Exact Computation of a Manifold Metric, via Shortest Paths on a graph

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

July 7, 2019

Abstract

In the machine learning setting, distances between two datapoints in a Euclidean point set are considered short if they are in the same data cluster - even if their Euclidean distance is long. A simple metric with this property is the Nearest neighbor metric. This metric and its close variants have been studied in the past by multiple researchers.

One key problem on manifold metrics, dating back for four centuries, is computing them exactly. The Nearest Neighbor metric is a manifold-based metric, and is defined as the infimum cost path over an uncountable number of paths that can go 'anywhere' on a continuous manifold. This makes computing the Nearest Neighbor metric exactly challenging, even for a fixed set of four points in two dimensions. In this paper, we overcome this challenge by equating the Nearest Neighbor metric to a shortest-path distance on a simple geometric graph, in all cases. This considerably strengthens the work of Cohen et al. We then exactly compute a generalization of the Nearest Neighbor metric, called the q-Nearest Neighbor metric, for small point sets. Our tools include conservative vector fields, Lipschitz extensions, minimum cost flows, and barycentric subdivisions, all applied to a geometric object we call the q-screw simplex.

The key geometric object in our proof, the q-screw simplex, was first discovered by John Von Neumann and Issai Schoenberg. This simplex is defined by taking n points on a line, applying the $1/q$ power to all pairwise distances, and isometrically embedding the resulting point set into Euclidean space. In this paper, we prove that the q-screw simplex has a deep connection to fractional Laplacians, a differential operator that appears in a wide variety of physics and mathematical settings. We use this connection to prove that the q-screw simplex embeds isometrically into Effective resistance distance, and that its volume and circumcenter can be approximated efficiently using a Laplacian determinant estimator and Laplacian system solver respectively. We also show that any finite l_1 metric raised to the $1/q$ power is isometrically embeddable into l_1, mirroring a famous theorem by Schoenberg on l_2.

Finally, we compute sparse spanners for the Nearest Neighbor metric and show links to existing problems ranging from persistent homology, geodesic approximation, density-based clustering, and Euclidean MST.

1 Introduction

A foundational hypothesis in non-linear dimension reduction and machine learning is that data can be represented as points in Euclidean space, and appropriate metrics on these points can be generated to solve a variety of problems including clustering, classification, regression, surface reconstruction, topological property inference, and more. In machine learning, two data points should intuitively be considered close if they are in the same data cluster, even if their Euclidean distance is far. This property is called the **density-sensitive** property.

Density-sensitive metrics are considered fundamental in the study of machine learning, and are implicitly central in celebrated machine learning methods such as k-NN graph methods, manifold learning, level-set methods, single-linkage clustering, and Euclidean MST-based clustering (See Appendix **??** for details). The construction of appropriate density-sensitive metrics is an active area of research in machine learning. We consider a simple density-sensitive metric with an underlying manifold structure, whose close variants have been studied by multiple researchers. This metric is called the Nearest Neighbor Metric. In this paper, we show how to compute the Nearest Neighbor metric exactly for any dimension, which solves one of the most important and challenging problem on any manifold-based metric.

To define the nearest neighbor metric, we first define the notion of a density-based distance. This is a slight variation of the original definition from [].

Definition 1.1. *Given a continuous cost function $c : \mathbb{R}^k \to \mathbb{R}$, we define the density-based cost of a path γ relative to c as:*

$$\ell_c(\gamma) = \int_0^1 c(\gamma(t))\|\gamma'(t)\|dt.$$

*Here, the path γ is defined as a continuous mapping $\gamma : [0,1] \to \mathbb{R}^k$. 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. We then define the **density-based distance** between two points $a, b \in \mathbb{R}^k$ as*

$$d_c(a,b) \inf_{\gamma\in \mathrm{path}(a,b)} \ell_c(\gamma)$$

PICTURE

Conceptually, the density-based cost of a path is the weighted path length, where each infinitesimal path piece is weighted with cost function c. Density-based distances have been notable in the machine learning setting for over a decade []. To build a density-sensitive metric from density-based distances, we would like a cost function c that is small when close to the data set, and large when far away. The Nearest Neighbor function is the most natural candidate, and has been traditionally used as a proximity measure between points and a data set in both the geometry and machine learning settings []. It has been used as such in Nearest Neighbor (and k-NN) classification, k-means/medians/center clustering, finite element methods, and any of the hundreds of methods that use Voronoi diagrams or Delaunay triangulation as intermediate data structures.

Definition 1.2. *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 $\mathbf{r}_P(z) = \min_{x\in p} \|x - z\|$. The **Nearest Neighbor Cost** of a path γ is $\ell_{2\mathbf{r}_P}$, which we will shorthand to ℓ_N. The **Nearest Neighbor Metric** between two points is defined as $\mathbf{d}_{2\mathbf{r}_P}$, which we short-hand as $\mathbf{d}_N$.*

The factor of 2 in the definition of the Nearest Neighbor Cost is a normalizing constant.

