Journal of Machine Learning Research 25 (2024) 1-65 Submitted 7/23; Published 6/24

Fermat Distances: Metric Approximation, Spectral Convergence, and Clustering Algorithms

Nicolás García Trillos GARCIATRILLO@WISC.EDU
Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Anna Little LITTLE@MATH.UTAH.EDU
Department of Mathematics, Utah Center for Data Science
University of Utah
Salt Lake City, UT 84112, USA

Daniel McKenzie DMCKENZIE@MINES.EDU
Department of Applied Mathematics and Statistics
Colorado School of Mines
Golden, CO 80401, USA

James M. Murphy JM.MURPHY@TUFTS.EDU
Department of Mathematics
Tufts University
Medford, MA 02155, USA

Editor: Sayan Mukherjee

Abstract

We analyze the convergence properties of *Fermat distances*, a family of density-driven metrics defined on Riemannian manifolds with an associated probability measure. Fermat distances may be defined either on discrete samples from the underlying measure, in which case they are random, or in the continuum setting, where they are induced by geodesics under a density-distorted Riemannian metric. We prove that discrete, sample-based Fermat distances converge to their continuum analogues in small neighborhoods with a precise rate that depends on the intrinsic dimensionality of the data and the parameter governing the extent of density weighting in Fermat distances. This is done by leveraging novel geometric and statistical arguments in percolation theory that allow for non-uniform densities and curved domains. Our results are then used to prove that discrete graph Laplacians based on discrete, sample-driven Fermat distances converge to corresponding continuum operators. In particular, we show the discrete eigenvalues and eigenvectors converge to their continuum analogues at a dimension-dependent rate, which allows us to interpret the efficacy of discrete spectral clustering using Fermat distances in terms of the resulting continuum limit. The perspective afforded by our discrete-to-continuum Fermat distance analysis leads to new clustering algorithms for data and related insights into efficient computations associated to density-driven spectral clustering. Our theoretical analysis is supported with numerical simulations and experiments on synthetic and real image data.

Keywords: clustering, data-driven metrics, discrete-to-continuum limits, Fermat distances, spectral graph theory

©2024 Nicolas García Trillos, Anna Little, Daniel Mckenzie, James M. Murphy.

License: CC-BY 4.0, see https://creativecommons.org/licenses/by/4.0/. Attribution requirements are provided at http://jmlr.org/papers/v25/23-0939.html.

1. Introduction

Data-driven metrics and related dimensionality reduction methods are a widely used tool in statistics, data science, and machine learning for analyzing point cloud data $\mathcal{X} \subset \mathbb{R}^D$. Seminal methods such as principal (Hotelling, 1933) and independent (Comon, 1994) component analysis, Laplacian eigenmaps (Belkin and Niyogi, 2003), diffusion maps (Coifman et al., 2005; Coifman and Lafon, 2006), Isomap (Tenenbaum et al., 2000), locally linear embedding (Roweis and Saul, 2000), and tSNE (van der Maaten and Hinton, 2008) often capture important structural properties in data (*e.g.,* cluster structure, concentration near low-dimensional sets) in a manner that is statistically efficient and robust to noise and outliers. In many cases, these methods first embed high-dimensional, noisy data into a low-dimensional Euclidean space so that Euclidean distances on the embedded points implicitly define a new metric. Working in these embedded spaces (or equivalently, analyzing with the induced metrics) has led to methods for unsupervised and semisupervised machine learning (Ng et al., 2002).

An alternative to this broad class of approaches is to consider weighted shortest path metrics. The idea is to construct a weighted graph associated to $\mathcal{X}$, and then use shortest path distances in this graph to learn important structures in $\mathcal{X}$. A particular class of density-weighted path metrics known as *Fermat distances* use powers of the Euclidean distance between points as weights (Bijral et al., 2011; Hwang et al., 2016; Chu et al., 2020; Little et al., 2022; Groisman et al., 2022; Fernández et al., 2023); in the discrete setting, we denote these distances as ℓ_p, where p is a parameter that determines the impact of data density. Fermat distances and related density-driven path metrics have been successfully applied to a range of problems in unsupervised and semisupervised machine learning (Vincent and Bengio, 2003; Bousquet et al., 2004; Sajama and Orlitsky, 2005; Chang and Yeung, 2008; Bijral et al., 2011; Moscovich et al., 2017; Alamgir and Von Luxburg, 2012; Mckenzie and Damelin, 2019; Little et al., 2020), as well as to topological data analysis for robust computation of persistent homology (Fernández et al., 2023) and in high-dimensional signal processing (Zhang and Murphy, 2021; Manousidaki et al., 2024). When data points are sampled from a compact Riemannian manifold $\mathcal{M}$, the discrete ℓ_p converges to a continuum metric $\mathcal{L}_p$, which can be interpreted as a density-weighted geodesic distance on $\mathcal{M}$.

Existing works in the literature establishing discrete-to-continuum convergence of general data-driven distances include Howard and Newman (2001); Díaz et al. (2016); Davis and Sethuraman (2019); Hwang et al. (2016); Bungert et al. (2022), among many others. The papers Díaz et al. (2016); Davis and Sethuraman (2019); Bungert et al. (2022), for example, discuss convergence of distances defined on random geometric graphs (RGG), either in the i.i.d. setting or for Poisson point processes. In the RGG setting, admissible paths between two points must consist of consecutive short range (as specified by a connectivity parameter) hops between data, in contrast to Fermat distances, where arbitrarily large hops between points are admissible. The results from Davis and Sethuraman (2019) are asymptotic, while the ones in Díaz et al. (2016); Bungert et al. (2022) provide high probability convergence rates in terms of the RGG's connectivity parameter. The results in Bungert et al. (2022), for example, discuss the convergence of the ratio between certain expectations of distances at different scales. When combined with concentration inequalities, this allows the authors to prove rates of convergence, in sparse settings, for a semisuper-

vised learning procedure known as Lipschitz learning. The works Kesten (1993); Howard and Newman (2001); Hwang et al. (2016); Groisman et al. (2022); Little et al. (2022); Fernández et al. (2023) study Fermat distances on point clouds and are the most relevant references for our metric approximation results. Note Hwang et al. (2016); Groisman et al. (2022); Fernández et al. (2023) only provide asymptotic convergence results, while Howard and Newman (2001); Little et al. (2022) assume a uniform density. We provide the first local quantitative convergence results for Fermat distance in the manifold setting with a general density.

While there are several mathematical objects that can be constructed over these finite data-driven metric spaces, in this paper we will discuss as particularly important examples the Laplacian operators that these metrics induce on a collection of data points sampled from a distribution over a smooth and compact manifold $\mathcal{M}$. Indeed, the metric ℓ_p can be used to define graph kernel functions which then lead to embeddings of $\mathcal{X}$ using, for example, the low-frequency eigenvectors of the graph Laplacian. These eigenvectors contain valuable information that can be used in machine learning tasks such as trend filtering, clustering, or dimensionality reduction. The difficulty in analyzing the induced Laplacians and their eigenvectors relies, particularly, on the fact that Fermat distances over data clouds are themselves random (in contrast to the more standard Euclidean metric). This difficulty has impeded full statistical and analytical understanding of the methods that utilize Fermat-based Laplacians.

Discrete-to-continuum convergence of Laplacian spectra on a manifold $\mathcal{M}$ is a well-developed area (Belkin and Niyogi, 2007; Burago et al., 2015; García Trillos et al., 2019), at least for random geometric graphs built with the Euclidean distance. The main problem is, given a finite sample $\mathcal{X} \subset \mathcal{M}$, understanding how the sample-based graph Laplacian induced from $\mathcal{X}$ converges as $|\mathcal{X}| \to \infty$ to an operator defined with respect to the intrinsic geometry of $\mathcal{M}$ (for example, the Laplace-Beltrami operator on $\mathcal{M}$). This allows one to characterize with high probability the behavior of the spectrum of the graph Laplacian, thereby ensuring good performance in downstream tasks such as clustering. Existing results typically focus on the cases in which the graph Laplacian is constructed using Euclidean distances or the intrinsic geodesic distance on $\mathcal{M}$. In the case of Fermat distances, there is a natural continuum analogue of ℓ_p on $\mathcal{M}$, which we shall denote $\mathcal{L}_p$. The metric $\mathcal{L}_p$ is in fact a geodesic distance function on $\mathcal{M}$ with respect to a certain density-dependent Riemannian metric parameterized by p. Unlike in the case of Euclidean distances, the underlying manifold geodesic $\mathcal{L}_p$ must itself be estimated from samples using ℓ_p, requiring new tools of analysis.

The computational requirements for computing Fermat distances has also been studied in the literature. Given n data points, Fermat distances can with high probability be computed in a k-nearest-neighbors (kNN) graph (Little et al., 2022; Chu et al., 2020; Groisman et al., 2022) as long as $k \sim \log(n)$. This implies all pairwise distances can be computed with the Floyd-Warshall algorithm with complexity $O(n^2 \log(n))$. Furthermore, the Fermat kNN's of all points can be computed with complexity $O((k^2 + CD)n \log(n))$, where C is a constant that depends exponentially on the intrinsic dimension of the data (Mckenzie and Damelin, 2019). This allows Fermat distance nearest neighbors (and consequently, graph Laplacians based on Fermat distance nearest neighbors) to be calculated relatively efficiently with quasi-linear complexity. Another quasi-linear approach is to compute Fermat

distances on a set of landmarks and then extend quantities of interest using Nystrom-based methods (Williams and Seeger, 2001; Ghojogh et al., 2020; Platt, 2005; Yu et al., 2012; Civril et al., 2006; Shamai et al., 2020).

1.1 Summary of Contributions

This paper makes several mathematical, statistical, and algorithmic contributions.

First, we establish precise local convergence rates of discrete Fermat distances to continuum Fermat distances defined with respect to a non-uniform density ρ on $\mathcal{M}$. These results are of independent interest for their connections to percolation theory and to a variety of applications in which machine learning tasks rely on the availability of a metric structure over a data point cloud.

Second, we develop spectral convergence results for graph Laplacians constructed with ℓ_p to operators on $\mathcal{M}$ with geodesic distance given by $\mathcal{L}_p$. These results leverage our metric convergence results and also new geometric results pertaining to the properties of $\mathcal{M}$ when endowed with $\mathcal{L}_p$. Importantly, our results quantify the impact of the underlying geometry of $\mathcal{M}$, and we calculate explicit constants whenever possible. The large sample spectral analysis of Fermat-based graph Laplacians is thus an important application of our metric approximation results.

Third, we suggest new spectral clustering algorithms using ℓ_p. These algorithms enjoy robustness with respect to cluster elongation, a regime where standard spectral clustering fails. We highlight the connection of Fermat Laplacians with a broad family of Laplacian normalizations. Our results provide geometric insight into the choice of normalization parameters and also suggest statistically and computationally efficient methods of practical implementation. We evaluate our proposed methods on image data, showing the impact of the key parameters in Fermat distance spectral clustering; see Figure 1 for an illustration.

1.1.1 Outline of Paper

The remainder of the paper is organized as follows. Section 2 provides background on Fermat distances and Laplace operators. Section 3 summarizes our main results. Section 4 discusses the Fermat Riemannian metric, including computation of geodesics and discrete-to-continuum metric approximation results. Section 5 utilizes the metric approximation to obtain spectral convergence results of Fermat graph Laplacians. Section 6 discusses some examples and numerical experiments supporting the paper. Section 7 concludes the article. The Appendix contains proofs of technical results and additional experiments. Code to reproduce numerical results is available at `https://github.com/JamesMMurphy11/FermatDistances`.

2. Background

Throughout the paper, we will use the notation in Table 1.

2.1 Fermat Distances

Discrete Fermat distances can be understood as classical shortest paths on graphs but with edge lengths penalized according to a parameter $p \in [1, \infty)$.

(a) Original Image

(b) Second eigenvector of Euclidean random walk Laplacian

(c) Second eigenvector of Fermat random walk Laplacian ($p = 2.5$)

Figure 1: A real image of a landscape from the Oljato-Monument Valley that is very elongated; patch-based features are computed which combine color and spatial information. The second eigenvector of the Euclidean random walk Laplacian is shown in (b) and the second eigenvector for the Fermat Laplacian with $p = 2.5$ is shown in (c). Despite the elongation, the Fermat Laplacian correctly segments background from foreground in a way the Euclidean Laplacian does not. Image available on unsplash courtesy of Robert Murray.

Definition 2.1 *For $p \in [1, \infty)$, $x, y \in \mathbb{R}^D$, and some finite set $\mathcal{X} \subset \mathbb{R}^D$, the* (discrete) p-weighted Fermat distance *from x to y is:*

$$\ell_p(x, y, \mathcal{X}) = \min_{\pi = \{x_i\}_{i=1}^T} \left(\sum_{i=1}^{T-1} \|x_i - x_{i+1}\|^p \right)^{\frac{1}{p}}, \tag{1}$$

where π is a path of points in $\mathcal{X} \cup \{x, y\}$ with $x_1 = x$ and $x_T = y$ and $\|\cdot\|$ is the Euclidean norm.

Note when there is no ambiguity about $\mathcal{X}$, we will drop it from the notation and simply write $\ell_p(x, y)$. The case $p \in (0, 1)$ was studied in Alamgir and Von Luxburg (2012) and shown to have counter-intuitive properties; we do not consider this parameter regime. While $\ell_p(x, y)$ depends on the point cloud $\mathcal{X}$, a population formulation is possible, as follows.

Notation	Definition
$\|\cdot\|$	Euclidean 2-norm
$\|\cdot\|$	Norm with respect to the Riemannian metric g
$\mathcal{M}$	Data manifold
m	Intrinsic dimension of data manifold
$\mathcal{X}$	Arbitrary point cloud in $\mathbb{R}^D$
$\mathcal{R}$	Reach of $\mathcal{M}$
R	Injectivity radius of $\mathcal{M}$
K	Bounds on sectional curvature of $\mathcal{M}$
ω_m	Volume of the m-dimensional unit ball
ν	Probability measure supported on $\mathcal{M}$
ρ	Density function associated to data measure ν
$\mathcal{X}_n$	Set of n i.i.d. samples from distribution ν
$H_{n\rho}$	Poisson point process on $\mathcal{M}$ with intensity $n\rho$
N_n	Poisson random variable with mean n
p	Density weight parameter
ℓ_p	Discrete Fermat distance with parameter p; see (1)
$\tilde{\ell}_p$	Normalized discrete Fermat distance with parameter p; see (4)
$\mathcal{L}_p$	Continuum Fermat distance with parameter p; see (2)
g	Arbitrary Riemannian metric
$\bar{g}$	Euclidean metric
g_p	Fermat Riemannian metric; see (3)
$\rho_p = \rho^p$	Density with respect to Fermat Riemannian metric
$d(x,y)$	Geodesic distance on $(\mathcal{M}, \bar{g})$
$d_g(x,y)$	Geodesic distance on $(\mathcal{M}, g)$
$\alpha = 2(p-1)/m$	Reweighting constant
μ	Percolation time constant, depending on p, m
$\Delta_{\Gamma^{\tilde{\ell},h}}$	Discrete random-walk Laplacian with metric $\tilde{\ell}_p^p$ and bandwidth h; see (6)
$\Delta_{2,p}$	Continuum random-walk Laplacian with metric $\mathcal{L}_p^p$; see Proposition 2.5
a	Diffusion maps parameter
η	Kernel function, generally taken as $\eta = \mathbb{1}_{[0,1]}$

Table 1: Notation used throughout the paper.

Definition 2.2 *Let $(\mathcal{M}, g)$ be a compact, m-dimensional Riemannian manifold and $\rho : \mathcal{M} \to \mathbb{R}_{>0}$ a continuous density function on $\mathcal{M}$. For $p \in [1, \infty)$ and $x, y \in \mathcal{M}$, the* (continuum) p-weighted Fermat distance *from x to y is:*

$$\mathcal{L}_p(x,y) = \left(\inf_{\gamma} \int_0^1 \frac{1}{\rho(\gamma(t))^{\frac{p-1}{m}}} \sqrt{g\left(\gamma'(t), \gamma'(t)\right)} dt \right)^{\frac{1}{p}}, \tag{2}$$

where $\gamma : [0,1] \to \mathcal{M}$ is a piecewise $\mathcal{C}^1$ path with $\gamma(0) = x, \gamma(1) = y$.

Note when $\mathcal{M}$ is embedded in $\mathbb{R}^D$ and g is the Euclidean metric, one simply has $\sqrt{g\left(\gamma'(t), \gamma'(t)\right)} = \|\gamma'(t)\|$. Suppose further that $\rho \in C^\infty(\mathcal{M})$. Then associated to $\mathcal{L}_p^p$ is a Riemannian metric, g_p, defined as:

$$g_{p,x}(X,Y) := \rho^{-\alpha} g_x(X,Y), \tag{3}$$

where $\alpha := 2(p-1)/m$ (Hwang et al., 2016). Indeed, one easily observes

$$\mathcal{L}_p^p(x,y) = \inf_\gamma \int_0^1 \sqrt{g_p(\gamma'(t),\gamma'(t))}\, dt,$$

i.e., the distance function associated to g_p is $\mathcal{L}_p^p$. This interpretation of $\mathcal{L}_p^p$ as a Riemannian metric on $\mathcal{M}$ leads us to consider ℓ_p^p rather than ℓ_p in the discrete setting.

Suppose $\mathcal{X}$ is sampled from a distribution ν with density ρ with respect to the volume form of $\mathcal{M}$. A natural question is to characterize the convergence of ℓ_p^p to $\mathcal{L}_p^p$ as $|\mathcal{X}|$ tends to infinity. Two concrete probabilistic models for $\mathcal{X}$ that are popular in the literature are:

- The *i.i.d. setting*, where we consider $\mathcal{X} = \mathcal{X}_n := \{x_1, \ldots, x_n\}$ to be a collection of i.i.d. draws from ν.
- The *Poisson point process (PPP) setting*, where we consider $\mathcal{X} = H_{n\rho}$ to be the realization of a Poisson point process on $\mathcal{M}$ with intensity $n\rho$.

In either case, one is naturally interested in the behavior of ℓ_p^p as n tends to infinity. In this paper we obtain convergence results for both PPP models as well as for the i.i.d. setting, the latter being the model that is most often assumed in statistical learning theory.

Remark 2.3 *The Poisson point process $H_{n\rho}$ and the i.i.d. model $\mathcal{X}_n$ are closely connected to each other. Indeed, from an infinite sequence $x_1, x_2, \ldots$ of i.i.d. samples from ν and from a Poisson random variable N_n with mean n, independent from the x_i, one can generate $H_{n\rho}$ by setting $H_{n\rho} = \{x_1, \ldots, x_{N_n}\}$. In particular, conditioned on N_n, $H_{n\rho}$ consists of N_n i.i.d. points drawn from ν.*

Informally, one of our main results (Theorem 4.3) states that, with high probability, the discrete Fermat distance (1) approximates locally the geodesic distance (2). We first derive a metric approximation result for PPPs on manifolds using results from percolation theory. We then extend the result to the i.i.d. case through a method typically known as *de-Poissonization*. We then explore some of the geometric implications of this convergence, including quantitative rates for the spectral convergence of normalized graph Laplacian operators induced by discrete Fermat distances toward weighted Laplace-Beltrami operators relative to the family of Riemannian metrics g_p, a result of interest for its direct application to manifold learning. From the structure of the resulting continuum Laplacian operators, we extract theoretical insights on the role that data density plays in clustering and how, by picking p appropriately, we can accentuate or suppress the effect of density on the resulting data partitioning.

This paper is not the first to address the discrete-to-continuum convergence of Fermat distances. Precise statement of such results requires an appropriate normalization, to account for the fact that $\ell_p \to 0^+$ as $n \to \infty$. It can be shown (Howard and Newman, 2001) that the correct scaling constant for ℓ_p is $n^{(p-1)/pm}$, so that the relevant convergence question is that of

$$\tilde{\ell}_p^p := n^{(p-1)/m} \ell_p^p \tag{4}$$

to $\mathcal{L}_p^p$.

Hwang et al. (2016) establish that for $\mathcal{M}$ compact and ρ continuous and bounded away from 0, there exists a constant $\mu > 0$, depending only on p and m, such that for all $b, \epsilon > 0$, there exists $\theta_0 > 0$ such that for n large enough,

$$\mathbb{P}\left(\sup_{\mathcal{L}_p^p(x,y)\geq b}\left|\frac{\tilde{\ell}_p^p(x,y)}{\mathcal{L}_p^p(x,y)} - \mu\right| > \epsilon\right) \leq \exp(-\theta_0 n^{1/(m+2p)}),$$

i.e., $\tilde{\ell}_p^p \to \mu\mathcal{L}_p^p$ in probability for $\mathcal{L}_p^p \geq b$. The uniform lower bound $b > 0$ was removed by Fernández et al. (2023), significantly improving the convergence results. However, this type of convergence in probability is inadequate for our purposes as we need precise characterization of ϵ and θ_0, at least locally.

Little et al. (2022) provide more precise convergence results, albeit in a limited setting. They show that if ρ is uniform and $\mathcal{M}$ is convex, compact, and has m-dimensional unit volume, then for n sufficiently large and for all points x, y sufficiently far from $\partial\mathcal{M}$, $\mathbb{E}\left[(\tilde{\ell}_p(x,y) - \mu^{\frac{1}{p}}\mathcal{L}_p(x,y))^2\right] \lesssim n^{-\frac{1}{m}}\log^2(n)$. While this result is precise, it applies only to uniform densities, and is thus inadequate for developing spectral convergence results for a large class of densities ρ.

2.2 Discrete Laplace Operators

Let $\mathcal{X} = \{x_1, \ldots, x_n\}$ be a finite set of points on $\mathcal{M} \subset \mathbb{R}^D$. Let $\eta : [0,\infty) \to [0,\infty)$ be any non-increasing function with support contained in $[0,1]$ and $\eta(1/2) > 0$. For ease of exposition in all that follows, we take $\eta = \frac{1}{\omega_m}\mathbf{1}_{[0,1]}$ where the normalizing factor ω_m denotes the volume of the unit ball in $\mathbb{R}^m$ and is chosen so that $\int_{\mathbb{R}^m}\eta(\|x\|)dx = 1$. Choosing a distance function $d_0(\cdot,\cdot)$ on $\mathcal{X}$ and a *bandwidth parameter* h, we define a weighted graph $\Gamma^{d_0,h} = (\mathcal{X}, W^{d_0,h})$ by setting

$$w_{ij}^{d_0,h} = \frac{1}{nh^m}\eta\left(\frac{d_0(x_i,x_j)}{h}\right). \tag{5}$$

There are several popular definitions of Laplacian operators on $\Gamma^{d_0,h}$; in this paper we focus on the *random-walk Laplacian*, defined for any $u : \mathcal{X} \to \mathbb{R}$ as

$$\left(\Delta_{\Gamma^{d_0,h}}u\right)(x_i) = \frac{2(m+2)}{h^2}\sum_j \frac{w_{ij}^{d_0,h}}{m_{d_0,h,i}}\left(u(x_i) - u(x_j)\right). \tag{6}$$

where $m_{d_0,h,i} := \sum_j w_{ij}^{d_0,h}$ is the *degree* of x_i; when the choices of metric d_0 and scaling parameter h are clear, we will simply write m_i. Up to a normalization constant, the random walk Laplacian can be expressed in matrix form as $\Delta_{\Gamma^{d_0,h}} = I - D^{-1}W$, where D is a diagonal matrix containing the degrees and W is the matrix of weights. As is well-known, (*e.g.,* García Trillos et al. (2019)), the eigenvalues of $\Delta_{\Gamma^{d_0,h}}$ may equivalently be expressed in terms of the corresponding *Dirichlet form*,

$$b^{d_0,h}(u,v) := \frac{m+2}{nh^2}\sum_{i,j} w_{ij}^{d_0,h}\left(u(x_i) - u(x_j)\right)\left(v(x_i) - v(x_j)\right). \tag{7}$$

Specifically,

$$\lambda_k(\Delta_{\Gamma^{d_0,h}}) = \min_{L_k\,:\,\dim(L_k)=k}\ \max_{u\in L_k\setminus\{0\}} \frac{b_\Gamma^{d_0,h}(u,u)}{\|u\|_{\mathbf{m}}^2},$$

where the minimization is over k-dimensional subspaces L_k and $\|u\|_{\mathbf{m}}^2 := \frac{1}{n}\sum_i m_i u_i^2$. When $d_0(x,y)$ is the Euclidean distance $\|x-y\|_2$, the eigenvectors of $\Delta_{\Gamma^{d_0,h}}$ with smallest eigenvalues (which we will hereafter refer to as "low frequency" in analogy with Fourier analysis) have important applications in unsupervised machine learning. Indeed for connected graphs, the second lowest frequency eigenvector (and first eigenvector with eigenvalue greater than 0) is closely connected to normalized graph cuts (Shi and Malik, 2000). The problem of finding a partition $(\mathcal{Z}_*, \mathcal{Z}_*^c)$ of $\mathcal{X}$ which minimizes

$$\mathrm{Ncut}(\mathcal{Z}) = \sum_{i\in\mathcal{Z},j\in\mathcal{Z}^c} W_{ij} \Big/ \min\left\{\sum_{i\in\mathcal{Z},j\in\mathcal{X}} W_{ij}, \sum_{i\in\mathcal{Z}^c,j\in\mathcal{X}} W_{ij}\right\} \tag{8}$$

is NP-hard. One can relax the hard cluster assignments (which correspond to integer constraints in optimization of the Rayleigh quotient) in (8), which leads to making partition assignments by thresholding the second lowest frequency eigenvector of Δ_Γ. This approach can be extended to more than 2 clusters by running K-means or Gaussian mixture modeling on a small number of the lowest frequency eigenvectors of Δ_Γ. These procedures—in which low-frequency Laplacian eigenfunctions are used as features in baseline clustering algorithms—are called *spectral clustering* and have been well-studied and extended in recent decades (Ng et al., 2002; Von Luxburg, 2007; Schiebinger et al., 2015; García Trillos et al., 2021).

2.2.1 Alternative Graph Laplacian Normalizations

In the previous section we introduced the random walk Laplacian matrix, but Laplacian matrices can in fact be categorized as a three parameter family of various density normalizations. In particular, given $j,q,r\in\mathbb{R}$ and an initial weight matrix $W\in\mathbb{R}^{n\times n}$ (for some $n\in\mathbb{N}$) together with its associated degree matrix D, let W_q be the weight matrix defined as $(W_q)_{ij} := W_{ij}/(D_i^q D_j^q)$ and D_q be its corresponding degree matrix. We then define the matrix:

$$L_{j,q,r} := \begin{cases} D_q^{\frac{1-j}{q-1}}(D_q - W_q)D_q^{-\frac{r}{q-1}}, & \text{if } q\neq 1,\\ D_q - W_q, & \text{if } q=1.\end{cases}$$

Some popular choices of (j,r,q) are $q=1$, which yields the standard unnormalized graph Laplacian; $(j,q,r)=(2,3,1)$, which yields the symmetric normalized Laplacian; and $(j,q,r)=(2,2,0)$, which yields the random walk Laplacian. We note moreover that the family of operators inducing *diffusion maps* (Coifman et al., 2005; Coifman and Lafon, 2006) is obtained by taking $(j,q,r)=(2(1-a),2(1-a),0)$ for $a\in[0,1]$.

In Hoffmann et al. (2022), the base weights W_{ij} over a data set $\mathcal{X}$ sampled from a distribution with density ρ over $\mathcal{M}$ are constructed as in (5) using the Euclidean distance as $d_0(\cdot,\cdot)$. The resulting graph Laplacians are closely related to the family of differential operators discussed in Remark 2.8.

2.3 Continuum Laplace Operators

There is also a corresponding *continuum* Laplacian that is not based on point clouds, but instead is defined at the population level. We recall the definition of the s-weighted Laplacian from Hein et al. (2007) (note we have changed the sign in Definition 2.4 for consistency with our notation).

