WITNESSED K-DISTANCE

LEONIDAS GUIBAS, DMITRIY MOROZOV, AND QUENTIN MÉRIGOT

Abstract. Distance functions to compact sets play a central role in several areas of computational geometry. Methods that rely on them are robust to the perturbations of the data by the Hausdorff noise, but fail in the presence of outliers. The recently introduced *distance to a measure* offers a solution by extending the distance function framework to reasoning about the geometry of probability measures, while maintaining theoretical guarantees about the quality of the inferred information. A combinatorial explosion hinders working with distance to a measure as an ordinary power distance function. In this paper, we analyze an approximation scheme that keeps the representation linear in the size of the input, while maintaining the guarantees on the inference quality close to those for the exact but costly representation.

1. Introduction

The problem of recovering the geometry and topology of compact sets from finite point samples has seen several important developments in the previous decade. Homeomorphic surface reconstruction algorithms have been proposed to deal with surfaces in $\mathbb{R}^3$ sampled without noise [1] and with moderate Hausdorff (local) noise [13]. In the case of submanifolds of a higher dimensional Euclidean space [20], or even for more general compact subsets [5], it is also possible, at least in principle, to compute the homotopy type from a Hausdorff sampling. If one is only interested in the homology of the underlying space, the theory of persistent homology [15] applied to Vietoris–Rips complexes provides an algorithmically tractable way to estimate the Betti numbers from a finite Hausdorff sampling [7].

All of these constructions share a common feature: they estimate the geometry of the underlying space by a union of balls of some radius r centered at the data points P. A different interpretation of this union is the r-sublevel set of the *distance function* to P, $\mathrm{d}_P : x \mapsto \min_{p\in P} \|x - p\|$. Distance functions capture the geometry of their defining sets, and they are stable to Hausdorff perturbations of those sets, making them well-suited for reconstruction results. However, they are also extremely sensitive to the presence of outliers (i.e., data points that lie far from the underlying set); all reconstruction techniques that rely on them fail even in presence of a single outlier.

To counter this problem, Chazal, Cohen-Steiner, and Mérigot [6] developed a notion of *distance function to a probability measure* that retains the properties of the (usual) distance important for geometric inference. Instead of assuming an underlying compact set that is sampled by the points, they assume an underlying probability measure μ from which the point sample P is drawn. The distance function d_{μ,m_0} to the measure μ depends on a mass parameter $m_0 \in (0, 1)$. This parameter acts as a smoothing term: a smaller m_0 captures the geometry of the support better, while a larger m_0 leads to better stability at the price of precision.

Crucially, the function d_{μ,m_0} is stable to the perturbations of the measure μ under the Wasserstein distance. Defined in Section 2.2, this distance evaluates the cost of the optimal way to transport one measure onto another, where we pay for the squared distance a unit mass travels. Consequently, the Wasserstein distance between the underlying measure μ and the uniform probability measure on the point set P is small even if P contains some outliers, since individual points support only a small fraction of the mass. The stability result ensures that distance function $\mathrm{d}_{\mathbf{1}_P,m_0}$ to the uniform probability measure $\mathbf{1}_P$ on P retains the geometric information contained in the underlying measure μ and its support.

Computing with distance functions to measures. In this article we address the computational issues related to this new notion. If P is a subset of $\mathbb{R}^d$ containing N points, and $m_0 = k/N$, we will denote the distance function to the uniform measure on P by $\mathrm{d}_{P,k}$. As observed in [6], the value of $\mathrm{d}_{P,k}$ at a given point x is easy to compute: it is the square root of the average squared distance from the point x to its k nearest neighbors in P. However, many geometric inference methods require a global representation of the sublevel sets of the function, i.e., the sets $\mathrm{d}_{P,k}^{-1}([0,r]) := \{x \in \mathbb{R}^d;\ \mathrm{d}_{P,k}(x) \leq r\}$. It turns out that the distance function $\mathrm{d}_{P,k}$ can be rewritten as a minimum

$$\mathrm{d}_{P,k}^2(x) = \min_{\bar{p}} \|x - \bar{p}\|^2 - w_{\bar{p}}, \tag{1}$$

where $\bar{p}$ ranges over the set of barycenters of k points in P (see Section 3). Computational geometry provides a rich toolbox to represent sublevel sets of such functions, for example, via weighted α-complexes [14].

The difficulty in applying these methods is that to get an equality in Equation (1) the minimum number of barycenters to store is the same as the number of sites in the order-k Voronoi diagram of P, making this representation unusable even for modest input sizes [9]. Our solution is to construct an approximation of the distance function $\mathrm{d}_{P,k}$, defined by the same equation as (1), but with $\bar{p}$ ranging over a smaller subset of barycenters. In this article, we study the quality of approximation given by a *linear-sized* subset — the *witnessed barycenters*, defined as the barycenters of any k points in P whose order-k Voronoi cell contains at least one of the sample points. The algorithmic simplicity of the scheme is appealing: we only have to find the $k-1$ nearest neighbors for each input point. We denote by $\mathrm{d}_{P,k}^{\mathrm{w}}$ and call *witnessed k-distance* the function defined by Equation (1), where $\bar{p}$ ranges over the witnessed barycenters.

Contributions. Our goal is to give conditions on the point cloud P under which the witnessed k-distance $\mathrm{d}_{P,k}^{\mathrm{w}}$ provides a good uniform approximation of the distance to measure $\mathrm{d}_{P,k}$. We first give a general multiplicative bound on the error produced by this approximation. However, most of our paper (Sections 4 and 5) analyzes the uniform approximation error, when P is a set of independent samples from a measure concentrated near a lower-dimensional subset of the Euclidean space.

(H) We assume that the "ground truth" is an unknown probability measure μ supported on a compact set K whose *dimension is bounded* by ℓ. This means that there exists a positive constant α_μ such that for every point x in K and every radius $r < \mathrm{diam}(K)$, one has $\mu(B(x,r)) \geq \alpha_\mu r^\ell$.

A prototypical example is the volume measure on a compact smooth ℓ-dimensional submanifold K, rescaled to be a probability measure. In this case, the constant α_μ can be lower-bounded as a function of the dimension, the diameter, the ℓ-volume and the minimum sectional curvature of K. This can be derived from the Bishop–Günther comparison theorem (e.g., see [6, Proposition 4.9]). This hypothesis on the dimension of μ ensures that the distance to the measure μ is close to the distance to the support K of μ, and lets us recover information about K.

Our first result asserts in a quantitative way that if the uniform measure to a point cloud P is a good Wasserstein approximation of μ, then the witnessed k-distance to P provides a good approximation of the distance to the underlying compact set K. We denote by $\|.\|_\infty$ the norm of uniform convergence *on* $\mathbb{R}^d$, that is $\|f\|_\infty := \sup_{x\in\mathbb{R}^d} |f(x)|$.

Witnessed Bound (Theorem 4.4). Let μ be a probability measure satisfying the hypothesis **(H)** and let K be its support. Consider the uniform measure $\mathbf{1}_P$ on a point cloud P, and set $m_0 = k/|P|$. Then,

$$\|\mathrm{d}^{\mathrm{w}}_{P,k} - \mathrm{d}_K\|_\infty \leq 3m_0^{-1/2}\,\mathrm{W}_2(\mu,\mathbf{1}_P) + 12\alpha_\mu^{-1/\ell} m_0^{1/\ell},$$

where $\mathrm{W}_2(\mu,\mathbf{1}_P)$ is the Wasserstein distance between measures μ and $\mathbf{1}_P$.

The above bound is only a constant times worse than a similar bound for the exact k-distance which was proven in [6]. In other words, under the hypothesis of this theorem the quality of the inference is not significantly decreased when replacing the exact k-distance by the witnessed k-distance.

In order to make the Witnessed Bound more explicit, we give in Section 5 an upper bound on the Wasserstein distance in the right-hand side of the inequality, when the point cloud P follows a certain sampling model. This model generalizes mixtures of isotropic Gaussians [12], and is similar to a model recently proposed for topological inference [21]. We assume that the point cloud P comes from N independent random samples of the underlying measure μ shifted by N isotropic random Gaussian vectors centered at the origin and with variance σ^2. We can then control the Wasserstein distance in the Witnessed Bound with high probability:

$$\lim_{N\to+\infty} \mathbb{P}(\mathrm{W}_2(\mu,\mathbf{1}_P) \leq 5\sigma) = 0$$

This result is stated more quantitatively in Corollary 5.6. We also include a similar result under a different model of noise in Corollary 5.4.

Outline. The relevant background appears in Section 2. We present our approximation scheme together with a general bound of its quality in Section 3. In Section 4, we give the proof of the Witnessed Bound. The convergence of the uniform measure on a point cloud sampled from a measure of low complexity appears in Section 5. We illustrate the utility of these two bounds with an example and a topological inference statement in our final Section 6.

2. Background

We begin by reviewing the relevant background on measure theory, Wasserstein distances, and distances to measures.

2.1. **Measure.** Let us briefly recap the few concepts of measure theory that we use. A *non-negative measure* μ on the space $\mathbb{R}^d$ is a map from (Borel) subsets of $\mathbb{R}^d$ to non-negative numbers, which is *additive* in the sense that $\mu\left(\cup_{i\in\mathcal{N}} B_i\right) = \sum_i \mu(B_i)$ whenever (B_i) is a countable family of disjoint (Borel) subsets. The *total mass* of a measure μ is $\mathrm{mass}(\mu) := \mu(\mathbb{R}^d)$. A measure μ with unit total mass is called a *probability measure*. The *support* of a measure μ, denoted by $\mathrm{spt}(\mu)$, is the smallest closed set whose complement has zero measure. The *expectation* or *mean* of μ is the point $\mathbb{E}(\mu) = \int_{\mathbb{R}^d} x \mathrm{d}\mu(x)$; the variance of μ is the number $\sigma_\mu^2 = \int_{\mathbb{R}^d} \|x - \mathbb{E}(\mu)\|^2 \mathrm{d}\mu(x)$.

