Nearest Neighbor Distances

Timothy Chu
CMU
tzchu@andrew.cmu.edu

Gary L. Miller
CMU
glmiller@cs.cmu.edu

Donald Sheehy
University of Connecticut
don.r.sheehy@gmail.com

September 25, 2017

Abstract

Some researchers have proposed using non-Euclidean metrics for clustering data points. Generally, the metric should recognize that two points in the same cluster are close, even if their Euclidean distance is far. Multiple proposals have been suggested, including the Edge-Squared Metric (a specific example of a graph geodesic) and the Nearest Neighbor Metric.

In this paper, we prove that the edge-squared and nearest-neighbor metrics are in fact equivalent. Previous best work showed that the edge-squared metric was a 3-approximation of the Nearest Neighbor metric. This paper represents one of the first proofs of equating a continuous metric with a discrete metric, using non-trivial discrete methods. Our proof uses the Kirszbraun theorem (also known as the Lipschitz Extension Theorem and Brehm's Extension Theorem), a notable theorem in functional analysis and computational geometry.

The results of our paper, combined with the results of Hwang, Damelin, and Hero, tell us that the Nearest Neighbor distance on i.i.d samples of a density is a reasonable constant approximation of a natural density-based distance function.

1 Intrinsic Distances

The foundational hypothesis supporting the fields of nonlinear dimensionality reduction [23, 4, 26], geometric inference [11], surface reconstruction [14], and topological data analysis [9] is that although points can be represented by vectors of real numbers, the intrinsic metric on those points can be highly non-euclidean.

Many algorithms across all these fields start by attempting to infer the intrinsic metric. The Isomap algorithm [23] is a paradigmatic example: pairwise distances are computed using shortest paths in either the k^{th} nearest neighbor graph or an ε-neighborhood graph. Then, multidimensional scaling gives the embedding of these "intrinsic" distances, which, with luck, approximate geodesic distances in the underlying sample space (i.e. the support of a distribution). Inspired by work on graph Laplacian convergence [15, 24], McQueen et al. [20] use ε-neighborhood graphs in their implementation of large-scale manifold learning algorithms. This approach, however, means that the approximating graph will be overconnected or underconnected for any measure in which the density varies, even for those adaptive samples [3, 14, 10] for which topologically precise reconstructions are computable.

In this paper, we show how two seemingly very different intrinsic metrics, the edge-squared metric and the nearest neighbor density-based distance are identical.

As the former is defined as the shortest path in a weighted graph, it can be computed exactly. These metrics or close variations thereof have appeared together in several previous works, with the discrete metric often used to *approximate* a continuous one [6, 13, 18]. It was not known, or perhaps even suspected, that these might actually be the same metric. This gives the first nontrivial example of a so-called density-based distances [21] that can be computed exactly.

Among intrinsic distance approximations, density-based distances have the advantage that they extend naturally to the entire Euclidean space. The density of a sample or a distribution is used to define a new Riemannian metric on $\mathbb{R}^d$ that scales space everywhere locally so that paths through high-density regions are shorter. Thus, shortest paths will be closer to geodesics if the points are sampled from a manifold, but no such strict hypothesis is necessary. One only needs a density or an approximation or a density estimate.

1.1 Preliminaries

The simplest, general form of a density-based distance built from a probability density f is

$$\mathbf{d}_f(x, y) := \inf_{\gamma} \int_0^1 f(\gamma(t))^{-\frac{1}{d}} \|\gamma'(t)\| dt,$$

where the infimum ranges over all piecewise smooth curves $\gamma : [0, 1] \to \mathbb{R}^d$ such that $\gamma(0) = x$ and $\gamma(1) = y$.

This metric is one of the simplest metrics that matches machine learning practitioner's intuitions about distance on points in the support of a distribution: distance between two points in a dense set should generally be considered shorter than distance through two points that are not connected by a dense set, even if their Euclidean distance is the same [2].

