and left-continuous, equality must in fact hold for every such m. In particular, one deduces that $\tilde{\mu}(\overline{\mathrm{B}}(x, \delta_{\mu,m_0}(x))) = m_0$, which means that all the mass of $\tilde{\mu}$ is in the closed ball $\overline{\mathrm{B}}(x, \delta_{\mu,m_0}(x))$ and $\tilde{\mu}(B(x, \delta_{\mu,m_0}(x))) = \mu(\mathrm{B}(x, \delta_{\mu,m_0}(x)))$. These property together with the inequality $\tilde{\mu} \leqslant \mu$ proves that $\tilde{\mu}$ is a minimizer iff it belongs to $\mathcal{R}_{\mu,m_0}(x)$.

To finish the proof (of *1.*), we should remark that the set of minimizer $\mathcal{R}_{\mu,m_0}(x)$ is never empty, *ie.* contains a measure μ_{x,m_0}. In the case where $\mu(\overline{\mathrm{B}}(x, \delta_{\mu,m}(x))) = m$, we simply let μ_{x,m_0} be the restriction of μ to the closed ball $\overline{\mathrm{B}}(x, \delta_{\mu,m_0}(x))$. When however the boundary of the ball carries too much mass, 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:

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

□

Semiconcavity of the squared distance. We remind the reader of following facts and definitions from convex analysis (see for instance [8]). A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be L-concave (resp. L-convex) if the function $x \mapsto L\|x\|^2 - f(x)$ is convex (resp. concave). If $f : \Omega \subseteq \mathbb{R}^d \to \mathbb{R}$ is a function, its subdifferential at a point x, denoted by $\partial_x f$ is the set of vectors v of $\mathbb{R}^d$ such that for all $h \in \mathbb{R}^d$ small enough, $f(x+h) \geqslant f(x) + \langle h|v\rangle$. The subdifferential $\partial_x f$ is a convex subset of $\mathbb{R}^d$. The function f is (locally) convex iff for any point x in Ω, $\partial_x f$ is not empty. A convex function f admits a derivative at x iff $\partial_x f = \{v\}$ is a singleton, in which case $\nabla_x f = v$. A convex function is derivable almost everywhere (hence, so is a semi-convex function).

Proposition III.2. *For any $x \in \mathbb{R}^d$ the subdifferential of the function $v_{\mu,m_0} : x \in \mathbb{R}^d \mapsto \|x\|^2 - \mathrm{d}^2_{\mu,m_0}$ at a point $x \in \mathbb{R}^d$ is*

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

As a consequence:

1. *v_{μ,m_0} is convex, and d^2_{μ,m_0} is semiconcave ;*
2. *d^2_{μ,m_0} is differentiable at a point $x \in \mathbb{R}^d$ iff the support of the restriction of μ to the sphere $\partial \mathrm{B}(x, \delta_{\mu,m_0}(x))$ contains at most 1 point;*
3. *d^2_{μ,m_0} is differentiable almost everywhere in $\mathbb{R}^d$, with gradient defined by*

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

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

Proof. For any two points x and y of $\mathbb{R}^d$, $\tilde{\mu}_{x,m_0}$ and $\tilde{\mu}_{y,m_0}$ in $\mathcal{R}_{\mu,m_0}(x)$ and $\mathcal{R}_{\mu,m_0}(y)$ respectively, we can use proposition III.1 to get the following sequence of equalities and inequalities:

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

with $V = \frac{2}{m_0} \int_{h \in \mathbb{R}^d} [x-h] \, \mathrm{d}\tilde{\mu}_{x,m_0}(h)$. We rewrite this as:

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

This proves that $2x - V$ is in the subdifferential of v at x; hence the function $v : x \mapsto \|x\|^2 - \mathrm{d}^2_{\mu,m_0}$ is convex, as announced. We now turn to the proof of the converse inclusion. The two sets

