Geometric Inference for Probability Measures

**Frédéric Chazal · David Cohen-Steiner ·
Quentin Mérigot**

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. It appears that the study of distance functions allows one 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 or 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 $\mathbb{R}^d$. These functions share many properties with classical distance functions, which make them suitable for inference purposes. In particular, by considering appropriate level sets of these distance functions, we show that it is possible to reconstruct offsets of sampled shapes with topological guarantees even in the presence of outliers. Moreover, in settings where empirical measures are considered, these functions can be easily evaluated, making them of particular practical interest.

F. Chazal (✉)
INRIA Saclay, Saclay, France
e-mail: frederic.chazal@inria.fr

D. Cohen-Steiner
INRIA Sophia-Antipolis, Nice, France

Q. Mérigot
Laboratoire Jean Kuntzmann, CNRS/Université de Grenoble I, Grenoble, France

1 Introduction

Extracting geometric and topological information from geometric data, such as three-dimensional (3D) point clouds obtained from laser scanners, is a requirement for many geometric processing and data analysis algorithms. The need for robust estimation of geometric invariants was recognized long ago in geometric processing, and such invariants have found applications in fields as different as shape matching, registration, and symmetry detection in 3D models, or more generally structure discovery, reconstruction, and meshing, to name just a few. More recently, it has become apparent that such geometric and topological quantities could also be used to analyze more general data sets, e.g., those of computational structural biology and large image databases. 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 into 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 those 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 centers lie in the point cloud. It has been recently proven in [4, 21] 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 nonsmooth compact sets by [7]. When one is only interested in the homology, the offset-based approach can be combined with persistent topology [14] to correctly infer the Betti numbers of the unknown set [3, 11, 24].

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, [13] 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 one to compute the Delaunay triangulation of C, and is hence impractical in higher dimensions. However, in this case efforts have been made to introduce lighter complexes [12]. For example, it has been proven that one can resort to Vietoris-Rips complexes and the theory of topological persistence to correctly infer the Betti numbers of offsets of C [5].

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 a *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 (to be described 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 [8, 9, 19]. 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. In other words, 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 number of outliers. To define this *certain number*, one can change the way point clouds are interpreted: they are not just 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 the Wasserstein distance, which quantifies the minimal cost of transporting one measure onto the other (cf. Sect. 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, the 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. The function 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 the main existing previous work on the subject [20], 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., comparisons between discrete and continuous representations, we cannot give the definitions and theorems only in the discrete case; we must also deal with the general case of probability measures.

2.1 Measure Theory

A measure μ on the space $\mathbb{R}^d$ is 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 $\mathbb{R}^d$, $\mu(\bigcup_{i \in \mathbb{N}} B_i) = \sum_i \mu(B_i)$. 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 $\mu_C(B) = \frac{1}{N} |B \cap C|$. Equivalently, it is the sum of N Dirac masses of weight $1/N$, centered at each point of C. Given a family of independent random points $X_1, \ldots, X_N$, 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 an *empirical measure*, and is 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. We start with the probability measure μ_M on a compact k-dimensional manifold $M \subseteq \mathbb{R}^d$ given by the rescaled uniform measure on M. Such measures can be combined, yielding a measure supported on a union of submanifolds of $\mathbb{R}^d$ with various intrinsic dimensions: $\nu = \sum_{i=1}^{\ell} \lambda_i \mu_{M_i}$, with $\sum_i \lambda_i = 1$. Finally, as a simple model of noise, this measure can be convolved with a Gaussian distribution: $\mu = \nu * \mathcal{N}(0, \sigma^2)$. This is the same as assuming that each sample that is drawn according to ν is known up to an independent 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 $i \in \{0, \ldots, \ell\}$ with probability λ_i, (ii) picking a random sample X_n uniformly distributed in M_i, and (iii) adding a random Gaussian vector with zero mean and variance σ^2 to X_n.