This kind of distance is also used in percolation theory and its generalization [17, 16, 18]s. In that field, another metric is also considered, the *power metric*. Given a point set $Q \in \mathbb{R}^d$ and number $p > 0$,

$$\mathbf{d}_p(x, y) = \min_{(q_0,\ldots,q_k)} \sum_{i=1}^{k} \|q_i - q_{i-1}\|^p,$$

where the minimum is over sequences of points $q_0, \ldots, q_k \in Q$ with $q_0 = x$ and $q_k = y$. Hwang et al. [18], it was shown that $\mathbf{d}_p$ converges (up to global scaling) to the geodesic distance in a Riemannian metric for sufficiently far points. We will call the power metric for $p = 2$, the *edge-squared metric*.

It was not known that $\mathbf{d}_2$ is identical (after rescaling) to a density-based distance $\mathbf{d}_N$, where N is the nearest neighbor density estimator $N(x) := (\min_{q \in Q} \|x - q\|)^{-d}$. That is, for all $x, y \in Q$,

$$\mathbf{d}_2(x, y) = 4\mathbf{d}_N(x, y).$$

This is particularly surprising given that for almost all pairs of points $x, y \in \mathbb{R}^d$, the shortest path in $\mathbf{d}_N$ is composed of hyperbolic arcs [13], yet for points in Q, the geodesic is a piecewise linear path through the points, allowing for a simple calculation of its length.

In previous work, it was shown that $\mathbf{d}_2$ is a 3-approximation to $\mathbf{d}_N$ [13]. The new result in this paper shows that no such approximation is necessary. They are the same. Recently, it was shown that $\mathbf{d}_N$ can be used to generalize the sampling conditions popular in surface reconstruction [10]. A similar definition with the opposite goal of avoiding high-density regions for robot motion planning was recently considered [1]. A variation of the density-avoiding approach was shown to yield poor results in machine learning applications [2].

2 Formal Definitions

For $x \in \mathbb{R}^d$, let $\|x\|$ denote the Euclidean norm. For a set of points $P \subset \mathbb{R}^d$, we can consider some other metrics as well. The *edge-squared metric* for $a, b \in P$ is

$$\mathbf{d}_2(a, b) = \min_{(p_0,\ldots,p_k)} \sum_{i=1}^{k} \|p_i - p_{i-1}\|^2,$$

where the minimum is over sequences of points $p_0, \ldots, p_k \in P$ with $p_0 = a$ and $p_k = b$.

A third metric on the points of P is called the *nearest neighbor metric* and is denoted $\mathbf{d}_N$. Before we can define it, we need a couple other definitions.

Given any finite set $P \subset \mathbb{R}^k$, there is a real-valued function $\mathbf{r}_P : \mathbb{R}^k \to \mathbb{R}$ defined as $\mathbf{r}_P(z) = \min_{x \in P} \|x - z\|$. A path is a continuous mapping $\gamma : [0, 1] \to \mathbb{R}^d$. Let $\mathrm{path}(a, b)$ denote the set of piecewise-C_1 paths from a to b. We will compute the lengths of paths relative to the distance function $\mathbf{r}_P$ as follows.

$$\ell(\gamma) := \int_0^1 \mathbf{r}_P(\gamma(t))\|\gamma'(t)\|dt.$$

By considering the velocity of γ, this definition is independent of the parameterization of the path. The nearest neighbor metric is then defined as

$$\mathbf{d}_N(a,b) := 4 \inf_{\gamma \in \mathrm{path}(a,b)} \ell(\gamma).$$

The factor of 4 normalizes the metrics. In particular, when P has only two points a and b, $\mathbf{d}_2(a,b) = \mathbf{d}_N(a,b)$. This reduces to a high school calculus exercise as the minimum path γ will be a straight line between the points and the nearest neighbor distance is

$$\mathbf{d}_N(a,b) = 4\int_0^1 \mathbf{r}_P(\gamma(t))\|\gamma'(t)\|dt = 8\int_0^{\frac{1}{2}} t\|a-b\|^2 dt = \|a-b\|^2 = \mathbf{d}_2(a,b).$$

This observation about pairs of points makes it easy to see that the nearest neighbor distance is never greater than the edge-squared distance as proven in the following lemma.