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

are both convex, and we have $\mathcal{D}_{\mu,m_0} \subseteq \partial_x v_{\mu,m_0}$. The subdifferential $\partial_x v_{\mu,m_0}$ coincides with the convex envelope of the set of limits $\lim_{x_n \to x} \nabla_{x_n} v_{\mu,m_0}$, where x_n is a sequence of points converging to x at which v_{μ,m_0} is differentiable (*cf. eg.* [8], Th 2.5.1). Hence we only need to prove that every such limit also belong to $\mathcal{D}_{\mu,m_0}(x)$. We let μ_n be the unique element in $\mathcal{R}_{\mu,m_0}(x_n)$; since v is differentiable at x_n, necessarily $\nabla_{x_n} v_{\mu,m_0} = 2x_n - 2/m_0 \int_h (x_n - h) \mathrm{d}\mu_n(h)$. Using Prokhorov theorem, we can extract a subsequence of n such that μ_n weakly converges to a measure μ_∞. This measure belong to $\mathcal{R}_{\mu,m_0}(x)$, and if one sets $D = 2x - 2/m_0 \int_h (x-h) \mathrm{d}\mu_\infty(h)$, by weak convergence of μ_n to μ_∞, one sees that $\nabla_{x_n} v_{\mu,m_0}$ converges to $D \in \mathcal{D}_{\mu,m_0}(x)$. This concludes the proof of this inclusion.

2. It is enough to remark that $\mathcal{R}_{\mu,m}(x)$ is a singleton iff the support of $\mu|_{\partial B(x,\delta_{\mu,m_0}(x))}$ is at most a single point.

3. From the two convex analysis facts reminded above, we know that $\mathrm{d}^2_{\mu,m_0}(y)$ is differentiable at almost every point of $x \in \mathbb{R}^d$ with gradient the only element of the subdifferential at that point.

4. The gradient of the (unsquared) 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. □

Distance-like functions. We call *distance-like* 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 $\lim_{\|x\|\to+\infty}\varphi(x) = +\infty$. Recall that a function $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ is called 1-semiconcave or 1-concave iff $\|x\|^2 - \varphi(x)$ is convex.

Proposition III.3. *Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a function whose square is 1-semiconcave. Then, there exists a closed set $K \subseteq \mathbb{R}^{d+1} = \mathbb{R}^d \times \mathbb{R}$ such that $\varphi^2(x) = \mathrm{d}^2_K(x)$, where x 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 $v' = v/2$. By 1-semiconcavity,

$$\psi_v : y \mapsto \varphi^2(x) - \|v'\|^2 + \|x - v' - y\|^2 \geqslant \varphi^2(y)$$

The function φ^2 is the lower envelope of all the ψ_v as defined above. Letting $y = x - v'$, we see that $\varphi^2(x) - \|v'\|^2 \geqslant 0$. This means that if we set $z = (x - v', (\varphi^2(x) - \|v'\|^2)^{1/2}) \in \mathbb{R}^{d+1}$, $\psi_v(x)$ is equal to the squared Euclidean distance between $(x, 0)$ and z in $\mathbb{R}^{d+1}$. Hence, φ 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. From the proof one also sees that *distance-like* functions are simply generalized power distances, with non-positive weights.

III.2 Stability of the distance function to a measure

Lipschitz stability of the map $\mu \mapsto \mathrm{d}_{\mu,m_0}$. The goal of this section is to prove that the map $\mu \to \mathrm{d}_{\mu,m_0}$ is $m_0^{-1/2}$–Lipschitz, where the space of measures is endowed with the Wasserstein distance of exponent 2 and the space of distance functions with the uniform norm $\|.\|_\infty$. We will use the fact that on the real line optimal transport planes are nothing but monotone rearrangements [21, Theorem 2.18]. This implies in particular the following result:

Theorem III.4. *If μ and ν are two measures with equal mass on $\mathbb{R}$, and F_μ and F_ν are their respective cumulative functions, then*

$$\mathrm{W}_2(\mu,\nu) = \left\|F_\mu^{-1} - F_\nu^{-1}\right\|_{\mathrm{L}^2([0,1])}$$

Lemma III.5. If μ and ν are two compactly supported probability measures on $\mathbb{R}^+$, then

$$\left|\int_0^{m_0} \delta^2_{\mu,m}(0)\mathrm{d}m - \int_0^{m_0} \delta^2_{\nu,m}(0)\mathrm{d}m\right| \leqslant \mathrm{W}_2(\mu,\nu)$$