The Nearest Neighbor metric, and density-based distances in general, are examples of manifold geodesics (see [] for details). Manifold geodesics are defined by embedding a point set into a continuous geometric manifold, and computing the infimum length path on the manifold structure between points. Within computer science, dozens of foundational papers in machine learning and surface reconstruction rely on manifold-based metrics to perform clustering, classification, regression, surface reconstruction, persistent homology, and more. Manifold geodesics predate computer science, and are the cornerstone of many fields of physics and mathematics. Exactly computing geodesics is fundamental to countless areas of physics including: the brachistochrone and minimal-drag-bullet problem of Bernoulli and Newton, exactly determining a particle's trajectory in classical physics (Hamilton's Principle of Least Action), computing the path of light through a non-homogeneous medium (Snell's law), finding the evolution of wave functions in quantum mechanics over time (Feynman path integrals), and determining the path of light in the presence of gravitational fields (General Relativity, Schwarzschild metric). In mathematics, manifold geodesics appear in nearly every branch of higher mathematics including differential equations, differential geometry, Lie theory, calculus of variations, algebraic geometry, and topology.

One of the most significant problems on any manifold geodesic is how to compute it exactly. Exact computation of manifold metrics is considered a fundamental problem in mathematics and physics, dating back for four centuries: entire fields of mathematics, including the celebrated calculus of variations, have arisen to tackle this []. Historically, mathematicians placed strong emphasis on exact computation as opposed to constant factor approximations. An algorithmic problem on manifold geodesics, with modern origins, is to $(1+\varepsilon)$ approximate these metrics efficiently on a computer. The core difficulty in the first problem is that geodesics are the minimum cost path out of an uncountable number of paths that can travel 'anywhere' on the manifold structure. This makes exactly computing these metrics challenging, even in the case of the Nearest Neighbor metric for just four fixed points in two dimensions (the authors are unaware of any easy method for this simplified task). The core tool for exactly computing manifold metrics, calculus of variations, is intractable on the nearest neighbor metric due to the metric's heavy dependence on the Voronoi diagram of the point set, which can be quite complicated for even five points in two dimensions (for more on this approach and its limitations, see []). Calculus of variations can show that the optimal nearest neighbor path is piecewise hyperbolic, but this is generally insufficient to exactly compute the nearest neighbor metric - there are point sets where there are many smooth, piecewise hyperbolic paths between two data points with different costs.

In this paper, we solve both problems: we exactly compute the Nearest Neighbor metric in all cases, and we $(1+\varepsilon)$ approximate it quickly. Our approach is based on conservative vector fields, Lipschitz extensions, and minimum cost flows on a graph. We combine these tools to prove that the nearest neighbor metric is exactly equal to a shortest path distance on a geometric graph, the so-called edge-squared metric, in all cases. This allows us to compute the nearest-neighbor metric exactly for any given point set in polynomial time, and it is the only known (non-trivial) density-based distance that can be computed discretely.

Definition 1.3. *Given points in Euclidean space, the* ***edge-squared graph*** *is the complete graph of Euclidean distances squared. The* ***edge-squared metric*** *is the shortest path distance between two points on this graph.*

Theorem 1.4. *The nearest neighbor metric and edge squared metric are equivalent for any compact point set in arbitrary dimension*

The exact equality is realized when the nearest neighbor path is piecewise linear, traveling straight from data point to data point. The edge squared metric has been previously studied by multiple researchers in machine learning and power-efficient wireless networks, but previously has only been linked to the nearest neighbor metric by a fairly weak 3-approximation. Exact equality is considered highly surprising for at least four reasons:

1. The optimal nearest neighbor path for two points not in the dataset is generally piecewise hyperbolic. This holds true even when the dataset is a single point, and was established by [] using tools in Riemannian surfaces and the complex plane. Meanwhile, Theorem **??** implies an optimal nearest neighbor path is piecewise linear when the start and end points are in the dataset!

2. There are simple and natural variants of the Nearest Neighbor metric, for which no analog of Theorem 1.6 is known nor suspected. These variants are known as the q-Nearest Neighbor metric, for $1 < q < 2$, and we will formally define these metrics later in the introduction. When $q = 2$, these metrics coincide with the Nearest Neighbor metric. gives us a natural suite of metrics that smoothly converge to the Nearest Neighbor metric, for which no theorem like Theorem 1.6 is known.

3. Even for just three points in a right triangle configuration, there exist an uncountable suite of optimal-cost paths between the two endpoints of the hypotenuse. Each path in this uncountable suite is piecewise hyperbolic, but, surprisingly, they all have the exact same cost as the edge-squared distance. In fact, the union of these paths is the entire right triangle. Thus, lowering the Nearest Neighbor function anywhere inside the triangle and using this function to build a density-based distance will necessarily break Theorem 1.6 on these points. This establishes that Theorem 1.6 is fairly tight, and won't work for any cost function meaningfully less than the Nearest Neighbor function on the right triangle point set.