Lemma 2.1. *For all $s, p \in P$, we have $\mathbf{d}_N(s,p) \le \mathbf{d}_2(s,p)$.*

Proof. Fix any points $s, p \in P$. Let $q_0, \ldots, q_k \in P$ be such that $q_0 = s$, $q_k = p$ and

$$\mathbf{d}_2(s,p) = \sum_{i=1}^{k} \|q_i - q_{i-1}\|^2.$$

Let $\psi_i(t) = tq_i + (1-t)q_{i-1}$ be the straight line segment from q_{i-1} to q_i. Observe that $\ell(\psi_i) = \|q_i - q_{i-1}\|^2/4$, by the same argument as in the two point case. Then, let ψ be the concatenation of the ψ_i and it follows that

$$\mathbf{d}_2(s,p) = 4\ell(\psi) \ge 4 \inf_{\gamma \in \mathrm{path}(s,p)} \ell(\gamma) = \mathbf{d}_N(s,p).$$

□

Note that the nearest neighbor distance for a fixed set of points can be viewed as a manifold distance.

3 The Equivalence of the Metrics

Let $P \subset \mathbb{R}^d$ be a set of n points. Pick any *source* point $s \in P$. Order the points of P as $p_1, \ldots, p_n$ so that

$$\mathbf{d}_2(s, p_1) \le \cdots \le \mathbf{d}_2(s, p_n).$$

This will imply that $p_1 = s$. It will suffice to show that for all $p_i \in P$, we have $\mathbf{d}_2(s,p_i) = \mathbf{d}_N(s,p_i)$. There are three main steps:

1. We first show that when P is a subset of the vertices of an axis-aligned box, $\mathbf{d} = \mathbf{d}_N$. In this case, shortest paths for $\mathbf{d}$ are single edges and shortest paths for $\mathbf{d}_N$ are straight lines.

2. We then show how to lift the points from $\mathbb{R}^d$ to $\mathbb{R}^n$ by a Lipschitz map m that places all the points on the vertices of a box and preserves $\mathbf{d}_2(s,p)$ for all $p \in P$.

3. Finally, we show how the Lipschitz extension of m is also Lipschitz as a function between nearest neighbor metrics. We combine these pieces to show that $\mathbf{d} \le \mathbf{d}_N$. As $\mathbf{d} \ge \mathbf{d}_N$ (Lemma 2.1), this will conclude the proof that $\mathbf{d} = \mathbf{d}_N$.

3.1 Boxes

Let Q be the vertices of a box in $\mathbb{R}^n$. That is, there exist some positive real numbers $\alpha_1, \ldots, \alpha_n$ such that each $q \in Q$ can be written as $q = \sum_{i \in I} \alpha_i e_i$, for some $I \subseteq [n]$.

Let the source s be the origin. Let $\mathbf{r}_Q : \mathbb{R}^n \to \mathbb{R}$ be the distance function to the set Q. Setting $r_i(x) := \min\{x_i, \alpha_i - x_i\}$ (a lower bound on the difference in the ith coordinate to a vertex of the box), it follows that

$$\mathbf{r}_Q(x) \geq \sqrt{\sum_{i=1}^{n} r_i(x)^2}. \tag{1}$$

Let $\gamma : [0,1] \to \mathbb{R}^n$ be a curve in $\mathbb{R}^n$. Define $\gamma_i(t)$ to be the projection of γ onto its ith coordinate. Thus,

$$r_i(\gamma(t)) = \min\{\gamma_i(t), \alpha_i - \gamma_i(t)\} \tag{2}$$

and

$$\|\gamma'(t)\| = \sqrt{\sum_{i=1}^{n} \gamma_i'(t)^2}. \tag{3}$$

We can bound the length of γ as follows.

