Distance Functions and Geodesics on Point Clouds*

Facundo Mémoli[†] Guillermo Sapiro

Electrical and Computer Engineering Department
University of Minnesota
Minneapolis, MN 55455
e-mail: {*memoli, guille*}*@ece.umn.edu*

Abstract

An new paradigm for computing intrinsic distance functions and geodesics on sub-manifolds of $\mathbb{R}^d$ given by point clouds is introduced in this paper. The basic idea is that, as shown here, intrinsic distance functions and geodesics on general co-dimension sub-manifolds of $\mathbb{R}^d$ can be accurately approximated by extrinsic Euclidean ones computed inside a thin offset band surrounding the manifold. This permits the use of computationally optimal algorithms for computing distance functions in Cartesian grids. We use these algorithms, modified to deal with spaces with boundaries, and obtain also for the case of intrinsic distance functions on sub-manifolds of $\mathbb{R}^d$, a computationally optimal approach. For point clouds, the offset band is constructed without the need to explicitly find the underlying manifold, thereby computing intrinsic distance functions and geodesics on point clouds while skipping the manifold reconstruction step. The case of point clouds representing noisy samples of a sub-manifold of Euclidean space is studied as well. All the underlying theoretical results are presented along with experimental examples for diverse applications and comparisons to graph-based distance algorithms.

1 Introduction

One of the most popular sources of point clouds are 3D shape acquisition devices, such as laser range scanners, with applications in geoscience, art (e.g., archival), medicine (e.g., prohestetics), manufacturing (from cars to clothes), and security (e.g., recognition), among other disciplines. These scanners provide in general raw data in the form of (noisy) unorganized point clouds representing surface samples. With the increasing popularity and very broad applications of this source of data, it is natural and important to work directly with this representation, without having to go to the intermediate step of fitting a surface to it (step that can add computational complexity and introduce errors). See for example [3, 7, 9, 13, 20, 25, 26] for a few of the recent works with this type of data. Note that point clouds can also be used as primitives for visualization, e.g., [4, 13, 27], as well as for editing [36].

Another important field where point clouds are found is in the representation of high-dimensional manifolds by samples (see for example [15, 19, 32]). This type of high-dimensional and general co-dimension data appears in almost all disciplines, from computational biology to image analysis to financial data. Due to the extremely high dimensions in this case, it is impossible to perform manifold reconstruction, and the work needs to be performed directly on the raw data, meaning the point cloud.

This paper addresses one of the most fundamental operations in the study and processing of sub-manifolds of Euclidean space: the computation of intrinsic distance functions and geodesics. We show that this can be done by working directly with the point cloud, without the need for reconstructing the underlying manifold. The results are valid for general dimensions and co-dimensions, and for manifolds with or without boundary. These results include the analysis of noisy point clouds obtained from sampling the manifold.

A number of key building blocks are part of the framework here introduced. The first one is based on the fact that distance functions intrinsic to a given sub-manifold of $\mathbb{R}^d$ can be accurately approximated by Euclidean distance functions computed in a thin offset band that surrounds this manifold. This concept was first introduced in [22], where convergence results were given for co-dimension one sub-manifolds of $\mathbb{R}^d$ (hyper-surfaces) without boundary. This result is reviewed in §2. In this paper, we first extend these results to general co-dimension and to deal with manifolds with or without boundary, §3. We also show that the approximation is true not only for the intrinsic distance function but also for the intrinsic geodesic. This is not a straightforward corollary, since geodesics are based on the gradient of the distance function, which contains singularities at the cut locus [35, 28]. The approximation of intrinsic distance functions (and geodesics) by extrinsic Euclidean ones permits to compute them using computationally optimal algorithms in Cartesian grids (as long as the discretization operation is permitted, memory wise, see §7 and §8). These

*This work was partially supported by ONR and NSF. The resarch of F.M. is also supported by CSIC-Uruguay.

†F.M. is also with the Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de la República, Uruguay.

algorithms are based on the fact that the distance function satisfies a Hamilton-Jacobi partial differential equation (see §2), for which consistent and fast algorithms have been developed in Cartesian grids [14, 30, 31, 34][1] (see [16] for extensions to triangular meshes and [33] for other Hamilton-Jacobi equations). That is, due to these results, we can use computationally optimal algorithms in Cartesian grids (with boundaries) also to compute distance functions, and from them geodesics, intrinsic to a given manifold, and in a computationally optimal fashion.

Once these basic results are available, we can then proceed and work with point clouds. The basic idea here is to construct the offset band directly from the point cloud and without the intermediate step of manifold reconstruction. This is addressed in §4 and §5 for noise-free points which are manifold samples, and in §6 for points considered to be noisy samples of the manifold. For this (random) cases, we explicitly compute the probability that the constructed offset band contains the underlying manifold. In the experimental section, §7, we present a number of important applications. These applications are given to show the importance of this novel computational framework, and are by no means exhaustive. Concluding remarks are provided in §8 where we also report the directions our research is taking.

We should note that to the best of our knowledge, the only additional work explicitly addressing the computation of distance functions and geodesics for point clouds is the one reported in [2, 32].[2] The comparison of performances in presence of noise between our framework and this one is given in [23], where we prove the advantages of our theory.[3]

2 Preliminary Results

We first introduce some basic notation that will be used throughout the article. For a compact and connected set $\Omega \in I\!\!R^d$, $d_\Omega(\cdot,\cdot)$ denotes the intrinsic distance between any two points of Ω, measured by paths constrained to be in Ω. We will also assume the convention that if $A \subset I\!\!R^d$ is compact, and x, y are not both in A then $d_A(x,y) = D$, for some constant $D \gg \max_{x,y\in A} d_A(x,y)$. Given a k-dimensional sub-manifold $\mathcal{M}$ of $I\!\!R^d$, $\Omega^h_{\mathcal{M}}$ denotes the set $\{x \in I\!\!R^d : d(\mathcal{M},x) \le h\}$ (here the distance $d(\cdot,\cdot)$ is the Euclidean one). This is basically an h-offset of $\mathcal{M}$. To state that the sequence of functions $\{f_n(\cdot)\}_{n\in Z\!\!\!Z^+}$ uniformly converges to $f(\cdot)$ as $n \uparrow \infty$, we frequently write $f_n \overset{n}{\rightrightarrows} f$. For a given event $\mathcal{E}$, $\mathbb{P}(\mathcal{E})$ stands for its probability of occurring. For a random variable (R.V. from now on) X, its expected value is denoted by $\mathbb{E}(X)$. We denote by $X \sim \mathbf{U}[A]$ that the R.V. X is *uniformly distributed* in the set A. For a function $f : \Omega \to I\!\!R$, and a subset A of Ω, $f|_A : A \to I\!\!R$ denotes the restriction of f to A. Given a point x on the complete manifold $\mathcal{S}$, $B_\mathcal{S}(x,r)$ will denote the (intrinsic) open ball of radius $r > 0$ centered at x, and $B(y,r)$ will denote the *Euclidean* ball centered at y of radius r.