4. This theorem holds for any compact point set, whether its n points in $n-1$ dimensional space or a finite union of compact geometric blobs in countably infinite dimension. The geometry of both can be quite complicated, and it is generally hard to prove these types of results on arbitrary point sets (TIM:CHECK AND MAKE SURE THE PROOF WORKS)

We can now tackle a second problem of interest for manifold geodesics, which is efficiently $(1+\varepsilon)$ approximating them. In this paper, we show that the nearest neighbor metric admits $(1+\epsilon)$ spanners computable in nearly-linear time, with linear size, for any point set in constant dimension. Remarkably, these spanners are significantly sparser and faster to compute than the theoretically optimal Euclidean spanners with the same approximation constant, and nearly match the sparsity of the best known Euclidean Steiner spanners. Moreover, if the point set comes from a well-behaved probability distribution in constant dimension (a foundational assumption in machine learning []), we show that the nearest neighbor metric has perfect 1-spanners of nearly linear size. The latter result is impossible for many non-density sensitive metrics, such as the Euclidean metric. Both results rely on Theorem 1.6, and significantly improve the Nearest Neighbor spanners of Cohen et al in [].

Theorem 1.6 and our spanner theorems solve two core problems of interest for the nearest neighbor metric: exactly computing it for any dimension, and approximating it quickly for both general point sets and point sets arising from a well-behaved probability distribution in constant dimension. This is the first work we know of that computes a manifold metric exactly without calculus of variations, and we hope that our tools can be useful for other metric computations and approximations.

Besides for this contribution, we also generalize the Nearest Neighbor Metric to the q-Nearest Neighbor metric (abbreviated q-NN for short), and exactly compute this metric for all small point sets for all $q > 2$. We do this by equating it to the q-edge power metrics, which we will define later. We then use Theorem 1.6 to compute the persistent homology of the Nearest Neighbor metric, a task important in computational geometry. Additionally, we study the behavior of the Nearest Neighbor metric when the points are drawn from a well-behaved distribution, as the number of points goes to infinity. This turns out to converge w.h.p. to an extremely nice, $1 + o(1)$-approximation of a beautiful geodesic defined on the underlying density previously studied by applied probability theorists. This strengthens the work of Hwang, Hero, and Damelin, who showed that the Nearest Neighbor metric converged to a $O(1)$-approximation of this beautiful geodesic. This geodesic is a beautiful and natural generalization of both Euclidean distances and a distance fundamental for clustering using level-set methods. We further show that q-edge power metrics (and thus, it is hoped, the q-Nearest Neighbor metrics) are natural generalizations of maximum-edge-length distances on Euclidean MSTs, which in turn are fundamental for celebrated clustering methods like single-linkage clustering []. This implies that the q-edge power metric, and the Nearest Neighbor metric, can be used to generalize popular methods in clustering.

Our final set of theorems regards the q-screw simplex, the core geometric object in our proof of Theorem 1.6 and its generalizations. The q-screw simplex was first discovered by John Von Neumann and Issai Schoenberg. It is defined by taking n points anywhere on a line, taking the $1/q$ power of the distances, and isometrically embedding the resulting distances into Euclidean space. The fact that such an embedding exists was the core contribution of Schoenberg and Von Neumann in **??**. The central role of q-screw simplices in our proofs motivates us to develop new theorems on the geometry of these objects. Surprisingly, we find that these simplices are useful for proving generalizations of Von Neumann's work, and are deeply related to spectral graph theory.

Isometric embedding is a topic of wide interest in the field of metric geometry, and has been studied for many decades. Von Neumann and Issai Schoenberg proved in their seminal work that any q-screw simplex is isometrically embeddable in l_2. We extend their work to prove a stronger result: the q-screw simplex isometrically embeds into the space of Effective Resistance metrics. Simple metrics like the square in l_2 are not isometrically embeddable into this class of metrics, and thus, most Euclidean metrics are not expected to isometrically embed into Effective Resistance distance. Isometric embedding into effective resistance metrics has been a popular question in spectral graph theory [], and this is the first result we know of where a geometric distance defined without an obvious underlying electrical network embeds isometrically into Effective Resistances. We prove this isometry by showing a deep link between q-screw simplices and a differential operator known as the fractional Laplacian, which has wide applications in fields including fractional quantum physics [], cell membrane biology [], financial mathematics [], Brownian motion [], differential equations [], semi-groups [], Fourier analysis [], and more []. This further allows us to show that fundamental geometric quantities like circumcenters (essential for Voronoi diagram construction) and volumes on the q-screw simplex can be determined using fundamental primitives on graph Laplacians, in this case Laplacian system solving and Laplacian determinant estimation respectively. The fractional Laplacian can be interpreted as a natural example of a geometric resistive graph, first introduced by Alman et. al. in []. We further conjecture that taking the q^{th} root of any tree metric is isometrically embeddable into effective resistance, which would imply that the Gomory Hu tree (and thus the inverse min-cut distance) embeds isometrically into effective resistances.

We also provide the first known closed form finite-dimensional embedding of the q-screw simplex into Euclidean space, when $q > 2$. The work of Von Neumann et. Al. proved the simplex's existence for $q > 1$ by embedding it into infinite dimensional Hilbert space using theorems from complex analysis, functional analysis, infinite dimensional Hilbert space theory, and Fourier analysis. Our embedding uses only elementary techniques of eigenvector computation on finite matrices. We hope that this embedding makes the work of Von Neumann and Schoenberg more accessible. We use our new embedding to state and prove a generalization of Von Neumann and Schoenbergs theorem (on the embeddability of the q-screw simplex):

Theorem 1.5. *For points $p_1, \dots p_n \in \mathbb{R}^n$ and any $q > 1$, the metric $D(p_i, p_j) = |p_i - p_j|_1^{1/q}$ is isometrically embeddable into l_1.*

This mirrors Schoenberg's famous theorem that any finite l_2 metric, raised to the $1/q$ power for $q > 1$, is isometrically embeddable in l_2.