Proof. Let us first remark that, since μ has no mass on $\mathbb{R}_*^-$,

$$\begin{aligned}\delta_{\mu,m}(0) &= \inf\{t\,;\,\mu(]-t,t[) > m\} \\ &= \inf\{t\,;\,\mu(]-\infty,t[) > m\} = F_\mu^{-1}(m)\end{aligned} \tag{III.1}$$

Using this equation and the previous theorem,

$$\begin{aligned}\left|\|\delta_{\mu,.}(0)\|_{\mathrm{L}^2(0,m_0)} - \|\delta_{\nu,.}(0)\|_{\mathrm{L}^2(0,m_0)}\right| &\leqslant \|\delta_{\mu,.}(0) - \delta_{\nu,.}(0)\|_{\mathrm{L}^2(0,m_0)} \\ &\leqslant \|\delta_{\mu,.}(0) - \delta_{\nu,.}(0)\|_{\mathrm{L}^2(0,1)} \\ &= \left\|F_\mu^{-1} - F_\nu^{-1}\right\|_{\mathrm{L}^2(0,1)} \\ &= \mathrm{W}_2(\mu,\nu)\end{aligned}$$

□

Theorem III.6 (distance function stability theorem). *If μ and ν are two probability measures on $\mathbb{R}^d$ and $m_0 > 0$, then $\|\mathrm{d}_{\mu,m_0} - \mathrm{d}_{\nu,m_0}\|_\infty \leqslant \frac{1}{\sqrt{m_0}}\mathrm{W}_2(\mu,\nu)$.*

Proof. For any point x in $\mathbb{R}^d$, we denote by d_x the distance function to x. We denote by μ^x the push forward of a measure μ by d_x. First remark that any transport plan π between μ and ν can also be pushed to obtain a transport plan $\pi_x := (\mathrm{d}_x, \mathrm{d}_x)_\#\pi$ between μ_x and ν_x. Using the triangle inequality we get:

$$\begin{aligned}\int_{(t,t')\in\mathbb{R}\times\mathbb{R}} |t-t'|^2 \,\mathrm{d}\pi_x(t,t') &= \int_{(g,h)\in\mathbb{R}^d\times\mathbb{R}^d} |\mathrm{d}_x(g) - \mathrm{d}_x(h)|^2 \,\mathrm{d}\pi(g,h) \\ &\leqslant \int_{(g,h)\in\mathbb{R}^d\times\mathbb{R}^d} \|g-h\|^2 \,\mathrm{d}\pi(g,h)\end{aligned}$$

This proves that $\mathrm{W}_2(\mu^x,\nu^x) \leqslant \mathrm{W}_2(\mu,\nu)$. We let μ_{x,m_0} and ν_{x,m_0} be as in Prop III.1.(2), *ie.* $\mathrm{d}^2_{\mu,m_0}(x) = \frac{1}{m_0}\int_{\mathbb{R}^+} t^2 \mathrm{d}\mu^x_{x,m_0}(t)$. Using the previous lemma, we get:

$$|\mathrm{d}_{\mu,m_0}(x) - \mathrm{d}_{\nu,m_0}(x)| \leqslant \frac{1}{\sqrt{m_0}}\mathrm{W}_2(\mu^x,\nu^x)$$

□

Distance functions to two measures with different masses. For simplicity, and also because the empirical measure is always a probability measure (we cannot estimate the mass from a set of samples), we have always supposed that the two measures have the same mass. In this paragraph we show how to obtain one-sided Lipschitz inequalities when one of the measure has a smaller mass than the other.

First remark that the distance function can be defined for any measure without change, whatever its mass (as soon as m_0 is smaller than this mass). One easily checks that the distance function decreases under the addition of another positive measure, *ie.* $\mathrm{d}_{\mu+\delta,m_0} \leqslant \mathrm{d}_{\mu,m_0}$.