Although the results we present are often more general, the typical probability measures we have in mind are of two kinds: (i) the probability measure defined by rescaling the volume form of a lower-dimensional submanifold of the ambient space and (ii) discrete probability measures that are obtained through noisy sampling of probability measures of the previous kind. For any finite set P with N points, recall that $\mathbf{1}_P$ is the uniform measure supported on P, i.e., the sum of Dirac masses centered at $p \in P$ with weight $1/N$.

2.2. **Wasserstein distance.** A natural way to quantify the distance between two measures is the *quadratic Wasserstein distance*. This distance measures the L^2-cost of transporting the mass of the first measure onto the second one. Note that it is possible to define Wasserstein distances with other exponents; for instance, the L^1 Wasserstein distance is commonly called the earth mover's distance [22]. A general study of this notion and its relation to the problem of optimal transport appear in [24]. We first give the general definition and then explain its interpretation when one of the two measures has finite support.

A *transport plan* between two measures μ and ν with the same total mass is a measure π on the product space $\mathbb{R}^d \times \mathbb{R}^d$ such that for every pair of subsets A, B of $\mathbb{R}^d$, $\pi(A \times \mathbb{R}^d) = \mu(A)$ and $\pi(\mathbb{R}^d \times B) = \nu(B)$. Intuitively, $\pi(A \times B)$ represents the amount of mass of μ contained in A that will be transported to B by π. The set of all transport plans between μ and ν is denoted by $\Gamma(\mu, \nu)$. The *cost* of a transport plan π is given by

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

Finally, the *Wasserstein distance* between μ and ν is the minimum cost of a transport plan between these measures, i.e.,

$$\mathrm{W}_2(\mu, \nu) = \min_{\pi \in \Gamma(\mu,\nu)} c(\pi).$$

Note that W_2 is indeed a distance function; in particular, it satisfies the triangle inequality on the space of probability measures on $\mathbb{R}^d$ with finite variance (cf. [24, Theorem 7.12]).

Discrete target measure. Consider the special case where the measure ν is supported on a finite set P. This means that ν can be written as $\sum_{p\in P} a_p \delta_p$, where δ_p is the unit Dirac mass at p. Moreover, $\sum_p a_p$ must equal the total mass of μ. Given a family of non-negative measures $(\mu_p)_{p\in P}$ such that $\mathrm{mass}(\mu_p) = a_p$ and $\mu = \sum_{p\in P} \mu_p$, one can define a transport plan π between μ and ν by

$$\pi = \sum_{p\in P} \mu_p \times \delta_p,$$

where $\mu_p \times \nu_p$ denotes the product measure. Moreover, all transport plans between μ and ν can be written in this way. The cost of the plan associated to a decomposition $(\mu_p)_{p\in P}$ is then

$$c(\pi) = \left(\sum_{p\in P} \left[\int_{\mathbb{R}^d} \|x-p\|^2 \mathrm{d}\mu_p(x) \right] \right)^{1/2}.$$

As before, $\mathrm{W}_2(\mu,\nu)$ is the minimum of this quantity over all transport plans.

Wasserstein noise. Two properties of the Wasserstein distances are particularly useful to us. Together, they show that the Wasserstein noise and sampling model generalize the commonly used empirical sampling with Gaussian noise model:

- Consider a probability measure μ and $f : \mathbb{R}^d \to \mathbb{R}$, the density of a probability measure centered at the origin, with finite variance $\sigma^2 := \int_{\mathbb{R}^d} \|x\|^2 f(x)\mathrm{d}x$. Denote by ν the result of the convolution of μ by f. Then, the quadratic Wasserstein distance between μ and ν is at most σ. This follows for instance from [24, Proposition 7.17].
- Let P denote a set of N points drawn independently from the measure ν. Suppose also that the ν has small tails, e.g., $\nu(\mathbb{R}^d \setminus \mathrm{B}(0,r)) \leq \exp(-cr^2)$ for some constant c. Then, the Wasserstein distance $\mathrm{W}_2(\nu, \mathbf{1}_P)$ between ν and the uniform probability measure on P converges to zero as N grows to infinity. Examples of such asymptotic convergence results are called "the uniform law of large numbers" and are common in statistics (see for instance [4] and references therein).

Using the notation and assumptions of the two items above, and using the triangle inequality for the Wasserstein distance, one has for any positive ε:

$$\mathbb{P}\big(\mathrm{W}_2(\mathbf{1}_P,\mu) > (1+\varepsilon)\sigma\big) \quad\leq\quad \mathbb{P}\big(\mathrm{W}_2(\mathbf{1}_P,\nu) > \varepsilon\sigma\big)$$

Consequently, the probability that $\mathrm{W}_2(\mathbf{1}_p,\mu)$ is at most $(1+\varepsilon)\sigma$ converges to 1 as N grows to infinity. If one assumes that μ satisfies **(H)**, Corollary 5.6 below gives a similar but more quantitative statement.

2.3. **Distance-to-measure and *k*-distance.**

In [6], the authors introduce a distance to a probability measure as a way to infer the geometry and topology of this measure in the same way the geometry and topology of a set is inferred from its distance function. Given a probability measure μ and a *mass parameter* $m_0 \in (0,1)$, they define a distance function d_{μ,m_0} which captures the properties of the ordinary distance function to a compact set that are used for geometric inference.

Definition 2.1. For any point x in $\mathbb{R}^d$, let $\delta_{\mu,m}(x)$ be the radius of the smallest ball centered at x that contains a mass at least m of the measure μ. The *distance to the measure* μ with parameter m_0 is defined by $\mathrm{d}_{\mu,m_0}(x) = m_0^{-1/2} \left(\int_{m=0}^{m_0} \delta_{\mu,m}(x)^2 \mathrm{d}m\right)^{1/2}$.

The parameter m_0 acts as a smoothing term: a smaller value captures the geometry of the support better, while a larger value leads to better stability at the price of precision. This balance is well captured by the inequality in Theorem 4.2 below. Given a point cloud P of N points, the measure of interest is the uniform measure $\mathbf{1}_P$ on P. When m_0 is a fraction k/N of the number of points (where k

is an integer), we call *k-distance* and denote by $\mathrm{d}_{P,k}$ the distance to the measure $\mathrm{d}_{\mathbf{1}_P,m_0}$. The value of $\mathrm{d}_{P,k}$ at a query point x is given by

$$\mathrm{d}^2_{P,k}(x) = \frac{1}{k} \sum_{p \in \mathrm{NN}^k_P(x)} \|x-p\|^2,$$

where $\mathrm{NN}^k_P(x) \subseteq P$ denotes the k nearest neighbors in P to the point $x \in \mathbb{R}^d$. (Note that while the k-th nearest neighbor itself might be ambiguous, on the boundary of an order-k Voronoi cell, the distance to the k-th nearest neighbor is always well defined, and so is $\mathrm{d}_{P,k}$.)

The most important property of the distance function d_{μ,m_0} is its stability, for a fixed m_0, under perturbations of the underlying measure μ. This property provides a bridge between the underlying (continuous) μ and the discrete measure $\mathbf{1}_P$. According to [6, Theorem 3.5], for any two probability measures μ and ν on $\mathbb{R}^d$,

$$\| \mathrm{d}_{\mu,m_0} - \mathrm{d}_{\nu,m_0} \|_\infty \leq m_0^{-1/2}\, \mathrm{W}_2(\mu,\nu), \tag{2}$$

where W_2 is the Wasserstein distance. The multiplicative constant $m_0^{-1/2}$ in this bound illustrates the fact that a larger m_0 leads to a more stable notion of distance, at the price of a less accurate fit of the data.

3. Witnessed k-distance

In this section we describe a simple scheme for approximating the distance to a uniform measure together with a general error bound. The main contribution of our work, presented in Section 4, is the analysis of the quality of this approximation when the input points come from a measure concentrated on a lower-dimensional subset of the Euclidean space.

3.1. k-Distance as a Power Distance.

Given a set of points $U = \{u_1, \dots, u_n\}$ in $\mathbb{R}^d$ with weights (w_u), we call the *power distance* to U the function pow_U obtained as the lower envelope of all the functions $\mathrm{d}^2_u : x \mapsto \|u-x\|^2 - w_u$, where u ranges over U. By Proposition 3.1 in [6], we can express the square of any distance to a measure as a power distance with non-positive weights. The following proposition recalls this property in the case of k-distance; it is equivalent to the well-known fact that the order-k Voronoi diagrams can be written as the power diagrams for a certain set of points and weights [3].

Proposition 3.1. *For any $P \subseteq \mathbb{R}^d$, denote by $\mathrm{Bary}^k(P)$ the set of barycenters of any subset of k points in P. Then*

$$\mathrm{d}^2_{P,k}(x) = \min\left\{ \|x-\bar{p}\|^2 - w_{\bar{p}};\ \bar{p} \in \mathrm{Bary}^k(P) \right\}, \tag{3}$$

where the weight of a barycenter $\bar{p} = \frac{1}{k}\sum_i p_i$ is given by $w_{\bar{p}} := -\frac{1}{k}\sum_i \|\bar{p}-p_i\|^2$.