2.1 Prelude

In [22], we presented a new approach for the computation of weighted intrinsic distance functions on hyper-surfaces. The key starting idea is that distance functions satisfy the (intrinsic) Eikonal equation, a particular case of the general class of Hamilton-Jacobi partial differential equations. Given $p \in \mathcal{S}$ (an hyper-surface in $I\!\!R^d$), we want to compute $d_\mathcal{S}(p,\cdot) : \mathcal{S} \to I\!\!R^+ \cup \{0\}$, the intrinsic distance function from every point on $\mathcal{S}$ to p. It is well known that the distance function $d_\mathcal{S}(p,\cdot)$ satisfies, in the viscosity sense (see [21]), the equation $(*) \begin{cases} \|\nabla_\mathcal{S} d_\mathcal{S}(p,x)\| = 1 \ \forall x \in \mathcal{S} \\ d_\mathcal{S}(p,p) = 0 \end{cases}$, where $\nabla_\mathcal{S}$ is the intrinsic differentiation (gradient). Instead of solving this intrinsic Eikonal equation on $\mathcal{S}$, we solve the corresponding extrinsic one in the offset band $\Omega^h_\mathcal{S}$, $\begin{cases} \|\nabla_x d_{\Omega^h_\mathcal{S}}(p,x)\| = 1 \ \forall x \in \Omega^h_\mathcal{S} \\ d_{\Omega^h_\mathcal{S}}(p,p) = 0 \end{cases}$, where $d_{\Omega^h_\mathcal{S}}(p,\cdot)$ is the Euclidean distance and therefore now the differentiation is the usual one.

Theorem 1 ([22]) *Let p and q be any two points on the smooth hyper-surface $\mathcal{S}$ (orientable, without boundary), then $\left| d_\mathcal{S}(p,q) - d_{\Omega^h_\mathcal{S}}(p,q) \right| \le C_\mathcal{S}\sqrt{h}$, for small enough h,*[4] *where $C_\mathcal{S}$ is a constant depending on the geometry of $\mathcal{S}$.*

This simplification of the intrinsic problem into an extrinsic one permits the use of the computationally optimal algorithms mentioned in the introduction. This makes computing intrinsic distances, and from them geodesics, as simple and computationally efficient as computing them in Euclidean spaces. Moreover, as detailed in [22], the approximation of the intrinsic distance $d_\mathcal{S}$ by the extrinsic Euclidean one $d_{\Omega^h_\mathcal{S}}$ is never less accurate than the numerical error of these algorithms. This was the initial motivation for developing this approach, there are currently no "fast marching" methods that can be used to deal with the discretization of equation $(*)$.

It is the purpose of the present work to extend this Theorem to deal with: (1) sub-manifolds of $I\!\!R^d$ of any codimension and possibly with boundary, (2) sub-manifolds of

[1] Tsitsiklis first described an optimal-control type of approach to solve the Hamilton-Jacobi equation, while independently Sethian and Helmsen both developed techniques based on upwind numerical schemes.

[2] In addition to studying the computation of distance functions on point clouds, [2, 32] address the important combination of this with multidimensional scaling for manifold analysis. Prior work on using geodesics and multidimensional scaling can be found in [29].

[3] While concluding this paper, we learned of a recent extension to Isomap reported in [12]. This paper is also mesh based, and follows the geodesics approach in Isomap with a novel neighborhood/connectivity approach and a number of interesting theoretical results and novel dimensionality estimation contributions. Further analysis of Isomap, as a dimensionality reduction technique, can be found in [8].

[4] "Small enough h" means that $h < 1/\max_i \kappa_i(\mathcal{S})$, where $\kappa_i(\mathcal{S})$ is the i-th principal curvature of $\mathcal{S}$. This guarantees having smoothness in $\partial\Omega^h_\mathcal{S}$, see [22].

$I\!R^d$ represented as point clouds, (3) random sampling of sub-manifolds of $I\!R^d$ in presence of noise, and (4) convergence of geodesic curves in addition to distance functions. We should note that Theorem 1 holds even when the metric is not the one inherited from $I\!R^d$, obtaining weighted distance functions, see [22]. Although we will not present these new results in such generality, this is a simple extension that will be reported elsewhere.

3 Sub-Manifolds of $I\!R^d$

We first extend Theorem 1 to more general manifolds (with boundary and higher co-dimension) and we deal not only with distance functions but also with geodesics. The first extension is important for the learning of high-dimensional manifolds from samples and for scanned open volumes. The extension to geodesics is important for path planning on surfaces and for finding special curves such as crests and valleys, see [22]. Theorem 2 below presents uniform convergence results for both distances and geodesics in $\Omega^h_{\mathcal{S}}$, under no conditions on $\partial\mathcal{S}$ except some smoothness. Theorem 3 and Corollary 2 provide very useful rate of convergence estimates (for the uniform convergence of $d_{\Omega^h_{\mathcal{S}}}$ towards $d_{\mathcal{S}}$), under convexity assumptions on $\partial\mathcal{S}$.