2.2 Wasserstein Distances

The definition of the Wasserstein W_p ($p \geq 1$) distance between probability measures relies on the notion of a transport plan between measures. It is related to the theory of *optimal transportation* (see, e.g., [26]). The Wasserstein distance W_1 is also known as the earth mover's distance; it has been used in vision by [22] and in image retrieval by [25] 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 $p \geq 1$, the p-cost of such a transport plan π is given by

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

This cost is finite if measures μ and ν both have finite *p-moments*, i.e., $\int_{\mathbb{R}^d} \|x\|^p \, d\mu(x) < +\infty$ and $\int_{\mathbb{R}^d} \|x\|^p \, d\nu(x) < +\infty$. 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 moving n points in C a large distance R from their original position. The Wasserstein distance between the uniform measures μ_C and μ is at most $(\frac{n}{N})^{\frac{1}{p}} R$. 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' could be as large as 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 fundamental 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 statements 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 $\sigma^p := \int_{\mathbb{R}^d} \|x\|^p \chi(x) \, dx$, the Wasserstein distance of order p between any probability measure μ and the convolved measure $\mu * \chi$ satisfies $W_p(\mu, \mu * \chi) \leq \sigma$. If one considers again the example given in the end of Sect. 2.1 of an empirical measure μ_N whose samples are drawn according to a "geometric" measure ν convolved with a Gaussian distribution $\mathcal{N}(0, \sigma^2)$, the combination of the two previous facts gives

$$\lim_{N \to +\infty} W_2(\mu_N, \mu) \leq \sigma \quad \text{with probability one.}$$

Similar bounds are also possible with convolution kernels that are not translation invariant, such as the ones defining the noise model used in [20]. This being said, we 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 following two properties:

1. d_K is 1-Lipschitz. For all x, y in $\mathbb{R}^d$, $|d_K(x) - d_K(y)| \leq \|x - y\|$.
2. d_K^2 is 1-semiconcave. This means that the map $x \in \mathbb{R}^d \mapsto d_K^2(x) - \|x\|^2$ is concave.

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 points where d_K is not differentiable, has zero d-volume. Semiconcavity is a stronger regularity property, because, due 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 [18] (Lemma 5.1), which is the main technical tool used in the topological inference results obtained by [7]. The semiconcavity of the squared distance function also plays a crucial role in geometric inference results such as [9] and [19].

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 $\varphi^2(x) = d_K^2(x)$, where a point x in $\mathbb{R}^d$ is identified with the point $(x, 0)$ in $\mathbb{R}^{d+1}$.*

Proof Since φ^2 is 1-semiconcave, we can write $\|x\|^2 - \varphi^2(x) = \sup_{i \in I} \langle v_i, x \rangle - a_i$ for some vectors $(v_i)_{i \in I}$ and reals $(a_i)_{i \in I}$, where I indexes the set of affine functions smaller than the l.h.s. Hence

$$\varphi^2(x) = \inf_{i \in I} \|x\|^2 - \langle v_i, x \rangle + a_i = \inf_{i \in I} \|x - v_i/2\|^2 + \left(a_i - \|v_i\|^2/4\right)$$

Since φ^2 is nonnegative, $a_i - \|v_i\|^2/4$ is also nonnegative for all $i \in I$. Letting K be the closure of the set of points $(v_i/2, \sqrt{(a_i - \|v_i\|^2/4)})$ for $i \in I$ concludes the proof. $\qquad\square$

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. Stated otherwise, the distance $d_K(x)$ is the minimum radius r such that the ball centered at x of radius r contains at least one 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 to mimic the definition above. Given a parameter $0 \le m < 1$, define the pseudo-distance $\delta_{\mu,m}$ by

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

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 that a generalization of the usual distance function to a compact set 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 for $\varepsilon = 0$, 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 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. The *distance function to μ with parameter m_0* is the function defined by

$$d_{\mu,m_0}^2 : \mathbb{R}^n \to \mathbb{R}^+, \quad x \mapsto \frac{1}{m_0} \int_0^{m_0} \delta_{\mu,m}(x)^2 \, dm.$$

As an example, let C be a point cloud with N points in $\mathbb{R}^d$, and let 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 fact, we obtain the following formula for the squared distance d_{μ,m_0}^2, where $m_0 = k_0/|C|$:

$$d_{\mu,m_0}^2(x) = \frac{1}{m_0} \int_0^{m_0} \delta_{\mu,m}(x)^2 = \frac{1}{m_0} \sum_{k=1}^{k_0} \frac{1}{N} \delta_{\mu,k/N}(x)^2$$

$$= \frac{1}{k_0} \sum_{p \in \mathrm{NN}_C^{k_0}(x)} \|p - x\|^2,$$

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

3.2 Equivalent Formulation

In this section, 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 an 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 $\mathbb{R}^d$, $\nu(B) \leq \mu(B)$. This is the same as requiring $\mu - \nu$ to be 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:*

$$d_{\mu,m_0}(x) = \min\left\{ m_0^{-1/2} \, W_2(m_0\delta_x, \nu); \, \nu \in \mathrm{Sub}_{m_0}(\mu) \right\}. \tag{1}$$