1.1 Contributions

Our paper has three main theorems.

Theorem 1.6. *Given a point set $P \in \mathbb{R}^d$, the edge-squared metric on P and the nearest-neighbor geodesic on P are always equivalent.*

Theorem 1.7. *For any set of points in $\mathbb{R}^d$ for constant d, there exists a $(1+\varepsilon)$ spanner of the edge-squared metric, with size $O\left(n\varepsilon^{-d/2}\right)$ computable in time $O\left(n \log n + n\varepsilon^{-d/2} \log \frac{1}{\varepsilon}\right)$. The $\log \frac{1}{\varepsilon}$ term goes away given a fast floor function.*

Theorem 1.8. *Suppose points P in Euclidean space are drawn i.i.d from a Lipschitz probability density bounded above and below by a constant, with support on a smooth, connected, compact manifold with intrinsic dimension d, and smooth boundary of bounded curvature. Then w.h.p. the k-NN graph of P for $k = O(2^d \ln n)$ and edges weighted with Euclidean distance squared, is a 1-spanner of the edge-squared metric on P.*

Theorem 1.6 considerably strengthens a result from in [**?**], which showed $\mathbf{d}_2$ is a 3-approximation of $\mathbf{d}_N$. Our theorem finds $\mathbf{d}_N$ exactly, and lets us compute the persistent homology of $\mathbf{d}_N$. $\mathbf{d}_N$ is defined on all points in space, and is thus a metric extension [**?**] of the edge-squared metric and of negative type distances [**?**] to the entire space.

Theorem 1.7 proves that a $(1+\varepsilon)$-spanner of the edge-squared metric with points in constant dimension is sparser and can be computed more quickly than the Euclidean spanners of Callahan and Kosaraju [**?**]. The latter spanners have $O(n\varepsilon^{-d})$ edges and are computable in $O(n \log n + n\varepsilon^{-d})$ time. To the authors' knowledge, these are the sparsest quickly-constructable Euclidean spanners in terms of ε dependence. Later works on spanners have focused on bounding diameter, degree, or total edge weight [**?**, **?**]. We give a size lower bound for $(1+\varepsilon)$-Euclidean spanners, which is close to the sparsity of our $(1+\varepsilon)$ spanner of the edge-squared metric. Previously, sparse spanners of the edge-squared metric were shown to exist in two dimensions via Yao graphs and Gabriel graphs [**?**].

Theorem 1.8 proves that a 1-spanner of the edge-squared metric can be found assuming points are samples from a probability density, by using a k-NN graph for appropriate k. Our result is tight when d is constant. This is not possible for Euclidean distance, as a 1-spanner is almost surely the complete graph. Without the probability density assumption, there are point sets in $\mathbb{R}^4$ where 1-spanners of the edge-squared metric require $\Omega(n^2)$ edges. Finally, we show that spanners of p-power metrics, which are edge-squared metrics but with powers of p instead of 2, generalize Euclidean spanners and Euclidean MSTs. p-power metrics were considered in [?].

2 Outline

Section ?? contains the proof of Theorem 1.6, equating the edge-squared metric and nearest-neighbor geodesic distance in all cases. We then compute the persistent homology of the nearest-neighbor geodesic distance. Section ?? outlines a proof of Theorem 1.7, and compares our spanner to new lower bounds on the sparsity of $(1+\varepsilon)$-spanners of the Euclidean metric. We outline a proof of Theorem 1.8 in Section ?? and discuss its implications.

Section ?? introduces the p-power metrics. We show that Euclidean spanners and Euclidean MSTs are special cases of p-power spanners. We show how clustering algorithms including k-means, level-set methods, and single linkage clustering, are special cases of clustering with p-power metrics.

Conclusions and open questions are in Section ??. Full proofs for Theorems 1.8, 1.7 are contained in the Appendix.

3 Equality between metrics

In this section, we prove our main theorems, Theorem ?? and ??. That is, we prove that the edge-squared metric exactly equals the nearest neighbor metric on any point set, and that the q-edge power metric equals the q-NN metric for any set of four points or less. We also conjecture that the q-edge power metric always equals the q-NN metric. Even though the number of points we handle for the general q-edge power metric is quite small, this still represents progress; exactly computing the NN or q-NN metric even for four points in two dimensions requires dealing with uncountably number of paths through space. Furthermore, we provide a discrete criterion for testing whether the q-NN metric equals the q-edge power metric, which may be useful for future work.

Our tools for proving both theorems will be use of min-cost flows generated from a conservative vector field. This is the core idea that lets us surmount the difficulties in dealing with uncountably many paths. We hope that our ideas may be more generally applicable to various metrics.