Theorem 2 ([23]) *Let $\mathcal{S}$ be a compact C^2 sub-manifold of $I\!R^d$ with (possibly empty) smooth boundary $\partial\mathcal{S}$. Let x, y be any two points in $\mathcal{S}$. Then we have: (1) Uniform convergence of distances: $d_{\Omega^h_{\mathcal{S}}}|_{\mathcal{S}\times\mathcal{S}}(\cdot,\cdot) \overset{h\downarrow 0}{\rightrightarrows} d_{\mathcal{S}}(\cdot,\cdot)$; (2) Convergence of geodesics: Let x and y be joined by a unique minimizing geodesic $\gamma_{\mathcal{S}} : [0,1] \to \mathcal{S}$ over $\mathcal{S}$, and let $\gamma_h : [0,1] \to \Omega^h_{\mathcal{S}}$ be a $\Omega^h_{\mathcal{S}}$-minimizing geodesic, then $\gamma_h \overset{h\downarrow 0}{\rightrightarrows} \gamma_{\mathcal{S}}$.*

We now present a uniform rate of convergence result for the distance in the band in the case $\partial\mathcal{S} = \emptyset$, and from this we deduce Corollary 2 below, which deals with the case $\partial\mathcal{S} \neq \emptyset$. This result generalizes the one presented in [22] because it allows for any codimension.

Theorem 3 ([23]) *Under the same hypotheses of the Theorem above, with $\partial\mathcal{S} = \emptyset$, we have that for small enough $h > 0$, $\max_{(x,y)\in\mathcal{S}\times\mathcal{S}} \left| d_{\Omega^h_{\mathcal{S}}}|_{\mathcal{S}\times\mathcal{S}}(x,y) - d_{\mathcal{S}}(x,y) \right| \leq C_{\mathcal{S}}\sqrt{h}$, where the constant $C_{\mathcal{S}}$ does not depend on h. Also, we have the "relative" rate of convergence bound, $1 \leq \sup_{\substack{x,y\in\mathcal{S} \\ x\neq y}} \frac{d_{\mathcal{S}}(x,y)}{d_{\Omega^h_{\mathcal{S}}}(x,y)} \leq 1 + C_{\mathcal{S}}\sqrt{h}$*

We immediately obtain the following Corollary which will be useful ahead.

Corollary 1 *Let $p \in \mathcal{S}$, and $r \leq H$, then $B(p,r) \cap \mathcal{S} \subseteq B_{\mathcal{S}}(p, r(1 + C_{\mathcal{S}}\sqrt{r}))$.*

Corollary 2 *($\partial\mathcal{S} \neq \emptyset$) Under certain smoothness conditions, and assuming $\mathcal{S}$ to be strongly convex (see [10]), we have for small enough $h > 0$ the same conclusions of Theorem 3 (rate of convergence).*

In this section we have extended the results in [22] to geodesics and distance functions in general codimension manifolds with or without (smooth) boundary, thereby covering all possible (constant co-dimension) manifolds in common shape, graphics, visualization, and learning applications. We are now ready to extend this to manifolds represented as point clouds.

4 Distance Functions and Geodesics on Point Clouds

We are now interested in making distance and geodesic computations on manifolds represented as point clouds, i.e. *sampled manifolds*. Let $\mathcal{P}_n \triangleq \{p_1, \ldots, p_n\}$ be a set of n different points sampled from the *compact* sub-manifold $\mathcal{S}$ and define $\Omega^h_{\mathcal{P}_n} \triangleq \bigcup_{i=1}^n B(p_i, h)$.[5]

Let h and $\mathcal{P}_n$ be such that $\mathcal{S} \subseteq \Omega^h_{\mathcal{P}_n}$. We then have $(\mathcal{S} \subseteq)\Omega^h_{\mathcal{P}_n} \subseteq \Omega^h_{\mathcal{S}}$. We want to consider $d_{\Omega^h_{\mathcal{P}}}(p,q)$ for any pair of points $p, q \in \mathcal{S}$ and prove some kind of proximity to the real distance $d_{\mathcal{S}}(p,q)$. The argument carries over easily since $d_{\Omega^h_{\mathcal{S}}}(p,q) \leq d_{\Omega^h_{\mathcal{P}_n}}(p,q) \leq d_S(p,q)$, hence $0 \leq d_{\mathcal{S}}(p,q) - d_{\Omega^h_{\mathcal{P}_n}}(p,q) \leq d_{\mathcal{S}}(p,q) - d_{\Omega^h_{\mathcal{S}}}(p,q)$, and the rightmost quantity can be bounded by $C_{\mathcal{S}}\, h^{1/2}$ (see §3) in the case that $\partial\mathcal{S}$ is either strongly convex of void. The key condition is $\mathcal{S} \subset \Omega_{\mathcal{P}^h_n}$, something that can obviously be coped with using the compactness of $\mathcal{S}$.[6] We can then state the following

Theorem 4 ([23]) (Uniform Convergence for Point Clouds) *Let $\mathcal{S}$ be a compact smooth submanifold of $I\!R^d$ possibly with boundary $\partial\mathcal{S}$. Then* **1. General Case***: Given $\varepsilon > 0$, there exists $h_\varepsilon > 0$, such that $\forall\ 0 < h \leq h_\varepsilon$ one can find finite $n(h)$ and a set of points $\mathcal{P}_{n(h)}(h) = \{p_1(h), \ldots, p_{n(h)}(h)\}$ sampled from $\mathcal{S}$ such that $\max_{p,q\in\mathcal{S}} \left(d_{\mathcal{S}}(p,q) - d_{\Omega^h_{\mathcal{P}_{n(h)}(h)}}(p,q) \right) \leq \varepsilon$;* **2. $\partial\mathcal{S}$ is either void or convex:** *For every sufficiently small $h > 0$ one can find finite $n(h)$ and a set of points $\mathcal{P}_{n(h)}(h) = \{p_1(h), \ldots, p_{n(h)}(h)\}$ sampled from $\mathcal{S}$ such that $\max_{p,q\in\mathcal{S}} \left(d_{\mathcal{S}}(p,q) - d_{\Omega^h_{\mathcal{P}_{n(h)}(h)}}(p,q) \right) \leq C_{\mathcal{S}}\sqrt{h}$.*

In practise, one must worry about both the number (n) of points and the radii (h) of the balls. Obviously, there is a tradeoff between both quantities. If we want to use only few points, in order to cover $\mathcal{S}$ with the balls we have to increase the value of the radius. Clearly, there exists a value H such that for values of h smaller than H we don't change the topology, see [1, 7, 12]. This implies that the number of points must be larger that a certain lower bound. This result

[5] The balls now used are defined with respect to the metric of $I\!R^d$, they are not intrinsic.