Definition 2.4 *Let $(\mathcal{M}, g)$ be a Riemannian manifold with probability measure ν and associated density ρ defined with respect to dVol. Let Δ be the Laplace-Beltrami operator on $(\mathcal{M}, g)$. For $s \in \mathbb{R}$, we define* the s^{th} weighted Laplacian Δ_s *as*

$$\Delta_s := -\Delta - \frac{s}{\rho} g^{ij}(\nabla_i \rho)\nabla_j = -\frac{1}{\rho^s} g^{ij}\nabla_i(\rho^s \nabla_j) = -\frac{1}{\rho^s}\mathrm{div}(\rho^s \nabla)\,.$$

Note that $s = 0$ gives the negative Laplace-Beltrami operator under the geometry determined by g; $s = 2$ is the continuous version of the normalized random walk Laplacian. The diffusion maps framework (Coifman and Lafon, 2006) considers the family of operators $s = 2(1 - a)$ for $a \in [0, 1]$. If, in the construction described in Definition 2.4, one uses the Fermat metric g_p as defined in (3) instead of g we arrive at a new operator defined on $\mathcal{M}$, which we denote as $\Delta_{s,p}$. Explicitly,

$$\Delta_{s,p} := -\frac{1}{\rho_p^s}\mathrm{div}_p(\rho_p^s \nabla_p)\,, \tag{9}$$

where div_p, ∇_p are the divergence and gradient in the geometry induced by g_p, and ρ_p is the density of the measure ν with respect to the volume form dVol_p. For the latter, recall that any Riemannian metric g has an associated volume form, dVol, defined with respect to any local coordinates $x^1, \ldots, x^m$ as $\mathrm{dVol} = \sqrt{\det(g)}dx^1 \cdots dx^m$, or simply $\mathrm{dVol} = \sqrt{\det(g)}dx$. From this, one sees that

$$\mathrm{dVol}_p := \sqrt{\det(g_p)}dx = \sqrt{\det(\rho^{-\alpha} g)}dx = \rho^{-m\alpha/2}\mathrm{dVol} = \rho^{1-p}\mathrm{dVol},$$

where the final equality comes from the definition of α. Recall that ρ is the density of the measure ν with respect to dVol. That is, for all $U \subset \mathcal{M}$,

$$\nu(U) = \int_U \rho \mathrm{dVol} = \int_U \rho \cdot \rho^{p-1}\left(\rho^{1-p}\mathrm{dVol}\right) = \int_U \rho^p \mathrm{dVol}_p.$$

Thus ρ_p, the density of ν with respect to dVol_p, is $\rho_p = \rho^p$.

To get a better intuition on how the parameters p, s affect the qualitative properties of the operators $\Delta_{s,p}$, specifically, properties of their spectra and implications for data clustering, it will be convenient to relate geometric quantities in $(\mathcal{M}, g_p)$ with those in $(\mathcal{M}, g)$; we will annotate the objects pertaining to the $(\mathcal{M}, g_p)$-geometry with the subscript p. First, gradients of smooth functions $f : \mathcal{M} \to \mathbb{R}$ under the different geometries are related according to $\nabla_p f(x) = \rho^\alpha(x)\nabla f(x), \quad x \in \mathcal{M}$. This standard identity is deduced by writing the first variation of f in a given direction as an inner product between the gradient of f (in each geometry) and the direction of variation. From the relations between gradients and volume forms for the two geometries g and g_p one can also deduce a relation between

the divergences of a smooth vector field under g and g_p. For this, we recall the integration by parts formulae:

$$\begin{aligned}\int_{\mathcal{M}} g(\nabla f, V)\mathrm{dVol} &= -\int_{\mathcal{M}} \mathrm{div}(V) f \mathrm{dVol},\\ \int_{\mathcal{M}} g_p(\nabla_p f, V)\mathrm{dVol}_p &= -\int_{\mathcal{M}} \mathrm{div}_p(V) f \mathrm{dVol}_p,\end{aligned} \tag{10}$$

which hold for all smooth scalar functions f and all smooth vector fields V. From the above, one can readily obtain the following relation: $\mathrm{div}_p(V) = \rho^{p-1}\mathrm{div}(\rho^{1-p}V)$. We can also relate the second order geometries of $(\mathcal{M}, g)$ and $(\mathcal{M}, g_p)$. In particular, the sectional curvatures of $(\mathcal{M}, g_p)$ can be controlled in terms of those of $(\mathcal{M}, g)$ as stated precisely in Theorem B.2 in Appendix B.

With the above identities we can now rewrite $\Delta_{s,p}$ in terms of differential operators in the geometry of $(\mathcal{M}, g)$.

Proposition 2.5 (Fermat s-Laplacian in Euclidean Coordinates) *When $\mathcal{M} \subseteq \mathbb{R}^D$ is an embedded manifold and $g = \overline{g} = \langle \cdot, \cdot \rangle$ is the Euclidean metric, we have:*

(a) $\Delta_{s,p} = -\rho^{\frac{2(p-1)}{m}} \left[\Delta + \left(p(s-1) + 1 + \dfrac{2(p-1)}{m}\right)\dfrac{\nabla\rho}{\rho}\cdot\nabla\right]$, *where Δ is the Laplace-Beltrami operator on $(\mathcal{M}, \overline{g})$.*

(b) When $s = 2$, $\Delta_{2,p} = -\frac{1}{\rho^{2p}}\mathrm{div}_p(\rho^{2p}\nabla_p) = -\rho^{\frac{2(p-1)}{m}}\left[\Delta + \left(p + 1 + \frac{2(p-1)}{m}\right)\frac{\nabla\rho}{\rho}\cdot\nabla\right]$ *is a random walk Laplacian.*

Proof To see (a), recall that the s-weighted Laplacian is $\Delta_{s,p} = -\dfrac{1}{\rho^{ps}}\mathrm{div}_p(\rho^{ps}\nabla_p)$. The desired formula for $\Delta_{s,p}$ follows by using the fact that $\nabla_p = \rho^\alpha \nabla$, $\mathrm{div}_p(\cdot) = \rho^{p-1}\mathrm{div}(\rho^{1-p}\,\cdot)$ (discussed in Section 4) and the product rule. Plugging in $s = 2$ yields (b). ■

The operator $\Delta_{s,p}$ is self-adjoint in an appropriately defined inner product space. For any $s \geq 0$ consider the function space:

$$L^2(\mathcal{M}, \rho^{ps}\mathrm{dVol}_p) := \left\{ f : \mathcal{M} \to \mathbb{R} \;\middle|\; \int_{\mathcal{M}} f^2 \rho^{ps}\mathrm{dVol}_p < \infty \right\}.$$

Define an inner product on this space by

$$\langle f, h\rangle_{p,s} := \int_{\mathcal{M}} fh\,\rho^{ps}\,\mathrm{dVol}_p = \int_{\mathcal{M}} fh\,\rho^{ps}\,\rho^{1-p}\mathrm{dVol} = \int_{\mathcal{M}} fh\,\rho^{p(s-1)+1}\mathrm{dVol}.$$

Also, consider the Dirichlet form $D_{s,p}$ defined as

$$\begin{aligned} D_{s,p}(f,h) &:= \int_{\mathcal{M}} g_p\left(\nabla_p f, \nabla_p h\right)\rho^{ps}\mathrm{dVol}_p \\ &= \int_{\mathcal{M}} \rho^\alpha g\left(\nabla f, \nabla h\right)\rho^{ps}\left[\rho^{1-p}\mathrm{dVol}\right] \\ &= \int_{\mathcal{M}} g\left(\nabla f, \nabla h\right)\rho^{1+p(s-1)+\alpha}\mathrm{dVol}. \end{aligned} \tag{11}$$

Using (10) and the first line in (11) we can immediately deduce the following.

Proposition 2.6 *In its domain of definition, $\Delta_{s,p}$ is self-adjoint with respect to the inner product $\langle \cdot, \cdot \rangle_{p,s}$. Moreover, $D_{s,p}(f,h) = \langle \Delta_{s,p} f, h \rangle_{p,s} = \langle f, \Delta_{s,p} h \rangle_{p,s}$, for all smooth f, h.*

Remark 2.7 *To intuitively interpret the role of ρ on the operator $\Delta_{s,p}$, it is helpful to see $\Delta_{s,p}$, at least when we consider $\mathcal{M} = \Omega$ to be a domain in Euclidean space for simplicity, as the generator of the following diffusion:*

$$dX_t := \rho^{\frac{2(p-1)}{m}} \left(p(s-1) + 1 + \frac{2(p-1)}{m} \right) \frac{\nabla \rho}{\rho} \, dt + \sqrt{2} \rho^{\frac{(p-1)}{m}} dB_t.$$

As $p > 1$, the scalar coefficients in front of the drift and diffusion terms in the above SDE are large at points of high density. This means that a particle moving according to this SDE explores connected regions with high density very rapidly while being drifted away from regions where ρ is small. Thus, low density barriers will be particularly difficult to cross, effectively inducing a separation of regions (clusters) of high density.

The problem of understanding for what choice of parameter scaling h does a discrete operator Δ_Γ built over samples from ν converge to Δ_s as $n \to \infty$ is by now a classical problem (Belkin and Niyogi, 2003; García Trillos et al., 2019; Wormell and Reich, 2021; Calder and García Trillos, 2022) when g is the metric induced by a Euclidean embedding of $\mathcal{M}$. One of our contributions in this paper is to analyze the convergence of the discrete graph Laplacian built from Fermat distances to its appropriate continuum analogue.

Remark 2.8 *It was informally argued in Hoffmann et al. (2022) that the graph Laplacian $L_{j,q,r}$ from Section 2.2.1, built with base weights as in (5) for i.i.d. points $x_1, \dots, x_n$ sampled from $\rho \, d\mathrm{Vol}$ and d_0 the Euclidean distance, approximates spectrally the family of differential operators*

$$\mathcal{L}_{j,q,r} f = -\frac{1}{\rho^j} \mathrm{div} \left(\rho^q \nabla \left(\frac{f}{\rho^r} \right) \right).$$

On the other hand, based on the relations between different geometric quantities under the geometries g and g_p discussed above, the operator $\Delta_{s,p}$ from (9) can be written as

$$\Delta_{s,p} f = -\frac{1}{\rho^{(s-1)p+1}} \mathrm{div}(\rho^{(s-1)p+1+\alpha} \nabla f),$$

where $\alpha = 2(p-1)/m$. This says that, in principle, the graph Laplacians discussed in Hoffmann et al. (2022) can be used to recover the spectrum of $\Delta_{s,p}$, provided one chooses $j = (s-1)p + 1$, $q = (s-1)p + 1 + \alpha$, and $r = 0$, in which case $\mathcal{L}_{j,q,0} = \Delta_{s,p}$. We will use this observation in our numerics Section 6.2 when comparing Fermat-based graph Laplacians with their degree-reweighted counterparts.

3. Set-up and Main Results

In this section, we introduce some more technical background and state our main results. Background on Riemannian geometry is given in Appendix A. We assume that $\mathcal{M}$ is normalized so that it has volume 1 with respect to its canonical measure dVol (*i.e.,* the measure given by the volume form associated to g).

3.1 Notation and Assumptions

We shall make the following standing assumptions on the manifold $\mathcal{M}$ and density ρ.

Assumption 3.1 *We assume $\mathcal{M} \subset \mathbb{R}^D$ is an embedded, smooth, compact manifold of dimension m and that ρ satisfies:*

(i) there exists $\beta > 0$ such that $\frac{1}{\beta} \leq \rho(x) \leq \beta$ for all $x \in \mathcal{M}$.

(ii) $\rho \in \mathcal{C}^\infty(\mathcal{M})$.

(iii) ρ is Lipschitz continuous: $|\rho(x) - \rho(y)| \leq L_1 d(x, y)$ for all $x, y \in \mathcal{M}$.

(iv) $\nabla \rho$ is Lipschitz continuous: $\|\nabla^2 \rho(x)\| \leq L_2$ for all $x \in \mathcal{M}$.

Fix $x \in \mathcal{M}$ and let $\log_x = \exp_x^{-1}$ be the inverse of the exponential map at x. Let $R = \text{inj}(\mathcal{M})$ be the injectivity radius, *i.e.,* the largest radius such that $\log_x$ is a diffeomorphism on $\mathcal{B}_x(R)$ for all x, where $\mathcal{B}_x(R)$ is a geodesic ball (in terms of geodesic distance d) of radius R centered at x. Let K be the maximal sectional curvature of $\mathcal{M}$, and let $\mathcal{R} = \text{reach}(\mathcal{M})$. For $v, w \in T_x\mathcal{M}$, let $J_x(v) = \sqrt{\det(g)}$ be the Jacobian in the normal coordinates induced by the exponential map.

For all $v, w \in B_0(R)$, we have:

$$1 - CmK\|v\|^2 \leq J_x(v) \leq 1 + CmK\|v\|^2, \tag{12}$$

$$\|v - w\| - CK\|v - w\|^3 \leq d(\exp_x(v), \exp_x(w)) \leq \|v - w\| + CK\|v - w\|^3. \tag{13}$$

For (12), see (1.34) in García Trillos et al. (2019), and for (13), see Proposition E.1 in Little et al. (2022). Furthermore (see Proposition 2 in García Trillos et al. (2019)), for all $x, y \in \mathcal{M}$ such that $\|x - y\| \leq \frac{\mathcal{R}}{2}$, we have

$$\|x - y\| \leq d(x, y) \leq \|x - y\| + \frac{8}{\mathcal{R}^2}\|x - y\|^3. \tag{14}$$

3.2 Statements of Main Results

In this section we state our main results, the first of which quantifies the discrepancy between Fermat distances at the continuum level and Fermat distances built from i.i.d. point clouds.

Theorem 3.2 (Local Metric Approximation for i.i.d. Point Clouds) *Let $x, y \in \mathcal{M}$. Fix $\epsilon \in (0, 1/(8p + 6))$ and let $\kappa = \frac{2\epsilon}{3} \min\{\frac{1}{m}, \frac{1}{p}\}$. Let Assumption 3.1 hold and suppose moreover that*

$$2(n\beta/2)^{-\frac{1}{m}\left(\frac{1}{3} - \epsilon\right)} \leq d(x, y) \leq C_{\mathcal{M},\rho,p} \quad , \quad m \geq 2\,.$$

Let $\mathcal{X}_n$ consist of n i.i.d. samples from ρ. Then for n large enough:

$$|\tilde{\ell}_p^p(x, y, \mathcal{X}_n) - \mu\mathcal{L}_p^p(x, y)| \leq C_1\mathcal{L}_p^p(x, y)^2 + C_2\mathcal{L}_p^p(x, y)^3$$

with probability at least $1 - C_\epsilon n \exp\left(-c_\epsilon (\frac{n}{4\beta})^\kappa\right)$, *where*

$$C_1 = \beta^{\frac{p-1}{m}} \left(\frac{5\mu}{2} \left(\frac{p-1}{m} \right) L_1 \beta + 1 \right) \ , \ C_2 = C_{p,m,\beta} \left(\mathcal{R}^{-2} + K(1 + L_1) + L_1^2 + L_2 \right) .$$

Here C_ϵ, c_ϵ *are constants depending on* ϵ, $C_{p,m,\beta}$ *is a constant depending on* p, m, β, *and* $C_{\mathcal{M},\rho,p}$ *is a constant depending on* ρ, p, *and the geometry of* $\mathcal{M}$.

Proof See Appendix G. ■

Theorem 3.2 follows from Theorem 4.3 (metric approximation for PPPs) and a de-Poissonization argument. As an application of this local distance approximation result, we obtain rates for the spectral convergence of Fermat-based graph Laplacians toward their continuum analogues. The graph Laplacians are built over realizations of an i.i.d. point cloud, as we state precisely below. For legibility, we write $\Gamma^{\tilde{\ell},h}$ instead of $\Gamma^{\tilde{\ell}_p^p,h}$ and $\Gamma^{\mathcal{L},h}$ instead of $\Gamma^{\mu\mathcal{L}_p^p,h}$ when p is clear from context. We also denote constants by C or c, with subscripts indicating the primitive quantities $(p, m, L_1 \dots)$ upon which they depend.

Theorem 3.3 *Let* $\Gamma^{\tilde{\ell},h}$ *denote the graph constructed using the discrete Fermat distance* $\tilde{\ell}_p^p$ *and bandwidth* $h = O(n^{-\frac{1}{3m}+\frac{\epsilon}{m}})$ *for some* $\epsilon \in (0, \frac{1}{8p+6})$ *over a point cloud of i.i.d. points* $\mathcal{X} = \{x_1, \dots, x_n\}$ *sampled from* ρ. *Let* $\Delta_{\Gamma^{\tilde{\ell},h}}$ *denote the random walk Laplacian of this graph (see* (6)*) and* $\lambda_k(\Gamma^{\tilde{\ell},h})$ *its* k*-th eigenvalue. Finally, let* $\Delta_{2,p}$ *denote the* 2*-weighted Fermat Laplacian:*

$$\Delta_{2,p} u = -\frac{1}{\rho_p^2} \mathrm{div}_p \left(\rho_p^2 \nabla_p u \right)$$

and let $\lambda_k(\Delta_{2,p})$ *denote its* k*-th eigenvalue. Then for* n *large enough*

$$\frac{\left| \lambda_k(\Delta_{\Gamma^{\tilde{\ell},h}}) - \lambda_k(\Delta_{2,p}) \right|}{|\lambda_k(\Delta_{2,p})|} \leq C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} + C_{\beta,p,m,L_1,L_2,K,\mathcal{R}} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}}$$

with probability at least $1 - C_\epsilon n^3 \exp\left(-c_\epsilon (\frac{n}{4\beta})^{\frac{\epsilon}{2p+1}} \min\{\frac{1}{m}, \frac{1}{p}\}\right)$.

Remark 3.4 *Note that the leading error term is* $O(n^{-\frac{1}{3m}+\frac{\epsilon}{m}})$. *In contrast, the leading error term in García Trillos et al. (2019) is* $O(n^{-\frac{1}{2m}})$. *This is explained by the fact that in Theorem 3.3 the dominating source of error comes from the approximation* $\tilde{\ell}_p^p(\cdot,\cdot) \approx \mu\mathcal{L}_p^p(\cdot,\cdot)$ *which decays more slowly than the* ∞*-optimal transport distance between the sample measure associated to* $\mathcal{X}$ *and population measure* ν*—the dominating source of error in García Trillos et al. (2019).*

We use Theorem 3.3 and combine it with some auxiliary estimates discussed in Sections 5.2 and 5.3 to prove the convergence of eigenvectors of $\Delta_{\Gamma^{\tilde{\ell},h}}$ toward eigenfunctions of $\Delta_{2,p}$. While said results can be derived for arbitrary eigenvectors, we will focus on the consistency of the Fiedler eigenvector (*i.e.,* the one with second smallest eigenvalue) for simplicity and refer the reader to works like Burago et al. (2015); García Trillos et al.

(2019) for extensions; see also Remark 5.12 below. We will further assume that $\Delta_{2,p}$'s Fiedler eigenvalue is simple. In order to compare the discrete and continuum eigenvectors, we use a *discretization map* that associates a function defined over the data set to every function defined over the manifold $\mathcal{M}$; this operator is introduced in (22). We prove that the difference between the discretization of the continuum Fiedler eigenfunction and its projection onto the graph-Laplacian Fiedler eigenspace tends to zero as $n \to \infty$; we also quantify the rate of convergence. Other notions of consistency for eigenvectors of discrete graph Laplacians toward manifold counterparts are possible and we refer the reader to García Trillos et al. (2019) for more details.

Theorem 3.5 *Let the assumptions and notation from Theorem 3.3 hold, and furthermore assume that the Fiedler eigenvalue $\lambda_2 = \lambda_2(\Delta_{2,p})$ of $\Delta_{2,p}$ is simple. Let f be a normalized eigenfunction with eigenvalue $\lambda_2(\Delta_{2,p})$ (which by assumption is unique up to sign). Then with probability at least $1 - C_\epsilon n^3 \exp\left(-c_\epsilon (\frac{n}{4\beta})^{\frac{\epsilon}{2p+1}} \min\{\frac{1}{m}, \frac{1}{p}\}\right)$, for n large enough,*

$$\begin{aligned}\|Pf - \psi_2^n\|_{\boldsymbol{m}}^2 \leq \frac{1}{\lambda_3 - \lambda_2}\left[C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} + C_{\beta,p,m,L_1,L_2,K,\mathcal{R}} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}}\right]\lambda_2 \\ + C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} + C_{m,\beta,p,L_1,L_2,K} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}},\end{aligned}$$

where $\lambda_3 = \lambda_3(\Delta_{2,p})$, P is the discretization map defined in (22), and ψ_2^n is one of the normalized (w.r.t. $\|\cdot\|_{\boldsymbol{m}}$) Fiedler eigenvectors of the graph Laplacian $\Delta_{\Gamma^{\tilde{\ell},h}}$.

Although Theorem 3.3 guarantees that for n large enough $\lambda_2(\Gamma^{\tilde{\ell},h})$ is also a simple eigenvalue, in the statement of Theorem 3.5 we talk about "one of the Fiedler eigenvectors of the graph Laplacian" since uniqueness only holds up to sign. In Section 5.4 we provide further remarks on the convergence stated in Theorem 3.5 and discuss other approaches in the literature to prove consistency of eigenvectors for Laplacian matrices built over data clouds.

4. Fermat Riemannian Metrics

4.1 Fermat Geodesics and Local Euclidean Equivalence

In this section we derive a concrete, local expression for geodesics in the Fermat distance on a flat domain (*i.e.,* an m-dimensional, open, connected subset of $\mathbb{R}^m$), and then establish a higher order local equivalence between $\mathcal{L}_p$ and Euclidean distance on a flat domain. These results will then be applied in the tangent plane of a general manifold to establish the local metric approximation results given in Section 4.2. To make clear that the domain lacks both intrinsic and extrinsic curvature, we denote it by Ω instead of $\mathcal{M}$.

Theorem 4.1 *Let $\Omega \subseteq \mathbb{R}^m$ be a m-dimensional, open, connected domain and assume (Ω, ρ) satisfies Assumption 3.1. Fix $y \in \Omega$ and let $\gamma_b(t)$ denote the unit speed geodesic with respect to g_p originating at y in the direction of unit vector $b \in \mathbb{R}^m$. Then:*

$$\begin{aligned}\gamma_b(t) =& \rho(y)^{\frac{\alpha}{2}} bt + \alpha\rho(y)^{\alpha-1}\left(\frac{1}{2}\langle b, \nabla\rho(y)\rangle b - \frac{1}{4}\nabla\rho(y)\right)t^2 \\ &+ (C_1' b + C_2' H(y) b + C_3' \nabla\rho(y)) t^3 + O(t^4)\,,\end{aligned} \tag{15}$$

where $H(y)$ denotes the Hessian matrix of ρ evaluated at y and

$$C_1' = \left(\frac{1}{3}\alpha^2 - \frac{1}{6}\alpha\right)\rho(y)^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho(y)\rangle^2 + \frac{\alpha}{6}\rho(y)^{\frac{3}{2}\alpha-1}\langle H(y)b, b\rangle - \frac{\alpha^2}{12}\rho(y)^{\frac{3}{2}\alpha-2}\langle\nabla\rho(y), \nabla\rho(y)\rangle,$$
$$C_2' = -\frac{\alpha}{12}\rho(y)^{\frac{3}{2}\alpha-1},$$
$$C_3' = \left(\frac{\alpha}{12} - \frac{\alpha^2}{6}\right)\rho(y)^{\frac{3}{2}\alpha-2}\langle\nabla\rho(y), b\rangle.$$

Proof See Appendix C. ■

As a numerical illustration of the local Fermat geodesics, consider the density function $\rho(x_1, x_2) = 1 + x_1$ for $(x_1, x_2) \in [-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$, so that $\nabla\rho = (1, 0)$, $\rho(0, 0) = 1$. We consider $p = 3$, and since $m = 2$ we have $\alpha = 2$. We solve the geodesic equations appearing in the proof of Theorem 4.1 with a numerical solver for a collection of initial directions b and for total time T in order to produce the Fermat geodesic balls appearing in Figure 2; note (15) gives a third order approximation of these curves, valid for small times. Intuitively, the $\mathcal{L}_p^p$ ball is elongated in the direction of the gradient because it costs less to travel in that direction.

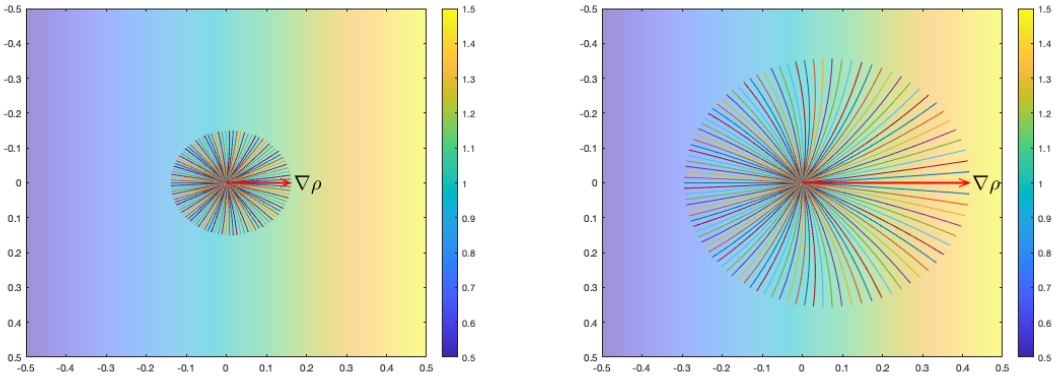


Figure 2: Fermat geodesic balls of radius $T = 0.15$ (left) and $T = 0.35$ (right). The $\mathcal{L}_p^p$ ball is wider in directions of high density because geodesics in that direction are less costly.

We can utilize the geodesic computation of Theorem 4.1 to establish a higher-order equivalence of Fermat distances with a density-warped Euclidean distance, depending both on ρ and $\nabla\rho$.

Theorem 4.2 *Let $\Omega \subseteq \mathbb{R}^m$ be an m-dimensional, open, connected domain and assume (Ω, ρ) satisfies Assumption 3.1. Then for $x, y \in \Omega$ with $x \neq y$ and unit vector $u = (y - x)/\|y - x\|$, we can relate Euclidean and Fermat distance as follows:*

$$\|y - x\| = \rho(x)^{\frac{p-1}{m}}\mathcal{L}_p^p(x, y) + \frac{1}{2}\left(\frac{p-1}{m}\right)\langle u, \nabla\rho(x)\rangle\,\rho(x)^{\frac{2(p-1)}{m}-1}\mathcal{L}_p^{2p}(x, y)$$

$$+ C\mathcal{L}_p^{3p}(x,y) + O(\mathcal{L}_p^{4p}(x,y))$$

for

$$C = \rho(x)^{\frac{3}{2}\alpha-2}\left[\frac{\alpha^2}{96}\langle\nabla\rho(x),\nabla\rho(x)\rangle + \left(\frac{7\alpha^2}{96}-\frac{\alpha}{12}\right)\langle u,\nabla\rho(x)\rangle^2 + \frac{\alpha}{12}\rho(x)\langle H(x)u,u\rangle\right].$$

Also:

$$\mathcal{L}_p^p(x,y) = \frac{1}{\rho(x)^{\frac{p-1}{m}}}\left(\|y-x\| - \frac{1}{2}\left(\frac{p-1}{m}\right)\left\langle u, \frac{\nabla\rho(x)}{\rho(x)}\right\rangle\|y-x\|^2\right) + O\left(\|y-x\|^3\right).$$

Proof See Appendix D. ■

4.2 Discrete-to-Continuum Fermat Distance Approximation

We now utilize the results of Section 4.1 to establish that, locally, the discrete, computable metric $\tilde{\ell}_p^p$ is well-approximated by the Fermat geodesic distance $\mathcal{L}_p^p$. Following is the main result of this section, which quantifies the metric approximation for PPPs; we work with PPPs in order to utilize results from percolation theory. Recall $H_{n\rho}$ denotes a nonhomogeneous Poisson point process on $\mathcal{M}$ with intensity $n\rho$, and $\tilde{\ell}_p^p(x,y,H_{n\rho}) = n^{\frac{p-1}{m}}\ell_p^p(x,y,H_{n\rho})$ denotes the discrete (normalized) Fermat distance computed in $H_{n\rho}\cup\{x,y\}$. Note although the number of points in $H_{n\rho}$ is random, $\mathbb{E}[|H_{n\rho}|] = n$.