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

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

In other words, 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:

$$d_{\mu,m_0}(x) = \frac{1}{\sqrt{m_0}} \mathrm{dist}_{W_2}\left(m_0\delta_x, \mathrm{Sub}_{m_0}(\mu) \right). \tag{2}$$

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 $\bar{B}(x, \delta_{\mu,m}(x))$, and whose restriction to the *open* ball $B(x, \delta_{\mu,m}(x))$ 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 distribution function $F_\nu : \mathbb{R}^+ \to \mathbb{R}$* of a measure ν on $\mathbb{R}^+$ is the nondecreasing 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 only if $F_\nu(t) \leq F_\mu(t)$ for all $t > 0$.

Proof Let us 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 $m_0\delta_x$ is given by

$$W_2^2(m_0\delta_x, \nu) = \int_{\mathbb{R}^d} \|h - x\|^2 \, d\nu(h).$$

Let $d_x : \mathbb{R}^d \to \mathbb{R}$ 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 $\mathbb{R}$, $\nu_x(I) = \nu(d_x^{-1}(I))$. Using the change-of-variable formula and the definition of the cumulative distribution function gives us

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

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

$$W_2^2(m_0 \delta_x, \nu) = \int_{\mathbb{R}^d} \|h - x\|^2 \, d\nu(h) \geq \int_0^{m_0} F_{\mu_x}^{-1}(m)^2 \, dm$$

$$= \int_0^{m_0} \delta_{\mu,m}(x)^2 \, dm = m_0 \, d_{\mu,m_0}^2(x). \tag{3}$$

The second equality holds because $F_{\mu_x}(t) = \mu(B(x, t))$, and thus $F_{\mu_x}^{-1}(m) = \delta_{\mu,m}(x)$. This proves that $d_{\mu,m_0}(x)$ 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 $m \leq m_0$, $F_{\nu_x}^{-1}(m) = F_{\mu_x}^{-1}(m)$. 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 $\nu(\bar{B}(x, \delta_{\mu,m_0}(x))) = m_0$; i.e., all the mass of ν is contained in the closed ball $\bar{B}(x, \delta_{\mu,m_0}(x))$ and $\tilde{\mu}(B(x, \delta_{\mu,m_0}(x))) = \mu(B(x, \delta_{\mu,m_0}(x)))$. Because ν is a submeasure of μ, this can be true 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 minimizers $\mathcal{R}_{\mu,m_0}(x)$ always contains 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

$$\mu_{x,m_0} = \mu|_{B(x, \delta_{\mu,m_0}(x))} + \left(m_0 - \mu\big(B\big(x, \delta_{\mu,m_0}(x)\big)\big)\right) \frac{\mu|_{\partial B(x, \delta_{\mu,m_0}(x))}}{\mu(\partial B(x, \delta_{\mu,m_0}(x)))}. \qquad \square$$

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,*

$$d_H\big(\mathrm{Sub}_{m_0}(\mu), \mathrm{Sub}_{m_0}(\mu')\big) \leq W_2\big(\mu, \mu'\big).$$

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 $\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \, d\pi(x, y) = \varepsilon^2$. 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,

$$W_2(\nu, \nu')^2 \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \, d\pi'(x, y) \leq \varepsilon^2.$$

This shows that $\mathrm{dist}(\nu, \mathrm{Sub}_{m_0}(\mu')) \leq \varepsilon$ for every submeasure $\nu \in \mathrm{Sub}_{m_0}(\mu)$. The same holds 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 $\left\| d_{\mu,m_0} - d_{\mu',m_0} \right\|_\infty \leq \frac{1}{\sqrt{m_0}} W_2(\mu, \mu')$.*

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

$$
\begin{aligned}
d_{\mu,m_0}(x) &= \frac{1}{\sqrt{m_0}} \mathrm{dist}_{W_2} \big(m_0 \delta_x, \mathrm{Sub}_{m_0}(\mu) \big) \\
&\leq \frac{1}{\sqrt{m_0}} \big(d_H \big(\mathrm{Sub}_{m_0}(\mu), \mathrm{Sub}_{m_0}(\mu') \big) + \mathrm{dist}_{W_2} \big(m_0 \delta_x, \mathrm{Sub}_{m_0}(\mu') \big) \big) \\
&\leq \frac{1}{\sqrt{m_0}} W_2(\mu, \mu') + d_{\mu',m_0}(x). \qquad\qquad \square
\end{aligned}
$$