[6] By compactness, given $h > 0$ we can find finite $N(h)$ and points $p_1, p_2, \ldots, p_{N(h)} \in \mathcal{S}$ such that $\mathcal{S} = \cup_{i=1}^{N(h)} B_{\mathcal{S}}(p_i, h)$. But since for $p \in \mathcal{S}$, $B_{\mathcal{S}}(p,h) \subset B(p,h) \cap \mathcal{S}$, and we also get $\mathcal{S} \subset \cup_{i=1}^{N(h)} B(p_i, h)$.

can be generalized to ellipsoids which can be locally adapted to the geometry of the point cloud [6], or from minimal spanning trees.

5 Random Manifold Sampling

In practise, we really do not have too much control over the way in which points are sampled by the acquisition device (e.g. scanner), or given by the learned sampled data. Therefore it is more realistic to make a probabilistic model of the situation and then try to conveniently estimate the probability of achieving a prescribed level of accuracy as a function of the number of points and the radii of the balls.

We now present a simple model for the current setting, while results for other models can be developed from the derivations below. Here we assume that the points in $\mathcal{P}_n$ are independently and identically sampled on the sub-manifold $\mathcal{S}$ in a uniform fashion,[7] we will write this as $p_i \sim \mathbf{U}[\mathcal{S}]$. For simplicity of exposition, we will restrict ourselves to the case when $\mathcal{S}$ has no boundary.[8] Also, we only deal with uniform i.i.d. sampling, other models for the sampling will be reported elsewhere.

We have to define the way in which we are going to measure accuracy. A possibility for such a measure is (for each $\varepsilon > 0$) $A_\varepsilon \triangleq \mathbb{P}\left(\max_{p,q\in\mathcal{S}}\left(d_{\mathcal{S}}(p,q) - d_{\Omega^h_{\mathcal{P}_n}}(p,q)\right) > \varepsilon\right)$. There is a potential problem with this way of testing accuracy, since we are assuming that when we use the approximate distance $d_{\Omega^h_{\mathcal{P}_n}}$, we will be evaluating it on $\mathcal{S}$. This might seem a bit awkward since we don't exactly know all the surface but just some points on it. Moreover, a more natural and real-problem-motivated approach would be to measure the discrepancy over $\mathcal{P}_n$ itself, over part of this set, or over another *trial* set of points $\mathcal{Q}_m$. However, since for any set of points $\mathcal{Q}_m \subset \mathcal{S}$ we have that $\left\{\max_{p,q\in\mathcal{Q}_m}\left(d_{\mathcal{S}}(p,q) - d_{\Omega^h_{\mathcal{P}_n}}(p,q)\right) > \varepsilon\right\} \subseteq \left\{\max_{p,q\in\mathcal{S}}\left(d_{\mathcal{S}}(p,q) - d_{\Omega^h_{\mathcal{P}_n}}(p,q)\right) > \varepsilon\right\}$, bounding A_ε suffices for dealing with any of the possibilities mentioned above. Notice that we are somehow considering $d_{\Omega^h_{\mathcal{P}_n}}$ to be defined for all pairs of points in $\mathcal{S}\times\mathcal{S}$, even if it might happen that $\mathcal{S}\cap\Omega^h_{\mathcal{P}_n} \neq \mathcal{S}$. In any case we extend $d_{\Omega^h_{\mathcal{P}_n}}$ to all of $\Omega^h_{\mathcal{S}} \times \Omega^h_{\mathcal{S}}$ by a large constant say $k \cdot diam\,(\mathcal{S})$, $k \gg 1$.

Let us define the events $\mathcal{E}_\varepsilon \triangleq \left\{\max_{p,q\in\mathcal{S}}\left(d_{\mathcal{S}}(p,q) - d_{\Omega^h_{\mathcal{P}_n}}(p,q)\right) > \varepsilon\right\}$, and $\mathcal{I}_{h,n} \triangleq \left\{\mathcal{S} \subseteq \Omega^h_{\mathcal{P}_n}\right\}$. Now, since $\mathcal{E}_\varepsilon = (\mathcal{E}_\varepsilon \cap \mathcal{I}_{h,n}) \cup (\mathcal{E}_\varepsilon \cap \mathcal{I}^c_{h,n})$, using the union bound and then Bayes rule we have $\mathbb{P}(\mathcal{E}_\varepsilon) \leq \mathbb{P}(\mathcal{E}_\varepsilon \,|\mathcal{I}_{h,n}) + \mathbb{P}\left(\mathcal{I}^c_{h,n}\right)$. It is clear now that we must find a convenient lower bound for the second term in the previous expression, the probability of covering all $\mathcal{S}$ with the union of balls. The first term can be easily dealt with using the convergence theorems presented in previous sections. We need a few lemmas.

Lemma 1 ([23]) *Let $x \in \mathcal{S}$ be a fixed point on $\mathcal{S}$ and $k = dim(\mathcal{S})$. Then under the hypotheses on $\mathcal{P}_n$ described above, there exists a constant $\omega_k > 0$ and a function $\theta_{\mathcal{S}}(\cdot)$ with $\lim_{h\downarrow 0}\frac{\theta_{\mathcal{S}}(h)}{h^{k+1}} = 0$ such that for small enough $h > 0$, $\mathbb{P}\left(\left\{x \notin \Omega^h_{\mathcal{P}_n} \cap \mathcal{S}\right\}\right) \leq \left(1 - \frac{\omega_k\, h^k + \theta_{\mathcal{S}}(h)}{\mu(\mathcal{S})}\right)^n$.*

Lemma 2 ([23]) *Under the hypotheses of the previous Lemma, let $\delta \in (0,h)$, then $\mathbb{P}\left(B_{\mathcal{S}}(x,\delta) \nsubseteq \Omega^h_{\mathcal{P}_n}\right) \leq \left(1 - \frac{\omega_k\,(h-\delta)^k + \theta_{\mathcal{S}}(h-\delta)}{\mu(\mathcal{S})}\right)^n$.*

Now, using compactness of $\mathcal{S}$ and an estimate of its $\frac{h}{2}$-covering number we can prove

Proposition 1 ([23]) *Let the set of hypotheses sustaining all of the previous lemmas hold. Let $([0,1)\ni)\, x_h \triangleq \frac{\omega_k (h/2)^k + \theta_{\mathcal{S}}(h/2)}{\mu(\mathcal{S})}$, where ω_k and $\theta_{\mathcal{S}}$ are given as in the proof of Lemma 1. Then $\mathbb{P}\left(\mathcal{S} \nsubseteq \Omega^h_{\mathcal{P}_n}\right) \leq \frac{(1-x_h)^n}{x_h}$.*[9]

We are now ready for the following convergence theorem.