Proof. Given k distinct points $p_1, \dots, p_k$ in P, denote their barycenter by $\bar{p}$ and consider the function $\mathrm{d}^2_{\bar{p}}(x) := \frac{1}{k}\sum_{1 \leq i \leq k} \|x-p_i\|^2$. An easy computation shows that

$$\mathrm{d}^2_{\bar{p}}(x) = \frac{1}{k} \sum_{1 \leq i \leq k} \|x-p_i\|^2 = \|x-\bar{p}\|^2 - w_{\bar{p}},$$

where the weight $w_{\bar{p}} = -\frac{1}{k}\sum_{1 \leq i \leq k} \|\bar{p}-p_i\|^2$. The proposition follows from the fact that $\mathrm{d}^2_{P,k}$ can be expressed as the minimum of all the functions $\mathrm{d}^2_{\bar{p}}$. □

In other words, the square of the k-distance function to P coincides exactly with the power distance to the set of barycenters $\mathrm{Bary}^k(P)$ with the weights defined above. From this expression, it follows that the sublevel sets of the k-distance $\mathrm{d}_{P,k}$ are finite unions of balls,

$$\mathrm{d}_{P,k}^{-1}([0,r]) = \bigcup_{\bar{p}\in \mathrm{Bary}^k(P)} \mathrm{B}(\bar{p},(r^2+w_{\bar{p}})^{1/2}).$$

Therefore, ignoring the complexity issues, it is possible to compute the homotopy type of this sublevel set by considering the weighted alpha-shape of $\mathrm{Bary}^k(P)$ [14], which is a subcomplex of the regular triangulation of the set of weighted barycenters.

From the proof of Proposition 3.1, we also see that the only barycenters that actually play a role in Equation (3) are the barycenters of k points of P whose order-k Voronoi cell is not empty. However, the dependence on the number of non-empty order-k Voronoi cells makes computation intractable even for moderately sized point clouds in the Euclidean space [9]. One way to avoid this difficulty is to replace the k-distance to P by an approximate k-distance, defined as in Equation (3), but where the minimum is taken over a smaller set of barycenters. Then, the question is: Given a point set P, can we replace the set of barycenters Bary_P^k in the definition of k-distance by a small subset B while controlling the approximation error $\|\,\mathrm{pow}_B^{1/2} - \mathrm{d}_{P,k}\|_\infty$?

Replacing the k-distance with another power distance is especially attractive since many geometric and topological inference methods relying on distance functions continue to hold when one of the functions is replaced by a good approximation *in the class of power distances*. When this is the case, and some sampling conditions are met, it is possible, for instance, to recover the homotopy type of the underlying compact set (see the Reconstruction Theorem in [6]).

3.2. **Approximating by witnessed k-distance.** We consider a subset of the barycenters suggested by the input data. The answer to our question is affirmative if we accept a multiplicative error.

Definition 3.2. For every point p in P, the barycenter of p and its $(k-1)$ nearest neighbors in P is called a *witnessed k-barycenter*. Let $\mathrm{Bary}_{\mathrm{w}}^k(P)$ be the set of all such barycenters. We get one witnessed barycenter for every point of the sampled point set, and define the *witnessed k-distance*,

$$\mathrm{d}_{P,k}^{\mathrm{w}}(x) = \Big(\min\{\|x-\bar{p}\|^2 - w_{\bar{p}} \mid \bar{p}\in \mathrm{Bary}_{\mathrm{w}}^k(P)\}\Big)^{1/2}.$$

Computing the set of all witnessed barycenters of a point set P requires only finding the $k-1$ nearest neighbors of every point in P. This search problem has a long history in computational geometry [2, 8, 17], and now has several practical implementations. Even a brute-force approach with the running time $\mathrm{O}(dn^2)$, where n is the number of points in P, is significantly better than the $\Omega(n^{\lfloor d/2\rfloor}k^{\lceil d/2\rceil})$ lower bound on the number of cells in order-k Voronoi diagrams [9]. (Note that this lower bound holds as $n/k \to \infty$, which is not the case in our problem; finding similar lower bounds when n/k is constant is an open problem.)

General error bound. Because the distance functions we consider are defined by minima, and $\mathrm{Bary}_{\mathrm{w}}^k(P)$ is a subset of $\mathrm{Bary}^k(P)$, the witnessed k-distance is never less than the exact k-distance. In the lemma below, we give a general multiplicative upper bound. This lemma does not assume any special property for the input point set P. However, even such a coarse bound can be used to estimate Betti numbers of sublevel sets of $\mathrm{d}_{P,k}$, using arguments similar to those in [7].

Lemma 3.3 (General Bound[1])**.** *For any finite point set $P \subseteq \mathbb{R}^d$ and $0 < k < |P|$,*

$$0 \leq \mathrm{d}^{\mathrm{w}}_{P,k} - \mathrm{d}_{P,k} \leq 2\,\mathrm{d}_P \leq 2\,\mathrm{d}_{P,k}\,.$$

Proof. Let y be a point in $\mathbb{R}^d$, and $\bar{p}$ the barycenter associated with an order-k Voronoi cell containing y, i.e., $\bar{p}$ is such that $\mathrm{d}_{P,k}(y) = \mathrm{d}_{\bar{p}}(y) := (\|x - \bar{p}\|^2 - w_{\bar{p}})^{1/2}$. Let us find a witnessed barycenter $\bar{q}$ close to $\bar{p}$. By definition, $\bar{p}$ is the barycenter of the k nearest neighbors $p_1, \dots, p_k$ of y in P. Let $x := p_1$ be the nearest neighbor of y, and $\bar{q}$ the barycenter witnessed by the point x. Then, $\mathrm{d}_P(y) = \|x - y\| \leq \mathrm{d}_{P,k}(y)$, and

$$\begin{aligned}\mathrm{d}^{\mathrm{w}}_{P,k}(y) \leq \mathrm{d}_{\bar{q}}(y) \leq \mathrm{d}_{\bar{q}}(x) + \|x - y\| &\leq \mathrm{d}_{\bar{p}}(x) + \|x - y\| \\ &\leq \mathrm{d}_{\bar{p}}(y) + 2\|x - y\| = \mathrm{d}_{P,k}(y) + 2\mathrm{d}_P(y).\end{aligned}$$

(We have repeatedly used the fact that power distance is a 1-Lipschitz function.) Since y was chosen arbitrarily, the claim follows. □

4. Approximation Quality

Let us briefly recall our hypotheses. There is an ideal, well-conditioned measure μ on $\mathbb{R}^d$ supported on an unknown compact set K. We also have a noisy version of μ, i.e., another measure ν with $\mathrm{W}_2(\mu, \nu) \leq \sigma$, and we suppose that our data set P consists of N points independently sampled from ν. In this section we give conditions under which the witnessed k-distance to P provides a good approximation of the distance to the underlying set K.

4.1. Dimension of a measure.

First, we make precise the main assumption **(H)** on the underlying measure μ, which we use to bound the approximation error made when replacing the exact by the witnessed k-distance. We require μ to be low dimensional in the following sense.

Definition 4.1. A measure μ on $\mathbb{R}^d$ is said to have *dimension at most ℓ with constant $\alpha_\mu > 0$* if the amount of mass contained in the ball $B(p, r)$ is at least $\alpha_\mu r^\ell$, for every point p in the support of μ and every radius r smaller than the diameter of this support. If μ is said to have *dimension at most ℓ*, this means that there exists a constant α_μ.

The important assumption here is that the lower bound $\mu(\mathrm{B}(p, r)) \geq \alpha r^\ell$ should be true for some positive constant α and for r smaller than a given constant R. The choice of $R = \mathrm{diam}(\mathrm{spt}(\mu))$ provides a normalization of the constant α_μ and slightly simplifies the statements of the results.

Let M be an ℓ-dimensional compact submanifold of $\mathbb{R}^d$, and $f : M \to \mathbb{R}$ a positive weight function on M with values bounded away from zero and infinity. Then, the dimension of the volume measure on M weighted by the function f is at most ℓ. A quantitative statement can be obtained using the Bishop–Günther comparison theorem; the bound depends on the maximum absolute sectional curvature of the manifold M, as shown in Proposition 4.9 in [6]. Note that the positive lower bound on the density is really necessary. For instance, the dimension of the standard

[1] The authors thank Daniel Chen for strengthening an earlier version of this bound.

Gaussian distribution $\mathcal{G}(0,1)$ on the real line is not bounded by 1, nor by any positive constant, because the density of this distribution decreases to zero faster than any function $r \mapsto 1/r^\ell$ as one moves away from the origin.

It is easy to see that if m measures $\mu_1, \dots, \mu_m$ have dimension at most ℓ, then so does their sum. Consequently, if (M_j) is a finite family of compact submanifolds of $\mathbb{R}^d$ with dimensions (d_j), and μ_j is the volume measure on M_j weighted by a function bounded away from zero and infinity, the dimension of the measure $\mu = \sum_{j=1}^m \mu_j$ is at most $\max_j d_j$.

4.2. **Bounds.** In the remainder of this section, we bound the error between the witnessed k-distance $\mathrm{d}^{\mathrm{w}}_{P,k}$ and the (ordinary) distance d_K to the compact set K. We start from a proposition from [6] that bounds the error between the exact distance to measure and d_K.

Theorem 4.2. *Let μ denote a probability measure with dimension at most ℓ, supported on a compact set K. Consider another measure ν, then for a mass parameter $m_0 \in (0,1)$,*

$$\|\mathrm{d}_{\nu,m_0} - \mathrm{d}_K\|_\infty \leq m_0^{-1/2}\, \mathrm{W}_2(\mu,\nu) + \alpha_\mu^{-1/\ell} m_0^{1/\ell},$$

where α_μ is the parameter in Definition 4.1.

Proof. Using the triangle inequality and Equation (2), one has

$$\begin{aligned}\|\mathrm{d}_{\nu,m_0} - \mathrm{d}_K\|_\infty &\leq \|\mathrm{d}_{\mu,m_0} - \mathrm{d}_{\nu,m_0}\|_\infty + \|\mathrm{d}_{\mu,m_0} - \mathrm{d}_K\|_\infty \\ &\leq m_0^{-1/2}\, \mathrm{W}_2(\mu,\nu) + \|\mathrm{d}_{\mu,m_0} - \mathrm{d}_K\|_\infty\end{aligned}$$

Then, from Lemma 4.7 in [6], $\|\mathrm{d}_{\mu,m_0} - \mathrm{d}_K\|_\infty \leq \alpha_\mu^{-1/\ell} m_0^{1/\ell}$, and the claim follows. □

To make this bound concrete, let us construct a simple example where the term corresponding to the Wasserstein noise and the term corresponding to the smoothing have the same order of magnitude.