$$\begin{aligned}
\ell(\gamma) &= \int_0^1 \mathbf{r}_Q(\gamma(t))\|\gamma'(t)\|dt && \text{[by definition]} \\
&\geq \int_0^1 \left(\sqrt{\sum_{i=1}^{n} r_i(\gamma(t))^2} \sqrt{\sum_{i=1}^{n} \gamma_i'(t)^2} \right) dt && \text{[by (1) and (3)]} \\
&\geq \sum_{i=1}^{n} \int_0^1 r_i(\gamma(t))\gamma_i'(t)dt && \text{[Cauchy-Schwarz]} \\
&= \sum_{i=1}^{n} \left(\int_0^{\ell_i} \gamma_i(t)\gamma_i'(t)dt + \int_{\ell_i}^1 (\alpha_i - \gamma_i(t))\gamma_i'(t)dt \right) && \text{[by (2) where } \gamma_i(\ell_i) = \alpha_i/2] \\
&= \sum_{i=1}^{n} 2 \int_0^{\ell_i} \gamma_i(t)\gamma_i'(t)dt && \text{[by symmetry]} \\
&= \sum_{i=1}^{n} \frac{\alpha_i^2}{4} && \text{[basic calculus]}
\end{aligned}$$

It follows that if γ is any curve that starts at s and ends at $p = \sum_{i=1}^{n} \alpha_i e_i$, then $\mathbf{d}_N(s,p) = \mathbf{d}_2(s,p)$.

3.2 Lifting the points to $\mathbb{R}^n$

Define a mapping $m : P \to \mathbb{R}^n$ so that $m(p_1) = 0$ and otherwise

$$m(p_i) = m(p_{i-1}) + \sqrt{\mathbf{d}_2(s,p_i) - \mathbf{d}_2(s,p_{i-1})}e_i, \tag{4}$$

where the vectors e_i are the standard basis vectors in $\mathbb{R}^n$.

Lemma 3.1. *For all $p_i, p_j \in P$, we have*

(i) $\|m(p_j) - m(p_i)\| = \sqrt{|\mathbf{d}_2(s,p_j) - \mathbf{d}_2(s,p_i)|}$, *and*

(ii) $\|m(s) - m(p_j)\|^2 \le \|m(p_i)\|^2 + \|m(p_i) - m(p_j)\|^2$.

Proof. Proof of (i). Without loss of generality, let $i \le j$.

$$\begin{aligned}
\|m(p_j) - m(p_i)\| &= \left\| \sum_{k=i+1}^{j} \sqrt{\mathbf{d}_2(s,p_k) - \mathbf{d}_2(s,p_{k-1})}e_k \right\| && \text{[from the definition of } m\text{]} \\
&= \sqrt{\sum_{k=i+1}^{j} (\mathbf{d}_2(s,p_k) - \mathbf{d}_2(s,p_{k-1}))} && \text{[expand the norm]} \\
&= \sqrt{\mathbf{d}_2(s,p_j) - \mathbf{d}_2(s,p_i)}. && \text{[telescope the sum]}
\end{aligned}$$

Proof of (ii). As $m(s) = 0$, it suffice to observe that

$$\begin{aligned}
\|m(p_j)\|^2 &= \mathbf{d}_2(s,p_j) && \text{[by } (i)\text{]} \\
&\le \mathbf{d}_2(s,p_i) + |\mathbf{d}_2(s,p_j) - \mathbf{d}_2(s,p_i)| && \text{[basic arithmetic]} \\
&= \|m_(p_i)\|^2 + \|m(p_i) - m(p_j)\|^2 && \text{[by } (i)\text{]}
\end{aligned}$$

□

We can now show that m has all of the desired properties.

Proposition 3.2. *Let $P \subset \mathbb{R}^d$ be a set of n points, let $s \in P$ be a designated source point, and let $m : P \to \mathbb{R}^n$ be the map defined as in* (4)*. Let $\mathbf{d}'$ denote the edge squared metric for the point set $m(P)$ in $\mathbb{R}^n$. Then,*

(i) m is 1*-Lipschitz as a map between Euclidean metrics,*

(ii) m maps the points of P to the vertices of a box, and

(iii) m preserves the edge squared distance to s, i.e. $\mathbf{d}'(m(s), m(p)) = \mathbf{d}_2(s,p)$ for all $p \in P$.