Theorem 5 ([23]) *Let $\mathcal{S}$ be a k-dimensional smooth compact submanifold of $\mathbb{R}^d$. Let $\mathcal{P}_n = \{p_1,\ldots,p_n\} \subseteq \mathcal{S}$ be a i.i.d. set of points such that $p_i \sim \mathbf{U}[\mathcal{S}]$ for $1 \leq i \leq n$. Then if $h = h_n$ is such that $h_n \downarrow 0$ and $h^k_n \gtrsim \frac{\ln n}{n}$ holds as $n \uparrow \infty$, we have that for any $\varepsilon > 0$, $\mathbb{P}\left(\max_{p,q\in\mathcal{S}}\left(d_{\mathcal{S}}(p,q) - d_{\Omega^h_{\mathcal{P}_n}}(p,q)\right) > \varepsilon\right) \overset{n\uparrow\infty}{\longrightarrow} 0$*

By simple considerations, one can see that the rate of convergence can be estimated by a constant times $\left(\left(\frac{\ln n}{n}\right)^{1/2k} + \frac{1}{\ln n}\right)$.

This concludes our study of distance functions on point clouds (sampled manifolds). We now turn to the even more real scenario where the points are considered to be contaminated by noise.

6 Noisy Sampling of Manifolds

We assume that we have some uncertainty on the actual position of the surface, and model this as if each point in the set of sampled points is modified by a (not yet random) perturbation of magnitude smaller than Δ. More explicitly, each p_i is given as $p_i = p + \zeta \times \vec{v}$ for some $\vec{v} \in S^{d-1}$, some p in $\mathcal{S}$ and $\Delta \geq \zeta \geq 0$. Then we can guarantee that the point p from which p_i comes can be found inside $B(p_i,\Delta)\cap\mathcal{S}$. We are again interested in comparing $d_{\Omega^h_{\mathcal{P}_n}} : \Omega^h_{\mathcal{P}_n} \to \mathbb{R}^+ \cup \{0\}$ with $d_{\mathcal{S}} : \mathcal{S} \to \mathbb{R}^+ \cup \{0\}$, but now

[7] This means that for any subset $A \subseteq \mathcal{S}$, and any $p_i \in \mathcal{P}_n$, $\mathbb{P}(p_i \in A) = \frac{\mu(A)}{\mu(\mathcal{S})}$

[8] Even if we elaborate on the modifications needed in our arguments we should say that the same corresponding considerations presented in [2] are still valid in our case.

[9] One can prove with little extra work that under the same conditions on n and h, $\mathbb{P}\left(d_{\mathcal{H}}(\mathcal{S},\Omega^h_{\mathcal{P}_n}) > \varepsilon\right) \to 0$, where $d_{\mathcal{H}}$ is the Hausdorff distance between sets.

these functions have different domains, therefore we must be careful in defining a meaningful way of relating them. If we consider $\mathcal{F}_{\mathcal{S}}^{\Delta} \triangleq \{f \; s.t. \; f : \Omega_{\mathcal{S}}^{\Delta} \to \mathcal{S}, \; f(p) \in B(p, \Delta) \cap \mathcal{S}\}$, we can compare, for some $f \in \mathcal{F}_{\mathcal{S}}^{\Delta}$, and $1 \leq i, j \leq n$, $d^h_{\Omega_{\mathcal{P}_n}}(p_i, p_j)$ with $d_{\mathcal{S}}(f(p_i), f(p_j))$. Note that as the magnitude of the perturbation goes to zero, $\mathcal{F}_{\mathcal{S}}^{\Delta} \ni f(p) \stackrel{\Delta \downarrow 0}{\longrightarrow} p$, for $p \in \Omega_{\mathcal{S}}^{\Delta}$. The next step is to write $\max_{1 \leq i,j \leq n} \left| d^h_{\Omega_{\mathcal{P}_n}}(p_i, p_j) - d_{\mathcal{S}}(f(p_i), f(p_j)) \right|$, the biggest error we have for our set of points. And finally, the next logical step is to look at the worst possible choice for f: $\mathcal{L}_{\mathcal{S}}(\mathcal{P}_n; \Delta, h) \triangleq \sup_{f \in \mathcal{F}_{\mathcal{S}}^{\Delta}} \max_{1 \leq i,j \leq n} \left| d_{\mathcal{S}}(f(p_i), f(p_j)) - d_{\Omega^h_{\mathcal{P}_n}}(p_i, p_j) \right|$.

We start by presenting deterministic bounds for this expression, and only later will we be more (randomically) greedy, and in the spirit of Theorem 5, prove, for $\varepsilon > 0$, a result of the form ($\mathcal{L}_{\mathcal{S}}(\mathcal{P}_n; \Delta, h)$ will be a RV) $\mathbb{P}\left(\mathcal{L}_{\mathcal{S}}(\mathcal{P}_n; \Delta, h) > \varepsilon\right) \stackrel{n \uparrow \infty}{\longrightarrow} 0$.