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, $f(x + h) \geq f(x) + \langle h | v \rangle$. 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 nonempty 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 $v_{\mu,m_0} : x \in \mathbb{R}^d \mapsto \|x\|^2 - d_{\mu,m_0}^2$ is convex, and its subdifferential at a point $x \in \mathbb{R}^d$ is given by*

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

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

$$
\begin{aligned}
d_{\mu,m_0}^2(y) &= \frac{1}{m_0} \int_{h \in \mathbb{R}^d} \|y - h\|^2 \, d\mu_{y,m_0}(h) \\
&\leq \frac{1}{m_0} \int_{h \in \mathbb{R}^d} \|y - h\|^2 \, d\mu_{x,m_0}(h)
\end{aligned}
$$

$$\leq \frac{1}{m_0} \int_{h \in \mathbb{R}^d} \|x - h\|^2 + 2\langle x - h | y - x \rangle + \|y - x\|^2 \, d\mu_{x,m_0}(h)$$

$$\leq d_{\mu,m_0}^2(x) + \|y - x\|^2 + \langle v | y - x \rangle,$$

where v is the vector defined by

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

The inequality can be rewritten as

$$\left(\|y\|^2 - d_{\mu,m_0}^2(y) \right) - \left(\|x\|^2 - d_{\mu,m_0}^2(x) \right) \geq \langle 2x - v | y - x \rangle$$

which shows that the vector $(2x - v)$ belongs to the subdifferential of v at x. It follows from the above-mentioned characterization of convex functions that v_{μ,m_0} is convex.

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

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

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 [10], 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,

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

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

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

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_{μ,m_0}^2 *is 1-semiconcave. Moreover:*

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

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

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

    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

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

Using the Cauchy–Schwarz inequality, we find the bound $\|\nabla_x d_{\mu,m_0}\| \leq 1$, which proves the statement. $\qquad\square$

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 [7] 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 small enough positive η. The regularity assumption on K is expressed as a lower bound on its μ-reach, which is a generalization of the classical notion of reach [15]. 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 this section, we show that the reconstruction results of [7] can be easily generalized to compare the sublevel sets of two uniformly close distance-like functions. It is also possible to adapt most of the topological and geometric inference results of [6, 8, 9] in a similar way.

4.1 Extending the Sampling Theory for Compact Sets

In this section we extend the sampling theory of [7] for compact sets to distance-like functions. We do not include all the results of the paper, but only those that are needed for the reconstruction theorem (Theorem 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 us to define a notion of gradient vector field $\nabla_x \varphi$ for φ, defined everywhere and satisfying $\|\nabla_x \varphi\| \leq 1$. Although not continuous, the vector field $\nabla \varphi$ is sufficiently regular to be integrated in a continuous locally Lipschitz flow [23] $\Phi^t : \mathbb{R}^d \to \mathbb{R}^d$. 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 $t > 0$, $\frac{d\gamma}{d\varepsilon}|_{t^-} = \nabla_{\gamma(t)} \varphi$. Moreover, if $\gamma : [a, b] \to \mathbb{R}^d$ is an arc-length parameterization of such a curve, we have

$$\varphi(\gamma(b)) = \varphi(\gamma(a)) + \int_a^b \|\nabla_{\gamma(t)} \varphi\| \, dt.$$

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 $\varphi^2(x + h) \leq \varphi^2(x) + 2\alpha \|h\| \varphi(x) + \|h\|^2$ 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 $\alpha \geq 0$ such that x is α-critical.