Corollary III.7. *Let μ be a measure, and ν another measure with* $\mathrm{mass}(\nu) \leqslant \mathrm{mass}(\mu)$. *We suppose that there exists a transport plan π defined on $\mathbb{R}^d \times \mathbb{R}^d$ such that $\mathrm{p}_{1\#}\pi = \nu$ and $\mathrm{p}_{2\#}\pi \leqslant \mu$ and with*

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

Then,

$$\mathrm{d}_{\mu,m_0} \leqslant \mathrm{d}_{\nu,m_0} + m_0^{-1/2}\varepsilon$$

Proof. In this case, $\mu' = \mathrm{p}_{2\#}\pi$ has the same mass as ν and is Wasserstein-close to ν. Moreover, $\mu - \mu'$ is a positive measure and $\mu = \mu' + (\mu - \mu')$. Hence, by the remark, $\mathrm{d}_{\mu,m_0} \leqslant \mathrm{d}_{\mu',m_0}$. Using the previous theorem, we get:

$$\mathrm{d}_{\mu,m_0} \leqslant \mathrm{d}_{\nu,m_0} + m_0^{-1/2}\mathrm{W}_2(\nu,\mu') \leqslant \mathrm{d}_{\nu,m_0} + m_0^{-1/2}\varepsilon$$

□

IV Applications

IV.1 Offset Reconstruction

Reconstruction from point clouds with outliers was one of the motivation for introducing the distance function to a measure. We show how to use distance functions to achieve these goals in three steps that are outlined in the sequel of this section. In the first step we study for a measure μ with compact support S the convergence of the function d_{μ,m_0} to the distance function d_S to S as m_0 goes to 0. In the second step we explain how the results in [4], which allow to compare the topology of the offsets of two Hausdorff-close compact sets, can be generalized to compare the sub-level sets of two uniformly-close distance-like functions. At last, we explain how the results of the first two steps can be combined to get reconstruction results from noisy point clouds with outliers. Note that a complete exposition of the application of the results of previous sections to reconstruction is beyond the scope of this paper and will be detailed in a future work.

Suppose we are given a manifold S, and the uniform measure μ on it. Using the generalized version of the results of [4] allows us to compare the topology of the sublevel sets of the distance d_S to the topology of the sublevel sets of the distance d_{μ,m_0}, when they are uniformly close.

Uniform convergence of d_{μ,m_0} ***to the distance function to its support.*** Let μ be a probability measure with compact support S. Our goal is to show that the distance functions to a measure d_{μ,m_0} converge uniformly to the distance to S as m_0 goes to 0. First remark that δ_{μ,m_0} is always larger than d_S. As a consequence it is sufficient to bound d_{μ,m_0} from above by $\mathrm{d}_S + \varepsilon(m_0)$ where $\varepsilon(m_0)$ tends to 0 as m_0 tends to 0.

Lemma IV.1. Suppose that given $m_0 > 0$, there exists a positive ε such that the μ-volume of ε-balls whose center is a point of S is uniformly bounded from below by m_0, *ie.* $\forall p \in S,\ \mu(\mathrm{B}(p,\varepsilon)) \geqslant m_0$. Then, $\|\mathrm{d}_{\mu,m_0} - \mathrm{d}_S\|_\infty$ is at most ε.

Proof. Let x be a point in $\mathbb{R}^d$, p a projection of x on S. By the 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 gives us the lemma. □

More generally, suppose there exists a non-decreasing positive function $f : \mathbb{R}^+ \to \mathbb{R}^+$ which uniformly bounds from below the μ-volume of balls whose center is in the support of μ :

$$\forall p \in S,\ \mu(\mathrm{B}(p,\varepsilon)) \geqslant f(\varepsilon)$$

Using the previous lemma, we get $\|\mathrm{d}_{\mu,m_0} - \mathrm{d}_S\|_\infty \leqslant \varepsilon$ provided that $m_0 \leqslant f(\varepsilon)$. Hence, $\lim_{m_0\to 0} \|\mathrm{d}_{\mu,m_0} - \mathrm{d}_S\|_\infty = 0$.

Lemma IV.2. If μ is a compactly-supported measure, then d_S is the uniform limit of d_{μ,m_0} as m_0 converges to 0.