6.1 Deterministic Setting

The idea is to prove that for some convenient function $\widehat{f} \in \mathcal{F}_{\mathcal{S}}^{\Delta}$, we can write $\mathcal{L}_{\mathcal{S}}(\mathcal{P}_n; \Delta, h) \leq \max_{1 \leq i,j \leq n} \left| d_{\mathcal{S}}(\widehat{f}(p_i), \widehat{f}(p_j)) - d_{\Omega^h_{\mathcal{P}_n}}(p_i, p_j) \right| + \lambda(h, \Delta)$, where $0 \leq \lambda(x, y) \stackrel{x,y \downarrow 0}{\longrightarrow} 0$. The natural candidate for $\widehat{f}$ is the orthogonal projection onto $\mathcal{S}$, $\Pi_{\mathcal{S}} : \Omega_{\mathcal{S}}^{H} \to \mathcal{S}$, whose properties are discussed in [23]. Then, we see that we can reduce everything to bounding $\max_{p,q \in \mathcal{S}} \left| d_{\mathcal{S}}(p, q) - d_{\Omega^h_{\mathcal{P}_n}}(p, q) \right|$. This is simple, since if $\mathcal{P}_n \subset \Omega_{\mathcal{S}}^{\Delta}$ then $\Omega^h_{\mathcal{P}_n} \subset \Omega_{\mathcal{S}}^{h+\Delta}$, and $d_{\mathcal{S}} \geq d_{\Omega^h_{\mathcal{P}_n}|_{\mathcal{S}}} \geq d_{\Omega^{h+\Delta}_{\mathcal{S}}|_{\mathcal{S}}}$, and finally from Theorem 3, $\|d_{\mathcal{S}} - d_{\Omega^h_{\mathcal{P}_n}}\|_{L^\infty(\mathcal{S})} \leq C_{\mathcal{S}}\sqrt{h + \Delta}$.

Let $\mathcal{S} \subset \Omega^h_{\mathcal{P}_n}$, $f \in \mathcal{F}_{\mathcal{S}}^{\Delta}$ and $1 \leq i, j \leq n$. Then, after using the triangle inequality a number of times we can write the bound, $|d_{\mathcal{S}}(f(p_i), f(p_j)) - d_{\Omega^h_{\mathcal{P}_n}}(p_i, p_j)| \leq 2 \sup_{f \in \mathcal{F}_{\mathcal{S}}^{\Delta}} \max_{p \in \mathcal{P}_n} d_{\mathcal{S}}(f(p), \Pi_{\mathcal{S}}(p)) + \max_{p,q \in \mathcal{S}} \left| d_{\mathcal{S}}(p, q) - d_{\Omega^h_{\mathcal{P}_n}}(p, q) \right| + \max_{p,q \in \mathcal{P}_n} \left| d_{\Omega^h_{\mathcal{P}_n}}(p, q) - d_{\Omega^h_{\mathcal{P}_n}}(\Pi_{\mathcal{S}}(p), \Pi_{\mathcal{S}}(q)) \right|$. The last term can be bounded by 2Δ, the one in the middle has already been discussed, hence we are left with the first one. Using Corollary 1, we find that since $f(p) \in B(\Pi_{\mathcal{S}}(p), 2\Delta) \cap \mathcal{S}$ then in fact $f(p) \in B_{\mathcal{S}}(\Pi_{\mathcal{S}}(p), 2\Delta(1 + C_{\mathcal{S}}\sqrt{\Delta}))$, and $d_{\mathcal{S}}(f(p), \Pi_{\mathcal{S}}(p)) \leq 2\Delta(1 + C_{\mathcal{S}}\sqrt{2}\sqrt{\Delta})$. Summing up, under the condition $\mathcal{S} \subset \Omega^h_{\mathcal{P}_n}$, we obtain the desired result: $\mathcal{L}_{\mathcal{S}}(\mathcal{P}_n; \Delta, h) \leq C_{\mathcal{S}}\sqrt{h + \Delta} + 2\Delta\left(2 + \sqrt{2}C_{\mathcal{S}}\sqrt{\Delta}\right)$.

6.2 Random Setting

Assume that $\{p_1, \ldots, p_n\}$ is a set of $i.i.d.$ random points such that each $p_i \sim \mathbf{U}[\Omega_{\mathcal{S}}^{\Delta}]$. At this time, we want to estimate the probability of having $\mathcal{S} \subseteq \Omega^h_{\mathcal{P}_n}$ It is easy to see that as a first "reality compliant" condition one should have that the noise level is not too big with respect to h. We will impose $h \geq \Delta$ for simplicity's sake.

Theorem 6 ([23]) *Let $\mathcal{S}$ be a k-dimensional smooth compact submanifold of $\mathbb{R}^d$. Let $\mathcal{P}_n = \{p_1, \ldots, p_n\}$ be such that $p_i \sim \mathbf{U}[\Omega_{\mathcal{S}}^{\Delta}]$ for $1 \leq i \leq n$. Then if $h = h_n$, $\Delta = \Delta_n$ are such that $\Delta_n \leq h_n$ and $h_n \downarrow 0$ and $\Delta_n^k \gtrsim \frac{\ln n}{n}$ as $n \uparrow \infty$, we have that for any $\varepsilon > 0$, $\mathbb{P}\left(\mathcal{L}_{\mathcal{S}}(\mathcal{P}_n; \Delta, h) > \varepsilon\right) \stackrel{n \uparrow \infty}{\longrightarrow} 0$.*

We have now concluded the analysis of the most general case for noisy sampling of manifolds. Note that although the results in this and in previous sections were presented for Euclidean balls, they can easily be extended to more general covering shapes (check Corollary 1 above), e.g. following [6, 15], or using minimal spanning trees, or from the local directions of the data [25]. Similarly, the results can be extended to other sampling or noise models.

7 Examples

We now present examples of distance functions and geodesics for point clouds, Figure 1 (first row), and use these computations to find intrinsic Voronoi diagrams, Figure 1 (second row), (see also [17, 18]).[10] We also present examples in high dimensions and use, following and extending [11], our results to compare manifolds given by point clouds. All these exercises were done to exemplify the importance of computing distance functions and geodesics on point clouds, and are by no means exhaustive.

The theoretical results presented in previous sections show that the intrinsic distance and geodesics can be approximated by the Euclidean ones computed in the band defined by the union of balls centered at the points of the cloud. The problem is then simplified to first computing this band, and then use well known computationally optimal techniques to compute the distances and geodesics inside this band, exactly as done in [22] for implicit surfaces. The band itself can be computed in several ways, and for the examples below we have used constant radii. Locally adaptive radii can be used, based for example on diameters obtained from minimal spanning trees. Automatic and local estimation of h defining $\Omega^h_{\mathcal{P}_n}$ was not pursued in this paper and is the subject of current research.