(ii) The *weak feature size* of φ at r is the maximum $r' > 0$ such that φ does not 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 [7] relies on two important observations. The first one is a consequence of a distance-like version of Grove's isotopy lemma [17, Proposition 1.8], which asserts that the topology of the sublevel sets of φ can only change when one passes critical values. As in [3, 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 $\|\varphi - \psi\|_\infty \leq \varepsilon$. Suppose also that $\mathrm{wfs}_\varphi(r) > 2\varepsilon$ and $\mathrm{wfs}_\psi(r) > 2\varepsilon$. Then, for every $0 < \eta \leq 2\varepsilon$, $\varphi^{r+\eta}$ and $\psi^{r+\eta}$ have the same homotopy type.*

Proof Let $\delta > 0$ be such that $\mathrm{wfs}_\varphi(r) > 2\varepsilon + \delta$ and $\mathrm{wfs}_\psi(r) > 2\varepsilon + \delta$. Since $\|\varphi - \psi\|_\infty \leq \varepsilon$, we have the following commutative diagram where each map is an inclusion.
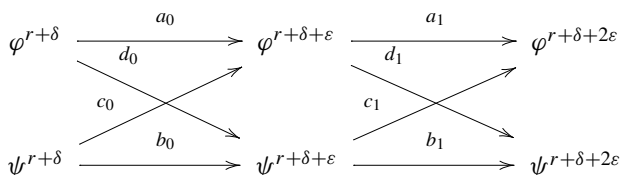
It follows from the isotopy lemma, Proposition 4.2, that the inclusions a_0, a_1, b_0, and b_1 are homotopy equivalences. Let s_0, s_1, r_0, and r_1 be homotopic inverses of a_0, a_1, b_0, and b_1, respectively. Now a straightforward computation shows that c_1 is a homotopy equivalence with homotopic inverse $r_1 \circ d_1 \circ s_1$:

$$c_1 \circ r_1 \circ d_1 \circ s_1 \cong c_1 \circ (r_1 \circ b_1) \circ d_0 \circ s_0 \circ s_1$$
$$\cong (c_1 \circ d_0) \circ s_0 \circ s_1$$
$$\cong a_1 \circ a_0 \circ s_0 \circ s_1 \cong \mathrm{id}_{\varphi^{r+\delta+2\varepsilon}}$$

Similarly, we get $r_1 \circ d_1 \circ s_1 \circ c_1 \cong \mathrm{id}_{\psi^{r+\delta+\varepsilon}}$ proving Proposition 4.3.   □

The second key observation made in [7] is that the critical points of a distance function are stable in a 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 $\|\varphi - \psi\|_\infty \leq \varepsilon$. For any α-critical point x of φ, there exists an α'-critical point x' of ψ with $\|x - x'\| \leq 2\sqrt{\varepsilon\varphi(x)}$ and $\alpha' \leq \alpha + 2\sqrt{\varepsilon/\varphi(x)}$.*

Proof The proof is very similar to the one of [7]. Let $\rho > 0$ and let γ be an integral curve of the flow defined by $\nabla\psi$, starting at x and parameterized by arc length. If γ reaches a critical point of ψ before length ρ, we are done. Assume that this is not the case. Then, with $y = \gamma(\rho)$, one has $\psi(y) - \psi(x) = \int_0^\rho \|\nabla_{\gamma(t)}\psi\| \, dt$. As a consequence, there exists a point $p(\rho)$ on the integral curve such that $\|\nabla_{p(\rho)}\varphi\| \leq \frac{1}{\rho}(\varphi(y) - \varphi(x))$.

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

$$\varphi(y)^2 \leq \varphi(x)^2 + 2\alpha \|x - y\| \varphi(x) + \|x - y\|^2 \quad \text{i.e.}$$

$$\varphi(y) \leq \varphi(x)\left(1 + 2\alpha\frac{\|x - y\|}{\varphi(x)} + \frac{\|x - y\|^2}{\varphi(x)^2}\right)^{1/2} \leq \varphi(x) + \alpha \|x - y\| + \frac{1}{2}\frac{\|x - y\|^2}{\varphi(x)}.$$

By combining, we get $\|\nabla_{p(\rho)}\varphi\| \leq \alpha + \frac{2\varepsilon}{\rho} + \frac{1}{2}\frac{\rho}{\varphi(x)}$. 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.   □

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

Proof 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,

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

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

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

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

Theorem 4.6 (Reconstruction) *Let φ, ψ be two ε-close distance-like functions, with* reach$_\alpha(\varphi) \geq R$ *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*

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

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

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

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 $2\varepsilon \leq R - \varepsilon(3 + 4/\alpha^2)$. The theorem follows.   □

Remark that in Definition 4.1 above 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 section, 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. Thus, 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 $d_S + o(m_0)$. 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 $B(p, r)$ centered at a point p of the support decreases with r. Let us state the following definitions:

(i) We say that a nondecreasing 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 $\varepsilon > 0$, $\mu(B(p, \varepsilon)) \geq f(\varepsilon)$.