Example. Consider the restriction μ of the Lebesgue measure to the ℓ-dimensional unit ball $K := \mathrm{B}(0,1)$, rescaled to become a probability measure by a factor $1/\operatorname{vol}^\ell \mathrm{B}(0,1)$. For a given mass parameter m_0, consider the second measure ν obtained by moving every bit of mass of μ in the ℓ-ball $\mathrm{B}(0, m_0^{1/\ell})$ to the closest point in the $(\ell-1)$-sphere $\mathrm{S}(0, m_0^{1/\ell})$; see Figure 1. By construction,

$$\begin{aligned}\mathrm{W}_2(\mu,\nu)^2 &= \int_{\mathrm{B}(0,m_0^{1/\ell})} (m_0^{1/\ell} - \|x\|)^2 \mathrm{d}\mu(x) \\ &= \frac{\operatorname{vol}^\ell \mathrm{B}(0,1)}{\operatorname{vol}^{\ell-1} \mathrm{S}(0,1)} \int_0^{m_0^{1/\ell}} r^{\ell-1} (m_0^{1/\ell} - r)^2 \mathrm{d}r \\ &= \ell \int_0^{m_0^{1/l}} \left(r^{\ell+1} - 2r^\ell m_0^{1/\ell} + r^{\ell-1} m_0^{2/\ell} \right) \mathrm{d}r \\ &= \frac{2m_0^{1+2/\ell}}{(\ell+1)(\ell+2)}\end{aligned}$$

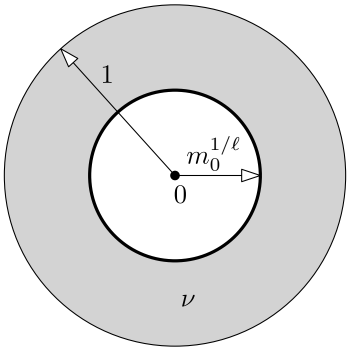


FIGURE 1. μ is the uniform measure on a ball of radius 1, $K = \mathrm{B}(0,1)$. ν is supported on the spherical shell with radii $m_0^{1/\ell}$ and 1. It is constructed by moving the mass of μ at every point in the ball $\mathrm{B}(0, m_0^{1/\ell})$ to the closest point in the sphere $\mathrm{S}(0, m_0^{1/\ell})$.

The distance $\mathrm{d}_{\nu,m_0}(0)$ of the origin to ν is easy to compute: the radius of the smallest ball centered at the origin with a mass m_0 of ν is exactly $m_0^{1/\ell}$. Hence,

$$\begin{aligned}\|\mathrm{d}_K - \mathrm{d}_{\nu,m_0}\|_\infty &\geq |\mathrm{d}_K(0) - \mathrm{d}_{\nu,m_0}(0)| \\ &= m_0^{1/\ell} = C_\ell m_0^{-1/2}\,\mathrm{W}_2(\mu,\nu).\end{aligned}$$

In other words, the two terms in the bound in Theorem 4.2 differ by a constant factor. □

In the previous theorem, when $\nu = \mathbf{1}_P$ is the uniform measure on a point cloud P and $m_0 = k/|P|$, we get the exact bound on the k-distance.

Corollary 4.3 (Exact Bound)**.** *Let μ denote a probability measure with dimension at most ℓ, supported on a compact set K. Consider the uniform measure $\mathbf{1}_P$ on a point cloud P, and set $m_0 = k/|P|$. Then*

$$\|\mathrm{d}_{P,k} - \mathrm{d}_K\|_\infty \leq m_0^{-1/2}\,\mathrm{W}_2(\mu, \mathbf{1}_P) + \alpha_\mu^{-1/\ell} m_0^{1/\ell},$$

where α_μ is the parameter in Definition 4.1.

In the main theorem of this section, the exact k-distance in Corollary 4.3 is replaced by the witnessed k-distance. Observe that the new error term is only a constant factor off from the old one.

Theorem 4.4 (Witnessed Bound)**.** *Let μ be a probability measure satisfying the dimension assumption and let K be its support. Consider the uniform measure $\mathbf{1}_P$ on a point cloud P, and set $m_0 = k/|P|$. Then,*

$$\|\mathrm{d}^{\mathrm{w}}_{P,k} - \mathrm{d}_K\|_\infty \leq 3 m_0^{-1/2}\,\mathrm{W}_2(\mu, \mathbf{1}_P) + 12\alpha_\mu^{-1/\ell} m_0^{1/\ell},$$

where α_μ is the parameter in Definition 4.1.

Before proving the theorem, we start with an auxiliary lemma showing that a measure ν close to a measure μ satisfying an upper dimension bound (as in Definition 4.1) remains concentrated around the support of μ.

Lemma 4.5 (Concentration)**.** *Let μ be a probability measure satisfying the dimension assumption, and let ν be another probability measure. Let m_0 be a mass parameter. Then, for every point p in the support of μ, $\nu(\mathrm{B}(p,\eta)) \geq m_0$, where $\eta = m_0^{-1/2}\,\mathrm{W}_2(\mu,\nu) + 4\alpha_\mu^{-1/\ell} m_0^{1/2+1/\ell}$.*

Proof. Let π be an optimal transport plan between ν and μ. For a fixed point p in the support K of μ, let r be the smallest radius such that $\mathrm{B}(p,r)$ contains at least $2m_0$ of mass μ. Consider now a submeasure μ' of μ of mass exactly $2m_0$ and whose support is contained in the ball $\mathrm{B}(p,r)$. This measure is obtained by transporting a submeasure ν' of ν by the optimal transport plan π. Our goal is to determine for what choice of η the ball $\mathrm{B}(p,\eta)$ contains a ν'-mass (and, therefore, a ν-mass) of at least m_0. We make use of Chebyshev's inequality for ν' to bound the mass of ν' *outside* of the ball $\mathrm{B}(p,\eta)$:

$$\begin{aligned} \nu'(\mathbb{R}^d \setminus \mathrm{B}(p,\eta)) &= \nu'(\{x \in \mathbb{R}^d;\ \|x-p\| \geq \eta\}) \\ &\leq \frac{1}{\eta^2}\int \|x-p\|^2 \mathrm{d}\nu'. \end{aligned} \tag{4}$$

Observe that the right-hand term of this inequality is exactly the squared Wasserstein distance between ν' and the Dirac mass $2m_0\delta_p$ divided by η^2. We bound this squared Wasserstein distance using the triangle inequality:

$$\begin{aligned} \int \|x-p\|^2 \mathrm{d}\nu' &= \mathrm{W}_2^2(\nu', 2m_0\delta_p) \\ &\leq (\mathrm{W}_2(\mu',\nu') + \mathrm{W}_2(\mu', 2m_0\delta_p))^2 \\ &\leq (\mathrm{W}_2(\mu,\nu) + 2m_0 r)^2. \end{aligned} \tag{5}$$

Combining equations (4) and (5), we get

$$\begin{aligned} \nu(\mathrm{B}(p,\eta)) \geq \nu'(\mathrm{B}(p,\eta)) &\geq \nu'(\mathbb{R}^d) - \nu'(\mathbb{R}^d \setminus \mathrm{B}(p,\eta)) \\ &\geq 2m_0 - \frac{(\mathrm{W}_2(\mu,\nu) + 2m_0 r)^2}{\eta^2}. \end{aligned}$$

By the lower bound on the dimension of μ, and the definition of the radius r, one has $r \leq (2m_0/\alpha_\mu)^{1/\ell}$. Hence, the ball $\mathrm{B}(p,\eta)$ contains a mass of at least m_0 as soon as

$$\frac{(\mathrm{W}_2(\mu,\nu) + \alpha_\mu^{-1/\ell} 2^{1+1/\ell} m_0^{1+1/\ell})^2}{\eta^2} \leq m_0.$$

This will be true, in particular, if η is larger than

$$\mathrm{W}_2(\mu,\nu) m_0^{-1/2} + 4\alpha_\mu^{-1/\ell} m_0^{1/2+1/\ell}. \qquad \square$$

Proof of the Witnessed Bound Theorem. Since the witnessed k-distance is a minimum over fewer barycenters, it is larger than the real k-distance. Using this fact and the Exact Bound Theorem one gets the lower bound:

$$\mathrm{d}^{\mathrm{w}}_{P,k} \geq \mathrm{d}_{P,k} \geq \mathrm{d}_K - \left(m_0^{-1/2}\,\mathrm{W}_2(\mu, \mathbf{1}_P) + \alpha_\mu^{-1/\ell} m_0^{1/\ell}\right).$$

For the upper bound, choose η as given by the previous Lemma 4.5 applied to the measure $\nu = \mathbf{1}_P$. Then, for every point y in K, the ball $\mathrm{B}(y,\eta)$ contains at least k points in the point cloud P. Let p_1 be one of these points, and $p_2,\ldots,p_k$ be the $(k-1)$ nearest neighbors of p_1 in P. The points $p_2,\ldots,p_k$ cannot be at a distance greater than 2η from p_1, and, consequently, cannot be at a distance greater than 3η from y. By definition, the barycenter $\bar{p}$ of the points $\{p_i\}$ is witnessed by p_1. Hence,

$$\mathrm{d}^{\mathrm{w}}_{P,k}(y) \leq \mathrm{d}_{\bar{p}}(y) := \left(\frac{1}{k} \sum_{i=1}^{k} \|y - p_i\|^2 \right)^{1/2} \leq 3\eta.$$

Since $\mathrm{d}^{\mathrm{w}}_{P,k}$ is 1-Lipschitz, we get $\mathrm{d}^{\mathrm{w}}_{P,k}(x) \leq 3\eta + \|y - x\|$. This inequality is true for every point y in K; minimizing over all such y, we obtain $\mathrm{d}^{\mathrm{w}}_{P,k}(x) \leq 3\eta + \mathrm{d}_K(x)$. Recall that $m_0 \leq 1$, as is $m_0^{1/2}$. To match the form of the bound in Corollary 4.3, we drop $m_0^{1/2}$ from the second term of η in the Concentration Lemma. Substituting the result into the last inequality, we complete the proof. □

5. Convergence under Empirical Sampling

One term remains moot in the bounds in Corollary 4.3 and Theorem 4.4, namely the Wasserstein distance $\mathrm{W}_2(\mu, \mathbf{1}_P)$. In this section, we analyze its convergence. The rate depends on the complexity of the measure μ, defined below. The moral of this section is that if a measure can be well approximated with few points, then it is also well approximated by random sampling.