Proof. Let $x_1, \ldots, x_n$ be a finite $\varepsilon/2$-sample of S – *ie.* $S \subseteq \cup_i \mathrm{B}(x_i, \varepsilon/2)$, and x_i is in S. By definition of the support of a measure, $\eta = \min_i \mu(\mathrm{B}(x_i, \varepsilon/2))$ 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 $\mu(\mathrm{B}(x,\varepsilon)) > 0$. □

This result, while proving the convergence of d_{μ,m_0} to d_S in a very general setting, does not give quantitative bounds on the speed of convergence.

The convergence speed of d_{μ,m_0} to d_S depends on the way the mass of μ contained within a ball $\mathrm{B}(x,r)$ ($x \in S$) decreases with r. If the measure μ has dimension at most $k > 0$, i.e. if there exists some constant C such that $\mu(\mathrm{B}(x,\varepsilon)) \geqslant C\varepsilon^k$ as soon as ε is small enough, then

$$\|\mathrm{d}_{\mu,m_0} - \mathrm{d}_S\| = O(m_0^{1/k}) \tag{IV.2}$$

When μ is the uniform probability measure on a k-dimensional compact submanifold S without boundary it has dimension at most k. In fact, we can give even more quantitative convergence speed estimates by using the Günter-Bishop theorem, which bounds the volume of intrinsic balls on S from below provided that its sectional curvature is upper bounded.

Theorem IV.3 (Günther-Bishop, [10, section 3.101]). *If the sectional curvatures of a Riemannian manifold M do not exceed δ, then for every $x \in M$,*

$$\mathrm{V}^k(\mathrm{B}_M(x,r)) \geqslant \mathrm{V}^k_\delta(r)$$

where $V^k_\delta(r)$ is the volume of a ball of radius r in the simply connected k-dimensional manifold with constant sectional curvature δ, provided that $r \leqslant \min(\mathrm{injrad}(M), \pi/\sqrt{\delta})$.

If R denote the minimum curvature radius of S, the sectional curvature above is bounded from above by $1/R^2$, and we can apply the previous result.

Proposition IV.4. *If S is a smooth k-dimensional submanifold of $\mathbb{R}^d$ whose curvature radius is at least R, and μ the uniform probability measure on S, then*

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

as soon as ε is smaller than the injectivity radius of S and πR. In particular, the dimension of μ is at most k.

Proof. 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 greater than $\mathrm{V}(\mathrm{B}_S(x,\varepsilon))/\mathrm{V}(S)$. Using the Günter-Bishop inequality we get the desired lower bound. □

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.

Extending the sampling theory for compact sets to distance-like functions. In this section we show how the sampling theory of [4] for distance functions to compact sets extends to the case of close distance-like functions. The approach in our setting follows the one of [4] and the proofs are very similar to the original ones. We refer the interested reader to [4] for more details.

Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a *distance-like* function. A point $x \in \mathbb{R}^d$ will be called α-critical (with $\alpha \in [0,1]$) if the inequality $\varphi^2(x+h) \leqslant \varphi^2(x)+\alpha\,\|h\|\,\varphi(x)+\|h\|^2$ is true for all $h \in \mathbb{R}^d$. A 0-critical point is simply called a critical point. This notion of α-critical point has been introduced for the distance function to a compact set in [4], where its geometric meaning is explained. The semi-concavity of φ allows to define a notion of gradient vector field $\nabla_x\varphi$ for φ such that for any $x \in \mathbb{R}^d$ $\|\nabla_x\varphi\|$ is the infimum of the $\alpha \geqslant 0$ such that x is α-critical. Although not continuous, the vector field $\nabla\varphi$ is sufficiently "regular" to be integrated in a continuous locally Lipschitz flow [17]. As a consequence many results for distance functions can be directly adapted to our setting. For example the Isotopy Lemma for distance functions [11, Prop. 1.8] can be proved for distance-like functions as well:

Proposition IV.5. *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 $r \in [r_1,r_2]$.*

As a consequence the topology of the sublevel sets of φ can only change when one passes critical values. In the same way, the Critical Point Stability Theorem for distance functions which is the key result of the sampling theory introduced in [4] generalizes almost verbatim to the case of distance-like functions:

Theorem IV.6. *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*

$$\|x-x'\| \leqslant 2\sqrt{\varepsilon\varphi(x)} \text{ and } \alpha' \leqslant \alpha + 2\sqrt{\varepsilon/\varphi(x)}$$

From this result, it is then possible to adapt most of the topological and geometric inference results of [4, 5, 6]. The statement and the proofs of these results is beyond the scope of this paper and will be done in a future work.