(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 $\|d_{\mu,m_0} - d_S\|_\infty < \varepsilon$ 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 $\mathbb{R}^d$, and p a projection of x on S, i.e., a point p such that $\|x - p\| = \mathrm{d}_S(x)$. By assumption, $\mu(\mathrm{B}(x, \mathrm{d}_S(x) + \varepsilon)) \geq \mu(\mathrm{B}(p, \varepsilon)) \geq m_0$. Hence, $\delta_{\mu,m_0}(x) \leq \mathrm{d}_S(x) + \varepsilon$. The function $m \mapsto \delta_{\mu,m}(x)$ being nondecreasing, we get $m_0 \mathrm{d}_S^2(x) \leq \int_0^{m_0} \delta_{\mu,m}^2(x)\, \mathrm{d}m \leq m_0(\mathrm{d}_S(x) + \varepsilon)^2$. Taking the square root of this expression proves the lemma. $\square$

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*

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

Proof (i) If S is compact, there exists a sequence $x_1, x_2, \ldots$ of points in S such that, for any $\varepsilon > 0$, $S \subseteq \bigcup_{i=1}^n B(x_i, \varepsilon/2)$ 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 an x_i such that $\|x - x_i\| \leq \varepsilon/2$. Hence, $\mathrm{B}(x_i, \varepsilon/2) \subseteq \mathrm{B}(x, \varepsilon)$, which means that $\mu(\mathrm{B}(x, \varepsilon)) \geq \eta(\varepsilon)$. (ii) This follows straightforwardly from the lemma. $\square$

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 let μ be the uniform probability measure on S. Then*

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

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. [16, Sect. 3.101]).

Theorem 4.10 (Günther–Bishop) *If the sectional curvatures of a Riemannian manifold M do not exceed δ, then for every $x \in M$, $\mathcal{H}^k(\mathrm{B}_M(x, r)) \geq \beta_{k,\delta}(r)$ 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 $1/R^2$, the Günther-Bishop theorem implies that, for any ε smaller than the injectivity radius of S and πR,

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

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

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

where β_k is the volume of the k-dimensional unit ball. Using $\sin(\alpha) \geq \frac{2}{\pi}\alpha$ 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 (e.g., the empirical measure given by a point cloud sampled according to μ), one has

$$\big\| d_K - d_{\mu',m_0} \big\|_\infty \leq \| d_K - d_{\mu,m_0} \|_\infty + \| d_{\mu,m_0} - d_{\mu',m_0} \|_\infty \tag{4}$$

$$\leq C(\mu)^{-1/k} m_0^{1/k} + \frac{1}{\sqrt{m_0}} \mathrm{W}_2\big(\mu,\mu'\big). \tag{5}$$

This inequality ensuring the closeness of d_{μ',m_0} to the distance function d_K for the sup-norm follows immediately from the stability theorem, Theorem 3.5, and Corollary 4.8. As expected, the choice of m_0 is a trade-off: a small m_0 leads to better approximation of the distance function to the support, while a large m_0 makes the distance functions to measures more stable. Equation (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 $\mathrm{reach}_\alpha(d_K) \geq R$ for some $R > 0$. Let μ' be another measure, and let 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*

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

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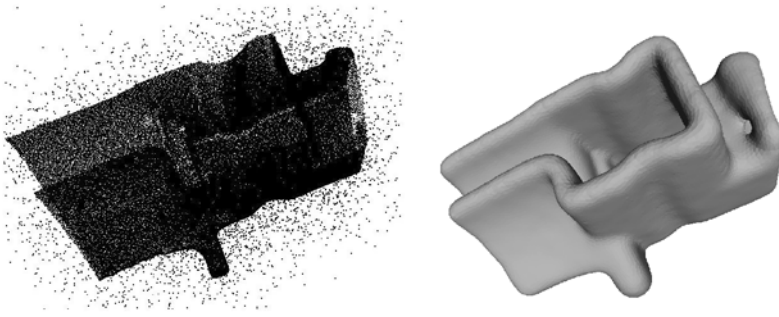
Fig. 1 *Left*: a point cloud sampled on a mechanical part to which 10% of outliers (uniformly sampled in a box enclosing the model) have been added. *Right*: the reconstruction of an isosurface of the distance function d_{μ_C, m_0} to the uniform probability measure on this point cloud