Theorem 4.3 (Local Metric Approximation for PPP) *Suppose $(\mathcal{M},\rho)$ satisfies Assumption 3.1. Let $x,y\in\mathcal{M}$, fix $\epsilon\in(0,1/(8p+6))$, and suppose that*

$$2(n\beta)^{-\frac{1}{m}\left(\frac{1}{3}-\epsilon\right)} \le d(x,y) \le C_{\mathcal{M},\rho,p}.$$

Then for n large enough:

$$|\tilde{\ell}_p^p(x,y,H_{n\rho}) - \mu\mathcal{L}_p^p(x,y)| \le C_1\mathcal{L}_p^p(x,y)^2 + C_2\mathcal{L}_p^p(x,y)^3$$

with probability at least $1 - C_\epsilon n\exp\left(-c_\epsilon(\frac{n}{2\beta})^{\frac{2\epsilon}{3}\min\{\frac{1}{m},\frac{1}{p}\}}\right)$, where

$$C_1 = \beta^{\frac{p-1}{m}}\left(\frac{5\mu}{2}\left(\frac{p-1}{m}\right)L_1\beta + 1\right)\ ,\ C_2 = C_{p,d,\beta}\left(\mathcal{R}^{-2} + K(1+L_1) + L_1^2 + L_2\right).$$

Here C_ϵ, c_ϵ are constants depending on ϵ, $C_{p,d,\beta}$ is a constant depending on p,d,β, and $C_{\mathcal{M},\rho,p}$ is a constant depending on ρ, p, and the geometry of $\mathcal{M}$.

4.2.1 Background and Outline for Proof of Theorem 4.3

We first establish some notation used in the proof of Theorem 4.3. Let $r = d(x,y)$; we assume r is upper bounded by a constant $C_{\mathcal{M},\rho,p}$ independent of n but depending on ρ,p, and the geometry of $\mathcal{M}$, but to simplify already long calculations we do not keep track of it. Let $u = \log_x(y)$ so that $\|u\| = r$.

Fix any $R > 0$ satisfying $4R < \mathcal{R}$ and define the function $\rho_x(z) := \rho(\exp_x(z))J_x(z)\mathbf{1}\{z \in B_0(R)\}$, where $\mathbf{1}\{z \in B_0(R)\}$ is 1 for $z \in B_0(R)$, and 0 otherwise. We note that $H_{n\rho_x} := \log_x(H_{n\rho} \cap \mathcal{B}_x(R))$ is a PPP on $B_0(R) \subseteq T_x\mathcal{M}$ with intensity $n\rho_x(z)$, since for $A \subseteq B_0(R)$,

$$\mathbb{E}[H_{n\rho_x}(A)] = \mathbb{E}[H_{n\rho}(\exp_x(A))] = n\int_{\exp_x(A)\subseteq\mathcal{M}} \rho(y)dv_y = n\int_{A\subseteq B_0(R)} \rho(\exp_x(z))J_x(z)dz\,,$$

where $H_{n\rho_x}(A)$ denotes the number of points of the PPP falling in A. The function ρ_x (thought of as the density ρ lifted to the tangent plane) implicitly defines a tangent plane continuum Fermat distance between x, y:

$$\mathcal{L}_p^p(0, u, \rho_x) = \inf_{\log_x\gamma\subseteq B_0(R)} \int \rho_x(\log_x\gamma(t))^{(1-p)/m}|(\log_x\gamma)'(t)|\ dt$$

where $\log_x\gamma$ is a path in $T_x\mathcal{M}$ connecting 0 and u. Let $\rho_{\min}, \rho_{\max}$ be the min/max of $\rho_x(z)$ over $B_0(2r)$. Let $H_{n\rho_{\min}}$ be the PPP obtained by replacing ρ_x with $\rho_{\min}$ on $B_0(2r)$ (no change outside of this ball), and similarly for $H_{n\rho_{\max}}$.

Note that we can couple the PPPs so that $H_{n\rho_{\min}} \subseteq H_{n\rho_x} \subseteq H_{n\rho_{\max}}$. Finally, let $\overline{H}_{n\rho_{\min}} = H_{n\rho_{\min}} \cap B_0(2r)$, $\overline{H}_{n\rho_{\max}} = H_{n\rho_{\max}} \cap B_0(2r)$. Note that $\overline{H}_{n\rho_{\min}}$, $\overline{H}_{n\rho_{\max}}$ are *homogeneous* PPPs on $B_0(2r)$ with intensities $n\rho_{\min}$, $n\rho_{\max}$, while $H_{n\rho_x}$ is inhomogeneous and harder to analyze directly.

Our metric approximation consists of a series of Lemmas:

1. Lemma E.1: $\tilde{\ell}_p^p(x, y, H_{n\rho}) = \tilde{\ell}_p^p(x, y, H_{n\rho} \cap \mathcal{B}_x(R))$ holds w.h.p. Intuitively, this means that optimal discrete Fermat paths between nearby points do not meander too much. This allows the paths on the manifold to be compared to those on the tangent plane.

2. Lemma E.2: $\tilde{\ell}_p^p(x, y, H_{n\rho} \cap \mathcal{B}_x(R)) \approx \tilde{\ell}_p^p(0, u, H_{n\rho_x})$. This means that optimal paths on the manifold are close to those on the tangent plane. The paths on the tangent plane are easier to analyze, because the domain is flat.

3. Lemma E.3: $\ell_p^p(0, u, H_{n\rho_{\min}}) = \ell_p^p(0, u, \overline{H}_{n\rho_{\min}})$ w.h.p., and same for $H_{n\rho_{\max}}$
This allows us to "trap" the Fermat distance for the nonhomogeneous PPP $H_{n\rho_x}$ as $\ell_p^p(0, u, \overline{H}_{n\rho_{\max}}) \leq \ell_p^p(0, u, H_{n\rho_x}) \leq \ell_p^p(0, u, \overline{H}_{n\rho_{\min}})$.

4. Lemma E.4: $\tilde{\ell}_p^p(0, u, H_{n\rho_x}) \approx \mu\mathcal{L}_p^p(0, u, \rho_x)$ w.h.p. This is a direct discrete-to-continuum metric approximation on the tangent plane.

5. Lemma E.5: $\mathcal{L}_p^p(0, u, \rho_x) \approx \mathcal{L}_p^p(x, y)$. This involves bounding the perturbation due to curvature for $\mathcal{L}_p^p$.

See Appendix E for precise statements and proofs of the above lemmas. An important tool in our analysis is the following result from percolation theory, which is a direct result of Theorem 2.2 in Howard and Newman (2001).

Proposition 4.4 *Let $H_{n\lambda}$ be a homogeneous PPP on $\mathbb{R}^m$ with intensity $n\lambda > 1$. Fix $q > 1$ and $\epsilon \in (0, 1/(8p + 6))$. Suppose*

$$\|u\| \geq (n\lambda)^{-\frac{1}{m}\left(\frac{1}{2q-1}-\epsilon\right)} \tag{16}$$

Then

$$\left|(n\lambda)^{\frac{(p-1)}{m}}\ell_p^p(0,u,H_{n\lambda})-\mu\|u\|\right|\leq\|u\|^q \tag{17}$$

with probability at least

$$1-C_\epsilon\exp\left(-c_\epsilon(n\lambda)^{\epsilon\left(\frac{2q-2}{2q-1}\right)\min\{\frac{1}{m},\frac{1}{p}\}}\right). \tag{18}$$

for constants C_ϵ, c_ϵ depending on ϵ.

Proof By Theorem 2.2 in Howard and Newman (2001), for every $\epsilon\in(0,\frac{1}{8p+6})$, $|\ell_p^p(0,z,H_1)-\mu\|z\||\leq\|z\|^{\frac{1}{2}+\epsilon}$ with probability at least $1-C_\epsilon\exp(-c_\epsilon\|z\|^{\epsilon\min\{1,\frac{m}{p}\}})$. Via $(n\lambda)^{\frac{1}{m}}u=z$, we obtain

$$\begin{aligned}
&|(n\lambda)^{\frac{p}{m}}\ell_p^p(0,u,H_{n\lambda})-\mu(n\lambda)^{\frac{1}{m}}\|u\||\leq\left((n\lambda)^{\frac{1}{m}}\|u\|\right)^{\frac{1}{2}+\epsilon}\\
\implies&|(n\lambda)^{\frac{1}{m}+\frac{p-1}{m}}\ell_p^p(0,u,H_{n\lambda})-\mu(n\lambda)^{\frac{1}{m}}\|u\||\leq\left((n\lambda)^{\frac{1}{m}}\|u\|\right)^{\frac{1}{2}+\epsilon}\\
\implies&|(n\lambda)^{\frac{p-1}{m}}\ell_p^p(0,u,H_{n\lambda})-\mu\|u\||\leq\frac{\|u\|^{\frac{1}{2}+\epsilon}}{(n\lambda)^{\frac{1}{m}(\frac{1}{2}-\epsilon)}}
\end{aligned} \tag{19}$$

with probability

$$1-C_\epsilon\exp\left(-c_\epsilon\left((n\lambda)^{\frac{1}{m}}\|u\|\right)^{\epsilon\min\{1,m/p\}}\right). \tag{20}$$

We observe that (16) implies $\|u\|\geq(n\lambda)^{\frac{-(1-2\epsilon)}{m(2q-1-2\epsilon)}}$; thus the right hand side of (19) can be bounded by

$$\frac{\|u\|^{\frac{1}{2}+\epsilon}}{(n\lambda)^{\frac{1}{m}(\frac{1}{2}-\epsilon)}}\leq\frac{\|u\|^{\frac{1}{2}+\epsilon}}{\|u\|^{-q+1/2+\epsilon}}=\|u\|^q.$$

Similarly $(n\lambda)^{1/m}\|u\|\geq(n\lambda)^{1/m}\left[(n\lambda)^{\frac{-(1-2\epsilon)}{m(2q-1-2\epsilon)}}\right]=(n\lambda)^{\frac{2q-2}{m(2q-1-2\epsilon)}}\geq(n\lambda)^{\frac{2q-2}{m(2q-1)}}$, assuming $n\lambda>1$. Substituting this into (20) yields (18). ■

Remark 4.5 *There is an interesting trade-off controlled by q. Taking q larger yields a faster rate of convergence in* (17)*, but one that only applies to larger u in* (16)*. As $q\to1^+$ we approach a very natural lower bound, $\|u\|\geq(n\lambda)^{-1/m}$, but the probability bound* (18) *becomes constant with respect to n, and the metric approximation error* (17) *becomes large. We shall focus exclusively on the $q=2$ case henceforth, as this choice leads to the best spectral convergence rates.*

Our PPP metric approximation result (Theorem 4.3) puts together the pieces discussed in the preceding sections to conclude that locally $\mu\mathcal{L}_p^p(x,y)$ is well approximated by $\tilde{\ell}_p^p(x,y,H_{n\rho})$:

Proof [Theorem 4.3] Let $\tilde{C}_1, \tilde{C}_2$ be as in Lemma E.4 and let $\mathcal{L}$ denote $\mathcal{L}_p^p(x,y)$. We have, for $\|u\| \leq C_{\mathcal{M},\rho,p}$,

$$
\begin{aligned}
&\tilde{\ell}_p^p(x, y, H_{n\rho}) \\
=&\tilde{\ell}_p^p(x, y, H_{n\rho} \cap \mathcal{B}_x(R)) \quad \text{(Lemma E.1, prob. } 1 - p_1) \\
=&\left(1 \pm Cp(mK + \mathcal{R}^{-2})\|u\|^2\right) \tilde{\ell}_p^p(0, u, H_{n\rho_x}) \quad \text{(Lemma E.2)} \\
=&\left(1 \pm Cp(mK + \mathcal{R}^{-2})\|u\|^2\right) \left(\mu \mathcal{L}_p^p(0, u, \rho_x) \pm \tilde{C}_1 \mathcal{L}^2 \pm \tilde{C}_2 \mathcal{L}^3\right) \quad \text{(Lemma E.4, prob. } 1 - p_2) \\
=&\left(1 \pm Cp(mK + \mathcal{R}^{-2})\|u\|^2\right) \left(\left(1 \pm Cp(mK + \mathcal{R}^{-2})\|u\|^2\right) \mu \mathcal{L} \pm \tilde{C}_1 \mathcal{L}^2 \pm \tilde{C}_2 \mathcal{L}^3\right) \quad \text{(Lemma E.5)} \\
=&\mu \mathcal{L} \pm \tilde{C}_1 \mathcal{L}^2 + \tilde{C}_2' \mathcal{L}^3,
\end{aligned}
$$

where $\tilde{C}_2' = C_{p,d,\beta}\left(\mathcal{R}^{-2} + K(1 + L_1) + L_1^2 + L_2\right)$ now includes the reach, $p_1 \leq p_2 = C_\epsilon n \exp\left(-c_\epsilon (\frac{n}{2\beta})^{\frac{2\epsilon}{3}} \min\{\frac{1}{m}, \frac{1}{p}$
and we have utilized $\|u\|^2 = \rho(x)^{\frac{2(p-1)}{m}} \mathcal{L}^2 + O(\mathcal{L}^3)$ (see proof of Lemma E.4). ■

We note Theorem 4.3 is closely related to existing results in the Fermat distance literature, especially Theorem 1 in Hwang et al. (2016) and Theorem 2.3 in Groisman et al. (2022). Our results are local (showing $\tilde{\ell}_p^p(x_i, x_j) \approx \mu \mathcal{L}_p^p(x_i, x_j)$ when x_i, x_j are close) but quantitative, giving a sharp third order estimate on the local deviations. In contrast, Hwang et al. (2016) and Groisman et al. (2022) provide macroscale/global convergence results, *i.e.*, x_i, x_j are arbitrary points in $\mathcal{M}$, but of an asymptotic nature. At a more technical level, we note that from the proof of (Hwang et al., 2016, Lemma 10), a result similar to Lemma E.2 can be deduced:

$$\tilde{\ell}_p^p(x, y, H_{n\rho} \cap \mathcal{B}_x(R)) = (1 \pm \delta)\, \tilde{\ell}_p^p(0, u, H_{n\rho_x}).$$

5. Continuum Limits of Graph Laplace Operators

In this section we show that *discrete graph Laplacians built from discrete Fermat distances* converge to *continuum Laplace operators built from continuum Fermat distances*. While this analysis is similar to the arguments of Burago et al. (2015) and García Trillos et al. (2019), new ideas are needed. Indeed:

1. In Burago et al. (2015) convergence of eigenvectors of $\Delta_{\Gamma^{d_g,h}}$ to eigenfunctions of Δ_0 associated to $(\mathcal{M}, g)$ is studied. Here, $\Gamma^{d_g,h}$ denotes the graph constructed using the kernel $\frac{1}{\omega_m}\mathbb{1}_{[0,1]}$, bandwidth parameter h, and the geodesic distance d_g. In this case the graph weights are degree-normalized and not chosen as in (5).

2. In García Trillos et al. (2019) convergence of eigenvectors of $\Delta_{\Gamma^{d_0,h}}$ to eigenfunctions of the random walk Laplacian Δ_2 associated to $(\mathcal{M}, g)$ is studied. Here, $\Gamma^{d_0,h}$ denotes the graph constructed using a kernel η satisfying certain mild assumptions, bandwidth parameter h, and Euclidean distance function $d_0(x, y) = \|x - y\|_2$.

Like García Trillos et al. (2019) our discrete and continuum operators are defined using two different metrics. Unlike García Trillos et al. (2019), the relationship between these two

metrics is not straightforward. Indeed, the key new ingredient is the refined approximation bounds $\mu\mathcal{L}_p^p(x,y) \approx \tilde{\ell}_p^p(x,y)$ of Section 4.2. Nonetheless, we follow the approach outlined in García Trillos et al. (2019). As many of the required technical lemmas can be applied with only minor modifications, instead of reproving them we indicate how the proof presented in García Trillos et al. (2019) should be modified.

We begin by summarizing some relevant definitions and notation. Let $\nu_n := \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ denote the *empirical measure* of $\mathcal{X}_n$. Let $T : \mathcal{M} \to \mathcal{X}$ denote a *transportation map* from ν to ν_n, *i.e.,* $T_\# \nu = \nu_n$ where the pushforward measure $T_\#\nu$ is defined via $T_\#\nu(U) = \nu(T^{-1}(U))$. We define

$$d_\infty^g(\nu,\nu_n) := \min_{T:\ T_\#\nu=\nu_n} \operatorname*{ess\,sup}_{x\in\mathcal{M}} d_g(x,T(x)), \tag{21}$$

Henceforth, we shall fix T to be a minimizer of (21). Later we shall use (García Trillos et al., 2019, Theorem 2) to quantify $d_\infty^g(\nu,\nu_n)$. For now, we note that we may assume $d_\infty^g(\nu,\nu_n) \ll h$. T induces a partition $\{U_i\}_{i=1}^n$ of $\mathcal{M}$ via $U_i = T^{-1}(x_i)$. Using this partition, we define the discretization operator:

$$\begin{aligned} &P : L^2(\mathcal{M},\nu) \to L^2(\mathcal{X},\nu_n), \\ &(Pf)(x_i) = n\int_{U_i} f(x)d\nu(x). \end{aligned} \tag{22}$$

Finally, we define a non-local energy

$$\tilde{E}_r(f) = \int_{\mathcal{M}}\int_{\mathcal{M}} \eta\left(\frac{d_g(x,y)}{r}\right)|f(y)-f(x)|^2\, d\nu(x)d\nu(y),$$

which will be used to approximate the continuous Dirichlet energy.

5.1 Technical Results for Kernels and Degrees

We first derive some supporting results regarding Fermat kernels and degrees; see Appendix F for the proofs. The following corollary applies Theorem 3.2 to kernels and is critical for what follows.

Corollary 5.1 *Let $\delta := 2\mu^{-1}C_1 h + 4\mu^{-1}C_2 h^2 \le \frac{1}{2}$, where C_1, C_2 are as in Theorem 3.2. Define $\widehat{h}_+ := h(1+\delta)$, $\widehat{h}_- := h(1-\delta)$. Then for n large enough, with probability at least $1 - C_\epsilon n^3 \exp\left(-c_\epsilon(\frac{n}{4\beta})^{\frac{\epsilon}{2p+1}}\min\{\frac{1}{m},\frac{1}{p}\}\right)$, we have for all $x_i, x_j \in \mathcal{X}$:*

$$\eta\left(\frac{\mu\mathcal{L}_p^p(x_i,x_j)}{\widehat{h}_-}\right) \le \eta\left(\frac{\tilde{\ell}_p^p(x_i,x_j)}{h}\right) \le \eta\left(\frac{\mu\mathcal{L}_p^p(x_i,x_j)}{\widehat{h}_+}\right), \tag{23}$$

where C_ϵ, c_ϵ are constants depending on ϵ, $\eta = \frac{1}{\omega_m}\mathbb{1}_{[0,1]}$ and $h \ge 4\mu\beta^{\frac{(p-1)}{m}}(n\beta/2)^{-\frac{1}{m}(\frac{1}{3}-\epsilon)}$.

The following lemma is a minor modification of (García Trillos et al., 2019, Lemma 18) and bounds degrees computed with respect to the (discrete) Fermat distance.

Lemma 5.2 *For C_1, C_2 as in Theorem 3.2 and K_p as in Theorem B.2:*

$$\left|m_{\tilde{\ell}_p^p,h,i} - \rho_p(x_i)\right| \le C\beta^p\left[\left(\frac{L_1}{\beta} + C_1 m\right)h + m(K_p + C_2\mu)h^2 + m\omega_m\frac{d_\infty^g(\nu,\nu_n)}{h}\right].$$

5.2 Analysis of Discrete Dirichlet Energies

Next, we consider the relationship between the discrete Dirichlet energies induced by d_g and d_0 respectively, where d_0 is any distance function approximating d_g as quantified by the following assumption:

Assumption 5.3 *Let d_g be the geodesic distance function on $(\mathcal{M}, g)$ while d_0 is any distance function on $\mathcal{X}_n$. Suppose that, for appropriate kernel sizes $h > 0$ there exists $\delta := \delta(h)$ and $\varepsilon_n \to 0^+$ such that for all $x_i, x_j \in \mathcal{X}_n$:*

$$\eta\left(\frac{d_g(x_i,x_j)}{\widehat{h}_-}\right) \leq \eta\left(\frac{d_0(x_i,x_j)}{h}\right) \leq \eta\left(\frac{d_g(x_i,x_j)}{\widehat{h}_+}\right)$$

holds with probability $1-\varepsilon_n$ where $\widehat{h}_+ = h(1+\delta)$ and $\widehat{h}_- = h(1-\delta)$.

For brevity, for $c = 0, g$ we write $w_{i,j}^{c,h}$ instead of $w_{i,j}^{d_c,h}$ and $b^{c,h}(\cdot)$ instead of $b^{d_c,h}(\cdot)$; see Section 2.2 for definitions.

Theorem 5.4 (General Dirichlet Energy Perturbation) *Suppose Assumption 5.3 holds for d_0, d_g and $h > 0$. Then, for all $u : \mathcal{X} \to \mathbb{R}$ and with probability $1-\varepsilon_n$,*

$$(1 - Cm\delta)\, b^{g,\widehat{h}_-}(u) \leq b^{0,h}(u) \leq (1 + Cm\delta)\, b^{g,\widehat{h}_+}(u).$$

Proof We simply compare the Dirichlet energies:

$$\begin{aligned} b^{g,\widehat{h}_-}(u) &= \frac{m+2}{n\widehat{h}_-^2} \sum_{i,j} w_{i,j}^{g,\widehat{h}_-} (u(x_i) - u(x_j))^2 \\ &= \frac{m+2}{n^2\widehat{h}_-^{m+2}} \sum_{i,j} \eta\left(\frac{d_g(x_i,x_j)}{\widehat{h}_-}\right) (u(x_i) - u(x_j))^2 \\ &\overset{(a)}{\leq} \left(\frac{1}{1-\delta}\right)^{m+2} \frac{m+2}{n^2 h^{m+2}} \sum_{i,j} \eta\left(\frac{d_0(x_i,x_j)}{h}\right) (u(x_i) - u(x_j))^2 \\ \Rightarrow (1-\delta)^{m+2}\, b^{g,\widehat{h}_-}(u) &\leq \frac{m+2}{nh^2} \sum_{i,j} w_{i,j}^{0,h} (u(x_i) - u(x_j))^2 = b^{0,h}(u), \end{aligned}$$

where we use Assumption 5.3 in (a). For δ small enough $(1-\delta)^{m+2} \leq 1 - Cm\delta$ and the stated bound follows. The proof of the upper bound is similar. ∎

Next, we relate the corresponding eigenvalues.

Theorem 5.5 *Suppose Assumption 5.3 holds for d_0, d_g and $h > 0$. Recall that $\lambda_k(\Delta_{\Gamma^{c,h}})$ denotes the k-th eigenvalue of the random walk Laplacian of the graph $\Gamma^{c,h} := (\mathcal{X}, W^{c,h})$ for $c = 0, g$. Then with probability $1-\varepsilon_n$*

$$\lambda_k(\Delta_{\Gamma^{0,h}}) \leq \left(1 + Cm\delta + C\beta L_1 h + C\beta^2 \frac{d_\infty^g(\nu,\nu_n)}{h} + C_{\beta,m,L_1} h^2 + C_{m,\beta,L_1}\delta h\right) \lambda_k(\Delta_{\Gamma^{g,\widehat{h}_+}}),$$

$$\lambda_k(\Delta_{\Gamma^{0,h}}) \geq \left(1 - Cm\delta - C\beta L_1 h - C\beta^2 \frac{d_\infty^g(\nu,\nu_n)}{h} - C_{\beta,m,L_1} h^2 - C_{m,\beta,L_1}\delta h\right) \lambda_k(\Delta_{\Gamma^{g,\widehat{h}_-}}),$$

where $d_\infty^g(\nu,\nu_n)$ is as defined in (21).

Proof As discussed in Section 2.2,

$$\lambda_k(\Delta_{\Gamma^{c,h'}}) = \min_{L_k} \max_{u \in L_k \setminus \{0\}} \frac{b^{c,h'}(u)}{\|u\|^2_{c,\mathbf{m}}} \quad \text{for } c = 0, g \text{ and } h' > 0.$$

First, we compare degrees, which we explicitly decorate with a value of h. From Assumption 5.3 we deduce that

$$\begin{aligned} m_{g,\widehat{h}_-,i} &:= \frac{1}{n\widehat{h}_-^m} \sum_{j=1}^n \eta\left(\frac{d_g(x_i,x_j)}{\widehat{h}_-}\right) \leq \left(\frac{1}{1-\delta}\right)^m \frac{1}{nh^m} \sum_{j=1}^n \eta\left(\frac{d_0(x_i,x_j)}{h}\right) \\ &\leq (1+Cm\delta)\,\frac{1}{nh^m} \sum_{j=1}^n \eta\left(\frac{d_0(x_i,x_j)}{h}\right) = (1+Cm\delta)\, m_{0,h,i}, \end{aligned} \tag{24}$$

and similarly $m_{g,\widehat{h}_+,i} \geq (1-Cm\delta)\, m_{0,h,i}$, where both hold with probability $1-\varepsilon_n$. Using (García Trillos et al., 2019, Lemma 18) we get

$$\begin{aligned} \max_{i=1,\dots,n} |m_{g,\widehat{h}_-,i} - m_{g,\widehat{h}_+,i}| &\leq \max_{i=1,\dots,n} |m_{g,\widehat{h}_-,i} - \rho(x_i)| + \max_{i=1,\dots,n} |m_{g,\widehat{h}_+,i} - \rho(x_i)| \\ &\leq CL_1 h + C\beta \frac{d^g_\infty(\nu,\nu_n)}{h} + C\beta m K h^2. \end{aligned} \tag{25}$$

for $d^g_\infty(\nu,\nu_n)$ as defined in (21). Also, letting $i^\star := \text{argmin}_{i=1,\dots,n} m_{g,\widehat{h}_\pm,i}$,

$$\begin{aligned} \min_{i=1,\dots,n} m_{g,\widehat{h}_\pm,i} &= \rho(x_{i^\star}) + \left(m_{g,\widehat{h}_\pm,i^\star} - \rho(x_{i^\star})\right) \\ &\geq \frac{1}{\beta} - CL_1 h - C\beta \frac{d^g_\infty(\nu,\nu_n)}{h} - C\beta m K h^2. \end{aligned}$$

So,

$$\begin{aligned} m_{0,h,i} &\geq \frac{1}{1+Cm\delta} m_{g,\widehat{h}_-,i} \quad \text{from (24)} \\ &\geq (1-Cm\delta)\left(m_{g,\widehat{h}_+,i} - m_{g,\widehat{h}_+,i} + m_{g,\widehat{h}_-,i}\right) \\ &\geq (1-Cm\delta)\left(m_{g,\widehat{h}_+,i} - \max_{i=1,\dots,n} |m_{g,\widehat{h}_-,i} - m_{g,\widehat{h}_+,i}|\right) \\ &\geq (1-Cm\delta)\left(m_{g,\widehat{h}_+,i} - CL_1 h - C\beta \frac{d^g_\infty(\nu,\nu_n)}{h} - C\beta m K h^2\right) \quad \text{from (25)} \\ &\geq (1-Cm\delta)\left(m_{g,\widehat{h}_+,i} - \frac{m_{g,\widehat{h}_+,i}}{\min_{i=1,\dots,n} m_{g,\widehat{h}_\pm,i}} \left(CL_1 h + C\beta \frac{d^g_\infty(\nu,\nu_n)}{h} + C\beta m K h^2\right)\right) \\ &\geq \left(1 - \left(Cm\delta + C\beta L_1 h + C\beta^2 \frac{d^g_\infty(\nu,\nu_n)}{h} + C_{\beta,m,L_1} h^2 + C_{m,\beta,L_1}\delta h\right)\right) m_{g,\widehat{h}_+,i} \\ &:= (1-E)\, m_{g,\widehat{h}_+,i} \end{aligned}$$

and similarly, $m_{0,h,i} \leq (1+E) m_{g,\widehat{h}_-,i}$, where we have ignored certain higher order terms. We now compare norms.

$$\|u\|^2_{0,\mathbf{m},h} = \sum_{i=1}^n m_{0,h,i} u_i^2 \leq \sum_{i=1}^n (1+E)\, m_{g,\widehat{h}_-,i} u_i^2 = (1+E)\|u\|^2_{g,\mathbf{m},\widehat{h}_-}$$

and similarly $\|u\|^2_{0,\mathbf{m},h} \geq (1-E)\|u\|^2_{g,\mathbf{m},\widehat{h}_+}$. Now fix any subspace L_k of dimension k.

$$\max_{u\in L_k\setminus\{0\}} \frac{b^{0,h}(u)}{\|u\|^2_{0,\mathbf{m},h}} \leq (1+E)\max_{u\in L_k\setminus\{0\}} \frac{b^{0,h}(u)}{\|u\|^2_{g,\mathbf{m},\widehat{h}_+}} \overset{(a)}{\leq} (1+E)(1+Cm\delta)\max_{u\in L_k\setminus\{0\}} \frac{b^{g,\widehat{h}_+}(u)}{\|u\|^2_{g,\mathbf{m},\widehat{h}_+}},$$

where we have used Theorem 5.4 in (a). But L_k was arbitrary, and so

$$\begin{aligned}
\lambda_k(\Delta_{\Gamma^{2,h}}) &= \min_{L_k\,:\,\dim(L_k)=k}\ \max_{u\in L_k\setminus\{0\}} \frac{b^{0,h}(u)}{\|u\|^2_{0,\mathbf{m},h}} \\
&\leq (1+E)(1+Cm\delta)\min_{L_k\,:\,\dim(L_k)=k}\ \max_{u\in L_k\setminus\{0\}} \frac{b^{g,\widehat{h}_+}(u)}{\|u\|^2_{g,\mathbf{m},\widehat{h}_+}} \\
&= (1+E)(1+Cm\delta)\lambda_k(\Delta_{\Gamma^{2,\widehat{h}_+}})\,.
\end{aligned}$$

Simplifying $(1+E)(1+Cm\delta)$, ignoring higher order terms *e.g.* $\delta h^2, \delta d^g_\infty(\nu,\nu_n)$, and adjusting constants as necessary yields the stated upper bound. A similar argument yields the lower bound. ■

5.3 Discrete-to-Continuum Convergence of Eigenvalues

With d_0 as in Section 5.2 we now show the eigenvalues of $\Delta_{\Gamma^{d_0,h}}$ converge to those of Δ_2 provided Assumption 5.3 holds. Later, we specialize this to $d_0 = \tilde{\ell}^p_p$ and $d_g = \mu\mathcal{L}^p_p$.