Definition 5.1. The *complexity* of a probability measure μ at a scale $\varepsilon > 0$ is the minimum cardinality of a finitely supported probability measure ν that ε-approximates μ in the Wasserstein sense, i.e., such that $\mathrm{W}_2(\mu, \nu) \leq \varepsilon$. We denote this number by $\mathcal{N}_\mu(\varepsilon)$.

Observe that this notion is very close to the *ε-covering number* of a compact set K, denoted by $\mathcal{N}_K(\varepsilon)$, which counts the minimum number of balls of radius ε needed to cover K. It is worth noting that if measures μ and ν are close — as are the measure μ and its noisy approximation ν in the previous section — and μ has low complexity, then so does the measure ν. The following lemma shows that measures satisfying the dimension assumption have low complexity. Its proof follows from a classical covering argument that appears, for example, in Proposition 4.1 of [18].

Lemma 5.2 (Dimension–Complexity). *Let K be the support of a measure μ of dimension at most ℓ with constant α_μ (as in Definition 4.1). Then, for every positive ε, $\mathcal{N}_\mu(\varepsilon) \leq 5^\ell/(\alpha_\mu \varepsilon^\ell)$.*

Combining this lemma with the theorem below, we get a bound on how well a measure satisfying an upper bound on its dimension is approximated by empirical sampling.

Theorem 5.3 (Convergence). *Let μ be a probability measure on $\mathbb{R}^d$ whose support has diameter at most D, and let P be a set of N points independently drawn from the measure μ. Then, for $\varepsilon > 0$,*

$$\mathbb{P}(\mathrm{W}_2(\mathbf{1}_P, \mu) \leq 2\varepsilon) \geq 1 - (2\mathcal{N}_\mu(\varepsilon) + 1) \exp\left(-\frac{2N\varepsilon^4}{D^4 \mathcal{N}_\mu(\varepsilon)^2} \right).$$

The proof of this theorem relies mainly on the following versions of Hoeffding's inequality. Given a sequence $(X_i)_{i \geq 0}$ of independent real-valued random variables with common mean x and such that $0 \leq X_i \leq m$, one has:

$$\mathbb{P}\left(\left|\frac{1}{N}(X_1 + \ldots + X_N) - x\right| \geq t\right) \leq 2\exp(-2t^2N/m^2), \tag{6}$$

$$\mathbb{P}\left(\frac{1}{N}(X_1 + \ldots + X_N) - x \geq t\right) \leq \exp(-2t^2N/m^2). \tag{7}$$

Proof. Let n be a fixed integer, and let ε be the minimum Wasserstein distance between μ and a measure $\bar{\mu}$ supported on (at most) n points. Let S be the support of the optimal measure $\bar{\mu}$, so that $\bar{\mu}$ can be decomposed as $\sum_{s\in S} \alpha_s \delta_s$ ($\alpha_s \geq 0$). Let π be an optimal transport plan between μ and $\bar{\mu}$; this is equivalent to finding a decomposition of μ as a sum of n non-negative measures $(\pi_s)_{s\in S}$ such that $\mathrm{mass}(\pi_s) = \alpha_s$, and

$$\sum_{s\in S} \int \|x - s\|^2 \mathrm{d}\pi_s(x) = \varepsilon^2 = \mathrm{W}_2(\mu, \bar{\mu})^2.$$

Drawing a random point X from the measure μ amounts to (i) choosing a random point s in the set S (with probability α_s) and (ii) drawing a random point X following the probability distribution π_s/α_s. Given N independent points $X_1, \ldots, X_N$ drawn from the measure μ, denote by $I_{s,N}$ the proportion of the (X_i) for which the point s was selected in step (i). Hoeffding's inequality (6) allows us to bound how far the proportion $I_{s,N}$ deviates from α_s: $\mathbb{P}(|I_{s,N} - \alpha_s| \geq \delta) \leq 2\exp(-2N\delta^2)$. If the sum of deviations for all points s exceeds δ, then at least one deviation exceeds δ/n; combining the inequalities and using the union bound yields

$$\mathbb{P}\left(\sum_{s\in S} |I_{s,N} - \alpha_s| \geq \delta\right) \leq 2n\exp(-2N(\delta/n)^2).$$

For every point s, denote by $\tilde{\pi}_s$ the distribution of the distances to s in the submeasure π_s, i.e., the measure on the real line defined by $\tilde{\pi}_s(I) := \pi_s(\{x \in \mathbb{R}^d; \|x - s\| \in I\})$ for every interval I. Define $\tilde{\mu}$ as the sum of the $\tilde{\pi}_s$; by the change of variable formula one has

$$\int_{\mathbb{R}} t^2 \mathrm{d}\tilde{\mu}(t) = \sum_s \int_{\mathbb{R}} t^2 \mathrm{d}\tilde{\pi}_s = \sum_s \int_{\mathbb{R}^d} \|x - s\|^2 \mathrm{d}\pi_s = \varepsilon^2.$$

Given a random point X_i sampled from μ, denote by Y_i the Euclidean distance between the point X_i and the point s chosen in step (i). By construction, the distribution of Y_i is given by the measure $\tilde{\mu}$; applying Hoeffding's inequality (7) to the sequence Y_i^2 yields

$$\mathbb{P}\left(\frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \varepsilon^2 \geq \eta^2\right) \leq \exp(-2N\eta^4/D^4).$$

In order to conclude, we need to define a transport plan from the empirical measure $\mathbf{1}_P = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i}$ to the finite measure $\bar{\mu}$. To achieve this, we order the points (X_i) by increasing distance Y_i; then transport every Dirac mass $\frac{1}{N}\delta_{X_i}$ to the corresponding point s in S until s is "full", i.e., the mass α_s is reached. The squared cost of this transport operation is at most $\frac{1}{N}\sum_{i=1}^{N} Y_i^2$. Then, distribute the remaining mass among the s points in any way; the squared cost of this step is at most $D^2 \sum_{s\in S} |I_{s,N} - \alpha_s|$. The total squared cost of this transport plan is the sum of these two costs. From what we have shown above, setting $\eta = \varepsilon$ and $\delta = \varepsilon^2/D^2$, one gets

$$\begin{aligned}\mathbb{P}(\mathrm{W}_2(\mathbf{1}_P, \mu) \leq 2\varepsilon) &\geq \mathbb{P}(\mathrm{W}_2^2(\mathbf{1}_P, \mu) \leq 3\varepsilon^2) \\ &\geq 1 - 2n \exp\left(-\frac{2N\varepsilon^4}{D^4 n^2}\right) - \exp\left(-\frac{2N\varepsilon^4}{D^4}\right) \\ &\geq 1 - (2n+1) \exp\left(-\frac{2N\varepsilon^4}{D^4 n^2}\right),\end{aligned}$$

where the last inequality follows since $n \geq 1$.

□

Sampling from a perturbation of the measure. A result similar to the Convergence Theorem follows when the samples are drawn not from the original measure μ, but from a "noisy" approximation ν. When the measure ν is also supported on a compact set, this follows directly from the Convergence Theorem.

Corollary 5.4 (Fixed Diameter Sampling)**.** *Let μ and ν be two probability measures on $\mathbb{R}^d$ whose support has diameter at most D and such that $\mathrm{W}_2(\mu, \nu) \leq \sigma$. Let P be a set of N points independently drawn from the measure ν. Then,*

$$\mathbb{P}(\mathrm{W}_2(\mathbf{1}_P, \mu) \leq 5\sigma) \geq 1 - (2\mathcal{N}_\mu(\sigma) + 1) \exp\left(-\frac{32N\sigma^4}{D^4 \mathcal{N}_\mu(\sigma)^2}\right).$$

Proof. First of all, observe that, by definition, the covering number $\mathcal{N}_\nu(\sigma + \delta)$ is upper bounded by $\mathcal{N}_\mu(\delta)$. We apply the previous theorem to the measure ν with $\varepsilon = (\sigma + \delta)$ to get

$$\begin{aligned}\mathbb{P}(\mathrm{W}_2(\mathbf{1}_P, \nu) \leq 2(\sigma + \delta)) &\geq 1 - (2\mathcal{N}_\nu(\sigma + \delta) + 1) \exp\left(-\frac{2N(\sigma+\delta)^4}{D^4 \mathcal{N}_\nu(\sigma+\delta)^2}\right) \\ &\geq 1 - (2\mathcal{N}_\mu(\delta) + 1) \exp\left(-\frac{32N\sigma^4}{D^4 \mathcal{N}_\mu(\delta)^2}\right).\end{aligned}$$

Setting $\delta = \sigma$, and applying the triangle inequality $\mathrm{W}_2(\mathbf{1}_P, \mu) \leq \mathrm{W}_2(\mathbf{1}_P, \nu) + \mathrm{W}_2(\nu, \mu)$ concludes the proof. □

Perturbations with non-compact support. In many cases the perturbed measure ν is not compactly supported, and the previous corollary does not apply. It is still possible to recover similar results under a stronger assumption than a simple bound on the Wasserstein distance between μ and ν.