5 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 one 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 [8]. 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 becomes 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 developed 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, N. Leekha, A simple algorithm for homeomorphic surface reconstruction, *Int. J. Comput. Geom. Appl.* **12**(1–2), 125–141 (2002).
2. F. Bolley, A. Guillin, C. Villani, Quantitative concentration inequalities for empirical measures on non-compact spaces, *Probab. Theory Relat.* **137**(3), 541–593 (2007).
3. F. Chazal, A. Lieutier, Stability and computation of topological invariants of solids in $\mathbb{R}^n$, *Discrete Comput. Geom.* **37**(4), 601–617 (2007).
4. F. Chazal, A. Lieutier, Smooth manifold reconstruction from noisy and non-uniform approximation with guarantees, *Comput. Geom. Theor. Appl.* **40**(2), 156–170 (2008).
5. F. Chazal, S.Y. Oudot, Towards persistence-based reconstruction in Euclidean spaces, in *Proc. 24th ACM Sympos. Comput. Geom.* (2008), pp. 232–241.
6. F. Chazal, D. Cohen-Steiner, A. Lieutier, B. Thibert, Stability of curvature measures, *Comput. Graph. Forum* **28**, 1485–1496 (2008) (proc. SGP 2009).

7. F. Chazal, D. Cohen-Steiner, A. Lieutier, A sampling theory for compact sets in Euclidean space, *Discrete Comput. Geom.* **41**(3), 461–479 (2009).
8. F. Chazal, D. Cohen-Steiner, A. Lieutier, Normal cone approximation and offset shape isotopy, *Comput. Geom. Theor. Appl.* **42**(6-7), 566–581 (2009).
9. F. Chazal, D. Cohen-Steiner, Q. Mérigot, Boundary measures for geometric inference, *Found. Comput. Math.* **10**, 221–240 (2010).
10. F.H. Clarke, *Optimization and Nonsmooth Analysis* (Wiley, New York, 1983).
11. D. Cohen-Steiner, H. Edelsbrunner, J. Harer, Stability of persistence diagrams, *Discrete Comput. Geom.* **37**(1), 103–120 (2007).
12. V. de Silva, G. Carlsson, Topological estimation using witness complexes, in *Symposium on Point-Based Graphics, ETH*, Zürich, Switzerland (2004).
13. H. Edelsbrunner, The union of balls and its dual shape, *Discrete Comput. Geom.* **13**, 415–440 (1995).
14. H. Edelsbrunner, J. Harer, *Computational Topology. An Introduction* (American Mathematical Society, Providence, 2010).
15. H. Federer, Curvature measures, *Trans. Am. Math. Soc.* **93**, 418–491 (1959).
16. S. Gallot, D. Hulin, J. Lafontaine, *Riemannian Geometry* (Springer, Berlin, 1990).
17. K. Grove, Critical point theory for distance functions, in *Proc. of Symposia in Pure Mathematics*, vol. 54 (1993).
18. A. Lieutier, Any open bounded subset of $\mathbb{R}^n$ has the same homotopy type as its medial axis, *Comput. Aided Geom. Des.* **36**(11), 1029–1046 (2004).
19. Q. Mérigot, M. Ovsjanikov, L. Guibas, Robust Voronoi-based curvature and feature estimation, in *Proc. SIAM/ACM Joint Conference on Geom. and Phys. Modeling* (2009), pp. 1–12.
20. P. Niyogi, S. Smale, S. Weinberger, A topological view of unsupervised learning from noisy data. Preprint (2008).
21. P. Niyogi, S. Smale, S. Weinberger, Finding the homology of submanifolds with high confidence from random samples, *Discrete Comput. Geom.* **39**(1), 419–441 (2008).
22. S. Peleg, M. Werman, 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).
23. A. Petrunin, Semiconcave functions in Alexandrov's geometry, in *Surveys in differential geometry*, vol. XI (International Press, Somerville, 2007), pp. 137–201.
24. V. Robins, Towards computing homology from finite approximations, *Topol. Proc.* **24**, 503–532 (1999).
25. Y. Rubner, C. Tomasi, L.J. Guibas, The earth mover's distance as a metric for image retrieval, *Int. J. Comput. Vis.* **40**(2), 99–121 (2000).
26. C. Villani, *Topics in Optimal Transportation* (American Mathematical Society, Providence, 2003).