Lemma 5.6 (cf. Lemma 13, part (ii) García Trillos et al. (2019)) *Suppose Assumption 5.3 holds. Then with probability $1-\varepsilon_n$, for all $f\in H^1(\mathcal{M})$,*

$$b^{d_0}_h(Pf) \leq \left(1 + C\beta L_1(1+\delta)h + Cm\frac{d^g_\infty(\nu,\nu_n)}{(1+\delta)h} + CmK_p(1+3\delta)h^2\right) D(f)\,.$$

where

$$D(f) := \int_{\mathcal{M}} g(\nabla f, \nabla f)\, dVol. \tag{26}$$

Proof We indicate how to modify the proof. Recalling σ_η in their notation is $\frac{1}{m+2}$:

$$b^{d_0}_h(Pf) \leq \frac{m+2}{h^{m+2}}\sum_i\sum_j \int_{U_i}\int_{U_j} \eta\left(\frac{d_0(x_i,x_j)}{h}\right)|f(y)-f(x)|^2\, d\nu(y)d\nu(x) \tag{27}$$

$$\leq \frac{m+2}{h^{m+2}}\sum_i\sum_j \int_{U_i}\int_{U_j} \eta\left(\frac{d_g(x_i,x_j)}{\widehat{h}^+}\right)|f(y)-f(x)|^2\, d\nu(y)d\nu(x) \tag{28}$$

$$\leq \frac{m+2}{h^{m+2}}\int_{\mathcal{M}}\int_{\mathcal{M}} \eta\left(\frac{(d_g(x,y)-2d^g_\infty(\nu,\nu_n))_+}{\widehat{h}^+}\right)|f(y)-f(x)|^2\, d\nu(y)d\nu(x) \tag{29}$$

$$= \frac{m+2}{h^{m+2}}\int_{\mathcal{M}}\int_{\mathcal{M}} \eta\left(\frac{d_g(x,y)}{\widehat{h}^+ + 2d^g_\infty(\nu,\nu_n)}\right)|f(y)-f(x)|^2\, d\nu(y)d\nu(x) \tag{30}$$

$$= \frac{m+2}{h^{m+2}} \tilde{E}_{\widehat{h}^+ + 2d^g_\infty(\nu,\nu_n)},$$

where (27) is shown in the proof of Lemma 13, part (ii) García Trillos et al. (2019), (28) follows from Assumption 5.3, (29) follows as the U_i have diameter at most $2d^g_\infty(\nu,\nu_n)$, and (30) follows from the identity $\eta\left(\frac{(t-s)_+}{r}\right) = \eta\left(\frac{t}{r+s}\right)$ for $r,t,s>0$, see (García Trillos et al., 2019, Lemma 10). From (García Trillos et al., 2019, Lemma 5):

$$\tilde{E}_{\widehat{h}^+ + 2d^g_\infty(\nu,\nu_n)}(f) \leq \left(1 + L_1\beta(\widehat{h}^+ + 2d^g_\infty(\nu,\nu_n))\right)\left(1 + CmK\left(\widehat{h}^+ + 2d^g_\infty(\nu,\nu_n)\right)^2\right) \left(\widehat{h}^+ + 2d^g_\infty(\nu,\nu_n)\right)^{m+2} \frac{D(f)}{m+2}.$$

Simplifying as in García Trillos et al. (2019) and using $d^g_\infty(\nu,\nu_n) \ll h < \widehat{h}^+ := h(1+\delta)$ yields the stated bound. ■

Proof [Proof of Theorem 3.3] By Corollary 5.1, as long as

$$h \geq 4\mu\beta^{\frac{(p-1)}{m}}(n\beta/2)^{-\frac{1}{m}(\frac{1}{3}-\epsilon)}, \tag{31}$$

Assumption 5.3 holds with $d_g = \mu\mathcal{L}^p_p$ and $d_0 = \tilde{\ell}^p_p$ where

$$\delta = 2\mu^{-1}C_1 h + 4\mu^{-1}C_2 h^2. \tag{32}$$

From Theorem 5.5,

$$\lambda_k(\Gamma^{\tilde{\ell},h}) - \lambda_k(\Gamma^{\mathcal{L},\widehat{h}_-}) \geq -E\lambda_k(\Gamma^{\mathcal{L},\widehat{h}_-}) \text{ and } \lambda_k(\Gamma^{\tilde{\ell},h}) - \lambda_k(\Gamma^{\mathcal{L},\widehat{h}_+}) \leq E\lambda_k(\Gamma^{\mathcal{L},\widehat{h}_+}) \tag{33}$$

where

$$E = Cm\delta + C\beta L_1 h + C\beta^2\frac{d^g_\infty(\nu,\nu_n)}{h} + C_{\beta,m,L_1}h^2 + C_{m,\beta,L_1}\delta h. \tag{34}$$

From (García Trillos et al., 2019, Theorem 4, part 1) we have

$$\lambda_k(\Gamma^{\mathcal{L},\widehat{h}_+}) - \lambda_k(\Delta_{2,p}) \leq \delta_1\lambda_k(\Delta_{2,p}), \tag{35}$$

where

$$\delta_1 = \tilde{C}\left[L_1\widehat{h}_+ + \frac{d^g_\infty(\nu,\nu_n)}{\widehat{h}_+} + \sqrt{\lambda_k(\Delta_{2,p})}d^g_\infty(\nu,\nu_n) + K_p\widehat{h}^2_+ + \|\mathbf{m}-\rho_p\|_\infty\right] \tag{36}$$

and $\tilde{C}$ depends only on m,β,p and L_1. We do not retain the term proportional to h^2/R^2 in (García Trillos et al., 2019, eq. (1.16)), as $\Gamma^{\mathcal{L},\widehat{h}_+}$ is constructed using geodesic distances between points, not Euclidean distances. More precisely, this discrepancy arises from using Lemma 5.6 in place of (García Trillos et al., 2019, Lemma 13). Similarly, from (García Trillos et al., 2019, Theorem 4, part 2) we have

$$\lambda_k(\Gamma^{\mathcal{L},\widehat{h}_-}) - \lambda_k(\Delta_{2,p}) \geq -\delta_2\lambda_k(\Delta_{2,p}), \tag{37}$$

where

$$\delta_2 = \tilde{C}\left[L_1\widehat{h}_- + \frac{d^g_\infty(\nu,\nu_n)}{\widehat{h}_-} + \sqrt{\lambda_k(\Delta_{2,p})}d^g_\infty(\nu,\nu_n) + K_p\widehat{h}_-^2 + \|\mathbf{m}-\rho_p\|_\infty\right]. \tag{38}$$

Combining (35) and (33) yields:

$$\begin{aligned}\lambda_k(\Gamma^{\tilde{\ell},h}) - \lambda_k(\Delta_{2,p}) &= \lambda_k(\Gamma^{\tilde{\ell},h}) - \lambda_k(\Gamma^{\mathcal{L},\widehat{h}_+}) + \lambda_k(\Gamma^{\mathcal{L},\widehat{h}_+}) - \lambda_k(\Delta_{2,p}) \\ &\leq E\lambda_k(\Gamma^{\mathcal{L},\widehat{h}_+}) + \delta_1\lambda_k(\Delta_{2,p}) \\ &\leq E(1+\delta_1)\lambda_k(\Delta_{2,p}) + \delta_1\lambda_k(\Delta_{2,p}) \\ &= (E+\delta_1+E\delta_1)\,(\Delta_{2,p}).\end{aligned} \tag{39}$$

Similarly, combining (37) and (33)

$$\lambda_k(\Gamma^{\tilde{\ell},h}) - \lambda_k(\Delta_{2,p}) \geq -\,(E+\delta_2+E\delta_2)\,\lambda_k(\Delta_{2,p})\,. \tag{40}$$

Using (32), (34), (36) and (38) the error terms in (39) and (40) can be expressed as

$$E+\delta_1+E\delta_1 = C_{\beta,p,L_1,m}h + C\beta^2\frac{d^g_\infty(\nu,\nu_n)}{h} + C_{\beta,m,L_1,L_2,p,K_p,\mathcal{R}}h^2 + O(h^3 + d^g_\infty(\nu,\nu_n)),$$

$$E+\delta_2+E\delta_2 = C_{\beta,p,L_1,m}h + C\beta^2\frac{d^g_\infty(\nu,\nu_n)}{h} + C_{\beta,m,L_1,L_2,p,K_p,\mathcal{R}}h^2 + O(h^3 + d^g_\infty(\nu,\nu_n)).$$

From (García Trillos et al., 2019, Theorem 2) we have

$$d^g_\infty(\nu,\nu_n) = O\left(\frac{\log(n)^{k_m}}{n^{1/m}}\right)$$

w.h.p, where $k_m = 3/4$ if $m = 2$ and $k_m = 1/m$ if $m \geq 3$. In García Trillos et al. (2019) contributions to the approximation error (δ_1 and δ_2 in our notation) from h and $d^g_\infty(\nu,\nu_n)$ are optimally balanced by setting $h = \sqrt{d^g_\infty(\nu,\nu_n)} = \mathcal{O}\left(\log(n)^{k_m}n^{-\frac{1}{2m}}\right)$. We cannot do so here as h must satisfy the lower bound of (31). Instead, we take $h = 4\mu\beta^{\frac{(p-1)}{m}}(n\beta)^{-\frac{1}{m}(\frac{1}{3}-\epsilon)} = O(n^{-\frac{1}{3m}+\frac{\epsilon}{m}})$. Noting that K_p may be bounded by an expression involving K, L_1, L, m, and p (see Appendix B) we may write $C_{\beta,m,L_1,L_2,p,K,\mathcal{R}}$ instead of $C_{\beta,m,L_1,L_2,p,K_p,\mathcal{R}}$. We now identify the leading and second-order terms in (39) and (40). Observe that

$$h^2 = O(n^{-\frac{2}{3m}+\frac{2\epsilon}{m}}) \gg d^g_\infty(\nu,\nu_n) \text{ and } h^2 \gg O(n^{-\frac{2}{3m}-\frac{\epsilon}{m}}) = \frac{d^g_\infty(\nu,\nu_n)}{h}\,.$$

retaining only terms proportional to h and h^2 yields the stated error bounds. ■

Remark 5.7 *The dependence of h on n is determined by Corollary 5.1, which is in turn determined by Proposition 4.4. One might enquire as to whether a better choice of q in Proposition 4.4 could result in a better error bound in Theorem 3.3. In fact it is straightforward (but tedious) to trace the dependency on q through Theorem 4.3 and Corollary 5.1 to the proof of Theorem 3.3, whence a straightforward computation reveals $q = 2$ is optimal.*

Remark 5.8 *Although we state the above theorem in terms of (discrete and continuous) Fermat distance, with superficial modifications it can be applied to any Riemannian distance $d_g(\cdot,\cdot)$ and any discrete approximation $d_0(\cdot,\cdot)$ satisfying Assumption 5.3.*

5.4 Discrete-to-Continuum Convergence of Eigenvectors

In this section we relate eigenvectors of the Fermat graph Laplacian $\Delta_{\Gamma^{\tilde{\ell},h}}$ to the eigenfunctions of the operator $\Delta_{2,p}$ by proving Theorem 3.5. The proof of this result relies on the relationship between Dirichlet energies that we established in previous sections as well as on the consistency of eigenvalues from Theorem 3.3.

Throughout this section let $\psi_1^n, \dots, \psi_n^n$ denote the eigenvectors of $\Delta_{\Gamma^{\tilde{\ell},h}}$ normalized according to $\|\cdot\|_{\mathbf{m}}^2$. Let f be an eigenfunction of $\Delta_{2,p}$ with eigenvalue $\lambda_2 := \lambda_2(\Delta_{2,p})$, normalized according to $\|\cdot\|^2_{L^2(\mathcal{M},\rho_p^2 \mathrm{dVol}_p)}$. Recall that, by assumption, this eigenfunction is unique up to sign. We shall use C decorated with subscripts to denote a constant depending on the indicated manifold or density related properties, but not on n.

Lemma 5.9 *Define $\mathbf{1}_{\mathcal{X}} : \mathcal{X} \to \mathbb{R}$ as $\mathbf{1}_{\mathcal{X}}(x_i) = 1$ for all $x_i \in \mathcal{X}$. Then*

$$\psi_1^n = \|\mathbf{1}_{\mathcal{X}}\|_{\boldsymbol{m}}^{-1} \mathbf{1}_{\mathcal{X}} \quad w.l.o.g \tag{41}$$

$$\langle Pf, \mathbf{1}_{\mathcal{X}} \rangle_{\boldsymbol{m}} \leq C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} + C_{m,\beta,p,L_1,L_2,K} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}} \tag{42}$$

$$\frac{1}{2}\beta^{-2p} \leq \|\mathbf{1}_{\mathcal{X}}\|_{\boldsymbol{m}}^2 \leq \frac{1}{2}\beta^{2p}. \tag{43}$$

Proof As is well-known, the smallest eigenvalue of $\Delta_{\Gamma^{\tilde{\ell},h}}$ is 0, and the associated eigenvector is a multiple of $\mathbf{1}_{\mathcal{X}}$. Up to sign, $\|\mathbf{1}_{\mathcal{X}}\|_{\mathbf{m}}^{-1}\mathbf{1}_{\mathcal{X}}$ is the unique multiple of $\mathbf{1}_{\mathcal{X}}$ that is normalized according to $\|\cdot\|_{\mathbf{m}}$, proving (41).

Using the definition of the operator P and recalling that $U_i = T^{-1}(x_i)$,

$$\begin{aligned}
\langle Pf, \mathbf{1}_{\mathcal{X}} \rangle_{\mathbf{m}} &= \frac{1}{n}\sum_{i=1}^n Pf(x_i) m_i \\
&= \sum_{i=1}^n m_i \int_{U_i} f(x)\rho_p(x)\mathrm{dVol}_p(x) \\
&= \sum_{i=1}^n \int_{U_i} f(x)\rho_p^2(x)\mathrm{dVol}_p(x) + \sum_{i=1}^n \int_{U_i} f(x)(m_i - \rho_p(x))\rho_p(x)\mathrm{dVol}_p(x) \\
&=: \sum_{i=1}^n \int_{U_i} f(x)\rho_p^2(x)\mathrm{dVol}_p(x) + c_n \\
&= c_n,
\end{aligned}$$

where we have used the fact that f is orthogonal to $\mathbf{1}_{\mathcal{M}}$—the first eigenfunction of $\Delta_{2,p}$—w.r.t. the inner product $\langle \cdot, \cdot \rangle_{L^2(\mathcal{M},\rho_p^2 \mathrm{dVol}_p)}$:

$$\sum_{i=1}^n \int_{U_i} f(x)\rho_p^2(x)\mathrm{dVol}_p(x) = \int_{\mathcal{M}} f(x)\rho_p^2(x)\mathrm{dVol}_p(x) = \langle f, \mathbf{1}_{\mathcal{M}} \rangle_{L^2(\mathcal{M},\rho_p^2 \mathrm{dVol}_p)} = 0.$$

In turn, the error term c_n can be bounded as

$$|c_n| \leq \|\mathbf{m} - \rho_p\|_\infty \int_{\mathcal{M}} |f(x)|\rho_p(x)\mathrm{dVol}_p \leq \beta^p \|f\|_{L^2(\mathcal{M},\rho_p^2 \mathrm{dVol}_p)} \|\mathbf{m} - \rho_p\|_\infty$$

$$= \beta^p \|\mathbf{m} - \rho_p\|_\infty \leq C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} + C_{m,\beta,p,L_1,L_2,K} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}} \tag{44}$$

where we have used Lemma 5.2, the value of h given in the assumptions, simplified as in the proof of Theorem 3.3, and ignored higher order terms (*i.e.* those $\ll n^{-\frac{2}{3m}+\frac{2\epsilon}{m}}$ such as terms proportional to $d^g_\infty(\nu,\nu_n)$). This proves (42). A similar computation reveals that

$$\|\mathbf{1}_\mathcal{X}\|^2_\mathbf{m} = \frac{1}{n}\sum_{i=1}^n m_i \geq \int_\mathcal{M} \rho_p^2 \mathrm{dVol}_p - \|\mathbf{m} - \rho_p\|_\infty \overset{(a)}{\geq} \frac{1}{2}\int_\mathcal{M} \rho_p^2 \mathrm{dVol}_p \geq \frac{1}{2}\beta^{-2p},$$

proving the lower bound of (43). The upper bound follows similarly. ■

Lemma 5.10 *Let $v : \mathcal{X} \to \mathbb{R}$ be the function $v := Pf - \langle Pf, \psi_1^n\rangle_\mathbf{m}\psi_1^n$ and let $\hat{v} := v/\|v\|_\mathbf{m}$ denote its normalization. Let $P_2 : \mathbb{R}^\mathcal{X} \to \mathbb{R}^\mathcal{X}$ denote the projection onto the eigenspace associated to $\lambda_2(\Gamma^{\tilde{\ell},h})$. Then:*

$$\|\hat{v} - P_2\hat{v}\|^2_\mathbf{m} \leq \frac{b_h^{\tilde{\ell}}(\hat{v}) - \lambda_2(\Gamma^{\tilde{\ell},h})}{\lambda_3(\Gamma^{\tilde{\ell},h}) - \lambda_2(\Gamma^{\tilde{\ell},h})} \tag{45}$$

$$\|v\|_\mathbf{m} \geq 1 - C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} - C_{m,\beta,p,L_1,L_2,K} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}} \tag{46}$$

Proof By construction, $\langle v, \psi_1^n\rangle_\mathbf{m} = 0$. So, expanding in the eigenbasis yields

$$\begin{aligned}\lambda_3(\Gamma^{\tilde{\ell},h})\|\hat{v} - P_2\hat{v}\|^2_\mathbf{m} &= \lambda_3(\Gamma^{\tilde{\ell},h})\sum_{i=3}^n (\langle \hat{v}, \psi_i^n\rangle_\mathbf{m})^2 \leq \sum_{i=3}^n \lambda_i(\Gamma^{\tilde{\ell},h})(\langle \hat{v}, \psi_i^n\rangle_\mathbf{m})^2 \\ &= b_h^{\tilde{\ell}}(\hat{v}) - \lambda_2(\Gamma^{\tilde{\ell},h})(\langle \hat{v}, \psi_2^n\rangle_\mathbf{m})^2.\end{aligned}$$

Because

$$\|\hat{v} - P_2\hat{v}\|^2_\mathbf{m} = \|\hat{v} - \langle \hat{v}, \psi_2^n\rangle_\mathbf{m}\psi_2^n\|^2_\mathbf{m} = 1 - (\langle \hat{v}, \psi_2^n\rangle_\mathbf{m})^2$$

it follows that

$$(\lambda_3(\Gamma^{\tilde{\ell},h}) - \lambda_2(\Gamma^{\tilde{\ell},h}))\|\hat{v} - P_2\hat{v}\|^2_\mathbf{m} + \lambda_2(\Gamma^{\tilde{\ell},h}) \leq b_h^{\tilde{\ell}}(\hat{v}),$$

and rearranging yields (45). As for (46),

$$\begin{aligned}\|v\|_\mathbf{m} &= \left\|Pf - \|\mathbf{1}_\mathcal{X}\|^{-1}_\mathbf{m}\langle Pf, \mathbf{1}_\mathcal{X}\rangle_\mathbf{m}\psi_1^n\right\|_\mathbf{m} \\ &\geq \|Pf\|_\mathbf{m} - |\langle Pf, \mathbf{1}_\mathcal{X}\rangle_\mathbf{m}|\,\|\mathbf{1}_\mathcal{X}\|^{-1}_\mathbf{m}\end{aligned}$$

From the proof of Lemma 13 in García Trillos et al. (2019) we know that

$$\begin{aligned}\left|\|Pf\|^2_\mathbf{m} - 1\right| &= \left|\|Pf\|^2_\mathbf{m} - \|f\|^2_{L^2(\mathcal{M},\rho_p^2\mathrm{dVol}_p)}\right| \\ &\leq C(\|\mathbf{m} - \rho_p\|_\infty + d^g_\infty(\nu,\nu_n)) + Cd^g_\infty(\nu,\nu_n)\lambda_2^{1/2}.\end{aligned} \tag{47}$$

Appealing to Lemma 5.2 and ignoring higher order terms, as was done in the proof of Lemma 5.9, we obtain

$$\|Pf\|^2_\mathbf{m} \geq 1 - C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} - C_{m,\beta,p,L_1,L_2,K} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}}. \tag{48}$$

Consequently we have

$$\|v\|_{\mathbf{m}} \geq 1 - C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} - C_{m,\beta,p,L_1,L_2,K} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}},$$

where the constants C_{β,p,m,L_1} and $C_{m,\beta,p,L_1,L_2,K}$ have changed to incorporate the bounds of (42) and (43) of Lemma 5.9. ■

Proof [Proof of Theorem 3.5] From Lemma 5.6, and the fact that $D(f) = \lambda_2$ as f is an eigenvector, we know that

$$b_h^{\tilde{\ell}}(Pf) \leq \left(1 + C\beta L_1 h + Cm\frac{d_\infty^g(\nu,\nu_n)}{h} + CmK_p h^2\right)\lambda_2. \tag{49}$$

where we have assumed, as in Corollary 5.1, that h is sufficiently small so that $\delta \leq \frac{1}{2}$ and adjusted the constants C accordingly. Let v be as in Lemma 5.10. By the invariance of $b_h^{\tilde{\ell}}(\cdot)$ to shifts, $b_h^{\tilde{\ell}}(v) = b_h^{\tilde{\ell}}(Pf)$. Appealing to (49) it follows that

$$\begin{aligned} b_h^{\tilde{\ell}}(v) = b_h^{\tilde{\ell}}(Pf) &\leq \left(1 + C\beta L_1 h + Cm\frac{d_\infty^g(\nu,\nu_n)}{h} + CmK_p h^2\right)\lambda_2 &(50)\\ &\leq \lambda_2 + C_{\beta,L_1}\lambda_2 n^{-\frac{1}{3m}+\frac{\epsilon}{m}} + C_{m,\beta,p,L_1,L_2,K}\lambda_2 n^{-\frac{2}{3m}+\frac{2\epsilon}{m}} &(51)\end{aligned}$$

where, as before, we have simplified by substituting in for h and ignoring higher order terms. Now observe that

$$\begin{aligned} \|\hat{v} - P_2\hat{v}\|_{\mathbf{m}}^2 &= \frac{1}{\|v\|_{\mathbf{m}}^2}\|v - P_2 v\|_{\mathbf{m}}^2 \\ &= \frac{1}{\|v\|_{\mathbf{m}}^2}\left\|Pf - P_2 Pf - \langle Pf, \psi_1^n\rangle_{\mathbf{m}}\psi_1^n\right\|_{\mathbf{m}}^2 \quad \text{as } P_2\psi_1^n = 0 \\ &\geq \frac{1}{2\|v\|_{\mathbf{m}}^2}\left(\|Pf - P_2 Pf\|_{\mathbf{m}}^2 - 2\left|\langle Pf, \mathbf{1}_{\mathcal{X}}\rangle\right|^2 \|\mathbf{1}_{\mathcal{X}}\|_{\mathbf{m}^{-2}}\right) \end{aligned}$$

and consequently

$$\begin{aligned} \|Pf - P_2 Pf\|^2 &\leq 2\|v\|_{\mathbf{m}}^2\|\hat{v} - P_2\hat{v}\|_{\mathbf{m}}^2 + 2\left|\langle Pf, \mathbf{1}_{\mathcal{X}}\rangle\right|^2\|\mathbf{1}_{\mathcal{X}}\|_{\mathbf{m}^{-2}} \\ &\overset{(a)}{\leq} \frac{2\|v\|_{\mathbf{m}}^2 b_h^{\tilde{\ell}}(\hat{v}) - 2\|v\|_{\mathbf{m}}^2\lambda_2(\Gamma^{\tilde{\ell},h})}{\lambda_3(\Gamma^{\tilde{\ell},h}) - \lambda_2(\Gamma^{\tilde{\ell},h})} + 2\left|\langle Pf, \mathbf{1}_{\mathcal{X}}\rangle\right|^2\|\mathbf{1}_{\mathcal{X}}\|_{\mathbf{m}^{-2}} \\ &\overset{(b)}{=} \frac{2b_h^{\tilde{\ell}}(v) - 2\|v\|_{\mathbf{m}}^2\lambda_2(\Gamma^{\tilde{\ell},h})}{\lambda_3(\Gamma^{\tilde{\ell},h}) - \lambda_2(\Gamma^{\tilde{\ell},h})} + 2\left|\langle Pf, \mathbf{1}_{\mathcal{X}}\rangle\right|^2\|\mathbf{1}_{\mathcal{X}}\|_{\mathbf{m}^{-2}} \end{aligned}$$

where (a) follows from Lemma 5.10 while (b) follows from the 2-homogeneity of $b_h^{\tilde{\ell}}(\cdot)$:

$$\|v\|_{\mathbf{m}}^2 b_h^{\tilde{\ell}}(\hat{v}) = b_h^{\tilde{\ell}}(\|v\|_{\mathbf{m}}\hat{v}) = b_h^{\tilde{\ell}}(v).$$

To conclude we bound terms by powers of n, ignoring higher order terms as we have done before. From the error estimates of Theorem 3.3

$$\lambda_2(\Gamma^{\tilde{\ell},h}) \leq \lambda_2\left(1 + C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} + C_{\beta,p,m,L_1,L_2,K_p,\mathcal{R}} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}}\right).$$

Bounds on $b_h^{\tilde{\ell}}(v)$, $\|v\|_{\mathbf{m}}^2$, $|\langle Pf, \mathbf{1}_{\mathcal{X}}\rangle|$, and $\|\mathbf{1}_{\mathcal{X}}\|_{\mathbf{m}^{-2}}$ follow from (51), Lemma 5.10 and Lemma 5.9 respectively, while bounds on $\left(\lambda_3(\Gamma^{\tilde{\ell},h}) - \lambda_2(\Gamma^{\tilde{\ell},h})\right)^{-1}$ follow from Lemma 5.11. Combining these bounds and ignoring higher order terms,

$$\begin{aligned}\|Pf - P_2(Pf)\|_{\mathbf{m}}^2 \leq & \frac{1}{\lambda_3 - \lambda_2}\left[C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} + C_{\beta,p,m,L_1,L_2,K,\mathcal{R}} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}}\right]\lambda_2 \\ & + C_{\beta,p,m,L_1} n^{-\frac{2}{3m}+\frac{2\epsilon}{3m}}.\end{aligned}$$