Proof. Proof of (i). To prove the Lipschitz condition, fix any $a, b \in P$ and bound the distance as follows.

$$\begin{aligned}
\|m(a) - m(b)\| &= \sqrt{|\mathbf{d}_2(s,a) - \mathbf{d}_2(s,b)|} && \text{[Lemma 3.1(i)]} \\
&\le \sqrt{\mathbf{d}_2(a,b)} && \text{[triangle inequality]} \\
&\le \|a - b\| && \left[\mathbf{d}_2(a,b) \le \|a-b\|^2 \text{ by the definition of } \mathbf{d}\right]
\end{aligned}$$

Proof of (ii). That m maps P to the vertices of a box is immediate from the definition. The box has side lengths $\|m_i - m_{i-1}\|$ for all $i > 1$ and $p_i = \sum_{k=1}^{i} \|m_k - m_{k-1}\| e_k$.
Proof of (iii). We can now show that the edge squared distance to s is preserved. Let $q_0, \ldots, q_k$ be the shortest sequence of points of $m(P)$ that realizes the edge-squared distance from $m(s)$ to $m(p)$, i.e., $q_0 = m(s)$, $q_k = m(p)$, and

$$\mathbf{d}'(m(s), m(p)) = \sum_{i=1}^{k} \|m(q_i) - m(q_{i-1})\|^2.$$

If $k > 1$, then Lemma 3.1(ii) implies that removing q_1 gives a shorter sequence. Thus, we may assume $k = 1$ and therefore, by Lemma 3.1(i),

$$\mathbf{d}'(m(s), m(p)) = \|m(s) - m(p)\|^2 = \mathbf{d}_2(s, p).$$

□

3.3 The Lipschitz Extension

Proposition 3.2 and the Kirszbraun theorem on Lipschitz extensions imply that we can extend m to a 1-Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}^n$ such that $f(p) = m(p)$ for all $p \in P$ [19, 25, 8].

Lemma 3.3. *The function f is also 1-Lipschitz as mapping from $\mathbb{R}^d \to \mathbb{R}^n$ with both spaces endowed with the nearest neighbor metric.*

Proof. We are interested in two distance functions $\mathbf{r}_P : \mathbb{R}^d \to \mathbb{R}$ and $\mathbf{r}_{f(P)} : \mathbb{R}^n \to \mathbb{R}$. Recall that each is the distance to the nearest point in P or $f(P)$ respectively.

$$\begin{aligned}
\mathbf{r}_{f(P)}(f(x)) &= \min_{q \in f(P)} \|q - f(x)\| && \text{[by definition]} \\
&= \min_{p \in P} \|f(p) - f(x)\| && [q = f(p) \text{ for some } p] \\
&\leq \min_{p \in P} \|p - x\| && [f \text{ is 1-Lipschitz}] \\
&= \mathbf{r}_P(x). && \text{[by definition]}
\end{aligned}$$

For any curve $\gamma : [0, 1] \to \mathbb{R}^d$ and for all $t \in [0, 1]$, we have $\|(f \circ \gamma)'(t)\| \leq \|\gamma'(t)\|$. It then follows that

$$\ell'(f \circ \gamma) = \int_0^1 \mathbf{r}_{f(P)}(f(\gamma(t))) \|(f \circ \gamma)'(t)\| dt \leq \int_0^1 \mathbf{r}_P(\gamma(t)) \|\gamma'(t)\| dt = \ell(\gamma), \tag{5}$$

where ℓ' denotes the length with respect to $\mathbf{r}_{f(P)}$. Thus, for all $a, b \in P$,

$$\begin{aligned}
\mathbf{d}_N(a, b) &= 4 \inf_{\gamma \in \mathrm{path}(a,b)} \ell(\gamma) && \text{[by definition]} \\
&\geq 4 \inf_{\gamma \in \mathrm{path}(a,b)} \ell'(f \circ \gamma) && \text{[by (5)]} \\
&\geq 4 \inf_{\gamma' \in \mathrm{path}(f(a), f(b))} \ell'(\gamma') && [\text{because } f \circ \gamma \in \mathrm{path}(f(a), f(b))] \\
&= \mathbf{d}_N(f(a), f(b)). && \text{[by definition]}
\end{aligned}$$