Shape reconstruction from noisy data. The previous results lead to shape reconstruction theorems from noisy data with outliers. To fit in the framework of this paper the shapes to reconstruct are 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, one has

$$\begin{aligned} \|d_K - d_{\mu',m_0}\|_\infty &\leqslant \|d_K - d_{\mu,m_0}\|_\infty + \|d_{\mu,m_0} - d_{\mu',m_0}\|_\infty && \text{(IV.3)} \\ &\leqslant O(m_0^{\frac{1}{k}}) + \frac{1}{\sqrt{m_0}} W_2(\mu,\mu') && \text{(IV.4)} \end{aligned}$$

This inequality insuring the closeness of d_{μ',m_0} to the distance function d_K for the sup-norm plays an important role to prove reconstruction results. It follows immediately from the stability theorem III.6 and the equation (IV.2). In [4], the notion of α-reach of K, $\alpha \in (0,1]$, is defined by $r_\alpha(K) = \inf\{d_K(x) > 0 : \|\nabla_x d_K\| < \gamma\}$ and it is proven that the offsets of compact sets with positive α-reach can be correctly reconstructed from sufficiently close approximations for the Hausdorff distance. The following theorem extends this result to our framework (see theorem 4.6 in [4]).

Theorem IV.7. *Let μ be a probability measure of dimension at most $k > 0$ with compact support $K \subset \mathbb{R}^d$ such that $r_\alpha(K) > 0$ for some $\alpha \in (0,1]$. For any $0 < \eta < r_\alpha(K)$, there exists a positive constants $m_1 = m_1(\mu,\alpha,\eta) > 0$ and $C = C(m_1) > 0$ such that:*
for any $m_0 < m_1$ and any probability measure μ' such that $W_2(\mu,\mu') < C\sqrt{m_0}$, the sublevel set $d^{-1}_{\mu',m_0}((-\infty,\eta])$ is homotopy equivalent to the offsets $d_K^{-1}([0,r])$ of K for $0 < r < r_\alpha(K)$.

Using the above inequality (IV.3) that bounds the sup-norm of $d_K - d_{\mu',m_0}$ the proof of this theorem follows almost verbatim the proof of theorem 4.6 in [4]. From the proof, it is possible to provide a more precise, but more technical, statement giving the bounds. It is also possible to prove that the offsets of K and the sublevel sets of d_{μ',m_0} are indeed isotopic as in [5]. Such results are beyond the scope of this paper and will be developed in a future work.

Niyogi, Smale and Weinberger [15] recently proved a related but different reconstruction result for point clouds sampled around a smooth manifold according to a distribution allowing some specific non local noise. Their approach consists in eliminating the outliers to obtain a subsample which is close to the smooth manifold for the Hausdorff distance so that the "classical" reconstruction results can be applied.

The figure IV.2 illustrates the reconstruction theorem IV.7 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.

IV.2 Distances to measures and 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. If C is a point cloud whose points are drawn with respect to a given probability measure with density, the density is estimated by