For n large ψ_2^n is simple, so $P_2(Pf) = \|P_2(Pf)\|_{\mathbf{m}}\psi_2^n$ and thus

$$\begin{aligned}\|Pf - \psi_2^n\|_{\mathbf{m}}^2 &\leq 2\|Pf - P_2(Pf)\|_{\mathbf{m}}^2 + 2\|P_2(Pf) - \psi_2^n\|_{\mathbf{m}}^2 \\ &= 2\|Pf - P_2(Pf)\|_{\mathbf{m}}^2 + 2\left(1 - \|P_2(Pf)\|_{\mathbf{m}}\right)^2 \\ &\leq 2\|Pf - P_2(Pf)\|_{\mathbf{m}}^2 + 2\left(1 - \|(Pf)\|_{\mathbf{m}} + \|Pf - P_2(Pf)\|_{\mathbf{m}}\right)^2 \\ &\leq 6\|Pf - P_2(Pf)\|_{\mathbf{m}}^2 + 4\left(1 - \|(Pf)\|_{\mathbf{m}}\right)^2 \\ &\leq 6\|Pf - P_2(Pf)\|_{\mathbf{m}}^2 + 4\left|1 - \|(Pf)\|_{\mathbf{m}}^2\right|\end{aligned}$$

where in the final line we have used the following easily verifiable inequality:

$$(a-b)^2 \leq |a^2 - b^2| \quad \text{for } a, b \geq 0. \tag{52}$$

Combining the above with (48) one obtains

$$\begin{aligned}\|Pf - \psi_2^n\|_{\mathbf{m}}^2 \leq & \frac{1}{\lambda_3 - \lambda_2}\left[C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} + C_{\beta,p,m,L_1,L_2,K,\mathcal{R}} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}}\right]\lambda_2 \\ & + C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} + C_{m,\beta,p,L_1,L_2,K} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}}.\end{aligned}$$

■

Lemma 5.11 *With notation as in the proof of Theorem 3.5,*

$$\begin{aligned}\frac{1}{\lambda_3(\Gamma^{\tilde{\ell},h}) - \lambda_2(\Gamma^{\tilde{\ell},h})} \leq & \frac{1}{\lambda_3 - \lambda_2} - \frac{\lambda_3}{(\lambda_3 - \lambda_2)^2}\left[C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} + C_{\beta,p,m,L_1,L_2,K,\mathcal{R}} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}}\right] \\ & - \frac{\lambda_2}{(\lambda_3 - \lambda_2)^2}\left[C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} + C_{\beta,p,m,L_1,L_2,K_p,\mathcal{R}} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}}\right].\end{aligned}$$

Proof Using the error estimates of Theorem 3.3 again and letting $\lambda_3 := \lambda_3(\Delta_{2,p})$,

$$\begin{aligned}\frac{1}{\lambda_3(\Gamma^{\tilde{\ell},h}) - \lambda_2(\Gamma^{\tilde{\ell},h})} &= \frac{1}{(\lambda_3(\Gamma^{\tilde{\ell},h}) - \lambda_3) + (\lambda_3 - \lambda_2) + (\lambda_2 - \lambda_2(\Gamma^{\tilde{\ell},h}))} \\ &\leq \frac{1}{\lambda_3 - \lambda_2} - \frac{\lambda_3}{(\lambda_3 - \lambda_2)^2}\left[C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} + C_{\beta,p,m,L_1,L_2,K,\mathcal{R}} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}}\right] \\ &\quad - \frac{\lambda_2}{(\lambda_3 - \lambda_2)^2}\left[C_{\beta,p,m,L_1} n^{-\frac{1}{3m}+\frac{\epsilon}{m}} + C_{\beta,p,m,L_1,L_2,K_p,\mathcal{R}} n^{-\frac{2}{3m}+\frac{2\epsilon}{m}}\right].\end{aligned}$$

■

Remark 5.12 *It is possible to consider alternative metrics to capture the convergence of graph Laplacian eigenvectors toward their continuum counterparts;* e.g., *see García Trillos et al. (2019). We also want to remark that extending the convergence result from Theorem 3.5 to higher eigenfunctions is straightforward but cumbersome as one needs to introduce additional notation and operators. The rationale of the proof is however the same and relies on the strong convexity of the discrete Raleigh quotient when restricted to eigenspaces. In this interpretation, the strong convexity constants are determined by the spectral gaps between consecutive eigenvalues. We refer the reader to García Trillos et al. (2019) for some related discussion.*

Remark 5.13 *It is interesting to observe that the proof of convergence of eigenvectors presented in Calder and García Trillos (2022), which ultimately produces faster rates of convergence than the ones based on energy considerations (*i.e., *the approach used in García Trillos et al. (2019); Burago et al. (2015) as well as in here) cannot be used in our setting, since the approach considered in Calder and García Trillos (2022) relies on pointwise consistency of graph Laplacians. Pointwise consistency results for Fermat-based graph Laplacians do not follow from similar considerations as in other works in the literature given that the discrete Fermat distances are random. Obtaining pointwise consistency results for Fermat-based graph Laplacians is thus left as an interesting open problem.*

5.5 Extensions and Discussion

It is possible to adapt the proofs of our theorems in Sections 5.2- 5.4 to obtain similar spectral convergence results (with rates) for more general normalizations of graph Laplacians such as the ones discussed in Section 2.2.1. Let the base weights $\tilde{W_{ik}}$ be defined according to $\tilde{W}_{ik} = w_{i,k}^{\tilde{\ell},h}$ as in (5) for $d_0 = \tilde{\ell}_p^p$ the discrete p-Fermat distance. We choose $r = 0, j = q = 1 - s/2$ in the construction from Section 2.2.1 and denote by $L_{p,s}$ the resulting normalized graph Laplacian. It is straightforward to see that $L_{p,s}$ is self-adjoint with respect to the inner product $\langle \cdot, \cdot \rangle_D$, where the matrix D is the degree matrix for the weights $W_{i,k} = \tilde{W}_{ik}/(\tilde{D}_i^{1-\frac{s}{2}} \tilde{D}_k^{1-\frac{s}{2}})$.

The (properly scaled) Dirichlet energy associated to the operator $L_{p,s}$ takes the form

$$\frac{2(m+2)}{n^2 h^{m+2}} \sum_{i,k} \frac{\eta(\frac{\tilde{\ell}_p^p(x_i,x_k)}{h})}{\tilde{D}_i^{1-\frac{s}{2}} \tilde{D}_k^{1-\frac{s}{2}}} \left(u(x_i) - u(x_k)\right)^2 .$$

Note that the only difference with the Dirichlet energy $b_h^{d_0}(u,u)$ from (7) is the extra normalization terms $\tilde{D}_i^{1-\frac{s}{2}}$, where

$$\tilde{D}_i := \frac{1}{nh^m} \sum_{k=1}^{n} \eta\left(\frac{\tilde{\ell}_p^p(x_i,x_k)}{h}\right).$$

Using Lemma (5.2), we can reduce our analysis to studying the Dirichlet energy

$$\frac{2(m+2)}{n^2 h^{m+2}} \sum_{i,k} \frac{\eta(\frac{\tilde{\ell}_p^p(x_i,x_k)}{h})}{\rho^{p(1-\frac{s}{2})}(x_i) \rho^{p(1-\frac{s}{2})}(x_k)} \left(u(x_i) - u(x_k)\right)^2 ,$$

which, in turn, can be related to the continuum Dirichlet energy

$$\int_{\mathcal{M}} |\nabla_p u|^2 \rho_p^s(x) \mathrm{dVol}_p(x),$$

after following very similar steps as in the proofs of our results in Sections 5.2 and 5.3.

The operators $L_{p,s}$ described above can thus be shown to converge spectrally toward the family of s-weighted operators (see Proposition 2.5)):

$$\Delta_{s,p} u = -\frac{1}{\rho_p^s} \mathrm{div}_p \left(\rho_p^s \nabla_p u \right).$$

See also the discussion in in Remark 2.8 and Section 6.

5.5.1 Dynamic Perspective and Comparison with Diffusion Maps

As we discuss below, every graph Laplacian operator $L_{p,s}$ induces a family of transformations of the original data set $\{x_1, \dots, x_n\}$ (which we recall is embedded in $\mathbb{R}^D$) into $\mathbb{R}^n$. The idea of using these transformations is to help "disentangle" the original data set $\mathcal{X}$ by finding a representation for it that more clearly reveals its intrinsic geometric structure. These transformations are in many instances effective preprocessing steps for tasks such as clustering or dimensionality reduction.

To introduce the family of transformations associated to $L_{p,s}$, let us denote by $Q_{s,p} := -L_{p,s}$ the transition rate matrix associated to the Laplacian $L_{p,s}$. Following Section 2 in Craig et al. (2022), we consider the family of evolution equations:

$$\begin{cases} \partial_t u_t(y) = \sum_{x \in \mathcal{X}} u_t(x) Q_{s,p}(x,y), \quad t > 0 \\ u_0 = u^0, \end{cases} \tag{53}$$

where each u_t is interpreted as a function $u_t : \mathcal{X} \to \mathbb{R}$; in the above, we use the notation $Q_{s,p}(x_i, x_j)$ to denote the ij entry of $Q_{s,p}$. Fixing a time horizon $T > 0$, for an arbitrary data point x_i we set $u^0 := \delta_{x_i}$ and define the transformation

$$x_i \mapsto u_T^{x_i} \in L^2(\{x_1, \dots, x_n\}) \cong \mathbb{R}^n, \tag{54}$$

where we have used the notation u^{x_i} to denote the solution to equation (53) at time T when its initial condition is δ_{x_i}, the function over $\mathcal{X}$ which is one at x_i and 0 everywhere else.

In what follows we discuss the similarities and differences between the family of transformations induced by the operators $L_{p,s}$ indexed by s for different values of p and the family of diffusion maps from Coifman and Lafon (2006). To compare these families, we use the connection between discrete and continuum operators that we developed in the previous sections and draw conclusions from the comparison between the continuum analogues of each of these families.

First, as in Section 2 in Craig et al. (2022), we can argue that the (discrete) evolution equation (53) has the following continuum counterpart:

$$\partial_t f_{t,p} = \mathrm{div}_p \left(\rho_p^{2s} \nabla_p \left(\frac{f_{t,p}}{\rho_p^{2s}} \right) \right),$$

where the function $f_{t,p}$ must be interpreted as a density function with respect to the volume form dVol_p. In terms of density functions and operators in the base geometry $(\mathcal{M}, g)$, the above equation can be rewritten as:

$$\partial_t f_t = \mathrm{div}(\rho^\alpha \nabla f_t) - ((2s-1)p+1)\mathrm{div}(f_t \rho^{\alpha-1}\nabla\rho) \tag{55}$$

where the function $f_t = f_{t,p}\rho^{1-p}$ is interpreted as a density function with respect to the base volume form dVol. If $\alpha > 0$, i.e. $p > 1$, the above can be written as

$$\partial_t f_t = \mathrm{div}(\rho^\alpha \nabla f_t) - \frac{((2s-1)p+1)}{\alpha}\mathrm{div}(f_t \nabla \rho^\alpha),$$

whereas for $\alpha = 0$ (which happens when $p = 1$) it can be written as

$$\partial_t f_t = \mathrm{div}(\nabla f_t) - 2s\mathrm{div}(f_t \nabla \log(\rho)).$$

Notice that regardless of the value of $\alpha \geq 0$ chosen, the second term on the right hand side of (55) is a "mean shift term", *i.e.,* it is a term that pushes the distribution f_t towards the local maxima of the function ρ (which are the same as the local maxima of $\log(\rho)$ or ρ^α for $\alpha > 0$). The first term on (55), on the other hand, is a non-homogeneous diffusion term with a diffusion coefficient that is larger at points with larger values of ρ. The larger the value of α, the more dramatic the difference between the diffusion rate at points with large values of ρ and at points with small values of ρ. Notice that α gets modulated by the value of p only, and in particular s does not play any role in determining it. As we will see below, being able to tune this effect in the diffusion term in (55) is one of the main differences between the family of maps represented by (55) and the diffusion maps of Coifman and Lafon (2006).

Let us recall that, as discussed in Section 4.2 in Craig et al. (2022), diffusion maps are connected to the following family of evolution equations:

$$\partial_t f_t = \mathrm{div}_{\mathcal{M}}\left(\rho^{2(1-a)}\nabla_{\mathcal{M}}\left(\frac{f_t}{\rho^{2(1-a)}}\right)\right) = \Delta_{\mathcal{M}} f_t - 2(1-a)\mathrm{div}_{\mathcal{M}}(f_t \nabla_{\mathcal{M}}\log(\rho)), \tag{56}$$

for $a \in (-\infty, 1]$. We see that (55) and (56) coincide when $s = 1 - a$ and $p = 1$. This is expected, since the distance function $\mathcal{L}_p$ is precisely the original metric on $\mathcal{M}$ when $p = 1$. In general, however, with the family of Fermat based Laplacians we have one extra degree of freedom that can be used to accelerate the diffusion in regions where the density ρ is larger, as was mentioned earlier. Numerical examples illustrating this effect are presented in Section 6. In settings like the desert horizon image from Figure 1, considering larger values of α in the maps (55) should be beneficial, as in that case the map (54) is expected to induce agglomeration of points at each of the clusters more quickly while still inducing the mean-shift force pushing clusters apart. However, we highlight that in order to translate our insights from the continuum model to the discrete setting, it is important to have enough samples to justify the approximation of the continuum model by the discrete one, and in this regard, larger values of α may require a larger number of samples (since the probabilities in Theorems 3.2, 3.3, and 3.5 deteriorate as p gets large).

Algorithm 1 Fermat Distance Spectral Clustering

Input: data points $x_1, \dots, x_n$, density parameter $p \geq 1$, normalization parameter $s \geq 0$, embedding dimension r, kernel scale h
Output: Laplacian $L_{p,s}$, FD-SC spectral embedding $[v_1, \dots, v_r] \in \mathbb{R}^{n \times r}$

$W_p(x_i, x_k) \leftarrow \eta(\ell_p(x_i, x_k)/h)$ ▷ Compute weights
$d_p(x_i) \leftarrow \sum_k W_p(x_i, x_k)$ ▷ Compute normalization factor
$W_{p,s}(x_i, x_k) \leftarrow \dfrac{W_p(x_i, x_k)}{d_p(x_i)^{(1-\frac{s}{2})} d_p(x_k)^{(1-\frac{s}{2})}}$ ▷ Compute normalized weight matrix

$D_{p,s}(x_i, x_i) \leftarrow \sum_k W_{p,s}(x_i, x_k)$ ▷ Compute degree matrix
$L_{p,s} \leftarrow D_{p,s}^{-1}(D_{p,s} - W_{p,s})$ ▷ Compute random walk Laplacian
$[v_1, \dots, v_r] \leftarrow$ bottom r eigenvectors of $L_{p,s}$

Algorithm 2 Degree Normalized Spectral Clustering

Input: data points $x_1, \dots, x_n$, parameters q, j, embedding dimension r, kernel scale h
Output: Laplacian $L_{q,j}$, DN-SC embedding $[v_1, \dots, v_r] \in \mathbb{R}^{n \times r}$

$W(x_i, x_k) \leftarrow \eta(\|x_i - x_k\|/h)$ ▷ Compute weights
$d(x_i) \leftarrow \sum_k W(x_i, x_k)$ ▷ Compute normalization factor
$W_q(x_i, x_k) \leftarrow \dfrac{W(x_i, x_k)}{d(x_i)^q d(x_k)^q}$ ▷ Compute normalized weight matrix

$D_q(x_i, x_i) \leftarrow \sum_k W_q(x_i, x_k)$ ▷ Compute degree matrix
$L_{q,j} \leftarrow D_q^{\frac{1-j}{q-1}}(D_q - W_q)$ ▷ Compute normalized Laplacian
$[v_1, \dots, v_r] \leftarrow$ bottom r eigenvectors of $L_{q,j}$

6. Numerical Simulations

The discrete Fermat distance spectral clustering algorithm is summarized in Algorithm 1. The algorithm depends on two key parameters: p, which determines the metric geometry, and s, which determines the Laplacian normalization. In Section 6.1 we explore the impact of these parameters. In Section 6.2 we explore the equivalence of Algorithm 1, where Fermat distances are computed explicitly, with Algorithm 2, a Euclidean spectral clustering algorithm with positive degree normalization.

6.1 Role of p and s in Fermat Spectral Clustering

In Figure 3, we consider an elongated data set with a density gap. There are two natural partitions: one that cuts "long" through the region of low density (which we call "Density Cut") and one that cuts "short" (which we call "Geometric Cut"). We see that as p increases, the density cut is eventually learned by Algorithm 1 by applying k-means to the spectral embedding to produce labels, which are then compared with the two natural partitions via an *accuracy* score that computes the proportion of points correctly assigned.

The behavior in p of Figure 3 corroborates our analysis that low p will favor geometry-driven cuts and large p density-driven cuts.

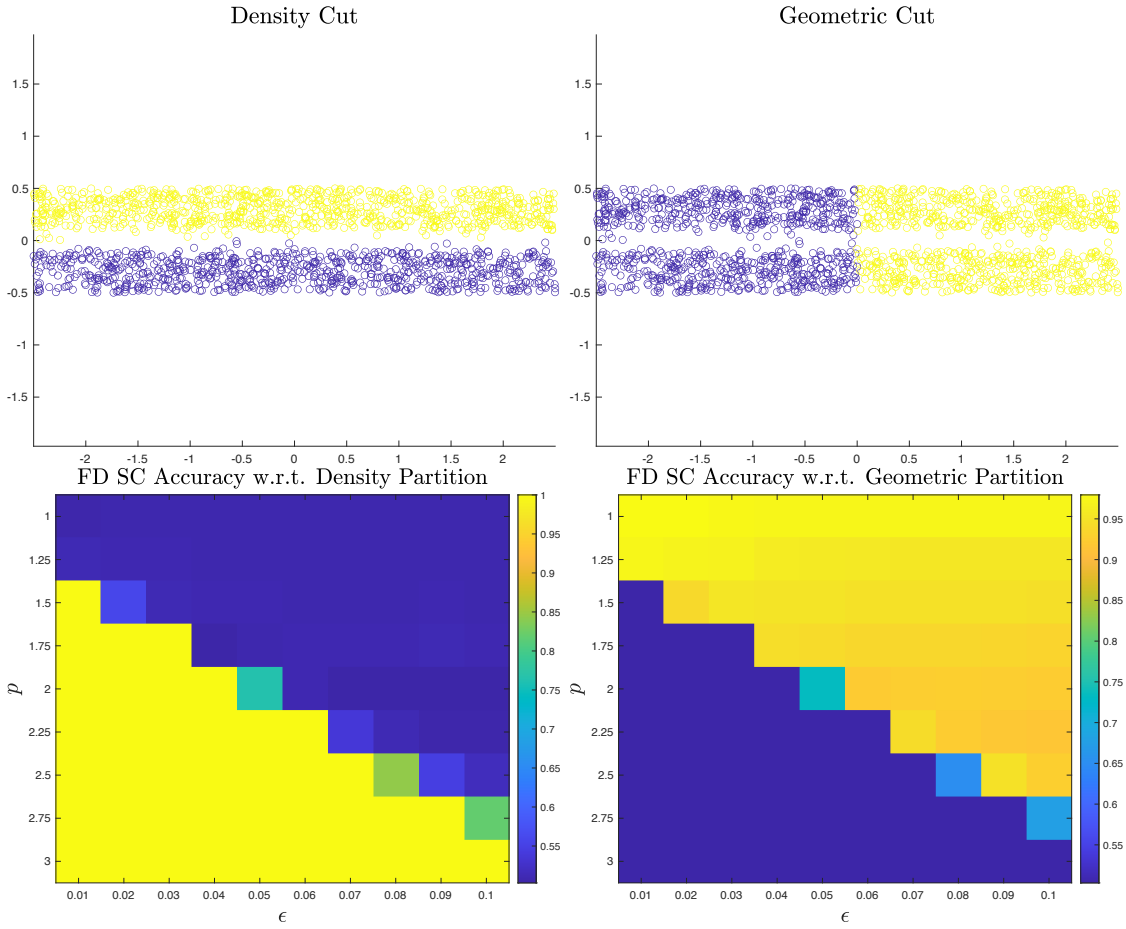
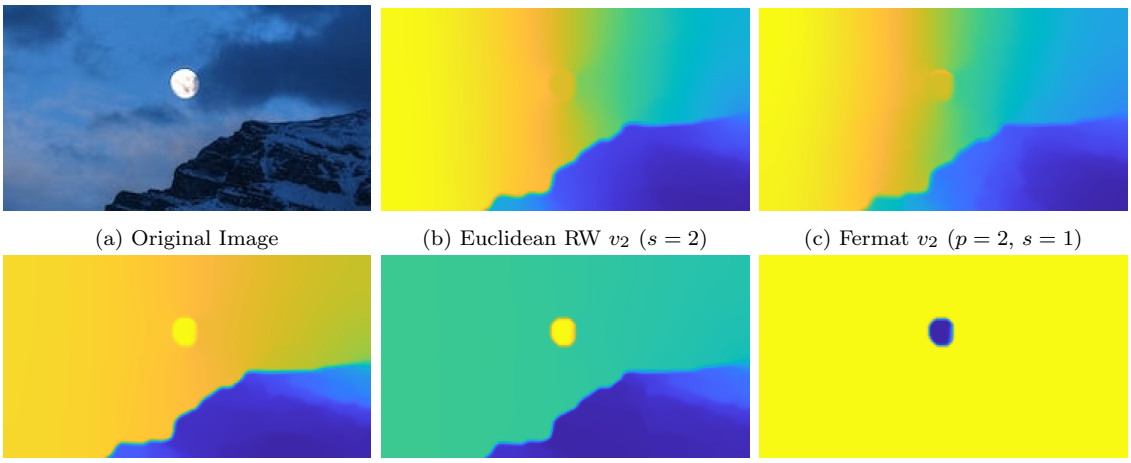


Figure 3: Fermat distance SC on elongated, density separated data. Two partitions are reasonable. As p increases, the learned clustering transitions from the geometric partition to the density partition.

In Figure 4, we demonstrate the impact of the normalization parameter s while keeping p fixed. As s increases, the moon—which is small but very separated from the rest of the pixels by intensity—is increasingly emphasized in the second Fermat eigenvector. Choosing a large s allows the Fiedler eigenvector to concentrate on a set of very small volume, while a smaller s prevents this behavior and recovers the more balanced cut of the horizon.

6.2 Asymptotic Equivalence with Degree Normalized Euclidean Laplacians

We compare to Hoffmann et al. (2022) by sampling data from a non-uniform density and constructing (i) a discrete Fermat graph Laplacian (*e.g.,* Algorithm 1) and (ii) a Euclidean distances graph Laplacian, with density rescaling so that in the large sample limit, the discrete Laplacians convergence to the same continuum operator (*e.g.,* Algorithm 2). As a baseline, we also compare with (iii) the random walk Laplacian build with Euclidean distances but without density rescaling.



(a) Original Image (b) Euclidean RW v_2 ($s = 2$) (c) Fermat v_2 ($p = 2$, $s = 1$)

(d) Fermat v_2 ($p = 2$, $s = 2$) (e) Fermat v_2 ($p = 2$, $s = 2.5$) (f) Fermat v_2 ($p = 2$, $s = 3$)

Figure 4: Moon example: effect of normalization. For small s, the Fermat v_2 cuts along the horizon, but for large s the cut produced by Fermat v_2 may yield very imbalanced clusters. Image available on unsplash courtesy of Kym MacKinnon.

We consider 2 example data sets in $\mathbb{R}^2$:

1. Data sampled from a ball of radius 1 with a density valley running down the vertical axis, specifically we consider for some $\tau > 0$ the density $\rho(x_1, x_2) \propto (\tau + x_1^2)^{-1} \cdot \mathbb{1}_{[0,1]}(x_1^2 + x_2^2)$.

2. A mixture of 2 Gaussians with uniform background noise, whose intensity is governed by a parameter $\tau > 0$.

Figures 5-6 illustrate the convergence of the Fermat Laplacian eigenvalues to their density normalized counterparts. The bottom row of these plots show comparisons between (i) and (ii). In general, smaller p and larger τ lead to larger regions of convergence. It is also relevant that the Fermat eigenvalues have larger variance, especially the higher frequency ones. Additional experiments appear in Appendix H.

This suggests that in practice and for sufficiently large sample size n, density-normalization provides a more computationally efficient (because Fermat distances do not need to be calculated) and statistically consistent alternative to constructing the Fermat distance graph Laplacian. Indeed, the complexity of computing a Fermat distances graph Laplacian, with each point connected to its k Fermat nearest neighbors, is $O((k^2 + CD)n\log(n))$ for a constant C that depends exponentially on the intrinsic dimension of the data (Mckenzie and Damelin, 2019). This is done via a modified Dijkstra's algorithm that leverages fast Euclidean nearest neighbor algorithms such as cover trees (Beygelzimer et al., 2006). In contrast, computing a k Euclidean nearest neighbors graph is essentially $O((k+CD)n\log(n))$. In practice, the cost of computing the Fermat nearest neighbor distances is significantly larger than computing the Euclidean nearest neighbors, even if they have the same asymptotic complexity in n (*i.e.,* k^2 is fixed to be relatively small compared to C and D). The spectral decomposition of the sparse Fermat distance Laplacian and density-reweighted Euclidean Laplacian have essentially the same complexity (since the sparsity level is the same),

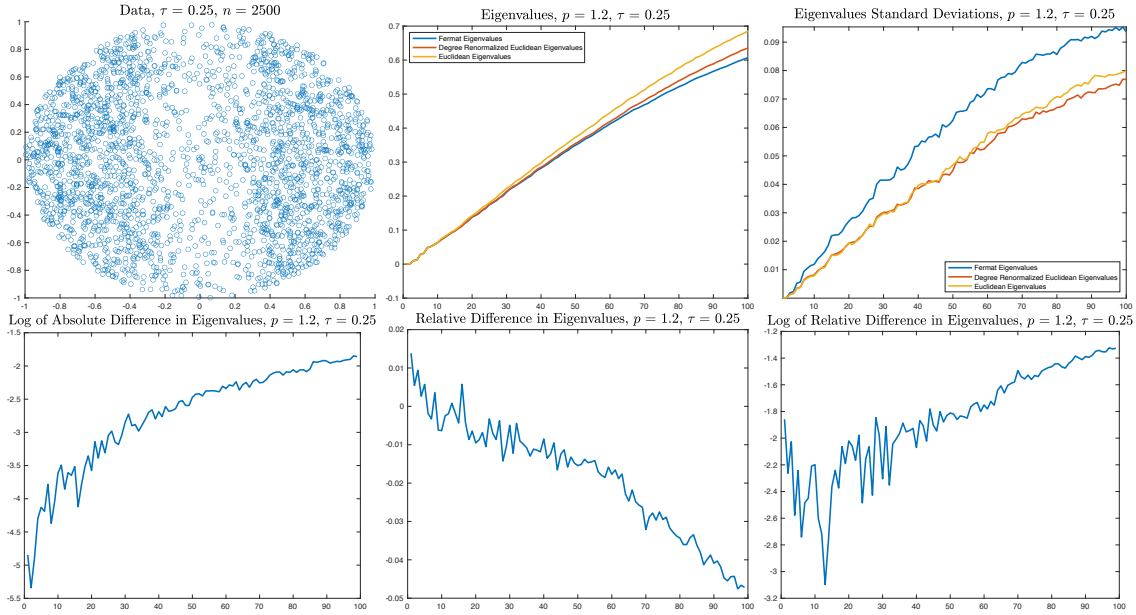


Figure 5: $p = 1.2$, $\tau = .25$. Runtime for Fermat Laplacian: 168.46 ± 10.70s. Runtime for Rescaled Euclidean Laplacian: $6.64 \pm .63$s.