As seen in Section ??, it suffices to prove our equalities on the q-screw simplex. As it turns out, the geometry of the 2-screw simplex has very elegant geometry. However, our techniques here may be appilcable to more general configurations of points.

3.1 Proof Strategy

Given points $x_0, x_1, \dots x_n$, our strategy is to build a graph G' with vertices $v_0, v_1, \dots v_{n-1}$, and v_S for any subset S of $\{0, 1, 2, 3, 4 \dots n-1\}$. Here, v_S is a vertex representing the circumcenter of the vertex set $\{x_s : s \in S\}$. Graph G' then has some edges (connected in a fashion to be detailed later), and costs on these edges. Graph G' is similar to the barycentric subdivision on a graph, used in [].

Tto lower bound any nearest neighbor path in our original point set, we will lower bound it by some unit flow (DFEINE) from v_0 to v_n in G'. However, note that the unit flow from v_0 to v_n is lower bounded by the shortest path (where lengths of edges are the same as their costs) from v_0 to

v_n in G'. If the shortest path in G' is equal to the q-edge power cost of going from x_0 to x_n, then we have completed our proof.

To lower bound the cost of the q-NN metric with a unit flow, we aim to build a potential function $\mathcal{C} : \mathbb{R}^n \to \mathbb{R}^{2^n}$ such that $\mathcal{C}(p)$ is a vector whose entries sum to one. Here, $\mathcal{C}(p)$ represents a convex combination of the vertices of G. We construct $\mathcal{C}$ such that $\mathcal{C}(v_S) = e_S$, where e_S represents the unit vector with 1 in the coordinate indexed by S, and 0 elsewhere. Additionally, $\mathcal{C}$ has the property that the q-NN cost of any path from $x \in \mathbb{R}^n$ to $x + \Delta(x) \in \mathbb{R}^n$ (with respect to point set $x_0, \ldots x_n$) is bounded below by the min cost flow on G satisfying demands $\mathcal{C}(x + \Delta(x)) - \mathcal{C}(x)$. If we can construct such a function $\mathcal{C}$, then the q-NN path from v_0 to v_n is thus lower bounded by the min cost unit flow from v_0 to v_n. Therefore, we have successfully lower bounded the q-NN cost of a path from x_0 to x_n with the shortest path in G.

Therefore, the remainder of the section is devoted to two tasks: first, constructing a G such that the shortest path in G is equal to the q-edge power distance from x_0 to x_n. Second, constructing a function $\mathcal{C}$ satisfying the properties listed above.

When $q = 2$, we can do this for any n, where $x_0, \ldots x_n$ form a 2-screw simplex. This will prove Theorem **??**. When $q > 2$, we can do this for any four points that are vertices of a q-screw simplex. This will prove Theorem **??**. We further conjecture our approach will work for any $q > 2$ and n.

3.2 Construction of G

Let G' be a graph on $\mathbb{R}^{2^n}$, with vertices v_S for all $S \subset \{0, 1, \ldots n-1\}$. Connect vertices v_S and v_T for $S, T \subset \{0, 1, \ldots n-1\}$ iff sets S and T differ by exactly one element. We assign this edge a cost of $|R_S^q - R_T^q|$, where R_S is the circumradius of points $\{x_s : s \in S\}$.

Lemma 3.1. *There exists a function B', such that for any path piece ϕ running from x and $x + \Delta(x)$, $q - NN(\phi) \geq MCF_{G'}(B'(x + \Delta(x)) - B'(x))$. Here, $MCF_{G'}(d)$ for $d \in \mathbb{R}^{2^n}$ represents the minimum cost flow on G' satisfying demands d on the vertices of G'.*

Lemma 3.2. *Let $S = s_1 < s_2 < \ldots s_{|S|}$, and $T = t_1 < \ldots \leq t_{|T|}$, where $S, T \subset \{0, 1, \ldots n-1\}$ and S and T differ by exactly one element. If*

$$|(R_S)^q - (R_T)^q| \geq |(s_{|S|} - s_1)/2^q - (t_{|T|} - t_1)/2^q|, \tag{1}$$

then the shortest path in G' equals the shortest path in G.

Note that this inequality is always an exact equality for $q = 2$. The proof of this lemma is combinatorial:

Proof. □

Therefore, if Equation 1 holds for any set of k points, and Lemma 3.1 and 3.2 are proven, then we've proven Theorem **??** for any k element point set.

Proof of Lemma 3.1

Proof of Lemma 3.2

Lemma 3.3. $SP_{G'}(v_0, v_n) \leq SP_G(x_0, x_n)$

Proof. The distance from v_0 to $v_{\{0,n-1\}}$ on the graph is equal to $u_{n-1} - u_0$ (NORMALIZE THINGS PROPERLY, INDEX THINGS PROPER, SET COMMON VARIABLE NAMES), and the distance from $v_{\{0,n-1\}}$ to v_{n-1} is the same. Thus, this path from v_0 to v_n equals $SP_G(x_0, x_n)$, as the latter is trivially equal to $u_{n-1} - u_0$.