To give a flavor of such a result, we consider the simple case where ν is a convolution of μ with an isotropic centered Gaussian distribution with variance σ^2, that is: $\nu = \mu * \mathcal{G}(0, (\sigma^2/d)\mathbf{I})$. We will make use of the following bound on the sum of squared norms of random Gaussian vectors.

Lemma 5.5. *Let $G_1, \ldots, G_N$ be i.i.d. isotropic Gaussian vectors with zero mean and covariance matrix $(\sigma^2/d)\mathbf{I}$. Then,*

$$\mathbb{P}\left(\left(\frac{1}{N}\sum_{i=1}^N \|G_i\|^2\right)^{1/2} \geq (1+\varepsilon)\sigma\right) \leq \exp\left(-\frac{1}{2}\varepsilon^2 N d\right).$$

Proof. By hypothesis, the vectors (G_i) can be written as $G_i = (\alpha Y_{i,1}, \ldots, \alpha Y_{i,d})$ where $(Y_{i,j})$ are $N \times d$ independent centered Gaussian random variables with variance 1, and $\alpha = \sigma/\sqrt{d}$. Define a random variable Z by

$$Z = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{d} \alpha^2 (Y_{i,j}^2 - 1).$$

Using Lemma 1 from [19], we can bound the tail of Z:

$$\mathbb{P}\left(Z \geq 2 \frac{\sqrt{x}\sigma^2}{\sqrt{Nd}} + 2 \frac{x\sigma^2}{Nd} \right) \leq \exp(-x).$$

Setting $x = \frac{1}{2}\varepsilon^2 Nd$ yields:

$$\begin{aligned}\mathbb{P}\left(\frac{1}{N} \sum_{i=1}^{N} \|G_i\|^2 \geq (1+\varepsilon)^2 \sigma^2 \right) &= \mathbb{P}\left(Z \geq 2\varepsilon\sigma^2 + \varepsilon^2\sigma^2 \right) \\ &\leq \mathbb{P}\left(Z \geq \sqrt{2}\varepsilon\sigma^2 + \varepsilon^2\sigma^2 \right) \leq \exp\left(-\frac{1}{2}\varepsilon^2 Nd \right). \ \square\end{aligned}$$

Corollary 5.6 (Gaussian Convolution Sampling)**.** *Let μ be a probability measure whose support has diameter at most D and ν be obtained by convolution of μ, $\nu = \mu * \mathcal{G}(0, (\sigma^2/d)\mathbf{I})$. Let Q be a set of N points drawn independently from the measure ν. Then,*

$$\mathbb{P}(\mathrm{W}_2(\mathbf{1}_Q, \mu) \leq 4\sigma) \geq 1 - \exp(-Nd/2) - (2\mathcal{N}_\mu(\sigma) + 1) \exp\left(-\frac{2N\sigma^4}{D^4 \mathcal{N}_\mu(\sigma)^2} \right).$$

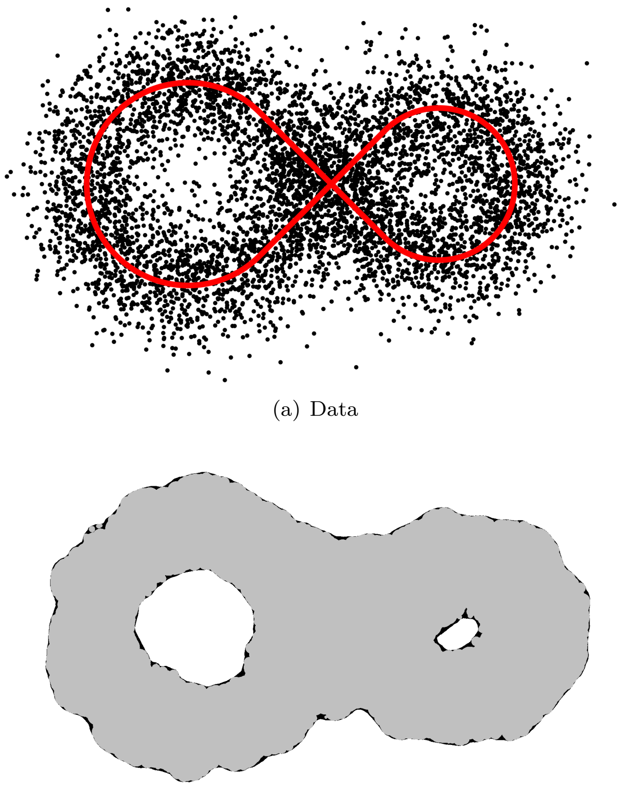
Proof. Let X be a random vector with distribution μ and G a random vector with distribution $\mathcal{G}(0, (\sigma^2/d)\mathbf{I})$. Then, by definition of the convolution, the vector $Y = X + G$ has distribution ν. We consider N independent copies of such vectors $(X_i, G_i)_i$, and set $Y_i = X_i + G_i$. The main difficulty is to bound the probability that the Wasserstein distance between the uniform probability measures on the point sets $P = \{X_1, \ldots, X_N\}$ and $Q = \{Y_1, \ldots, Y_N\}$ exceeds $(1+\varepsilon)\sigma$. The following inequality is an immediate consequence of the previous lemma:

$$\mathbb{P}(\mathrm{W}_2(\mathbf{1}_P, \mathbf{1}_Q) \geq (1+\varepsilon)\sigma) \leq \exp\left(-\frac{1}{2}\varepsilon^2 Nd \right).$$

We set $\varepsilon = 1$ in this inequality and apply Theorem 5.3 and the triangle inequality to get:

$$\begin{aligned}\mathbb{P}(\mathrm{W}_2(\mathbf{1}_Q, \mu) \leq 4\sigma) &\geq \mathbb{P}(\mathrm{W}_2(\mathbf{1}_P, \mathbf{1}_Q) < 2\sigma) \cdot \mathbb{P}(\mathrm{W}_2(\mathbf{1}_P, \mu) \leq 2\sigma) \\ &\geq 1 - \exp(-Nd/2) - (2\mathcal{N}_\mu(\sigma) + 1) \exp\left(-\frac{2N\sigma^4}{D^4 \mathcal{N}_\mu(\sigma)^2} \right). \ \square\end{aligned}$$

Remark. We note that the result of Corollary 5.6 can be extended to more general models of noise. The crucial point is to be able to control the Wasserstein distance between $\mathbf{1}_P$ and $\mathbf{1}_Q$, where P and Q are point sets obtained by sampling N points from μ and ν. Estimates of this kind can be obtained, for instance, if there exist random variables X and Y with distributions μ and ν such that the random variable $Z = \|X - Y\|^2$ is sub-exponential, i.e., $\mathbb{P}(Z \geq t) \leq \exp(-ct)$. We refer the interested reader, for example, to Proposition 5.16 in [23].



(a) Data

(b) Sublevel sets

Figure 2. (a) 6000 points sampled from a sideways figure 8 (displayed on top of the point set), with circle radii $R_1 = \sqrt{2}$ and $R_2 = \sqrt{9/8}$. The points are sampled from the uniform measure on the figure-8, convolved with the Gaussian distribution $\mathcal{G}(0, \sigma^2)$, where $\sigma = .45$. (b) r-sublevel sets of the witnessed (in gray) and exact (additional points in black) k-distances with mass parameter $m_0 = 50/6000$, and $r = .24$.

6. Discussion

We illustrate the utility of the Witnessed Bound Theorem with an example and an inference statement. Figure 2(a) shows 6000 points drawn from the uniform distribution on a sideways figure-8, convolved with a Gaussian distribution. The ordinary distance function has no hope of recovering geometric information out of these points since both loops of the figure-8 are filled in. In Figure 2(b), we show the sublevel sets of the distance to the uniform measure on the point set, both the witnessed k-distance and the exact k-distance. Both functions recover the topology of figure-8; the bits missing from the witnessed k-distance smooth out the boundary of the sublevel set, but do not affect the image at large.

Complexes. Since we are working with the power distance to a weighted point set (the witnessed barycenters U), we can employ different simplicial complexes commonly used in the computational geometry literature [16]. Recall that an (abstract) simplex is a subset of some universal set, in our case the witnessed barycenters U; a simplicial complex is a collection of simplices, where each subset of every simplex belongs to the collection.

The simplest construction is the Čech complex [16, Section III.2]. It contains a simplex if the balls defined by the points at the power distance r from the witnessed barycenters intersect:

$$\check{\mathrm{C}}_r(U) = \left\{ \sigma \subseteq U \mid \bigcap_{u \in \sigma} \mathrm{B}(u, (r^2 + w_u)^{1/2}) \neq \emptyset \right\}.$$

A closely related geometric construction, the weighted alpha complex, is defined by clipping these balls using the power diagram of the witnessed barycenters, see [14]. By the Nerve Theorem [16], both the Čech complex $\check{\mathrm{C}}_r(U)$ and the alpha complex are homotopy equivalent to the sublevel sets of the power distance to U, $\mathrm{pow}_U^{-1}(-\infty, r]$.

In many applications, points are given only through their pairwise distances, rather than explicit coordinates. For this reason and because of its computational simplicity, the Vietoris–Rips complex is a popular choice. This complex is defined as the flag (or clique) complex of the 1-skeleton of the Čech complex. Simply put, a simplex σ belongs to the Vietoris–Rips complex iff all its edges belong to the Čech complex, i.e.,

$$\mathrm{VR}_r(U) = \left\{ \sigma \subseteq U \mid \{u, v\} \in \check{\mathrm{C}}_r(U) \text{ for all } u, v \in \sigma \right\}.$$

In the case of the witnessed k-distance, the pairwise distances between the input points suffice for the construction of the Vietoris–Rips complex on the witnessed barycenters; we give the details in Appendix A.