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

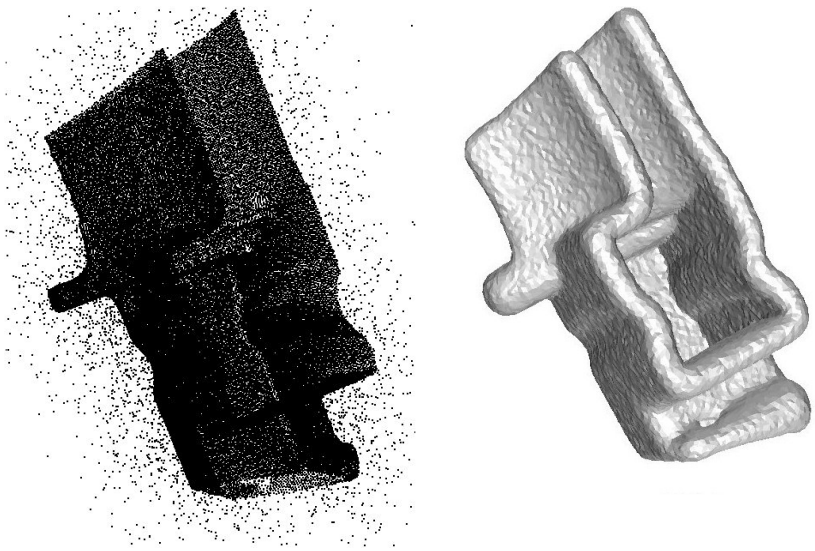

Figure IV.2: On the left, a point cloud sampled on a mechanical part (`blade`) to which 10% of outliers have been added — the outliers are uniformly distributed in a box enclosing the original point cloud. On the right, the reconstruction of an isosurface of the distance function d_{μ_C,m_0} to the uniform probability measure on this point cloud (obtained by using the CGAL Surface Mesher).

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

For some of the applications where density estimators are used (eg. mean shift, see below), distance functions could also be used as well. The advantages of the distance function d_{μ,m_0} over kNN density estimators are multiple. First of all, this distance is well defined even when the underlying probability measure does not have a density, eg. if it is concentrated on a lower-dimensional subset. Second, the distance is always stable with respect to Wasserstein perturbations of the data. Third, the distance function d_{μ,m_0}, being 1-semiconcave, is much more regular than the distance δ_{μ,m_0}, as illustrated by Figure IV.2.

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.

IV.3 Mean-Shift Methods using Distance Functions

Kernel-based mean-shift clustering. Mean-shift clustering [9] 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