High Dimensional Data: We now present a simple example for high dimensional data. We embedded a circle of radius 15 in $\mathbb{R}^5$, and use a grid of size $34 \times 4 \times 4 \times 4 \times 34$ (with uniform spacing $\Delta x = 1$) such that each of the sample points is of the form $p_i = 15\left(\cos(\frac{2\pi i}{N}), 0, 0, 0, \sin(\frac{2\pi i}{N})\right) + (17, 2, 2, 2, 17)$, for $1 \leq i \leq N$. We then used our approach to compute the (approximate) distance function d_h in a band in $\mathbb{R}^5$, and then, the error $e_{ij} = |d_{\mathcal{S}}(p_i, p_j) - d_h(p_i, p_j)|$ for $i, j \in \{1, \ldots, N\}$. In our experiments we used $h =$

[10] All the figures in this paper are in color. VRML files corresponding to these examples can be found at mountains.ece.umn.edu/∼guille/pc.htm.

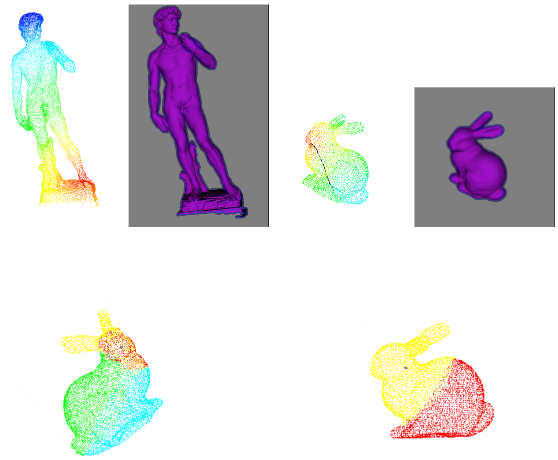

Figure 1: First row left: Intrinsic distance function for a point cloud. A point is selected in the head of the David, and the intrinsic distance is computed following the framework here introduced. The point cloud is colored according to their intrinsic distance to the selected point, going from bright red (close) to dark blue (far). The offset band, given by the union of balls, is shown next to the distance figure. First row right: Same as before, with a geodesic curve between two selected points. Second row: Voronoi diagram for point clouds. Four points (left) and two points (right) are selected on the cloud, and the point cloud is divided (colored) according to their geodesic distance to these four points. Note that this is a surface Voronoi, based on geodesics computed with our proposed framework, not an Euclidean one. *(This is a color figure). Datasets are courtesy of the Digital Michelangelo Project.*

$2.5 > \Delta x\sqrt{5}$.[11] We randomly sampled 500 points from the $N = 1000$ points used to construct the union of balls to build the 500×500 error matrix $((e_{ij}))$. We found $max_{ij}\{e_{ij}\} = 2.0275$, that is a 4.3% L_∞-error. In Figure 2 we show the histogram of all the (500^2) entries of $((e_{ij}))$. We should also note that when following the dimensionality reduction approach in [32], with the geodesic distance computation here proposed, the correct dimensionality of the circle was obtained.

Object Recognition: The goal of this application is to use our framework to compare manifolds given by point clouds. The comparison is done in an intrinsic way, that is, isometrically (bending) invariant. This application is motivated by [11], where they use geodesic distances (computed using a graph based approach) to compare $3D$ triangulated surfaces. In contrast with [11], we compare point clouds using our framework (which, is not only based in the original raw data, but it is also, as shown in [23], more robust to noise than mesh approaches as those of [11] and valid in any dimensions), and use a different procedure/similarity metric between the manifolds (in particular, we directly compare the distance matrices obtained from the point clouds, see [23] for details). As an example, this metric is used here to compare, in a bending invariant fashion, 5 human artificial models, 3 of them are bendings of a man and 2 bendings of a woman, see figure 3. The results of this cross comparison with the metric suggested in [23] are presented in Table 1 below.

[11] For a discussion on how to make a preliminary estimation of the value of h see [22].

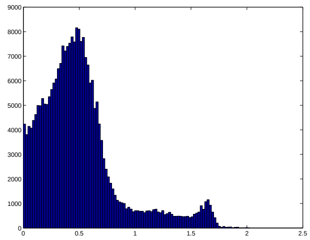

Figure 2: Histogram for the error in the case of a circle embedded in $I\!R^5$.

	M2	**M3**	**M5**	**W2**	**W3**
M2	$*$	0.0514	0.0570	0.4690	0.4853
M3	$*$	$*$	0.0206	0.4701	0.4859
M5	$*$	$*$	$*$	0.4702	0.4862
W2	$*$	$*$	$*$	$*$	0.2639
W3	$*$	$*$	$*$	$*$	$*$

Table 1: Cross comparisons for the point cloud human models using the error measure in [23]. *Datasets are courtesy of J. Leifman.*

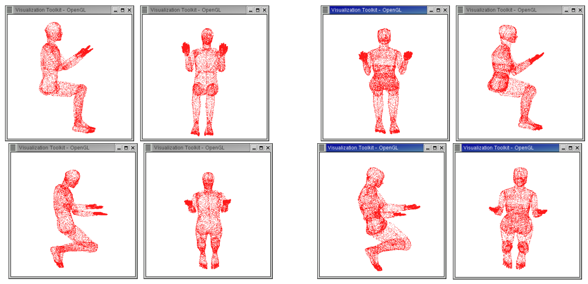

Figure 3: Left: MAN models. From top to bottom (two views of each model): MAN2 and MAN3. Right: Same for WOMAN models

8 Concluding Remarks

In this paper, we have shown how to compute distance functions and geodesics intrinsic to a generic manifold defined by a point cloud, without the intermediate step of manifold reconstruction. The basic idea is to use well developed computational algorithms for computing Euclidean distances in an offset band surrounding the manifold, and use these to approximate the intrinsic distance. The underlying theoretical results were complemented by experimental illustrations.

As mentioned in the introduction, an alternative technique to compute geodesic distances was introduced in [2, 32] (see also [12]). In contrast with our work, the effects of noise were not addressed in [2, 12]. Moreover, as one can see from considerations in [23], our framework is more robust to noise. We should note that the memory requirements of the current way of implementing our framework are large, and this needs to be addressed for very high dimensions. In particular, we are interested in direct ways of computing distances inside regions defined by union of balls, without the need to use the Hamilton-Jacobi approach.