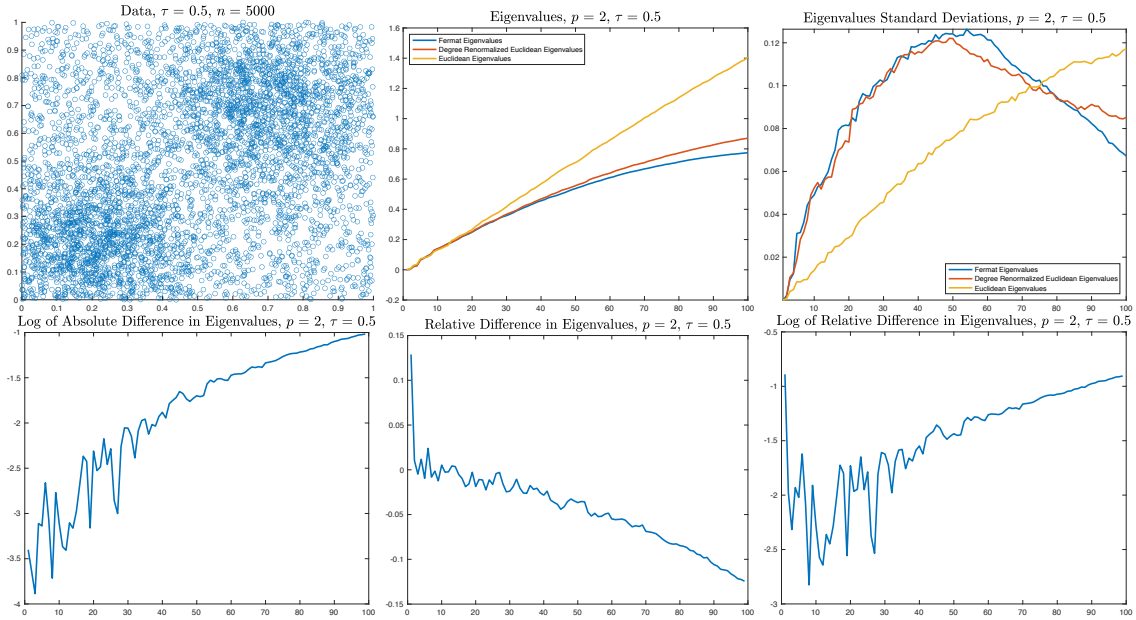


Figure 6: $p = 2$, $\tau = .5$. Runtime for Fermat Laplacian: 271.85 ± 96.75s. Runtime for Rescaled Euclidean Laplacian: $1.45 \pm .23$s.

so the benefit of avoiding Fermat Laplacians is at the level of avoiding expensive distance calculations that suffer from relatively high statistical variance (Little et al., 2022).

7. Conclusions and Future Work

By computing Fermat geodesics and applying percolation results in the manifold tangent plane, we have developed the first quantitative local Fermat metric convergence results in the literature. We apply these results to develop a continuum limit theory for Fermat graph Laplacians, and prove convergence results of the eigenvalues and eigenvectors of the discrete operators to those of their continuum analogues. An interesting consequence of this analysis is that it establishes the similarity of (i) Fermat Laplacians and (ii) density-reweighted Euclidean Laplacians in the large sample limit. The geometric framework we develop thus leads to new computational schemes that leverage the theoretical benefits of Fermat spectral clustering (for example robustness with respect to cluster elongation as illustrated in Figure 1) without the need for onerous calculation of pairwise Fermat distances.

In future work, it is of interest to investigate whether our local metric convergence results can be extended to apply globally, whether it is possible to obtain point-wise consistency for Fermat graph Laplacians, how our results are impacted by noise, and also to theoretically investigate the role of the normalization parameter s, which we have fixed in this work.

Acknowledgements: NGT was supported by NSF DMS-2005797. AL thanks NSF DMS-2136198 and NSF DMS-2309570. JMM acknowledges support from NSF DMS-1912737, DMS-1924513, DMS-2309519, and DMS-2318894.

References

M. Alamgir and U. Von Luxburg. Shortest path distance in random k-nearest neighbor graphs. In *ICML*, pages 1251–1258, 2012.

M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

M. Belkin and P. Niyogi. Convergence of Laplacian eigenmaps. In *NIPS*, pages 129–136, 2007.

A. L. Besse. *Einstein manifolds*. Springer Science & Business Media, 2007.

A. Beygelzimer, S. Kakade, and J. Langford. Cover trees for nearest neighbor. In *ICML*, pages 97–104, 2006.

A. Bijral, N. Ratliff, and N. Srebro. Semi-supervised learning with density based distances. In *UAI*, pages 43–50, 2011.

N. Boumal. An introduction to optimization on smooth manifolds. *Available online, May*, 3, 2020.

O. Bousquet, O. Chapelle, and M. Hein. Measure based regularization. In *NIPS*, pages 1221–1228, 2004.

L. Brewin. Riemann normal coordinate expansions using cadabra. *Classical and Quantum Gravity*, 26(17):175017, 2009.

L. Bungert, J. Calder, and T. Roith. Ratio convergence rates for Euclidean first-passage percolation: Applications to the graph infinity Laplacian. *arXiv preprint arXiv:2210.09023*, 2022.

D. Burago, S. Ivanov, and Y. Kurylev. A graph discretization of the Laplace–Beltrami operator. *Journal of Spectral Theory*, 4(4):675–714, 2015.

J. Calder and N. García Trillos. Improved spectral convergence rates for graph Laplacians on ε-graphs and k-NN graphs. *Applied and Computational Harmonic Analysis*, 60:123–175, 2022.

H. Chang and D.-Y. Yeung. Robust path-based spectral clustering. *Pattern Recognition*, 41(1):191–203, 2008.

T. Chu, G. Miller, and D. Sheehy. Exact computation of a manifold metric, via Lipschitz embeddings and shortest paths on a graph. In *SODA*, pages 411–425, 2020.

A. Civril, M. Magdon-Ismail, and E. Bocek-Rivele. SSDE: Fast graph drawing using sampled spectral distance embedding. In *International Symposium on Graph Drawing*, pages 30–41. Springer, 2006.

R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.

R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner, and S. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences*, 102(21):7426–7431, 2005.

P. Comon. Independent component analysis, a new concept? *Signal processing*, 36(3): 287–314, 1994.

K. Craig, N. García Trillos, and D. Slepčev. Clustering dynamics on graphs: From spectral clustering to mean shift through Fokker–Planck interpolation. In N. Bellomo, J. A. Carrillo, and E. Tadmor, editors, *Active Particles, Volume 3: Advances in Theory, Models, and Applications*, pages 105–151. Birkhäuser, 2022.

E. Davis and S. Sethuraman. Approximating geodesics via random points. *The Annals of Applied Probability*, 29(3):1446 – 1486, 2019.

J. Díaz, D. Mitsche, G. Perarnau, and X. Pérez-Giménez. On the relation between graph distance and Euclidean distance in random geometric graphs. *Advances in Applied Probability*, 48(3):848 – 864, 2016.

X. Fernández, E. Borghini, G. Mindlin, and P. Groisman. Intrinsic persistent homology via density-based metric learning. *Journal of Machine Learning Research*, 24(75):1–42, 2023.

N. García Trillos, M. Gerlach, M. Hein, and D. Slepčev. Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace–Beltrami operator. *Foundations of Computational Mathematics*, pages 1–61, 2019.

N. García Trillos, F. Hoffmann, and B. Hosseini. Geometric structure of graph Laplacian embeddings. *Journal of Machine Learning Research*, 22(63):1–55, 2021.

B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley. Multidimensional scaling, sammon mapping, and isomap: Tutorial and survey. *arXiv preprint arXiv:2009.08136*, 2020.

P. Groisman, M. Jonckheere, and F. Sapienza. Nonhomogeneous Euclidean first-passage percolation and distance learning. *Bernoulli*, 28(1):255–276, 2022.

M. Hein, J.-Y. Audibert, and U. Von Luxburg. Graph Laplacians and their convergence on random neighborhood graphs. *Journal of Machine Learning Research*, 8(6), 2007.

F. Hoffmann, B. Hosseini, A. A. Oberai, and A. M. Stuart. Spectral analysis of weighted laplacians arising in data clustering. *Applied and Computational Harmonic Analysis*, 56: 189–249, 2022.

H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

C. Howard and C. Newman. Geodesics and spanning trees for Euclidean first-passage percolation. *Annals of Probability*, pages 577–623, 2001.

S. Hwang, S. Damelin, and A. Hero. Shortest path through random points. *The Annals of Applied Probability*, 26(5):2791–2823, 2016.

H. Kesten. On the speed of convergence in first-passage percolation. *The Annals of Applied Probability*, pages 296–338, 1993.

W. Kühnel. *Differential geometry*, volume 77. American Mathematical Soc., 2015.

J. M. Lee. *Riemannian manifolds: an introduction to curvature*, volume 176. Springer Science & Business Media, 2006.

A. Little, M. Maggioni, and J. Murphy. Path-based spectral clustering: Guarantees, robustness to outliers, and fast algorithms. *Journal of Machine Learning Research*, 21(6): 1–66, 2020.

A. Little, D. McKenzie, and J. M. Murphy. Balancing geometry and density: Path distances on high-dimensional data. *SIAM Journal on Mathematics of Data Science*, 4(1):72–99, 2022.

A. Manousidaki, A. Little, and Y. Xie. Clustering and visualization of single-cell rna-seq data using path metrics. *PLOS Computational Biology*, 20(5):e1012014, 2024.

D. Mckenzie and S. Damelin. Power weighted shortest paths for clustering Euclidean data. *Foundations of Data Science*, 1(3):307, 2019.

A. Moscovich, A. Jaffe, and B.Nadler. Minimax-optimal semi-supervised regression on unknown manifolds. In *AISTATS*, pages 933–942, 2017.

A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *NIPS*, pages 849–856, 2002.

J. Platt. Fastmap, metricmap, and landmark mds are all Nyström algorithms. In *AISTATS*, pages 261–268. PMLR, 2005.

S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

Sajama and A. Orlitsky. Estimating and computing density based distance metrics. In *ICML*, pages 760–767, 2005.

G. Schiebinger, M. Wainwright, and B. Yu. The geometry of kernelized spectral clustering. *The Annals of Statistics*, 43(2):819–846, 2015.

G. Shamai, M. Zibulevsky, and R. Kimmel. Efficient inter-geodesic distance computation and fast classical scaling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(1):74–85, 2020.

J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

M. Skorski. Chain rules for Hessian and higher derivatives made easy by tensor calculus. *arXiv preprint arXiv:1911.13292*, 2019.

J. Tenenbaum, V. D. Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

L. van der Maaten and G. Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

R. Vershynin. High-dimensional probability, volume 47 of cambridge series in statistical and probabilistic mathematics. *Cambridge University Press, Cambridge*, 7:14–22, 2018.

P. Vincent and Y. Bengio. Density-sensitive metrics and kernels. In *Snowbird Learning Workshop*, 2003.

U. Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

C. Williams and M. Seeger. Using the Nyström method to speed up kernel machines. In *NIPS*, pages 682–688, 2001.

C. L. Wormell and S. Reich. Spectral convergence of diffusion maps: Improved error bounds and an alternative normalization. *SIAM Journal on Numerical Analysis*, 59(3):1687–1734, 2021.

H. Yu, X. Zhao, X. Zhang, and Y. Yang. Isomap using Nyström method with incremental sampling. *Advances in Information Sciences & Service Sciences*, 4(12), 2012.

S. Zhang and J. Murphy. Hyperspectral image clustering with spatially-regularized ultrametrics. *Remote Sensing*, 13(5):955, 2021.

Appendix A. Preliminaries from Differential Geometry

In this appendix, we shall use the Einstein summation convention: when an index appears in an upper and lower position, it means summation over that index. In contrast to the main body, we shall use grad (resp. Hess) to denote the Riemannian gradient (resp. Hessian) so that ∇ (resp. H) can be reserved for Euclidean gradients (resp. Hessians) computed in the ambient space or on a tangent space.

Definition A.1 *Let* $\operatorname{dist}(x, \mathcal{M}) = \min_{y\in\mathcal{M}} \|x - y\|$. *The* reach *of* $\mathcal{M} \subset \mathbb{R}^D$ *is*

$$\mathcal{R} := \sup\{t > 0 : \forall x \in \mathbb{R}^D \text{ with } \operatorname{dist}(x, \mathcal{M}) \leq t, \exists! y \in \mathcal{M} \text{ s.t. } \operatorname{dist}(x, \mathcal{M}) = \|x - y\|\}.$$

As $\mathcal{R}$ depends only on the embedding, not on the specific metric chosen on $\mathcal{M}$, the reach of $(\mathcal{M}, g_p)$ equals the reach of $(\mathcal{M}, g)$.

Definition A.2 *For any linearly independent* $X, Y \in T_x\mathcal{M}$ *the* sectional curvature*, with respect to* g*, of the plane spanned by* X, Y *is defined as*

$$K_x(X, Y) = \frac{\mathbf{R}_x(X, Y, Y, X)}{g_x(X, X)g_x(Y, Y) - g_x(X, Y)^2},$$

where $\mathbf{R}$ *is the Riemannian curvature tensor.*

We are interested in $(\mathcal{M}, g)$ with globally bounded sectional curvature:

$$|K_x(X,Y)| \leq K \quad \text{for all } x \in \mathcal{M},\ X, Y \in T_x\mathcal{M}.$$

Lemma 1 *Suppose that $\rho \in C^2\mathcal{M}$. Fix any $x \in \mathcal{M}$ and consider the function*

$$\begin{aligned} \rho_x^* &: T_x\mathcal{M} \to \mathbb{R} \\ \rho_x^*(u) &= \rho(\exp_x(u)) J_x(u) \end{aligned} \tag{57}$$

where $J_x = \sqrt{\det(g_x)}$. Then $\|\nabla \rho_x^*(0)\| = \|\nabla \rho(x)\|$ and

$$\|H\rho_x^*(0)\| \lesssim \|H\rho(x)\| + mK\rho(x) + Kd^{\frac{3}{2}}\|\nabla\rho(x)\|.$$

Proof Starting from the normal coordinate expansion of the metric (see, *e.g.* Brewin (2009)), taking determinants, and using $\sqrt{1+x} = 1 + \frac{1}{2}x + O(x^2)$ yields:

$$J_x(z) = \sqrt{\det(g_x)} = 1 - \frac{1}{6} z^i z^j R_{ij} - \frac{1}{12} z^i z^j z^k \nabla_i R_{jk} + O(\|z\|^4) \tag{58}$$

where R_{ij} is the Ricci curvature tensor, expressed in normal coordinates. As all sectional curvatures are bounded by K, $R_{ij} \leq (m-1)Kg_{ij}$. Differentiating (58):

$$\|\nabla J_x(z)\| \lesssim (m-1)K\|z\| + O(\|z\|^2) \quad , \quad \|HJ_x(z)\| \lesssim (m-1)K + O(\|z\|).$$

As $J_x(0) = 1, \nabla J_x(0) = 0$ differentiating (57) yields

$$\nabla \rho_x^*(0) = J_x(0)\nabla\rho_x^*(\exp_x(u)) = \nabla\rho(\exp_x(z))|_{z=0} = \nabla\rho(x) \cdot \mathrm{Jac}(\exp_x(0)) = \nabla\rho(x).$$

Using $\mathrm{Jac}(\exp_x(0)) = \mathrm{Id}$. Taking norms yields the first claim. From the product rule for Hessians:

$$H\rho_x^* = J_x(H\rho \circ \exp_x) + (\nabla\rho \circ \exp_x)^T \nabla J_x + (\nabla J_x)^T \nabla\rho \circ \exp_x + \rho \circ \exp_x(HJ_x)$$

evaluating at 0 and recalling $J_x(0) = 1, \nabla J_x(0) = 0$

$$H\rho_x^*(0) = H\rho \circ \exp_x(0) + \rho(x) HJ_x(0).$$

Since $\|HJ_x(0)\| \lesssim (m-1)K$, it remains to bound $H\rho_x(0)$. From Skorski (2019) we have the following chain rule for Hessians:

$$H(\rho \circ \exp_x) = (\mathrm{Jac}\exp_x)^T \cdot H\rho(\exp_x) \cdot (\mathrm{Jac}\exp_x) + \sum_{k=1}^{D} \frac{\partial \rho}{\partial x^k} \cdot H(\exp_x^k)$$

where $\exp_x^k$ is the k^{th} coordinate of $\exp_x$ and $x_1, \ldots, x_D$ are full-dimensional Euclidean coordinates. Since $\|H(\exp_x^k)\| \lesssim Km$, evaluating at $z = 0$ and taking norms gives:

$$\|H\rho \circ \exp_x(0)\| \leq \|H\rho(x)\| + CKm\|\nabla\rho(x)\|_1 \leq \|H\rho(x)\| + CKm^{\frac{3}{2}}\|\nabla\rho(x)\|$$

(here we have used the fact that we can choose coordinates so that $\nabla\rho(x)$ is zero except in the first m coordinates, and once again we have ignored the lower order term). Thus we obtain:

$$\|H\rho_x^*(0)\| \lesssim \|H\rho(x)\| + mK\rho(x) + Km^{\frac{3}{2}}\|\nabla\rho(x)\|.$$

■

Appendix B. Fermat Sectional Curvature

In this section we bound the sectional curvatures of $(\mathcal{M}, g_p)$ in terms of those of $(\mathcal{M}, g)$. This will follow from standard results on conformally equivalent metrics.

Lemma B.1 *Suppose ρ is a density function on $(\mathcal{M}, g)$ satisfying Assumptions 3.1. Let $\varphi = \frac{\alpha}{2}\log(\rho)$ where $\alpha = 2(p-1)/m$. Then:*

1. $\|\operatorname{grad}\varphi\| \leq \beta\left(\frac{p-1}{m}\right) L_1$.
2. $\|\operatorname{Hess}\varphi\| \leq \beta\left(\frac{p-1}{m}\right)(\beta L_1^2 + L_2)$.

Proof We compute:

$$\operatorname{grad}\varphi = \frac{\alpha}{2}\frac{1}{\rho}\operatorname{grad}\rho = \frac{\alpha}{2\rho}\operatorname{grad}\rho = \frac{p-1}{m\rho}\operatorname{grad}\rho.$$

From *e.g.* (Boumal, 2020, Proposition 10.43), we obtain $\|\operatorname{grad}\rho\| \leq L_1$. Hence:

$$\|\operatorname{grad}\varphi\| = \frac{p-1}{m\rho}\|\operatorname{grad}\rho\| \leq \beta\left(\frac{p-1}{m}\right) L_1.$$

For part 2, a simple calculation reveals

$$\operatorname{Hess}\varphi = -\frac{p-1}{d\rho^2}\operatorname{grad}\rho \otimes \operatorname{grad}\rho + \frac{p-1}{d\rho}\operatorname{Hess}\rho.$$

By assumption $\|\operatorname{Hess}\rho\| \leq L_2$, and $\|\operatorname{grad}\rho \otimes \operatorname{grad}\rho\| = \|\operatorname{grad}\rho\|^2 \leq L_1^2$, so

$$\|\operatorname{Hess}\varphi\| \leq \frac{p-1}{m\rho^2}L_1^2 + \frac{p-1}{m\rho}L_2 \leq \beta\left(\frac{p-1}{m}\right)(\beta L_1^2 + L_2).$$

■

Theorem B.2 *Assume ρ satisfies Assumption 3.1 and $(\mathcal{M}, g)$ has sectional curvature bounded by K. Then $(\mathcal{M}, g_p)$ has sectional curvature bounded by K_p where*

$$K_p := \beta^\alpha\left(K + \frac{3\beta^2(p-1)^2L_1^2}{m^2} + \frac{2\beta^2(p-1)L_1^2}{m} + \frac{\beta(p-1)L_2}{m}\right).$$

Proof Let $\mathbf{R}$ (resp. $\mathbf{R}_p$) denote the $(0,4)$ Riemannian curvature tensor of $(\mathcal{M}, g)$ (resp. $(\mathcal{M}, g_p)$). Let $\phi := -\frac{\alpha}{2}\log\rho$. From (Kühnel, 2015, pp. 345) (see also (Besse, 2007, pp. 58)) we have the identity:

$$e^{2\phi}\mathbf{R}_p = \mathbf{R} - \frac{1}{2}\langle\operatorname{grad}\varphi, \operatorname{grad}\varphi\rangle g \bullet g + (\operatorname{Hess}\varphi) \bullet g + (\operatorname{grad}\varphi)^2 \bullet g$$

where $\bullet$ is the Kulkarni-Nomizu product between symmetric two tensors defined as

$$A \bullet B(X, Y, Z, W) = A(X, Z)B(Y, W) + A(Y, W)B(X, Z)$$

$$- A(X,W)B(Y,Z) - A(Y,Z)B(X,W),$$

see (Kühnel, 2015, Def. 8.20), and $(\operatorname{grad}\varphi)^2$ is the symmetric (0,2) tensor defined as

$$(\operatorname{grad}\varphi)^2(X,Y) = (X\varphi)(Y\varphi).$$

By the definition of sectional curvature,

$$K_{p,x}(X,Y) = \frac{\mathbf{R}_{p,x}(X,Y,Y,X)}{g_{p,x}(X,X)g_{p,x}(Y,Y) - g_{p,x}(X,Y)^2}$$

for any linearly independent $X, Y \in T_x\mathcal{M}$. It suffices to consider only orthonormal X, Y, in which case $K_{p,x}(X,Y) = \mathbf{R}_{p,x}(X,Y,Y,X)$. As g and g_p are conformally equivalent, X, Y are orthonormal with respect to g if and only if they are orthonormal with respect to g_p. Thus if X, Y are g_p orthonormal, we also have $K_x(X,Y) = \mathbf{R}_x(X,Y,Y,X)$. Assuming orthonormality, some straightforward calculations reveal

$$\begin{aligned}
&g \bullet g(X,Y,Y,X) = -2,\\
&(\operatorname{grad}\varphi)^2 \bullet g(X,Y,Y,X) = -\left(g(\operatorname{grad}\varphi, X)\right)^2 - \left(g(\operatorname{grad}\varphi, Y)\varphi\right)^2,\\
&(\operatorname{Hess}\varphi) \bullet g(X,Y,Y,X) = -\operatorname{Hess}\varphi(X,X) - \operatorname{Hess}\varphi(Y,Y).
\end{aligned}$$

Now,

$$\begin{aligned}
|K_{p,x}(X,Y)| &= |\mathbf{R}_{p,x}(X,Y,Y,X)|\\
&\le e^{2\varphi}\left(|\mathbf{R}(X,Y,Y,X)| + \left|\frac{1}{2}\langle\operatorname{grad}\varphi, \operatorname{grad}\varphi\rangle g \bullet g(X,Y,Y,X)\right|\right.\\
&\left. + |(\operatorname{Hess}\varphi) \bullet g(X,Y,Y,X)| + \left|(\nabla\varphi)^2 \bullet g(X,Y,Y,X)\right|\right)\\
&\le e^{2\varphi}\left(K + \frac{1}{2}\|\operatorname{grad}\varphi\|^2 |-2| + \right.\\
&\left. + |\operatorname{Hess}\varphi(X,X)| + |\operatorname{Hess}\varphi(Y,Y)| + \left(g(\operatorname{grad}\varphi, X)\right)^2 + \left(g(\operatorname{grad}\varphi, Y)\right)^2\right)\\
&\overset{(a)}{\le} (\rho^\alpha)\left(K + \frac{(p-1)^2\beta^2L_1^2}{m^2} + 2\beta\left(\frac{p-1}{m}\right)(\beta L_1^2 + L_2) + \frac{2(p-1)^2\beta^2L_1^2}{m^2}\right)\\
&= \rho^\alpha\left(K + \frac{3(p-1)^2\beta^2L_1^2}{m^2} + \frac{2(p-1)\beta^2L_1^2}{m} + \frac{(p-1)\beta L_2}{m}\right)\\
&\le \beta^\alpha\left(K + \frac{3(p-1)^2\beta^2L_1^2}{m^2} + \frac{2(p-1)\beta^2L_1^2}{m} + \frac{(p-1)\beta L_2}{m}\right).
\end{aligned}$$

■

where in (a) we use various bounds from Lemma B.1.

Appendix C. Fermat Geodesics

This appendix contains the proof of Theorem 4.1.

Theorem 4.1 *Let $\Omega \subseteq \mathbb{R}^m$ be a m-dimensional, open, connected domain and assume (Ω, ρ) satisfies Assumption 3.1. Fix $y \in \Omega$ and let $\gamma_b(t)$ denote the unit speed geodesic with respect to g_p originating at y in the direction of unit vector $b \in \mathbb{R}^m$. Then:*

$$\begin{aligned}\gamma_b(t) =&\rho(y)^{\frac{\alpha}{2}} bt + \alpha\rho(y)^{\alpha-1}\left(\frac{1}{2}\langle b, \nabla\rho(y)\rangle b - \frac{1}{4}\nabla\rho(y)\right)t^2 \\ &+(C_1' b + C_2' H(y)b + C_3'\nabla\rho(y))t^3 + O(t^4)\,,\end{aligned} \tag{15}$$

where $H(y)$ denotes the Hessian matrix of ρ evaluated at y and

$$\begin{aligned}C_1' &= \left(\frac{1}{3}\alpha^2 - \frac{1}{6}\alpha\right)\rho(y)^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho(y)\rangle^2 \\ &\quad + \frac{\alpha}{6}\rho(y)^{\frac{3}{2}\alpha-1}\langle H(y)b, b\rangle - \frac{\alpha^2}{12}\rho(y)^{\frac{3}{2}\alpha-2}\langle\nabla\rho(y), \nabla\rho(y)\rangle, \\ C_2' &= -\frac{\alpha}{12}\rho(y)^{\frac{3}{2}\alpha-1}, \\ C_3' &= \left(\frac{\alpha}{12} - \frac{\alpha^2}{6}\right)\rho(y)^{\frac{3}{2}\alpha-2}\langle\nabla\rho(y), b\rangle.\end{aligned}$$

Proof We compute the geodesic equations for (Ω, g_p) for geodesics starting at some $y \in \Omega$. Without loss of generality, assume $y = 0$ and for notational convenience, let ρ_0, $\nabla\rho_0$, and H_0 denote $\rho(y) = \rho(0) = \rho_0$, $\nabla\rho(y) = \nabla\rho(0) = \nabla\rho_0$, and $H(y) = H(0) = H_0$, respectively; also let g^p_{ij} denote $(g_p)_{ij}$. We have global coordinates on Ω which are just the standard coordinates (x^i). The function $x(t) = (x^1(t), \dots, x^m(t))$ is a g_p geodesic if and only if it satisfies for all $k = 1, \dots, d$ the geodesic equation

$$\ddot{x}^k(t) + \dot{x}^i(t)\dot{x}^j(t)\Gamma^k_{ij}(x(t)) = 0,$$

where

$$\Gamma^k_{ij} = \frac{1}{2}g_p^{k\ell}(\partial_i g^p_{j\ell} + \partial_j g^p_{i\ell} - \partial_\ell g^p_{ij})$$

are the Christoffel symbols associated to g_p (Lee, 2006). Now, $g^p_{ij} = \rho^{-\alpha}g_{ij}$, where $g_{ij} = \langle\partial_i, \partial_j\rangle = \delta_{ij}$ is the Euclidean metric tensor on $\mathbb{R}^m$. In addition, since g_{ij} is the identity matrix, $g_p^{ij} = (g^p_{ij})^{-1} = \rho^\alpha(g_{ij})^{-1} = \rho^\alpha\langle\partial_i, \partial_j\rangle = \rho^\alpha\delta^{ij}$ and $\partial_i g^p_{j\ell} = \partial_i(\rho^{-\alpha}g_{j\ell}) = \delta_{j\ell}\partial_i\rho^{-\alpha}$. Thus:

$$\begin{aligned}\Gamma^k_{ij} &= \frac{1}{2}g_p^{k\ell}(\partial_i g^p_{j\ell} + \partial_j g^p_{i\ell} - \partial_\ell g^p_{ij}) \\ &= \frac{1}{2}\rho^\alpha\delta^{k\ell}\left(\delta_{j\ell}\partial_i\rho^{-\alpha} + \delta_{i\ell}\partial_j\rho^{-\alpha} - \delta_{ij}\partial_\ell\rho^{-\alpha}\right) \\ &= \frac{1}{2}\rho^\alpha\left(\delta^k_j\partial_i\rho^{-\alpha} + \delta^k_i\partial_j\rho^{-\alpha} - \delta_{ij}\delta^{k\ell}\partial_\ell\rho^{-\alpha}\right).\end{aligned}$$

So, for k fixed:

$$\begin{aligned}\dot{x}^i\dot{x}^j\Gamma^k_{ij} &= \dot{x}^i\dot{x}^j\frac{1}{2}\rho^\alpha\left(\delta^k_j\partial_i\rho^{-\alpha} + \delta^k_i\partial_j\rho^{-\alpha} - \delta_{ij}\delta^{k\ell}\partial_\ell\rho^{-\alpha}\right) \\ &= \frac{1}{2}\rho^\alpha\dot{x}^i\dot{x}^k\partial_i\rho^{-\alpha} + \frac{1}{2}\rho^\alpha\dot{x}^k\dot{x}^j\partial_j\rho^{-\alpha} - \frac{1}{2}\rho^\alpha\dot{x}^i\dot{x}_i\delta^{k\ell}\partial_\ell\rho^{-\alpha}\end{aligned}$$

$$= \frac{1}{2}\rho^\alpha \left(\dot{x}^i \partial_i \rho^{-\alpha} + \dot{x}^j \partial_j \rho^{-\alpha}\right) \dot{x}^k - \frac{1}{2}\rho^\alpha \dot{x}^i \dot{x}_i \delta^{k\ell} \partial_\ell \rho^{-\alpha}$$
$$= \rho^\alpha \langle \dot{x}, \nabla \rho^{-\alpha} \rangle \dot{x}^k - \frac{1}{2}\rho^\alpha \langle \dot{x}, \dot{x} \rangle \partial_k \rho^{-\alpha}.$$

Where in the last line we have switched to vector notation and identified $d\rho^{-\alpha}$ with $\nabla \rho^{-\alpha}$ (which is valid as the metric is the Euclidean one). Thus, in vector notation the geodesic equation becomes:

$$\ddot{x} + \rho^\alpha \langle \dot{x}, \nabla \rho^{-\alpha} \rangle \dot{x} - \frac{1}{2}\rho^\alpha \langle \dot{x}, \dot{x} \rangle \nabla \rho^{-\alpha} = 0. \tag{59}$$

For a unit Euclidean norm vector $b \in \mathbb{R}^m$, consider the initial value problem

$$x(0) = 0,$$
$$\dot{x}(0) = \rho_0^{\frac{\alpha}{2}} b,$$

which has a unique geodesic solution $\gamma_b(t)$ in some local neighborhood (Lee, 2006). Note $|\dot{x}(0)|_p = \rho_0^{-\frac{\alpha}{2}} \|\rho_0^{\frac{\alpha}{2}} b\| = \|b\| = 1$ where $|\cdot|_p$ denotes the norm with respect to g_p, so $\gamma_b(t)$ is the unit speed geodesic in the direction of b. Because ρ is assumed $\mathcal{C}^\infty$ and bounded away from 0, we have that each g_{ij}^p is C^∞, hence the Christoffel symbols are $\mathcal{C}^\infty$. This implies the geodesic $\gamma_b(t)$ is also $\mathcal{C}^\infty$, and hence we can Taylor expand it about $t = 0$ to obtain:

$$\gamma_b(t) = \rho_0^{\frac{\alpha}{2}} bt + vt^2 + qt^3 + O(t^4),$$
$$\dot{\gamma}_b(t) = \rho_0^{\frac{\alpha}{2}} b + 2vt + 3qt^2 + O(t^3),$$
$$\ddot{\gamma}_b(t) = 2v + 6qt + O(t^2),$$

for some vectors v, q depending on b, ρ, which can be computed from the leading order terms in (59).