□

Thus, it suffices to prove that

$$SP_{G'}(v_0, v_n) \geq SP_G(x_0, x_n)$$

3.3 Construction of B'

In this section, we prove we Lemma 3.1. We construct a function $\mathcal{C} : \mathbb{R}^n \to \mathbb{R}^{2^n}$ assigning every point in Euclidean space to a vector, representing a convex combination of vertices in G'.

Define $\overline{p}, \overline{x}, \overline{R}, \overline{B}$ here. Points $\overline{p}$ numbered from 0 to $k-1$. Here, $\overline{x}$ is numbered 1 through $k-1$. B is 0 through $k-1$ indexed.

LEGEND: p are points in the screw simplex, u are initial points on the line generating the screw simplex, x is the point inside the screw simplex, and v are the vertices of the graph.

Our general strategy is to create $\mathcal{C}$ piecewise, where we define it on each Voronoi subdivision bounded by the vertices $p_{a_0 a_1 \ldots a_t}$ for all $0 \leq t \leq k$. By the construction of these points, $p_{a_0} p_{a_0 a_1}$ is perpendicular to $p_{a_0 a_1} p_{a_0 a_1 a_2}$, and so forth.

To simplify our notation, we will denote $p_{a_0 a_1 \ldots a_t}$ as $\overline{p}_t$, for any $0 \leq t \leq k$. This is to simplify our notation. Thus, we can now say the line $\overline{p}_i \overline{p}_{i+1}$ is perpendicular to $\overline{p}_{i+1} \overline{p}_{i+2}$. These lines define a natural coordinate axis. Thus, for any $\overline{x}$ in the convex hull of $\overline{p}_i$, we can write $\overline{x}$ in coordinates $(\overline{x}_0, \overline{x}_1, \ldots \overline{x}_k)$, where the i^{th} coordinate axis is parallel to $\overline{p}_i \overline{p}_{i+1}$. Now we define $\overline{B}$ the i^{th} coordinate of $\overline{B}$:

$$\overline{B}(\overline{x})_i = \frac{\left(\sum_{s=1}^{i} \overline{x}_s{}^2\right)^{q/2} - \left(\sum_{s=1}^{i-1} \overline{x}_s{}^2\right)^{q/2}}{\overline{R}_i^q - \overline{R}_{i-1}^q} - \sum_{i<j<k} \overline{B}(\overline{x})_j \tag{2}$$

for all $1 \leq i \leq k-1$, and

$$\overline{B}(\overline{x})_0 = 1 - \sum_{1 \leq j < k} \overline{B}(\overline{x})_j$$

The key feature about $\overline{B}$ is that

$$\sum_{j=i}^{k} \overline{B}(\overline{x})_j = \frac{\left(\sum_{s=1}^{i} \overline{x}_s{}^2\right)^{q/2} - \left(\sum_{s=1}^{i-1} \overline{x}_s{}^2\right)^{q/2}}{\overline{R}_i^q - \overline{R}_{i-1}^q}$$

for all $i > 0$, and for $i = 0$ the LHS evaluates to 1.

Since we defined this function piecewise, we need to check that this function is piecewise continuous.

Lemma 3.4. *If $\overline{x}$ is on a face of $\overline{p}_0, \ldots \overline{p}_n$, then $\overline{B}(\overline{x})$ has non-zero coordinates only on that face. Furthermore, the coordinates depend only on SOMETHING.*

Proof. ∎

Now we are ready to prove our core lemma:

Lemma 3.5. *For x and $x + \Delta(x)$ in the convex hull of $\overline{p}_1, \ldots \overline{p}_k$, the distance:*

$$\|x - \overline{p}_1\|^{q-1} \cdot \|\Delta(x)\| \geq F\left(\overline{B}(x + \Delta(x)) - \overline{B}(x)\right)$$

where $F(d)$ is the unique cost of a flow satisfying demand $d \in \mathbb{R}^k$ on $\overline{G}$.

Here, the left hand side represents the q-NN cost of a path piece from x to $x + \Delta(x)$, and the right hand side is the unique cost of the induced flow on graph G', with the restriction that the flow is only nonzero on the vertices $\overline{p}_i\overline{p}_{i+1}$ for any $0 \leq i < k$. This flow is unique since we forced our flow to be non-zero only on the edges $\overline{p}_i\overline{p}_{i+1}$, which form a line graph; and for any set of demands on vertices of a line, there is a unique flow satisfying those demands.

Proof. For any edge $\overline{p}_i\overline{p}_{i+1}$, the cost of a flow (satisfying some set of demands whose sum is 0) on that edge is the absolute value of the sum of the demands on vertices $\overline{p}_{i+1}\overline{p}_{i+2}, \ldots \overline{p}_k$, multiplied by the cost of the edge from $\overline{p}_i$ to $\overline{p}_{i+1}$. This quantity comes out to be:

$$\left(\overline{R}_{i+1}^q - \overline{R}_i^q\right) \sum_{j=i+1} \overline{B}(\overline{x})_j \tag{3}$$

$$= \left(\sum_{s=1}^{i} \overline{x}_s{}^2\right)^{q/2} - \left(\sum_{s=1}^{i-1} \overline{x}_s{}^2\right)^{q/2}. \tag{4}$$