□

Theorem 3.4. *For any point set $P \subset \mathbb{R}^d$, the edge squared metric $\mathbf{d}$ and the nearest neighbor metric $\mathbf{d}_N$ are identical.*

Proof. Fix any pair of points s and p in P. Define the Lipschitz mapping m and its extension f as in (4). Let $\mathbf{d}'$ and $\mathbf{d}'_N$ denote the edge-squared and nearest neighbor metrics on $f(P)$ in $\mathbb{R}^n$.

$$\begin{aligned}
\mathbf{d}_2(s,p) &= \mathbf{d}'(m(s), m(p)) && \text{[Proposition 3.2(iii)]} \\
&= \mathbf{d}'_N(m(s), m(p)) && \text{[}f(P)\text{ are vertices of a box]} \\
&\leq \mathbf{d}_N(s,p) && \text{[Lemma 3.3]}
\end{aligned}$$

We have just shown that $\mathbf{d} \leq \mathbf{d}_N$ and Lemma 2.1 states that $\mathbf{d} \geq \mathbf{d}_N$, so we conclude that $\mathbf{d} = \mathbf{d}_N$ as desired. ☐

4 Conclusions and Open Questions

The paper in many ways has generated more interesting open question than it has solved.

We has shown the the Nearest Neighbor metric and the Edge squared metric are equivalent. Is the power metric $\mathbf{d}_3(x, y)$ equivalent to $\mathbf{d}_{N_2}(x, y)$ where $N_2(x) := (\min_{q \in Q} \|x - q\|)^{-2d}$? We do not even know if is true for 2D.

In 2D we know that the edge squared metric has a linear size 1-spanner, namely, the Gabriel graph or the Delaunay Triangulation. It is also know that in 3D the Gabriel graph may contain a quadratic number of edges [12]. In general, the 1-spanner for the edge squared metric must be quadratic in general. Does the edge cubed or edge fourth power metric in 3D have a linear size 1-spanner? Does a linear-size $(1 + \epsilon)$-spanner for the edge squared metric?

One reason we consider this problem is because it's a widely studied problem in machine learning to find a sparse graph representation of a high-dimensional data set. While ad-hoc methods like k-NN graphs and ϵ-graphs do in a pinch, there graphs have known limitations, including the fact that k and ϵ must be hand-tuned. It's known that variants of the k-NN graph such as the unweighted k-NN graph has negative properties [2].

If a linear size spanner exists for edge cubed metrics, or if a sparse spanner exists for edge squared metrics, then we could obtain a sparse graph that captures our underlying metric extremely well, which we hope has a variety of applications for machine learning from clustering to dimension reduction to Isomap.

In addition, there are many other estimators one may use to estimate the density function. In particular, the second nearest neighbor estimator is quite popular. Is there a simple discrete metric that is equivalent to it?

It is known that a Euclidean $(1 + \epsilon)$-spanner can be made into a $(1 + \epsilon)^2$-spanner for the edge squared metric. Are there better spanners for the edge squared metric? Arya et. al. provided great bounds for Euclidean $(1 + \epsilon)$ spanners [5], assuming constant dimension. Can we do better?

References

[1] Pankaj K. Agarwal, Kyle Fox, and Oren Salzman. An efficient algorithm for computing high quality paths amid polygonal obstacles. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1179–1192, 2016.

[2] Morteza Alamgir and Ulrike von Luxburg. Shortest path distance in random k-nearest neighbor graphs. In *Proceedings of the 29th International Conference on Machine Learning*, 2012.

[3] Nina Amenta and Marshall Bern. Surface reconstruction by Voronoi filtering. *Discrete & Computational Geometry*, 22:481–504, 1999.

[4] D.C. Wunsch A. Zinovyev A.N. Gorban, B. Kégl, editor. *Principal Manifolds for Data Visualization and Dimension Reduction*, volume 58 of *LNCSE*. Springer, 2007.