We have the following Taylor expansions for $\rho, \nabla\rho, \rho^\alpha, \nabla\rho^{-\alpha}$:

$$\begin{aligned}
\rho(x) &= \rho_0 + \langle \nabla\rho_0, x \rangle + \frac{1}{2} x^T H_0 x + O(x^3), \\
\nabla\rho(x) &= \nabla\rho_0 + H_0 x + O(x^2), \\
\rho(x)^\alpha &= \rho_0^\alpha + \alpha\rho_0^{\alpha-1} \langle \nabla\rho_0, x \rangle + O(x^2), \\
\nabla\rho(x)^{-\alpha} &= -\alpha\rho(x)^{-\alpha-1} \nabla\rho(x) \\
&= -\alpha(\rho_0 + \langle \nabla\rho_0, x \rangle + O(x^2))^{-\alpha-1} (\nabla\rho_0 + H_0 x + O(x^2)) \\
&= -\alpha\rho_0^{-\alpha-1} (1 + \rho_0^{-1} \langle \nabla\rho_0, x \rangle + O(x^2))^{-\alpha-1} (\nabla\rho_0 + H_0 x + O(x^2)) \\
&= -\alpha\rho_0^{-\alpha-1} (1 - (\alpha+1)\rho_0^{-1} \langle \nabla\rho_0, x \rangle + O(x^2)) (\nabla\rho_0 + H_0 x + O(x^2)) \\
&= -\alpha\rho_0^{-\alpha-1} (\nabla\rho_0 - (\alpha+1)\rho_0^{-1} \langle \nabla\rho_0, x \rangle \nabla\rho_0 + H_0 x + O(x^2)) \\
&= -\alpha\rho_0^{-\alpha-1} \nabla\rho_0 + \alpha(\alpha+1)\rho_0^{-\alpha-2} \langle \nabla\rho_0, x \rangle \nabla\rho_0 - \alpha\rho_0^{-\alpha-1} H_0 x + O(x^2).
\end{aligned}$$

Substituting $x = \gamma_b(t)$, we obtain for some vectors z, w the following Taylor expansions in terms of t:

$$\rho^\alpha = \rho_0^\alpha + \alpha\rho_0^{\frac{3}{2}\alpha-1} \langle \nabla\rho_0, b \rangle t + O(t^2)$$

$$
\begin{aligned}
&:= \rho_0^\alpha + zt + O(t^2),\\
\nabla \rho^{-\alpha} &= -\alpha \rho_0^{-\alpha-1}\nabla\rho_0 + \alpha(\alpha+1)\rho_0^{-\frac{\alpha}{2}-2}\langle \nabla\rho_0, b\rangle \nabla\rho_0 t - \alpha\rho_0^{-\frac{\alpha}{2}-1} H_0 b t + O(t^2)\\
&:= -\alpha\rho_0^{-\alpha-1}\nabla\rho_0 + wt + O(t^2),\\
\langle \dot\gamma_b, \nabla\rho^{-\alpha}\rangle &= \langle \rho_0^{\frac{\alpha}{2}} b + 2vt, -\alpha\rho_0^{-\alpha-1}\nabla\rho_0 + wt\rangle + O(t^2)\\
&= -\alpha\rho_0^{-\frac{\alpha}{2}-1}\langle b, \nabla\rho_0\rangle + \rho_0^{\frac{\alpha}{2}}\langle b, w\rangle t - 2\alpha\rho_0^{-\alpha-1}\langle v, \nabla\rho_0\rangle t + O(t^2),\\
\langle \dot\gamma_b, \dot\gamma_b\rangle &= \langle \rho_0^{\frac{\alpha}{2}} b + 2vt, \rho_0^{\frac{\alpha}{2}} b + 2vt\rangle + O(t^2)\\
&= \rho_0^\alpha + 4\rho_0^{\frac{\alpha}{2}}\langle b, v\rangle t + O(t^2).
\end{aligned}
$$

To solve for v, we only need to compute the leading order term in the geodesic equation (59). We have:

$$
\begin{aligned}
\rho^\alpha \langle \dot\gamma_b, \nabla\rho^{-\alpha}\rangle \dot\gamma_b &= \rho_0^\alpha(-\alpha\rho_0^{-\frac{\alpha}{2}-1}\langle b, \nabla\rho_0\rangle)\rho_0^{\frac{\alpha}{2}} b + O(t)\\
&= -\alpha\rho_0^{\alpha-1}\langle b, \nabla\rho_0\rangle b + O(t)
\end{aligned}
$$

as well as

$$
\begin{aligned}
-\frac{1}{2}\rho^\alpha\langle \dot\gamma_b, \dot\gamma_b\rangle\nabla\rho^{-\alpha} &= -\frac{1}{2}\rho_0^\alpha\,(\rho_0^\alpha)\left(-\alpha\rho_0^{-\alpha-1}\nabla\rho_0\right) + O(t)\\
&= \frac{\alpha}{2}\rho_0^{\alpha-1}\nabla\rho_0 + O(t).
\end{aligned}
$$

Since $\ddot\gamma_b(t) = 2v + O(t)$, plugging into (59) gives:

$$
2v - \alpha\rho_0^{\alpha-1}\langle b, \nabla\rho_0\rangle b + \frac{\alpha}{2}\rho_0^{\alpha-1}\nabla\rho_0 + O(t) = 0.
$$

Since the above must hold for arbitrarily small t, we obtain

$$
v = \alpha\rho_0^{\alpha-1}\left(\frac{1}{2}\langle b, \nabla\rho_0\rangle b - \frac{1}{4}\nabla\rho_0\right).
$$

We have thus established that the unit geodesic in direction b has form:

$$
\gamma_b(t) = \rho_0^{\frac{\alpha}{2}} bt + \alpha\rho_0^{\alpha-1}\left(\frac{1}{2}\langle b, \nabla\rho_0\rangle b - \frac{1}{4}\nabla\rho_0\right)t^2 + O(t^3)\,.
$$

We can now compute q to obtain a higher order expansion; we need to compute the linearization of each term in the geodesic equation. We have three terms:

$$
\begin{aligned}
&\text{(I)} := \ddot\gamma_b = 2v + 6qt + O(t^2)\\
&\text{(II)} := \rho^\alpha\langle \dot\gamma_b, \nabla\rho^{-\alpha}\rangle\dot\gamma_b\\
&\quad = (\rho_0^\alpha + zt)(-\alpha\rho_0^{\frac{-\alpha}{2}-1}\langle b, \nabla\rho_0\rangle + \rho_0^{\frac{\alpha}{2}}\langle b, w\rangle t - 2\alpha\rho_0^{-\alpha-1}\langle v, \nabla\rho_0\rangle t)(\rho_0^{\frac{\alpha}{2}} b + 2vt) + O(t^2)\\
&\quad = -\alpha\rho_0^{\alpha-1}\langle b, \nabla\rho_0\rangle b - \alpha\rho_0^{-1} z\langle b, \nabla\rho_0\rangle bt + \rho_0^{\frac{3}{2}\alpha} b\left[\rho_0^{\frac{\alpha}{2}}\langle b, w\rangle t - 2\alpha\rho_0^{-\alpha-1}\langle v, \nabla\rho_0\rangle t\right]\\
&\qquad - 2\alpha\rho_0^{\frac{\alpha}{2}-1}\langle b, \nabla\rho_0\rangle vt + O(t^2)
\end{aligned}
$$

$$
\begin{aligned}
&= -\alpha\rho_0^{\alpha-1}\langle b, \nabla\rho_0\rangle b - \alpha\rho_0^{-1}z\langle b, \nabla\rho_0\rangle bt + \rho_0^{2\alpha}\langle b, w\rangle bt - 2\alpha\rho_0^{\frac{\alpha}{2}-1}\langle v, \nabla\rho_0\rangle bt \\
&\quad - 2\alpha\rho_0^{\frac{\alpha}{2}-1}\langle b, \nabla\rho_0\rangle vt + O(t^2) \\
&= -\alpha\rho_0^{\alpha-1}\langle b, \nabla\rho_0\rangle b + \left(-\alpha\rho_0^{-1}z\langle b, \nabla\rho_0\rangle + \rho_0^{2\alpha}\langle b, w\rangle - 2\alpha\rho_0^{\frac{\alpha}{2}-1}\langle v, \nabla\rho_0\rangle\right) bt \\
&\quad - 2\alpha\rho_0^{\frac{\alpha}{2}-1}\langle b, \nabla\rho_0\rangle vt + O(t^2) \\
&= -\alpha\rho_0^{\alpha-1}\langle b, \nabla\rho_0\rangle b + \left(-\alpha\rho_0^{-1}z\langle b, \nabla\rho_0\rangle + \rho_0^{2\alpha}\langle b, w\rangle - 2\alpha\rho_0^{\frac{\alpha}{2}-1}\langle v, \nabla\rho_0\rangle\right) bt \\
&\quad - 2\alpha\rho_0^{\frac{\alpha}{2}-1}\langle b, \nabla\rho_0\rangle - \alpha\rho_0^{\alpha-1}\left(\frac{1}{4}\nabla\rho_0 - \frac{1}{2}\langle b, \nabla\rho_0\rangle b\right) t + O(t^2) \\
&= -\alpha\rho_0^{\alpha-1}\langle b, \nabla\rho_0\rangle b + C_1 bt + \frac{\alpha^2}{2}\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle\nabla\rho_0 t + O(t^2)
\end{aligned}
$$

where

$$
\begin{aligned}
C_1 &= -\alpha\rho_0^{-1}z\langle b, \nabla\rho_0\rangle + \rho_0^{2\alpha}\langle b, w\rangle - 2\alpha\rho_0^{\frac{\alpha}{2}-1}\langle v, \nabla\rho_0\rangle - \alpha^2\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle^2 \\
&= -\alpha^2\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle^2 + \rho_0^{2\alpha}\left(-\alpha(-\alpha-1)\rho_0^{-\frac{\alpha}{2}-2}\langle\nabla\rho_0, b\rangle^2 - \alpha\rho_0^{-\frac{\alpha}{2}-1}\langle H_0 b, b\rangle\right) \\
&\quad - 2\alpha\rho_0^{\frac{\alpha}{2}-1}\left(-\frac{\alpha}{4}\rho_0^{\alpha-1}\langle\nabla\rho_0, \nabla\rho_0\rangle + \frac{\alpha}{2}\rho_0^{\alpha-1}\langle b, \nabla\rho_0\rangle^2\right) - \alpha^2\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle^2 \\
&= -2\alpha^2\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle^2 + (\alpha^2+\alpha)\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle^2 - \alpha\rho_0^{\frac{3}{2}\alpha-1}\langle H_0 b, b\rangle \\
&\quad + \frac{\alpha^2}{2}\rho_0^{\frac{3}{2}\alpha-2}\langle\nabla\rho_0, \nabla\rho_0\rangle - \alpha^2\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle^2 \\
&= (-2\alpha^2+\alpha)\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle^2 - \alpha\rho_0^{\frac{3}{2}\alpha-1}\langle H_0 b, b\rangle + \frac{\alpha^2}{2}\rho_0^{\frac{3}{2}\alpha-2}\langle\nabla\rho_0, \nabla\rho_0\rangle.
\end{aligned}
$$

Similarly:

$$
\begin{aligned}
\text{(III)} &:= -\frac{1}{2}\rho^\alpha\langle\dot\gamma_b, \dot\gamma_b\rangle\nabla\rho^{-\alpha} \\
&= -\frac{1}{2}\left(\rho_0^\alpha + zt\right)\left(\rho_0^\alpha + 4\rho_0^{\frac{\alpha}{2}}\langle b, v\rangle t\right)\left(-\alpha\rho_0^{-\alpha-1}\nabla\rho + wt\right) + O(t^2) \\
&= -\frac{1}{2}\left(-\alpha\rho_0^{\alpha-1}\nabla\rho_0 + \rho_0^{2\alpha}wt - 4\alpha\rho_0^{\frac{\alpha}{2}-1}\langle b, v\rangle\nabla\rho_0 t - \alpha\rho_0^{-1}z\nabla\rho_0 t\right) + O(t^2) \\
&= \frac{\alpha}{2}\rho_0^{\alpha-1}\nabla\rho_0 - \frac{1}{2}\rho_0^{2\alpha}wt + 2\alpha\rho_0^{\frac{\alpha}{2}-1}\langle b, v\rangle\nabla\rho_0 t + \frac{\alpha}{2}\rho_0^{-1}z\nabla\rho t + O(t^2) \\
&= \frac{\alpha}{2}\rho_0^{\alpha-1}\nabla\rho_0 - \frac{1}{2}\rho_0^{2\alpha}\left((\alpha^2+\alpha)\rho_0^{-\frac{\alpha}{2}-2}\langle\nabla\rho_0, b\rangle\nabla\rho_0 - \alpha\rho_0^{-\frac{\alpha}{2}-1}H_0 b\right) t \\
&\quad 2\alpha\rho_0^{\frac{\alpha}{2}-1}\langle b, v\rangle\nabla\rho_0 t + \frac{\alpha}{2}\rho_0^{-1}(\alpha\rho_0^{\frac{3}{2}\alpha-1})\langle\nabla\rho_0, b\rangle\nabla\rho_0 t + O(t^2) \\
&= \frac{\alpha}{2}\rho_0^{\alpha-1}\nabla\rho_0 + \left(-\frac{\alpha}{2} - \frac{\alpha^2}{2}\right)\rho_0^{\frac{3}{2}\alpha-2}\langle\nabla\rho_0, b\rangle\nabla\rho_0 t + \frac{\alpha}{2}\rho_0^{\frac{3}{2}\alpha-1}H_0 bt \\
&\quad + 2\alpha\rho_0^{\frac{\alpha}{2}-1}\langle b, v\rangle\nabla\rho_0 t + \frac{\alpha^2}{2}\rho_0^{\frac{3}{2}\alpha-2}\langle\nabla\rho_0, b\rangle\nabla\rho_0 t + O(t^2) \\
&= \frac{\alpha}{2}\rho_0^{\alpha-1}\nabla\rho_0 + C_2\nabla\rho_0 t + \frac{\alpha}{2}\rho_0^{\frac{3}{2}\alpha-1}H_0 bt + O(t^2)
\end{aligned}
$$

where

$$\begin{aligned} C_2 &= -\frac{\alpha}{2}\rho_0^{\frac{3}{2}\alpha-2}\langle \nabla\rho_0, b\rangle + 2\alpha\rho_0^{\frac{\alpha}{2}-1}\langle b, v\rangle \\ &= -\frac{\alpha}{2}\rho_0^{\frac{3}{2}\alpha-2}\langle \nabla\rho_0, b\rangle + 2\alpha\rho_0^{\frac{\alpha}{2}-1}\left(-\frac{\alpha}{4}\rho_0^{\alpha-1}\langle \nabla\rho_0, b\rangle + \frac{\alpha}{2}\rho_0^{\alpha-1}\langle b, \nabla\rho_0\rangle\right) \\ &= \left(\frac{\alpha^2}{2} - \frac{\alpha}{2}\right)\rho_0^{\frac{3}{2}\alpha-2}\langle \nabla\rho_0, b\rangle. \end{aligned}$$

The linear terms in the geodesic equation must sum to zero, and we obtain:

$$\begin{aligned} 6qt + C_1 bt + \frac{\alpha^2}{2}\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle\nabla\rho_0 t + C_2\nabla\rho_0 t + \frac{\alpha}{2}\rho_0^{\frac{3}{2}\alpha-1}H_0 bt = 0 \\ \implies \quad 6q + C_1 b + C_3\nabla\rho_0 + \frac{\alpha}{2}\rho_0^{\frac{3}{2}\alpha-1}H_0 b = 0 \end{aligned}$$

where

$$C_3 = C_2 + \frac{\alpha^2}{2}\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle = \left(\alpha^2 - \frac{\alpha}{2}\right)\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle.$$

We thus obtain:

$$\begin{aligned} q &= -\frac{1}{6}C_1 b - \frac{\alpha}{12}\rho_0^{\frac{3}{2}\alpha-1}H_0 b - \frac{C_3}{6}\nabla\rho_0, \\ &= C_1' b + C_2' H_0 b + C_3'\nabla\rho_0, \end{aligned}$$

where

$$\begin{aligned} C_1' &= \left(\frac{1}{3}\alpha^2 - \frac{1}{6}\alpha\right)\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle^2 + \frac{\alpha}{6}\rho_0^{\frac{3}{2}\alpha-1}\langle Hb, b\rangle - \frac{\alpha^2}{12}\rho_0^{\frac{3}{2}\alpha-2}\langle \nabla\rho_0, \nabla\rho_0\rangle, \\ C_2' &= -\frac{\alpha}{12}\rho_0^{\frac{3}{2}\alpha-1}, \\ C_3' &= \left(\frac{\alpha}{12} - \frac{\alpha^2}{6}\right)\rho_0^{\frac{3}{2}\alpha-2}\langle \nabla\rho_0, b\rangle. \end{aligned}$$

In summary we have the following geodesic expansion depending on $b, \rho_0, \nabla\rho_0, H_0$:

$$\gamma_b(t) = \rho_0^{\frac{\alpha}{2}} bt + \alpha\rho_0^{\alpha-1}\left(\frac{1}{2}\langle b, \nabla\rho_0\rangle b - \frac{1}{4}\nabla\rho_0\right)t^2 + (C_1' b + C_2' H_0 b + C_3'\nabla\rho_0)t^3 + O(t^4)\,.$$

■

Appendix D. Local Euclidean Equivalence

This appendix contains the proof of Theorem 4.2.

Theorem 4.2 *Let $\Omega \subseteq \mathbb{R}^m$ be an m-dimensional, open, connected domain and assume (Ω, ρ) satisfies Assumption 3.1. Then for $x, y \in \Omega$ with $x \neq y$ and unit vector $u = (y - x)/\|y - x\|$, we can relate Euclidean and Fermat distance as follows:*

$$\begin{aligned}\|y - x\| = \rho(x)^{\frac{p-1}{m}} \mathcal{L}_p^p(x, y) + \frac{1}{2}\left(\frac{p-1}{m}\right) \langle u, \nabla\rho(x)\rangle \, \rho(x)^{\frac{2(p-1)}{m}-1} \mathcal{L}_p^{2p}(x, y) \\ + C\mathcal{L}_p^{3p}(x, y) + O(\mathcal{L}_p^{4p}(x, y))\end{aligned}$$

for

$$C = \rho(x)^{\frac{3}{2}\alpha-2}\left[\frac{\alpha^2}{96}\langle\nabla\rho(x), \nabla\rho(x)\rangle + \left(\frac{7\alpha^2}{96} - \frac{\alpha}{12}\right)\langle u, \nabla\rho(x)\rangle^2 + \frac{\alpha}{12}\rho(x)\,\langle H(x)u, u\rangle\right].$$

Also:

$$\mathcal{L}_p^p(x, y) = \frac{1}{\rho(x)^{\frac{p-1}{m}}}\left(\|y - x\| - \frac{1}{2}\left(\frac{p-1}{m}\right)\left\langle u, \frac{\nabla\rho(x)}{\rho(x)}\right\rangle \|y - x\|^2\right) + O\left(\|y - x\|^3\right).$$

Proof Without loss of generality assume $x = 0$, and let $\epsilon = \|y\|$. Now consider $B(0, \epsilon)$, a Euclidean ball of radius ϵ about 0. As long as ϵ is not too large, each point on $\partial B(0, \epsilon)$ is on a unique $\mathcal{L}_p^p$ geodesic curve leaving the origin; let γ_b be the geodesic which goes through y (recall it is unit speed in the direction of unit vector b). Note in general $b \neq u_y = \frac{y}{\|y\|}$, but these vectors are close for ϵ small.

For notational brevity we denote $\mathcal{L}_p^p(x, y), \rho(0), \nabla\rho(0), H(0)$ by $\mathcal{L}, \rho_0, \nabla\rho_0, H_0$ throughout the proof. Note since the geodesic is unit speed, γ_b reaches y at time $\mathcal{L}$, *i.e.,* $y = \gamma_b(\mathcal{L})$. By Theorem 4.1, we have

$$\gamma_b(\mathcal{L}) = \rho_0^{\frac{\alpha}{2}} b\mathcal{L} + \alpha\rho_0^{\alpha-1}\left(\frac{1}{2}\langle b, \nabla\rho_0\rangle b - \frac{1}{4}\nabla\rho_0\right)\mathcal{L}^2 + (C_1' b + C_2' H_0 b + C_3'\nabla\rho_0)\mathcal{L}^3 + O(\mathcal{L}^4). \tag{60}$$

Thus for

$$C_4 = \frac{\alpha^2}{16}\rho_0^{2\alpha-2}\langle\nabla\rho_0, \nabla\rho_0\rangle + 2\rho_0^{\frac{\alpha}{2}}\left(C_1' + C_2'\langle H_0 b, b\rangle + C_3'\langle\nabla\rho_0, b\rangle\right)$$

we have

$$\begin{aligned}\epsilon &= \|\gamma_b(\mathcal{L})\| \\ &= \sqrt{\rho_0^\alpha\mathcal{L}^2 - 2\alpha\rho_0^{\frac{3}{2}\alpha-1}\left(\frac{1}{4}\langle b, \nabla\rho_0\rangle - \frac{1}{2}\langle b, \nabla\rho_0\rangle\right)\mathcal{L}^3 + C_4\mathcal{L}^4 + O(\mathcal{L}^5)} \\ &= \sqrt{\rho_0^\alpha\mathcal{L}^2 + \frac{1}{2}\alpha\rho_0^{\frac{3}{2}\alpha-1}\langle b, \nabla\rho_0\rangle\mathcal{L}^3 + C_4\mathcal{L}^4 + O(\mathcal{L}^5)} \\ &= \rho_0^{\frac{\alpha}{2}}\mathcal{L}\sqrt{1 + \frac{1}{2}\alpha\rho_0^{\frac{\alpha}{2}-1}\langle b, \nabla\rho_0\rangle\mathcal{L} + \rho_0^{-\alpha}C_4\mathcal{L}^2 + O(\mathcal{L}^3)} \\ &= \rho_0^{\frac{\alpha}{2}}\mathcal{L}\left(1 + \frac{1}{4}\alpha\rho_0^{\frac{\alpha}{2}-1}\langle b, \nabla\rho_0\rangle\mathcal{L} + \left(\frac{C_4}{2}\rho_0^{-\alpha} - \frac{1}{32}\alpha^2\rho_0^{\alpha-2}\langle b, \nabla\rho_0\rangle^2\right)\mathcal{L}^2 + O(\mathcal{L}^3)\right)\end{aligned}$$

$$= \rho_0^{\frac{\alpha}{2}}\mathcal{L} + \frac{1}{4}\alpha\rho_0^{\alpha-1}\langle b, \nabla\rho_0\rangle\mathcal{L}^2 + \left(\frac{C_4}{2}\rho_0^{-\frac{\alpha}{2}} - \frac{1}{32}\alpha^2\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle^2\right)\mathcal{L}^3 + O(\mathcal{L}^4)$$

so that

$$\epsilon = \rho_0^{\frac{\alpha}{2}}\mathcal{L} + \frac{1}{4}\alpha\rho_0^{\alpha-1}\langle b, \nabla\rho_0\rangle\mathcal{L}^2 + C_4'\mathcal{L}^3 + O(\mathcal{L}^4) \tag{61}$$

for

$$\begin{aligned}
C_4' &= \frac{1}{2}\rho_0^{-\frac{\alpha}{2}}\left(\frac{\alpha^2}{16}\rho_0^{2\alpha-2}\langle\nabla\rho_0, \nabla\rho_0\rangle + 2\rho_0^{\frac{\alpha}{2}}(C_1' + C_2'\langle H_0 b, b\rangle + C_3'\langle\nabla\rho_0, b\rangle)\right) \\
&\quad - \frac{1}{32}\alpha^2\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle^2 \\
&= \frac{\alpha^2}{32}\rho_0^{\frac{3}{2}\alpha-2}\langle\nabla\rho_0, \nabla\rho_0\rangle + (C_1' + C_2'\langle H_0 b, b\rangle + C_3'\langle\nabla\rho_0, b\rangle) - \frac{1}{32}\alpha^2\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle^2 \\
&= -\frac{5\alpha^2}{96}\rho_0^{\frac{3}{2}\alpha-2}\langle\nabla\rho_0, \nabla\rho_0\rangle + \left(\frac{13}{96}\alpha^2 - \frac{\alpha}{12}\right)\rho_0^{\frac{3}{2}\alpha-2}\langle b, \nabla\rho_0\rangle^2 + \frac{\alpha}{12}\rho_0^{\frac{3}{2}\alpha-1}\langle H_0 b, b\rangle\,.
\end{aligned}$$

We now relate b with u_y to obtain an expansion independent of b. Combining (60) and (61), we obtain:

$$\begin{aligned}
u_y &= \frac{\rho_0^{\frac{\alpha}{2}}\mathcal{L}\left(b + \alpha\rho_0^{\frac{\alpha}{2}-1}(\frac{1}{2}\langle b, \nabla\rho_0\rangle - \frac{1}{4}\nabla\rho_0)\mathcal{L} + O(\mathcal{L}^2)\right)}{\rho_0^{\frac{\alpha}{2}}\mathcal{L}\left(1 + \frac{\alpha}{4}\rho_0^{\frac{\alpha}{2}-1}\langle b, \nabla\rho_0\rangle\mathcal{L} + O(\mathcal{L}^2)\right)} \\
&= b + \frac{\alpha}{4}\rho_0^{\frac{\alpha}{2}-1}\langle b, \nabla\rho_0\rangle b\mathcal{L} - \frac{\alpha}{4}\rho_0^{\frac{\alpha}{2}-1}\nabla\rho_0\mathcal{L} + O(\mathcal{L}^2)\,.
\end{aligned}$$

Since $b = u_y + O(\mathcal{L})$, we obtain:

$$b = u_y - \frac{\alpha}{4}\rho_0^{\frac{\alpha}{2}-1}\langle u_y, \nabla\rho_0\rangle u_y\mathcal{L} + \frac{\alpha}{4}\rho_0^{\frac{\alpha}{2}-1}\nabla\rho_0\mathcal{L} + O(\mathcal{L}^2)\,.$$