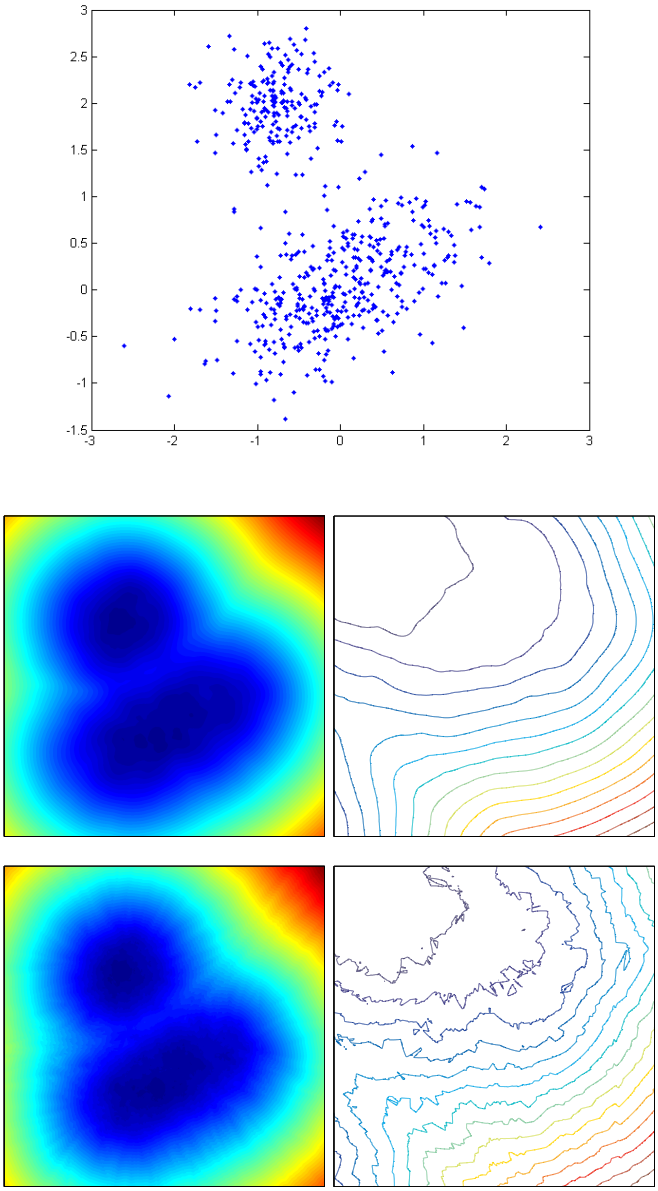

Figure IV.3: The distance functions to the measure $\mu = \frac{1}{600}\sum_{i=1}^{600}\delta_{p_i}$ associated to a 600 points 2D data set $P = \{p_1, \cdots p_{600}\}$ sampled according two gaussians (top figure). The four bottom figures represent the distance functions $\mathrm{d}_{\mu,1/30}$ (top) and $\delta_{\mu,1/30}$ (bottom). The right figures representing the details of some level sets of $\mathrm{d}_{\mu,1/30}$ (top) and $\delta_{\mu,1/30}$ (bottom) illustrate the difference of regularity between the two distance functions.

by:

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

Figure IV.4: 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.

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

$$x_0 = x \text{ and } x_{i+1} = \frac{\sum_{p\in C} K\left(\frac{p-x}{h}\right) p}{\sum_{p\in C} K\left(\frac{p-x}{h}\right)}$$

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:

$$x_0 = x \text{ and } x_{i+1} = x_i - \frac{1}{2}\nabla_{x_i} \mathrm{d}^2_{\mu,m_0}$$

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 IV.8. *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. III.1.(1).
2. Since d^2_{μ,m_0} is 1-concave,

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

Now, if we set $y = x_t$,

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

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

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 has been proved for classical mean-shift, when the kernel K has convex and monotonically decreasing profile [9, Theorem 1] and the second one when K is Gaussian [9, Theorem 2] (in which case it is not convex). However, we are not aware of any choice of kernel such that the resulting mean-shift scheme satisfies these two properties simultaneously.

References

[1] A. D. Alexandrov. Almost everywhere existence of second differentials of convex functions and convex surfaces connected with it. *Leningrad State Univ. Ann. (Uchenye Zapiski) Math. Ser.*, 6:3–35, 1939.

[2] F. Aurenhammer, F. Hoffmann, and B. Aronov. Minkowski-Type Theorems and Least-Squares Clustering. *Algorithmica*, 20(1):61–76, 1998.

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

[4] F. Chazal, D. Cohen-Steiner, and A. Lieutier. A Sampling Theory for Compacts in Euclidean Space. *Proc. 22nd Annu. ACM Sympos. Comput. Geom*, 2006.

[5] F. Chazal, D. Cohen-Steiner, and A. Lieutier. Normal cone approximation and offset shape isotopy. *Computational Geometry: Theory and Applications*, 2008.

[6] F. Chazal, D. Cohen-Steiner, A. Lieutier, and B. Thibert. Stability of Curvature Measures. 2008.

[7] F. Chazal, D. Cohen-Steiner, and Q. Mérigot. Stability of boundary measures. *Arxiv preprint arXiv:0706.2153*, 2007.

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

[9] 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.

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

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

[12] A. Lieutier. Any open bounded subset of $\mathbb{R}^n$ has the same homotopy type as its medial axis. *Computer-Aided Design*, 36(11):1029–1046, 2004.

[13] Q. Merigot, M. Ovsjanikov, and L. Guibas. Robust Voronoi-based Curvature and Feature estimation. *European Workshop on Computational Geometry*, 2009.

[14] F. Morgan. *Geometric Measure Theory: A Beginner's Guide.* Academic Press, 1988.

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

[16] S. Peleg, M. Werman, and H. Rom. A unified approach to the change of resolution: space and gray-level. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):739–742, 1989.

[17] A. Petrunin. Semiconcave functions in Alexandrov's geometry. *Preprint*.

[18] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding, 2000.

[19] Y. Rubner, C. Tomasi, and L. Guibas. The Earth Mover's Distance as a Metric for Image Retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

[20] J. Tenenbaum, V. Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction, 2000.

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

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