As $\Delta(x)$ goes to 0, the change in Expression 4 is

$$\left(q\overline{x}_0 \left(\sum_{s=1}^{i} \overline{x}_s^2\right)^{q/2-1} q\overline{x}_0 \left(\sum_{s=1}^{i-1} \overline{x}_s^2\right)^{q/2-1}\right) \Delta(\overline{x})_0 + \left(q\overline{x}_1 \left(\sum_{s=1}^{i} \overline{x}_s^2\right)^{q/2-1} q\overline{x}_1 \left(\sum_{s=1}^{i-1} \overline{x}_s^2\right)^{q/2-1}\right) \Delta(\overline{x})_1 + \ldots + \left(q\overline{x}_{i-1}\right. \tag{5}$$

To simplify our notation, let $a_{i,j}$ denote

$$q\overline{x}_j \left(\sum_{s=0}^{i} \overline{x}_s^2\right) 3\overline{x}_j \left(\sum_{s=0}^{i-1} \overline{x}_s^2\right)$$

Now we sum this across all i to get the overall cost, and group by $\Delta(\overline{x})_i$ for fixed i. This gives us:

$$a_{0,0}\Delta(\overline{x})_0 + a_{1,0}\Delta(\overline{x}_0) + \ldots + a_{k,0}\Delta(\overline{x}_0) + \ldots a1, 1\Delta(\overline{x}_1) + \ldots + a_{k,1}\Delta(\overline{x}_1) + \ldots + ak, k\Delta\overline{x}_k.$$

This sum is upper bounded by:

$$|a_{0,0} + a_{1,0} + ldots + ak, 0||\Delta(\overline{x}_0)| + \ldots |a_{k,k}||\Delta(\overline{x}_k)|$$

Which equals

$$|q\overline{x}_0 \left(\sum_{s=1}^{k} \overline{x}_s^2)^{q/2-1}\right) |\Delta(\overline{x})_0 + |q\overline{x}_1 \left(\sum_{s=1}^{k} \overline{x}_s^2)^{q/2-1}\right) |\Delta(\overline{x})_1 \ldots + |q\overline{x}_k \left(\sum_{s=1}^{k} \overline{x}_s^2)^{q/2-1}\right) |\Delta(\overline{x})_k$$

This expression, by Cauchy Schwarz, is upper bounded by

$$\sqrt{\sum_{s=0}^{k} \Delta(\overline{x})_s^2} \cdot \left(q\sqrt{\sum_{s=0}^{k} \overline{x}_s^{q-1}}\right)$$

Which is exactly the q-NN distance.

□

Note that this function is piecewise continuous on the boundary. Therefore, we have shown that the q-NN cost of any path piece is less than the min-cost flow on G satisfying $\mathcal{C}\left(x+\Delta(x)\right) - \mathcal{C}\left(x\right)$ for infinitesimal $\Delta(x)$, as desired.

So far, the only property our flow construction used is that the points $x_0, x_1, \ldots x_n$ have Voronoi subdivisions defined by $\overline{p}_{a_0}$, $\overline{p}_{a_0a_1}$, $\ldots \overline{p}_{a_0a_1\ldots a_k}$, for some $a_0, ldotsa_k \subset \{0, 1, \ldots n\}$. (DOES THIS WORK FOR ANY GEOMETRY, OR DO I NEED THE INTERIOR CIRCUMCENTER PROPERTY?).

Thus, we have proven a core lemma:

Lemma 3.6.
The q-NN distance between two points in a point set is lower bounded by the shortest path between the two corresponding points in G. Here, G is constructed as in Definition **??**

We now prove the following two lemmas, completing our proof of Theorem 1.6 and Theorem **??** respectively.

Lemma 3.7. *Let G be the edge-squared graph (DEFINE), and let G be defined as in Definition* **??** *for $q = 2$. The shortest path in G is the same as the shortest path in G, when the initial point set generating G and G is a 2-screw simplex.*

Lemma 3.8. *Let $q > 2$. Let G be the q-edge power graph, and G be defined as in Definition* **??** *(MAKE SURE THE DEFINITION IS Q DEPENDENT). The shortest path in G is the same as the shortest path in G, when the initial point set generating G and G is a q-screw simplex with 4 points.*

Combined with Theorem **??**, Lemmas **??** and **??** prove Theorems 1.6 and **??** respectively. Moreover, we make the following conjecture, which we have some computational evidence for (See Appendix **??** for details):

Conjecture 3.9. *For $q > 2$, let G and G be defined as in Lemma 3.8. Then the shortest path in G is the same as the shortest path in G.*

If this were true, it would prove that the q-edge power metric and the q-NN metric were equal for all $q > 2$.

Now, our proof will proceed in two parts. First, we construct a function $B' : \mathbb{R}^n \to \mathbb{R}^{2^n}$ assigning every point in Euclidean space to a vector, representing a convex combination of vertices in G'. We build B' such that $B'(v_S) = e_S$, where e_S is the unit vector with 1 in the dimension indexed by S. Next, we lower bound the Nearest-Neighbor cost of an infinitesimal path piece. If the infinitesimal path piece runs from $x \in \mathbb{R}^n$ to $x + \Delta(x)$, where both x and $x + \Delta(x)$ are in the same circumcentric subdivision (DEFINE), then we lower bound it with some flow on the graph G' satisfying demands $B'(x + \Delta(x)) - B'(x)$.