[5] Sunil Arya, Gautam Das, David M. Mount, Jeffrey S. Salowe, and Michiel Smid. Euclidean spanners: Short, thin, and lanky. In *Proceedings of the Twenty-seventh Annual ACM Symposium on Theory of Computing*, STOC '95, pages 489–498, New York, NY, USA, 1995. ACM.

[6] Avleen Singh Bijral, Nathan D. Ratliff, and Nathan Srebro. Semi-supervised learning with density based distances. In Fabio Gagliardi Cozman and Avi Pfeffer, editors, *UAI*, pages 43–50. AUAI Press, 2011.

[7] Avleen Singh Bijral, Nathan D. Ratliff, and Nathan Srebro. Semi-supervised learning — wikipedia, the free encyclopedia, 2017.

[8] Ulrich Brehm. Extensions of distance reducing mappings to piecewise congruent mappings on m. *Journal of Geometry*, 16(1):187–193, 1981.

[9] Gunnar Carlsson. Topology and data. *Bull. Amer. Math. Soc.*, 46:255–308, 2009.

[10] Nicholas J. Cavanna and Donald R. Sheehy. Adaptive metrics for adaptive samples. In *Proceedings of the Canadian Conference on Computational Geometry*, 2016.

[11] Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11:733–751, 2011.

[12] Bernard Chazelle, Herbert Edelsbrunner, Leonidas J. Guibas, John E. Hershberger, Raimund Seidel, and Micha Sharir. Selecting heavily covered points. *SIAM J. Comput.*, 23(6):1138–1151, 1994.

[13] Michael B. Cohen, Brittany Terese Fasy, Gary L. Miller, Amir Nayyeri, Donald R. Sheehy, and Ameya Velingker. Approximating nearest neighbor distances. In *Proceedings of the Algorithms and Data Structures Symposium*, 2015.

[14] T. K. Dey. *Curve and Surface Reconstruction : Algorithms with Mathematical Analysis*. Cambridge University Press, 2007.

[15] Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg. Graph laplacians and their convergence on random neighborhood graphs. *Journal of Machine Learning Research*, 8:1325–1368, 2007.

[16] C. Douglas Howard and Charles M. Newman. Special invited paper: Geodesics and spanning tees for euclidean first-passage percolaton. *Ann. Probab.*, 29(2):577–623, 04 2001.

[17] C. Douglas Howeard and Charles M. Newman. Euclidean models of first-passage percolation. *Probab. Theory Relat. Fields*, 108, 1997.

[18] Sung Jin Hwang, Steven B. Damelin, and Alfred O. Hero III. Shortest path through random points. *Ann. Appl. Probab.*, 26(5):2791–2823, 10 2016.

[19] M. Kirszbraun. Über die zusammenziehende und lipschitzsche transformationen. *Fundamenta Mathematicae*, 22(1):77–108, 1934.

[20] James McQueen, Marina Meilă, Jacob VanderPlas, and Zhongyue Zhang. Megaman: Scalable manifold learning in python. *Journal of Machine Learning Research*, 17:1–5, 2016.

[21] Sajama and Alon Orlitsky. Estimating and computing density based distance metrics. In *ICML '05*, pages 760–767, New York, NY, USA, 2005. ACM.

[22] Prashant Sridhar. An experimental study into spectral and geometric approaches to data clustering. Master's thesis, Carnegie Mellon University, October 2015. CMU CS Tech Report CMU-CS-15-149.

[23] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[24] Daniel Ting, Ling huang, and Michael Jordan. An analysis of the convergence of graph laplacians. In *Proceedings of the 27th International Conference on Machine Learning*, pages 1079–1086, 2010.

[25] F. A. Valentine. A lipschitz condition preserving extension for a vector function. *American Journal of Mathematics*, 67(1):83–93, 1945.

[26] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[27] Pascal Vincent and Yoshua Bengio. Density sensitive metrics and kernels. In *Snowbird Workshop*, 2003.