Note that the Vietoris–Rips complex $\mathrm{VR}_r(U)$ does not, in general, have the homotopy type of $\mathrm{pow}_U^{-1}(-\infty, r]$. It is, however, possible to prove inference results for homology given an *interleaving property*, i.e., there exists a constant $\alpha \geq 1$ such that $\check{\mathrm{C}}_r(U) \subseteq \mathrm{VR}_r(U) \subseteq \check{\mathrm{C}}_{\alpha r}(U)$. The inclusion $\check{\mathrm{C}}_r(U) \subseteq \mathrm{VR}_r(U)$ always holds, simply by definition. However, the second inclusion does not necessarily hold if the weights are positive, as the following example demonstrates.

Example. Consider the weighted point set U made of the three vertices (u, v, w) of an equilateral triangle with unit side length and weights $w_u = w_v = w_w = 1/4$. Then, for any non-negative r, the Vietoris–Rips complex $\mathrm{VR}_r(U)$ contains the triangle, while the Čech complex $\check{\mathrm{C}}_r(U)$ contains this triangle only as soon as $(r^2 + 1/4)^{1/2} \geq 1/\sqrt{3}$, i.e., $r \geq 1/\sqrt{12}$. In this case, there is no α such that the inclusion $\mathrm{VR}_r(U) \subseteq \check{\mathrm{C}}_{\alpha r}(U)$ holds for every positive r.

On the other hand, the following lemma shows that when the weights $(w_u)_{u \in U}$ are non-positive, the inclusion $\mathrm{VR}_r(U) \subseteq \check{\mathrm{C}}_{2r}(U)$ always holds. This property lets us extend the usual homology inference results from Vietoris–Rips complexes to the (weighted) Vietoris–Rips complexes associated with the witnessed k-distance.

Lemma 6.1. *If U is a point cloud with non-positive weights,* $\mathrm{VR}_r(U) \subseteq \check{\mathrm{C}}_{2r}(U)$.

Proof. Let u, v be two weighted points such that the balls $\mathrm{B}(u, (r^2 + w_u)^{1/2})$ and $\mathrm{B}(v, (r^2 + w_v)^{1/2})$ intersect. Let ℓ denote the Euclidean distance between u and v.

By hypothesis, we know that one of the two radii is at least $\ell/2$. Suppose $w_u > w_v$; in this case, $(r^2 + w_u)^{1/2} \geq \ell/2$. Since the weights are non-positive, we also know that $r \geq \ell/2$. Using these two facts, we deduce

$$(2r)^2 + w_u = 3r^2 + (r^2 + w_u) \geq 3\ell^2/4 + \ell^2/4 = \ell^2.$$

This means that the point v belongs to the ball $\mathrm{B}(u, ((2r)^2 + w_u)^{1/2})$.

Now, choose a simplex σ in the Vietoris–Rips complex $\mathrm{VR}_r(U)$. Let v be its vertex with the smallest weight. By the previous paragraph, we know that v belongs to every ball $\mathrm{B}(u, (2r)^{1/2} + w_u)$, for every $u \in \sigma$. Therefore, all these balls intersect, and, by definition, σ belongs to the Čech complex $\check{\mathrm{C}}_{2r}(U)$. □

Inference. Suppose we are in the conditions of the hypothesis **(H)**. Additionally, we assume that the support K of the original measure μ has a *weak feature size* larger than R. This means that the distance function d_K has no critical value in the interval $(0, R)$. A consequence of this hypothesis is that all the offsets $K^r = \mathrm{d}_K^{-1}[0, r]$ of K are homotopy equivalent for $r \in (0, R)$. Suppose again that we have drawn a set P of N points from a nearby measure μ. The following theorem combines the results of Sections 4 and 5.

Theorem 6.2 (Approximation)**.** *Suppose that μ is a measure satisfying hypothesis* **(H)***, supported on a compact set K of diameter at most D, and ν is another measure with $\mathrm{W}_2(\mu, \nu) \leq \sigma$. Let P be a set of N points independently sampled from ν.*

(D) *If the diameter of the support of ν does not exceed D, then*

$$\|\mathrm{d}^{\mathrm{w}}_{P,k} - \mathrm{d}_K\|_\infty \leq 15 m_0^{-1/2}\sigma + 12 m_0^{1/\ell}\alpha_\mu^{-1/\ell}$$

with probability at least

$$1 - (2\mathcal{N}_\mu(\sigma) + 1)\exp\left(-\frac{32N\sigma^4}{D^4\mathcal{N}_\mu(\sigma)^2}\right).$$

(G) *If ν is a convolution of μ with a Gaussian, $\nu = \mu * \mathcal{G}(0, (\sigma^2/d)\mathbf{I})$, then*

$$\|\mathrm{d}^{\mathrm{w}}_{P,k} - \mathrm{d}_K\|_\infty \leq 12 m_0^{-1/2}\sigma + 12 m_0^{1/\ell}\alpha_\mu^{-1/\ell}$$

with probability at least

$$1 - \exp(-Nd/2) - (2\mathcal{N}_\mu(\sigma) + 1)\exp\left(-\frac{2N\sigma^4}{D^4\mathcal{N}_\mu(\sigma)^2}\right).$$

In both statements, $\mathcal{N}_\mu(\sigma)$ is the complexity of measure μ, as in Definition 5.1, and α_μ is the parameter in Definition 4.1.

The standard argument [10] shows that the Betti numbers of the compact set K can be inferred from the function $\mathrm{d}^{\mathrm{w}}_{P,k}$, which is defined only from the point sample P, as long as $e = 3m_0^{-1/2}\,\mathrm{W}_2(\mu, \mathbf{1}_P) + 12 m_0^{1/\ell}\alpha_\mu^{-1/\ell}$ is less than $R/4$. Indeed, denoting by K^r and U^r the r-sublevel sets of the functions d_K and $\mathrm{d}^{\mathrm{w}}_{P,k}$, the sequence of inclusions

$$K^0 \subseteq U^e \subseteq K^{2e} \subseteq U^{3e} \subseteq K^{4e}$$

holds with high probability. By assumption, the function d_K has no critical values in the range $(0, 4e) \subseteq (0, R)$. Therefore, the rank of the image on the homology induced by inclusion $\mathsf{H}(U^e) \to \mathsf{H}(U^{3e})$ is equal to the Betti numbers of the set K. In the language of persistent homology [15], the persistent Betti numbers $\beta^{(e,3e)}$ of

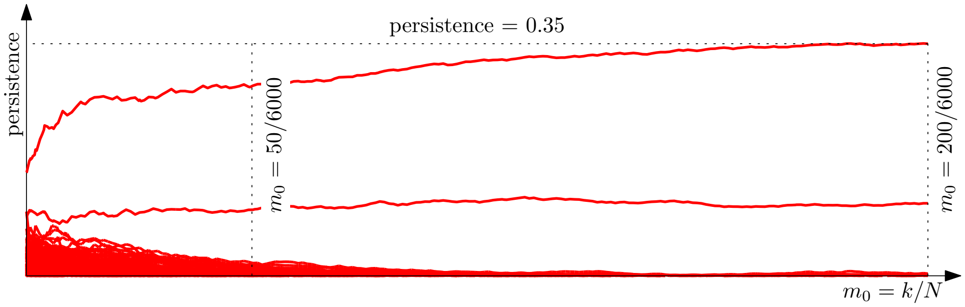


Figure 3. (PL-approximation of the) 1-dimensional persistence vineyard of the witnessed k-distance function. Topological features of the space, obscured by noise for low values of m_0, stand out as we increase the mass parameter.

the function $\mathrm{d}^{\mathrm{w}}_{P,k}$ are equal to the Betti numbers of the set K. Computationally, we can construct the sublevel sets U^e as either the Čech complex or the alpha shape.

Using the interleaving of Vietoris–Rips and Čech complexes, proved in Lemma 6.1, we can recover the Betti numbers from the Vietoris–Rips complex if $e < R/9$ [7]. From the following diagram of inclusions and homotopy equivalences

$$\begin{tikzcd}
& \check{\mathrm{C}}_e \arrow[r, phantom, "\subseteq"] \arrow[d, phantom, sloped, "\simeq"] & \mathrm{VR}_e \arrow[r, phantom, "\subseteq"] & \check{\mathrm{C}}_{2e} \arrow[rr, phantom, "\subseteq"] \arrow[d, phantom, sloped, "\simeq"] & & \check{\mathrm{C}}_{4e} \arrow[r, phantom, "\subseteq"] \arrow[d, phantom, sloped, "\simeq"] & \mathrm{VR}_{4e} \arrow[r, phantom, "\subseteq"] & \check{\mathrm{C}}_{8e} \arrow[d, phantom, sloped, "\simeq"] & \\
K_0 \arrow[r, phantom, "\subseteq"] & U^e \arrow[rr, phantom, "\subseteq"] & & U^{2e} \arrow[r, phantom, "\subseteq"] & K^{3e} \arrow[r, phantom, "\subseteq"] & U^{4e} \arrow[rr, phantom, "\subseteq"] & & U^{8e} \arrow[r, phantom, "\subseteq"] & K^{9e},
\end{tikzcd}$$

it follows that the map on homology $\mathsf{H}(\mathrm{VR}_e(U)) \to \mathsf{H}(\mathrm{VR}_{4e}(U))$ has the same rank as the homology of the space K.

Choice of the mass parameter. The language of persistent homology also suggests a strategy for choosing a mass parameter m_0 for the distance to a measure — a question not addressed by the original paper [6]. For every mass parameter m_0, the p-dimensional *persistence diagram* $\mathrm{Pers}_p(\mathrm{d}_{\mu,m_0})$ is a set of points $\{(b_i(m_0), d_i(m_0))\}_i$ in the extended plane $(\mathbb{R} \cup \{\infty\})^2$. Each of these points represents a homology class of dimension p in the sublevel sets of d_{μ,m_0}; $b_i(m_0)$ and $d_i(m_0)$ are the values at which it is born and dies. The distance to measure $\mathrm{d}_{\mathbf{1}_P,m_0}$ depends continuously on m_0 and, by the Stability Theorem [10], so do its persistence diagrams. Thus, one can use the vineyards algorithm [11] to track their evolution. Figure 3 illustrates such a construction for the point set in Figure 2 and the witnessed k-distance. It displays the evolution of the persistence $(d_1(m_0) - b_1(m_0))$ of each of the 1-dimensional homology classes as m_0 varies. This graph highlights the choices of the mass parameter that expose the two prominent classes (corresponding to the two loops of the figure-8).