We are currently working on the use of this framework to create multiresolution representations of point clouds (in collaboration with N. Dyn, see also [3, 7, 9, 26]), to further perform object recognition for larger libraries, and to compute basic geometric characteristics of the underlying manifold, all this of course without reconstructing the manifold (see [24] for recent results on normal computations for 2D and 3D noisy point clouds). Further applications of our framework for high dimensional data are also currently being addressed, beyond the preliminary (toy) results reported in §7. Of particular interest in this direction is the combination of this work with the one developed by Coifman and colleagues and the recent one in [12].

References

[1] N. Amenta, S. Choi, and R. Kolluri, "The power crust, unions of balls, and the medial axis transform," *Computational Geometry: Theory and Applications* **19**, pp. 127-153, 2001.

[2] M. Bernstein, V. de Silva, J. Langford, and J. Tenenbaum, "Graph approximations to geodesics on embedded manifolds," `http://isomap.stanford.edu/BdSLT.ps`

[3] J-D. Boissonnat and F. Cazals, "Coarse-to-fine surface simplification with geometric guarantees," in A. Chalmers and T-M. Rhyne, Editors, *EUROGRAPHICS '01*, Manchester, 2001.

[4] M. Botsch, A. Wiratanaya, and L. Kobbelt, "Efficient high quality rendering of point sampled geometry," *EUROGRAPHICS Workshop on Rendering*, 2002.

[5] I. Chavel, *Riemannian Geometry: A Modern Introduction*, Cambridge University Press, 1993.

[6] R. Coifman, personal communications.

[7] T. K. Dey, J. Giesen, and J. Hudson, "Decimating samples for mesh simplification," *Proc. 13th Canadian Conference on Computational Geometry*, pp. 85-88, 2001.

[8] D. L. Donoho and C. Grimes, "When does ISOMAP recover the natural parametrization of families of articulated images?," *Technical Report 2002-27: Department of Statistics, Stanford University*, 2002.

[9] N. Dyn, M. S. Floater, and A. Iske, "Adaptive thinning for bivariate scattered data," *Journal of Computational and Applied Mathematics* **145(2)**, pp. 505-517, 2002.

[10] M. doCarmo, *Riemannian Geometry.* Birkhäuser Boston, 1992.

[11] A. Elad (Elbaz) and R. Kimmel, "Bending invariant representations for surfaces," *Proc. of CVPR'01*, Hawaii, Dec. 2001.

[12] J. Giesen and U, Wagner, "Shape dimension and intrinsic metric from samples of manifolds with high co-dimension," *ACM Symposium on Computational Geometry*, San Diego, June 2003.

[13] M. Gross et al, *Point Based Computer Graphics*, EUROGRAPHICS Lecture Notes, 2002.

[14] J. Helmsen, E. G. Puckett, P. Collela, and M. Dorr, "Two new methods for simulating photolithography development in 3D," *Proc. SPIE Microlithography* **IX**, pp. 253, 1996.

[15] P. W. Jones, "Rectifable sets and the traveling salesman problem," *Invent. Math.* **102**, pp. 1-15, 1990.

[16] R. Kimmel and J. A. Sethian, "Computing geodesic paths on manifolds," *Proc. National Academy of Sciences* **95:15**, pp. 8431-8435, 1998.

[17] R. Kunze, F. E. Wolter, and T. Rausch, "Geodesic Voronoi diagrams on parametric surfaces," *IEEE*, 1997.

[18] G. Leibon and D. Letscher, "Delaunay triangulations and Voronoi diagrams for Riemannian manifolds," *Computational Geometry 2000*, Hong Kong, 2000.

[19] G. Lerman, "How to partition a low-dimensional data set into disjoint clusters of different geometric structure," preprint, 2000.

[20] L. Linsen and H. Prautzsch, "Local versus global triangulations," *EUROGRAPHICS '01*, 2001.

[21] C. Mantegazza and A.C. Menucci, "Hamilton-Jacobi equations and distance functions on Riemannian manifolds," preprint, 1999.

[22] F. Mémoli and G. Sapiro, "Fast computation of weighted distance functions and geodesics on implicit hyper-surfaces," *Journal of Computational Physics* **173:2**, pp. 730-764, 2001.

[23] F. Mémoli and G. Sapiro, "Distance functions and geodesics on point clouds," *IMA Preprint (www.ima.umn.edu)*, May 2003.

[24] N. J. Mitra and A. Nguyen, "Estimating surface normals in noisy point cloud data," *ACM Symposium on Computational Geometry*, San Diego, June 2003.

[25] M. Pauly and M. Gross, "Spectral processing of point-sampled geometry," *ACM SIGGRAPH*, pp. 379-386, 2001

[26] M. Pauly, M. Gross, and L. Kobbelt, "Efficient simplification of point-sampled surfaces," *IEEE Visualization 2002*.

[27] S. Rusinkiewicz and M. Levoy, "QSplat: A multiresolution point rendering system for large meshes," *Computer Graphics (SIGGRAPH 2000 Proceedings)*, 2000.

[28] T. Sakai, *Riemannian Geometry*, AMS Translations of Mathematical Monographs **149**, 1996.

[29] E. Schwartz, A. Shaw, and E. Wolfson, "A numerical solution to the generalized mapmaker's problem: Flattening nonconvex polyhedral surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11:9**, pp. 1005 -1008, 1989.

[30] J. Sethian, "Fast marching level set methods for three-dimensional photolithography development," *Proc. SPIE International Symposium on Microlithography*, Santa Clara, California, March, 1996.

[31] J. A. Sethian, "A fast marching level-set method for monotonically advancing fronts," *Proc. Nat. Acad. Sci.* **93:4**, pp. 1591-1595, 1996.

[32] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, pp. 2319-2323, December 2000.

[33] R. Tsai, L.-T. Cheng, S. Osher, and H.-K. Zhao, "Fast sweeping algorithms for a class of Hamilton-Jacobi equations," *UCLA-CAM Report*, 2001.

[34] J. N. Tsitsiklis, "Efficient algorithms for globally optimal trajectories," *IEEE Transactions on Automatic Control* **40** pp. 1528-1538, 1995.

[35] F. E. Wolter, "Cut loci in bordered and unbordered Riemannian manifolds," *Doctoral Dissertation, Technische Universität Berlin*, 1985.

[36] M. Zwicker, M. Pauly, O. Knoll, and M. Gross, "PointShop 3D: An interactive system for point-based surface editing," *SIGGRAPH*, 2002.