If we can find a cost function B' with these properties, we can integrate over all the infinitesimal path pieces to show that the p-NN cost of a path from x_0 to x_{n-1} is lower bounded by some flow on G'. This is a unit flow from x_0 to x_n. If we furthermore have that the shortest path on G' equals the shortest path on G, then we have:

$$NN(path) \geq Q' \geq SP_{G'}(v_0, v_n) = SP_G(x_0, x_n).$$

Here, Q' represents the cost of some flow from v_0 to v_n on G'. Here, $SP_{G'}(v_0, v_n)$ is the shortest path from v_0 to v_n on G' and $SP_G(x_0, x_n)$ is the shortest path from x_0 to x_n on the q-edge power graph of the q-screw simplex.

Our proof then consists of two parts: the first is finding a graph G' and showing $SP_{G'}(v_0, v_n) = SP_G(x_0, x_n)$ (STANDARDIZE INDICES). The second is finding a function B' satisfying our desired properties listed in our strategy. In the remainder of this section, we build our graph G' and establish

a discrete, sufficient-but-not-necessary criterion for when $SP_{G'}(v_0, v_n) = SP_G(x_0, x_n)$. Then we build a function B' and prove that $NN(path) \geq Q'$, where the path runs from v_0 to v_1.

We then show that this necessary-but-not-sufficient criterion holds for any 5 point q-screw simplex, thereby showing that the q-edge power metric equals the q-NN metric for all 5 point sets. We further conjecture that this criterion holds for any n points, but the authors are currently unable to prove it.

Therefore, we have provided a sufficient but not necessary criterion for which the q-edge power metric equals the q-NN Metric on the q-screw simplex. Furthermore

Then we establish a discrete criterion in which $G' \geq G$. We then prove this criterion holds for all sets of 5 points in arbitrary dimension, and conjecture that it holds for all sets of n points in arbitrary dimension.

4 Fractional Laplacian

In this section, we prove that the q-screw simplex distances arise as effective resistance of the Fractional Laplacian, for powers $s = -1/2 - 1/q$, when $q > 2$.

Preliminaries: Fractional Laplacian.

We present two definitions: the first definition is based on taking the limit of graph Laplacians raised to the fractional power (which are known in folklore to be graph Laplacians themselves), and the second definition is based on taking fractional powers of the Laplacian differential operator when the latter is written in terms of its Eigenvectors, the Fourier bases.

In this work, we build on Von Neumann and Schoenbergs proof of embeddability of the q-screw simplex. Their work (slightly simplified by the authors of this paper) shows that the q-screw simplex for $q > 1$ can be embedded in infinite dimensional Hilbert space, by the embedding $f : mathbbR \to L_2$ defined as:

$$f(x) = \frac{e^{i\omega x}}{\omega^{1/2+1/q}} \tag{6}$$

Where $f(x)$ is a function in the variable ω.

The proof of their embedding hinges on the following remarkable integral formula:

$$\|x_1 x_2\|_2^{1/q} = \frac{sin^2(\omega(x_1 - x_2))}{\omega^{1+2/q}},$$

the left hand side of which is the norm for Equation 6. This integral formula is a classical integral formula [], and can be proven using Jordans integration theorem from complex analysis [**?**, **?**].

Notice that the step function S_x, which is 1 between $-x$ and x and 0 elsewhere, can be written in Fourier bases as:

$$S(x) = frace^{i\omega x}\omega. \tag{7}$$

Equation 7 is a classic result, dating back to the earliest days of functional analysis. However, this formulation compared with Equation 6 practically invites us to use the Fractional Laplacian, when viewed through the Eigenvector lens in Section **??**.

Therefore, we can write Equation 6 as:

$$f(x) = \Delta^{1/4-1/(2q))} Step = \Delta^{1/4-1/(2q)}(\Delta^{-1/2}\delta_{-x} - \delta_x) = \Delta^{-1/4-1/(2q)}(\delta_{-x} - \delta_x)$$

In the above expression, δ_x represents the Dirac Delta function. DEFINE Δ in this case!!!! Here, the second part of the equation is a standard manipulation in differential equations, as $\Delta^{1/2}$ is conceptually similar to the integral operator. For more on manipulations with this fractional Laplacian operator, see **??**.

And thus $|x - y|^{2/q} = ||f(x)||_2^2$ can be written as:

$$(\delta_{-x} - \delta_x)^T \cdot \Delta^{-1/2-1/q} \cdot (-\delta_{-x}\delta_x) \tag{8}$$

Which is an effective resistance distance. This can be seen since $\Delta^{-1/2-1/q}$ can be written as the limit of fractional graph Laplacians (which are in turn graph Laplacians, by Lemma **??**). Given a finite screw simplex, our distance is thus the limit of the Schur complement of these graph Laplacians onto a finite point set, which is the limit of a sequence of graph Laplacians. It can be seen easily that such graph Laplacians must converge, and the limit of this convergence is a graph Laplacian whose effective resistance distance are the screw simplex distances. This proves Theorem **??**.