Acknowledgements

The authors would like to thank the anonymous referees for their insightful feedback. This work has been partly supported by a grant from the French ANR, ANR-09-BLAN-0331-01, NSF grants FODAVA 0808515, CCF 1011228, and NSF/NIH grant 0900700. Quentin Mérigot would also like to acknowledge the support of the Fields Institute during the revision of this article. Dmitriy Morozov would like to acknowledge the support at LBNL of the DOE Office of Science, Advanced Scientific Computing Research, under award number KJ0402-KRD047, under contract number DE-AC02-05CH11231.

References

[1] N. Amenta and M. Bern. Surface reconstruction by Voronoi filtering. *Discrete and Computational Geometry*, 22(4):481–504, 1999.

[2] S. Arya and D. Mount. Computational geometry: proximity and location. *Handbook of Data Structures and Applications*, pages 63.1–63.22, 2005.

[3] F. Aurenhammer. A New Duality Result Concerning Voronoi Diagrams. *Discrete and Computational Geometry*, 5(1):243–254, 1990.

[4] 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.

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

[6] F. Chazal, D. Cohen-Steiner, and Q. Mérigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11:733–751, 2011.

[7] F. Chazal and S. Oudot. Towards persistence-based reconstruction in Euclidean spaces. *Proceedings of the ACM Symposium on Computational Geometry*, pages 232–241, 2008.

[8] K. Clarkson. Nearest-neighbor searching and metric space dimensions. *Nearest-Neighbor Methods for Learning and Vision: Theory and Practice*, pages 15–59, 2006.

[9] K. Clarkson and P. Shor. Applications of random sampling in computational geometry, II. *Discrete and Computational Geometry*, 4:387–421, 1989.

[10] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete and Computational Geometry*, 37(1):103–120, 2007.

[11] D. Cohen-Steiner, H. Edelsbrunner, and D. Morozov. Vines and vineyards by updating persistence in linear time. In *Proceedings of the ACM Symposium on Computational Geometry*, pages 119–126, 2006.

[12] S. Dasgupta. Learning mixtures of Gaussians. In *Proceedings of the IEEE Symposium on Foundations of Computer Science*, page 634, 1999.

[13] T. Dey and S. Goswami. Provable surface reconstruction from noisy samples. *Computational Geometry*, 35(1-2):124–141, 2006.

[14] H. Edelsbrunner. The union of balls and its dual shape. *Discrete and Computational Geometry*, 13:415–440, 1995.

[15] H. Edelsbrunner and J. Harer. Persistent homology — a survey. *Surveys on Discrete and Computational Geometry. Twenty Years Later*, pages 257–282, 2008.

[16] H. Edelsbrunner and J. Harer. *Computational Topology.* American Mathematical Society, 2010.

[17] P. Indyk. Nearest neighbors in high-dimensional spaces. *Handbook of Discrete and Computational Geometry*, pages 877–892, 2004.

[18] B. Kloeckner. Approximation by finitely supported measures. Preprint (arXiv:1003.1035), 2010.

[19] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The annals of Statistics*, 28(5):1302–1338, 2000.

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

[21] P. Niyogi, S. Smale, and S. Weinberger. A topological view of unsupervised learning from noisy data. *SIAM Journal on Computing*, 40(4):646–663, 2011.

[22] 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.

[23] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *Arxiv preprint arXiv:1011.3027*, 2010.

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

Appendix A. Pairwise Distances

A valuable property of the Vietoris–Rips complex is that the pairwise distances between the points suffice for its construction. We (re-)construct this property for the Vietoris–Rips complex on the witnessed barycenters. To do so, we need the weights of the barycenters as well as their pairwise distances in terms of the pairwise distances between the points of P.

Intersection criteria. Suppose we are given two weighted barycenters $\bar{p} = (1/k)\sum_{i=1}^k p_i$ and $\bar{q} = (1/k)\sum_{i=1}^k q_i$. We start by finding the intersection point between the line $(\bar{p}\bar{q})$ and the bisector of the power cells of $\bar{p}$ and $\bar{q}$. The point $x_t = (1-t)\bar{p} + t\bar{q}$ belongs to this bisector if and only if:

$$\begin{aligned}\|x_t - \bar{p}\|^2 - w_{\bar{p}} = \|x_t - \bar{q}\|^2 - w_{\bar{q}} &\iff t^2\|\bar{p} - \bar{q}\|^2 - w_{\bar{p}} = (1-t)^2\|\bar{p} - \bar{q}\|^2 - w_{\bar{q}} \\ &\iff 2t = 1 + \frac{w_{\bar{p}} - w_{\bar{q}}}{\|\bar{p} - \bar{q}\|^2}.\end{aligned}$$

The two balls $\mathrm{B}(\bar{p}, (r^2 + w_{\bar{p}})^{1/2})$ and $\mathrm{B}(\bar{q}, (r^2 + w_{\bar{q}})^{1/2})$ intersect if and only if the point x_t belongs to one of them, in which case it also belongs to the other. With the value of t that we found, this is equivalent to

$$\begin{aligned}\|x_t - \bar{p}\|^2 \le r^2 + w_{\bar{p}} &\iff t^2\|\bar{p} - \bar{q}\|^2 - w_{\bar{p}} \le r^2 \\ &\iff \frac{1}{4}\left(1 + \frac{w_{\bar{p}} - w_{\bar{q}}}{\|\bar{p} - \bar{q}\|^2}\right)^2 \|\bar{p} - \bar{q}\|^2 - w_{\bar{p}} \le r^2.\end{aligned}$$

Consequently, one can determine whether a segment $\{\bar{p}, \bar{q}\}$ belongs to the Vietoris–Rips complex of the witnessed barycenters with parameter r by knowing only the weights of the barycenters and their pairwise distances. In the next two paragraphs, we show how to express these quantities in terms of the pairwise distances between the data points.

Vertex weights. For a barycenter $\bar{p} = \frac{1}{k}(p_1 + \ldots + p_k)$ of k distinct points of P,

$$\begin{aligned}-w_{\bar{p}} &= \frac{1}{k}\sum_{i=1}^k \|\bar{p} - p_i\|^2 = \frac{1}{k}\sum_{i=1}^k \Big\|\frac{1}{k}\sum_{j=1}^k p_j - p_i\Big\|^2 = \frac{1}{k}\sum_{i=1}^k \Big\|\frac{1}{k}\sum_{j=1}^k (p_j - p_i)\Big\|^2 \\ &= \frac{1}{k^3}\sum_{i=1}^k\sum_{j=1}^k\sum_{l=1}^k \langle p_j - p_i | p_l - p_i\rangle \\ &= \frac{1}{2k^3}\sum_{i=1}^k\sum_{j>i}^k\sum_{l>j}^k (\|p_i - p_j\|^2 + \|p_i - p_l\|^2 + \|p_j - p_j\|^2) \\ &= \frac{k-2}{2k^3}\sum_{i=1}^k\sum_{j>i}^k \|p_i - p_j\|^2.\end{aligned}$$

The second to last equality is obtained by considering every triangle $\triangle(p_i, p_j, p_l)$ and observing that

$$\begin{aligned}0 &= (p_i - p_j + p_j - p_l + p_l - p_i)^2 \\ &= \|p_i - p_j\|^2 + \|p_j - p_l\|^2 + \|p_l - p_i\|^2 \\ &\quad +2\langle p_i - p_j | p_j - p_l\rangle + 2\langle p_i - p_j | p_l - p_i\rangle + 2\langle p_j - p_l | p_l - p_i\rangle,\end{aligned}$$

and the last equality comes from observing that each edge appears in $(k-2)$ triangles.

Barycenter distances. It remains to express the distance $\|\bar{p}-\bar{q}\|^2$ between the barycenters in terms of the pairwise distances between the points $\{p_1,\dots,p_k,q_1,\dots,q_k\}$.

$$\|\bar{p}-\bar{q}\|^2 = \|\frac{1}{k}\left(\sum q_i - \sum p_i\right)\|^2 = \frac{1}{k^2}\|\sum(q_i-p_i)\|^2 = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\langle(q_i-p_i)|(q_j-p_j)\rangle.$$

Rewriting

$$\langle(q_i-p_i)|(q_j-p_j)\rangle = \langle(q_i-p_i)|(q_j-p_i+p_i-p_j)\rangle = \langle(q_i-p_i)|(q_j-p_i)\rangle - \langle(q_i-p_i)|(p_j-p_i)\rangle$$

we get dot products between vectors with the same base point, which we express in terms of the areas of their respective triangles:

$$\langle(q_i-p_i)|(p_j-p_i)\rangle^2 = \|q_i-p_i\|^2\|p_j-p_i\|^2\cos^2\theta = \|q_i-p_i\|^2\|p_j-p_i\|^2 - 4S^2,$$

where S is the area of the triangle $\triangle(p_i,q_i,p_j)$. We compute it from the pairwise distances using Heron's formula, $S^2 = s(s-a)(s-b)(s-c)$, where s is the semiperimeter, and a,b,c are the lengths of the sides of the triangle.

E-mail address: guibas@cs.stanford.edu

Department of Computer Science, Stanford University

E-mail address: dmitriy@mrzv.org

Lawrence Berkeley National Laboratory

E-mail address: quentin.merigot@imag.fr

Laboratoire Jean Kuntzmann, Université de Grenoble and CNRS