Plugging the above into (61), we obtain:

$$\epsilon = \rho_0^{\frac{\alpha}{2}}\mathcal{L} + \frac{\alpha}{4}\rho_0^{\alpha-1}\langle u_y, \nabla\rho_0\rangle\mathcal{L}^2 + C_5'\mathcal{L}^3 + O(\mathcal{L}^4)\,, \tag{62}$$

where

$$\begin{aligned}
C_5' &= C_4' - \frac{\alpha^2}{16}\rho_0^{\frac{3}{2}\alpha-2}\langle u_y, \nabla\rho_0\rangle^2 + \frac{\alpha^2}{16}\rho_0^{\frac{3}{2}\alpha-2}\langle\nabla\rho_0, \nabla\rho_0\rangle \\
&= \rho_0^{\frac{3}{2}\alpha-2}\left[\frac{\alpha^2}{96}\langle\nabla\rho_0, \nabla\rho_0\rangle + \left(\frac{7\alpha^2}{96} - \frac{\alpha}{12}\right)\langle u_y, \nabla\rho_0\rangle^2 + \frac{\alpha}{12}\rho_0\langle H_0 u_y, u_y\rangle\right]\,,
\end{aligned}$$

and we obtain the first statement in the theorem. Rearranging (62) yields:

$$O(\mathcal{L}^3) + \frac{\alpha}{4}\rho_0^{\frac{\alpha}{2}-1}\langle u_y, \nabla\rho_0\rangle\mathcal{L}^2 + \mathcal{L} - \rho_0^{-\frac{\alpha}{2}}\epsilon = 0\,. \tag{63}$$

We now solve for $\mathcal{L}$, which is the root of (63) satisfying $\mathcal{L} \sim \epsilon$ as $\epsilon \to 0$. Expanding $\mathcal{L}(\epsilon) = c_0 + c_1\epsilon + c_2\epsilon^2 + \ldots$, plugging into (63), and solving for the coefficients c_i, one obtains that any root satisfying $c_0 = 0$ has form:

$$\begin{aligned}\mathcal{L} &= \rho_0^{-\frac{\alpha}{2}}\epsilon - \frac{\alpha}{4}\rho_0^{-\frac{\alpha}{2}-1}\langle u_y, \nabla\rho_0\rangle\epsilon^2 + O(\epsilon^3) \\ &= \rho_0^{-\frac{(p-1)}{m}}\|y\| - \frac{1}{2}\left(\frac{p-1}{m}\right)\rho_0^{-\frac{(p-1)}{m}-1}\langle u_y, \nabla\rho_0\rangle \cdot \|y\|^2 + O(\|y\|^3),\end{aligned}$$

which proves the second theorem statement. ■

Appendix E. Metric Approximation

Lemma E.1 (Fermat Paths are Local) *Choose $R > 0$ such that $4R \leq \mathcal{R}$ and suppose $x, y \in \mathcal{B}_z(R)$. Then $\tilde{\ell}_p^p(x, y, H_{n\rho}) = \tilde{\ell}_p^p(x, y, H_{n\rho} \cap \mathcal{B}_z(R))$ with probability $1 - \exp\left(-cn^{\frac{1}{m+2p}}\right)$.*

Proof This is shown in the proof of (Hwang et al., 2016, Lemma 10), see pp. 2807, using the conclusion of (Hwang et al., 2016, Corollary 9). ■

Lemma E.2 (Curvature Perturbation for Discrete Metric) *Suppose $d(x, y) \leq C_{\mathcal{M}}$. Then*

$$\tilde{\ell}_p^p(x, y, H_{n\rho} \cap \mathcal{B}_x(R)) = \left(1 \pm Cp(K + \mathcal{R}^{-2})d(x, y)^2\right)\tilde{\ell}_p^p(0, u, H_{n\rho_x}) \ .$$

Proof Define $u := \log_x(y)$. Note $0 = \log_x(x)$ and $d(x, y) = \|\log_x(x) - \log_x(y)\| = \|u\|$. Let $\pi = \{0 = \log_x(x_0), \log_x(x_1), \ldots, \log_x(x_L) = u\}$ be the optimal path for $\tilde{\ell}_p^p(0, u, H_{n\rho_x})$. By (13),

$$\begin{aligned}\|x_{i+1} - x_i\| \leq d(x_{i+1}, x_i) &\leq \|\log_x(x_{i+1}) - \log_x(x_i)\| + CK\|\log_x(x_{i+1}) - \log_x(x_i)\|^3 \\ &\leq \|\log_x(x_{i+1}) - \log_x(x_i)\|\left(1 + CK\|\log_x(x) - \log_x(y)\|^2\right) \\ &= \|\log_x(x_{i+1}) - \log_x(x_i)\|\left(1 + CKd(x, y)^2\right).\end{aligned}$$

Thus:

$$\begin{aligned}\tilde{\ell}_p^p(x, y, H_{n\rho} \cap \mathcal{B}_x(R)) &\leq n^{\frac{p-1}{m}}\sum\|x_{i+1} - x_i\|^p \\ &\leq n^{\frac{p-1}{m}}\sum\|\log_x(x_{i+1}) - \log_x(x_i)\|^p\left(1 + CKd(x, y)^2\right)^p \\ &\leq (1 + CpKd(x, y)^2 + O(d(x, y)^4))\tilde{\ell}_p^p(0, u, H_{n\rho_x}) \ .\end{aligned}$$

Now let $\pi = \{x = x_0, x_1, \ldots, x_L = y\}$ be the optimal path for $\tilde{\ell}_p^p(x, y, H_{n\rho} \cap \mathcal{B}_x(R))$. Note (13) implies

$$\|\log_x(x_{i+1}) - \log_x(x_i)\| = d(x_{i+1}, x_i) \pm CKd(x_{i+1}, x_i)^3 + O(d(x_{i+1}, x_i)^5) \, .$$

Combining the above with (14) gives

$$\begin{aligned}\|\log_x(x_{i+1}) - \log_x(x_i)\| &= \|x_{i+1} - x_i\| \left(1 \pm C(K + \mathcal{R}^{-2})\|x_{i+1} - x_i\|^2 + O(\|x_{i+1} - x_i\|^4)\right) \\ &\leq \|x_{i+1} - x_i\| \left(1 \pm C(K + \mathcal{R}^{-2})\|x - y\|^2 + O(\|x - y\|^4)\right) \\ &\leq \|x_{i+1} - x_i\| \left(1 \pm C(K + \mathcal{R}^{-2})d(x,y)^2 + O(d(x,y)^4)\right) ,\end{aligned}$$

since the optimality of π ensures $\|x_{i+1} - x_i\| \leq \|x - y\|$. We thus obtain:

$$\begin{aligned}\tilde{\ell}_p^p(0, u, H_{n\rho_x}) &\leq n^{\frac{p-1}{m}} \sum \|\log_x(x_{i+1}) - \log_x(x_i)\|^p \\ &\leq n^{\frac{p-1}{m}} \sum \|x_{i+1} - x_i\|^p \left(1 + C(K + \mathcal{R}^{-2})d(x,y)^2 + O(d(x,y)^4)\right)^p \\ &= n^{\frac{p-1}{m}} \sum \|x_{i+1} - x_i\|^p \left(1 + Cp(K + \mathcal{R}^{-2})d(x,y)^2 + O(d(x,y)^4)\right) \\ &= \left(1 + Cp(K + \mathcal{R}^{-2})d(x,y)^2 + O(d(x,y)^4)\right) \tilde{\ell}_p^p(x, y, H_{n\rho} \cap \mathcal{B}_x(R)).\end{aligned}$$

For $d(x,y) \leq C_{\mathcal{M}}$, we can remove the fourth order term by increasing the constant on the second order term, and we obtain

$$\frac{\tilde{\ell}_p^p(0, u, H_{n\rho_x})}{1 + Cp(K + \mathcal{R}^{-2})d(x,y)^2} \leq \tilde{\ell}_p^p(x, y, H_{n\rho} \cap \mathcal{B}_x(R)) \leq \left(1 + Cp(K + \mathcal{R}^{-2})d(x,y)^2\right) \tilde{\ell}_p^p(0, u, H_{n\rho_x})$$

which proves the lemma. ■

Lemma E.3 (Locality of Homogeneous Paths) *Suppose*

$$\mathcal{C}_{p,m,\beta,K} \geq \|u\| \geq (\frac{n}{2\beta})^{-\frac{1}{m}(\frac{1}{3} - \epsilon)}.$$

Then $\ell_p^p(0, u, H_{n\rho_{\min}}) = \ell_p^p(0, u, \overline{H}_{n\rho_{\min}})$ *with probability at least* $1 - C_\epsilon n \exp\left(-c_\epsilon (\frac{n}{2\beta})^{\frac{2\epsilon}{3} \min\{\frac{1}{p}, \frac{1}{m}\}}\right)$, *and same for* $H_{n\rho_{\max}}$.

Proof The proof is similar to that of (Hwang et al., 2016, Thm. 7). For completeness, we reprove this lemma here. Suppose, for the sake of contradiction, that the optimal path leaves $B_0(2r)$ where $r := \|u\|$. Then $\ell_p^p(0, \partial B_0(2r), H_{n\rho_{\min}}) \leq \ell_p^p(0, u, H_{n\rho_{\min}})$, where

$$\ell_p^p(0, \partial B_0(2r), H_{n\rho_{\min}}) = \min_{|v|=2r} \ell_p^p(0, v, H_{n\rho_{\min}}).$$

Applying Proposition 4.4 yields

$$(n\rho_{\min})^{\frac{p-1}{m}} \ell_p^p(0, \partial B_0(2r), H_{n\rho_{\min}}) \leq (n\rho_{\min})^{\frac{p-1}{m}} \ell_p^p(0, u, H_{n\rho_{\min}}) \leq \mu\|u\| + \|u\|^2 = \mu r + r^2 \tag{64}$$

with probability at least $1 - C_\epsilon \exp\left(-c_\epsilon (n\rho_{\min})^{\frac{2\epsilon}{3} \min\{\frac{1}{p}, \frac{1}{m}\}}\right)$ if $\|u\| \geq (n\rho_{\min})^{-\frac{1}{m}(\frac{1}{3} - \epsilon)}$.

On the other hand, let $\delta = 3rn^{-\frac{1}{m}}$. Then, by (Vershynin, 2018, Cor. 4.2.13), we may find a δ-net of points $v_1, \ldots, v_n$, as the covering number of $S^{m-1}(2r)$ is upper-bounded by $\left(\frac{3r}{\delta}\right)^m$. Again by Proposition 4.4:

$$(n\rho_{\min})^{\frac{p-1}{m}} \ell_p^p(0, v_i, H_{n\rho_{\min}}) \geq \mu\|v_i\| - \|v_i\|^2 = 2\mu r - 4r^2 \quad \text{for } i = 1, \ldots, n$$

$$\Rightarrow \min_{i=1,\dots,n} (n\rho_{\min})^{\frac{p-1}{m}} \ell_p^p(0, v_i, H_{n\rho_{\min}}) \geq \mu\|v_i\| - \|v_i\|^2 = 2\mu r - 4r^2$$

with probability at least $1 - C_\epsilon n \exp\left(-c_\epsilon (n\rho_{\min})^{\frac{2\epsilon}{3}\min\{\frac{1}{p},\frac{1}{m}\}}\right)$, via a union bound. But also

$$\begin{aligned} \min_{|v|=2r} \ell_p^p(0, v, H_{n\rho_{\min}}) &\geq \min_{i=1,\dots,n} \ell_p^p(0, v_i, H_{n\rho_{\min}}) - \delta^p \\ \Rightarrow (n\rho_{\min})^{\frac{p-1}{m}} \ell_p^p(0, \partial B_0(2r), H_{n\rho_{\min}}) &\geq 2\mu r - 4r^2 - \frac{3^p r^p g_{\min}^{\frac{p-1}{m}}}{n^{\frac{1}{m}}}. \end{aligned} \tag{65}$$

Combining (64) and (65) yields

$$5r^2 + \frac{3^p r^p g_{\min}^{\frac{p-1}{m}}}{n^{\frac{1}{m}}} \geq \mu r.$$

This yields a contradiction for $r = \|u\| \leq \mathcal{C}_{p,m,\beta,K}$. For n sufficiently large, we have $\mathcal{C}_{p,m,\beta,K} > (\frac{n}{2\beta})^{-\frac{1}{m}(\frac{1}{3}-\epsilon)}$. ∎

Lemma E.4 (Discrete to Continuum Approximation in the Tangent Plane) *Fix $\epsilon \in (0, 1/(8p+6))$ and suppose*

$$\left(\frac{n}{2\beta}\right)^{-\frac{1}{m}(\frac{1}{3}-\epsilon)} \leq \|u\| \leq \mathcal{C}_{p,m,\beta,K}\,.$$

For notational convenience define $\mathcal{L} = \mathcal{L}_p^p(0, u, \rho_x)$. Then we have

$$|\tilde{\ell}_p^p(0, u, H_{n\rho_x}) - \mu\mathcal{L}| \leq \tilde{C}_1\mathcal{L}^2 + \tilde{C}_2\mathcal{L}^3 + O(\mathcal{L}^4)$$

with probability at least $1 - C_\epsilon n \exp\left(-c_\epsilon (\frac{n}{2\beta})^{\frac{2\epsilon}{3}\min\{\frac{1}{m},\frac{1}{p}\}}\right)$, where

$$\tilde{C}_1 := \beta^{\frac{p-1}{m}} \left(\frac{5\mu}{2}\left(\frac{p-1}{m}\right)\beta L_1 + 1\right)\,,\ \tilde{C}_2 := C_{p,d,\beta}\left(K(1+L_1) + L_1^2 + L_2\right)\,.$$

Proof Note throughout the proof we let $C_{p,d,\beta}$ be a constant depending on p, d, β whose value may change line to line. We know $\ell_p^p(0, u, H_{n\rho_{\min}}) = \ell_p^p(0, u, \overline{H}_{n\rho_{\min}})$ and $\ell_p^p(0, u, H_{n\rho_{\max}}) = \ell_p^p(0, u, \overline{H}_{n\rho_{\max}})$ w.h.p. by Lemma E.3. Since we can couple the PPPs so that $H_{n\rho_{\min}} \subseteq H_{n\rho_x} \subseteq H_{n\rho_{\max}}$, we obtain:

$$\begin{aligned} &\ell_p^p(0, u, \overline{H}_{n\rho_{\max}}) \leq \ell_p^p(0, u, H_{n\rho_x}) \leq \ell_p^p(0, u, \overline{H}_{n\rho_{\min}}) \\ \implies &n^{\frac{(p-1)}{m}} \ell_p^p(0, u, \overline{H}_{n\rho_{\max}}) \leq n^{\frac{(p-1)}{m}} \ell_p^p(0, u, H_{n\rho_x}) \leq n^{\frac{(p-1)}{m}} \ell_p^p(0, u, \overline{H}_{n\rho_{\min}}). \end{aligned}$$

By applying Proposition 4.4 with $q = 2$:

$$\ell_p^p(0, u, \overline{H}_{\rho_{\min} n}) = \frac{1}{(n\rho_{\min})^{\frac{(p-1)}{m}}} \left(\mu\|u\| \pm \|u\|^2\right) \tag{66}$$

for $\|u\| \geq (n\rho_{\min})^{-\frac{1}{m}(\frac{1}{3}-\epsilon)}$ with probability at least $1 - C_\epsilon \exp\left(-c_\epsilon (n\rho_{\min})^{\frac{2\epsilon}{3}\min\{\frac{1}{m},\frac{1}{p}\}}\right)$. From Theorem 4.2, we have:

$$\begin{aligned} \|u\| &= \rho_x(0)^{\frac{p-1}{m}}\mathcal{L} \pm \frac{1}{2}\left(\frac{p-1}{m}\right)\rho_x(0)^{\frac{2(p-1)}{m}-1}\|\nabla\rho_x(0)\|\mathcal{L}^2 + C\mathcal{L}^3 + O(\mathcal{L}^4) \\ \implies \|u\|^2 &= \rho_x(0)^{\frac{2(p-1)}{m}}\mathcal{L}^2 \pm \left(\frac{p-1}{m}\right)\rho_x(0)^{\frac{3(p-1)}{m}-1}\|\nabla\rho_x(0)\|\mathcal{L}^3 + O(\mathcal{L}^4) \end{aligned}$$

where $|C| \leq C_{p,d,\beta}(\|\nabla\rho_x(0)\|^2 + \|H\rho_x(0)\|)$. Returning to (66) we obtain:

$$\begin{aligned} & n^{\frac{(p-1)}{m}}\ell_p^p(0,u,\overline{H}_{n\rho_{\min}}) \\ =& \frac{1}{\rho_{\min}^{\frac{(p-1)}{m}}}\left(\mu\|u\| \pm \|u\|^2\right) \\ \leq& \mu\rho_{\min}^{\frac{-(p-1)}{m}}\left(\rho_x(0)^{\frac{p-1}{m}}\mathcal{L} + \frac{1}{2}\left(\frac{p-1}{m}\right)\rho_x(0)^{\frac{2(p-1)}{m}-1}\|\nabla\rho_x(0)\|\mathcal{L}^2 + \mu^{-1}\rho_x(0)^{\frac{2(p-1)}{m}}\mathcal{L}^2 + \tilde{C}\mathcal{L}^3 + O(\mathcal{L}^4)\right) \\ \leq& \left(\frac{\rho_x(0)}{\rho_{\min}}\right)^{\frac{p-1}{m}}\left(\mu\mathcal{L} + \frac{\mu}{2}\left(\frac{p-1}{m}\right)\rho_x(0)^{\frac{(p-1)}{m}-1}\|\nabla\rho_x(0)\|\mathcal{L}^2 + \rho_x(0)^{\frac{(p-1)}{m}}\mathcal{L}^2 + \tilde{C}\mathcal{L}^3 + O(\mathcal{L}^4)\right) \end{aligned}$$

where

$$\tilde{C} \leq C_{p,d,\beta}\left(\|\nabla\rho_x(0)\| + \|\nabla\rho_x(0)\|^2 + \|H\rho_x(0)\|\right).$$

Since $|\rho_x(0) - \rho_{\min}| \leq 2\nabla\rho_x(0)\|u\| + 2\|H\rho_x(0)\| \cdot \|u\|^2 + O(\|u\|^3)$, we have:

$$\begin{aligned} \frac{\rho_x(0)}{\rho_{\min}} &\leq \frac{\rho_x(0)}{\rho_x(0) - 2\|\nabla\rho_x(0)\| \cdot \|u\| - 2\|H\rho_x(0)\| \cdot \|u\|^2 + O(\|u\|^3)} \\ &\leq 1 + \frac{2\|\nabla\rho_x(0)\| \cdot \|u\|}{\rho_x(0)} + \left(\frac{2\|H\rho_x(0)\|}{\rho_x(0)} + \frac{4\|\nabla\rho_x(0)\|^2}{\rho_x(0)^2}\right)\|u\|^2 + O(\|u\|^3) \\ \implies \left(\frac{\rho_x(0)}{\rho_{\min}}\right)^{\frac{p-1}{m}} &\leq 1 + 2\left(\frac{p-1}{m}\right)\rho_x(0)^{-1}\|\nabla\rho_x(0)\| \cdot \|u\| \\ &\quad + C_{p,d,\beta}\left(\|\nabla\rho_x(0)\|^2 + \|H\rho_x(0)\|\right)\|u\|^2 + O(\|u\|^3) \end{aligned}$$

which gives

$$\begin{aligned} & \left(\frac{\rho_x(0)}{\rho_{\min}}\right)^{\frac{p-1}{m}} \\ =& 1 + 2\left(\frac{p-1}{m}\right)\rho_x(0)^{\frac{p-1}{m}-1}\|\nabla\rho_x(0)\|\mathcal{L} + C_{p,d,\beta}\left(\|\nabla\rho_x(0)\|^2 + \|H\rho_x(0)\|\right)\mathcal{L}^2 + O(\mathcal{L}^3). \end{aligned}$$

Thus:

$$\begin{aligned} & n^{\frac{(p-1)}{m}}\ell_p^p(0,u,\overline{H}_{n\rho_{\min}}) \\ \leq& \left(1 + 2\left(\frac{p-1}{m}\right)\rho_x(0)^{\frac{p-1}{m}-1}\|\nabla\rho_x(0)\|\mathcal{L} + C_{p,d,\beta}\left(\|\nabla\rho_x(0)\|^2 + \|H\rho_x(0)\|\right)\mathcal{L}^2 + O(\mathcal{L}^3)\right) \end{aligned}$$

$$\times \left(\mu\mathcal{L} + \frac{\mu}{2}\left(\frac{p-1}{m}\right)\rho_x(0)^{\frac{(p-1)}{m}-1}\|\nabla\rho_x(0)\|\mathcal{L}^2 + \rho_x(0)^{\frac{(p-1)}{m}}\mathcal{L}^2 + \tilde{C}\mathcal{L}^3 + O(\mathcal{L}^4)\right)$$
$$= \mu\mathcal{L} + \left(\frac{5\mu}{2}\left(\frac{p-1}{m}\right)\rho_x(0)^{\frac{(p-1)}{m}-1}\|\nabla\rho_x(0)\| + \rho_x(0)^{\frac{p-1}{m}}\right)\mathcal{L}^2 + \tilde{C}_2\mathcal{L}^3 + O(\mathcal{L}^4)$$

where

$$\tilde{C}_2 \leq C_{p,d,\beta}\left(\|\nabla\rho_x(0)\| + \|\nabla\rho_x(0)\|^2 + \|H\rho_x(0)\|\right).$$

A similar argument shows that

$$n^{\frac{(p-1)}{m}}\ell_p^p(0,u,\overline{H}_{n\rho_{\max}})$$
$$\geq \mu\mathcal{L} - \left(\frac{5\mu}{2}\left(\frac{p-1}{m}\right)\rho_x(0)^{\frac{(p-1)}{m}-1}\|\nabla\rho_x(0)\| + \rho_x(0)^{\frac{p-1}{m}}\right)\mathcal{L}^2 - \tilde{C}_2\mathcal{L}^3 + O(\mathcal{L}^4),$$

and we obtain

$$|\tilde{\ell}_p^p(0,u,H_{n\rho_x}) - \mu\mathcal{L}_p^p(0,u,\rho_x)| \leq \rho_x(0)^{\frac{p-1}{m}}\left(\frac{5\mu}{2}\left(\frac{p-1}{m}\right)\frac{\|\nabla\rho_x(0)\|}{\rho_x(0)} + 1\right)\mathcal{L}^2 + \tilde{C}_2\mathcal{L}^3 + O(\mathcal{L}^4).$$

Note $\rho_x(0) = \rho(x)$. To finish the proof we use bounds for $\|\nabla\rho_x(0)\|$, $\|H\rho_x(0)\|$ shown in Lemma 1 (note $\tilde{C}_2$ depends on $H\rho_x(0)$). Altogether we have:

$$|\tilde{\ell}_p^p(0,u,H_{n\rho_x}) - \mu\mathcal{L}_p^p(0,u,\rho_x)| \leq \rho(x)^{\frac{p-1}{m}}\left(\frac{5\mu}{2}\left(\frac{p-1}{m}\right)\frac{\|\nabla\rho(x)\|}{\rho(x)} + 1\right)\mathcal{L}^2 + \tilde{C}\mathcal{L}^3 + O(\mathcal{L}^4)$$

with

$$\tilde{C} \leq C_{p,d,\beta}\left(K + K\|\nabla\rho(x)\| + \|\nabla\rho(x)\|^2 + \|H\rho(x)\|\right).$$

Since $\rho_{\min} \geq \frac{1}{\beta}(1 - CmK\|u\|^2) \geq \frac{1}{2\beta}$ as long as $\|u\|$ is small enough, the above holds for $\mathcal{C}_{p,m,\beta,K} \geq \|u\| \geq (\frac{n}{2\beta})^{-\frac{1}{m}(\frac{1}{3}-\epsilon)}$ with probability at least $1 - C\exp\left(-c(\frac{n}{2\beta})^{\frac{2\epsilon}{3}\min\{\frac{1}{m},\frac{1}{p}\}}\right)$. Bounding $\|\nabla\rho(x)\| \leq L_1$ and $\|H\rho(x)\| \leq L_2$ concludes the proof. ∎

Lemma E.5 (Curvature Perturbation, Continuum Metric) *For $\|u\| \leq C_{\mathcal{M},\rho}$,*

$$\mu\mathcal{L}_p^p(0,u,\rho_x) = (1 \pm CpK\|u\|^2)\mu\mathcal{L}_p^p(x,y).$$

Proof We first note that for $\|u\| \leq C_{\mathcal{M},\rho}$, the optimal $\mathcal{L}_p$ path $\gamma \in \mathcal{M}$ stays inside $\mathcal{B}_x(2r)$ and that the optimal path for $\mathcal{L}_p^p(0,u,\rho_x)$ stays inside $B_0(2r)$ (see for example Theorem 4.2 and Lemma 2.2 from Little et al. (2022)). We thus have:

$$\mathcal{L}_p^p(0,u,\rho_x) = \inf_{\log_x\gamma\in B_0(2r)}\int \rho_x(\log_x\gamma(t))^{(1-p)/m}|(\log_x\gamma)'(t)|\,dt$$
$$= \inf_{\gamma\in\mathcal{B}_x(2r)}\int \rho(\gamma(t))^{(1-p)/m}J_x(\log_x\gamma(t))^{(1-p)/m}\,|(\log_x\gamma)'(t)|\,dt$$

$$
\begin{aligned}
&= \inf_{\gamma \in \mathcal{B}_x(2r)} \int \rho(\gamma(t))^{(1-p)/m} J_x(\log_x \gamma(t))^{(1-p)/m} \, |\log_x'(\gamma(t))| \cdot |\gamma'(t)| \, dt \\
&= \inf_{\gamma \in \mathcal{B}_x(2r)} \int \rho(\gamma(t))^{(1-p)/m} (1 \pm CmK\|u\|^2)^{(1-p)/m} \cdot (1 \pm CK\|u\|^2) \cdot |\gamma'(t)| \, dt \\
&= (1 \pm CpK\|u\|^2) \inf_{\gamma \in \mathcal{B}_x(2r)} \int \rho(\gamma(t))^{(1-p)/m} |\gamma'(t)| \, dt \\
&= (1 \pm CpK\|u\|^2) \inf_{\gamma \in \mathcal{M}} \int \rho(\gamma(t))^{(1-p)/m} |\gamma'(t)| \, dt \\
&= (1 \pm CpK\|u\|^2) \mathcal{L}_p^p(x, y).
\end{aligned}
$$

For the bound on $|\log_x'(\gamma(t))|$, see 1.34 of García Trillos et al. (2019). ■

Appendix F. Fermat Kernels and Degrees

Corollary 5.1 *Let $\delta := 2\mu^{-1}C_1 h + 4\mu^{-1}C_2 h^2 \leq \frac{1}{2}$, where C_1, C_2 are as in Theorem 3.2. Define $\widehat{h}_+ := h(1+\delta)$, $\widehat{h}_- := h(1-\delta)$. Then for n large enough, with probability at least $1 - C_\epsilon n^3 \exp\left(-c_\epsilon (\frac{n}{4\beta})^{\frac{\epsilon}{2p+1}} \min\{\frac{1}{m}, \frac{1}{p}\}\right)$, we have for all $x_i, x_j \in \mathcal{X}$:*

$$
\eta\left(\frac{\mu \mathcal{L}_p^p(x_i, x_j)}{\widehat{h}_-}\right) \leq \eta\left(\frac{\widetilde{\ell}_p^p(x_i, x_j)}{h}\right) \leq \eta\left(\frac{\mu \mathcal{L}_p^p(x_i, x_j)}{\widehat{h}_+}\right), \tag{23}
$$

where C_ϵ, c_ϵ are constants depending on ϵ, $\eta = \frac{1}{\omega_m} \mathbb{1}_{[0,1]}$ and $h \geq 4\mu\beta^{\frac{(p-1)}{m}} (n\beta/2)^{-\frac{1}{m}(\frac{1}{3}-\epsilon)}$.

Proof We prove the result by defining $\xi := 2\mu\beta^{\frac{(p-1)}{m}} (n\beta/2)^{-\frac{1}{m}(\frac{1}{3}-\epsilon)} \leq \frac{1}{2}h$ and considering the following 3 cases for any fixed x_i, x_j:

1. $\mu\mathcal{L}_p^p < \xi$
2. $\xi \leq \mu\mathcal{L}_p^p \leq 2h$
3. $\mu\mathcal{L}_p^p > 2h$

Case 1: Assume $\mu\mathcal{L}_p^p < \xi$. We want to show that all kernels evaluate to 1, *i.e.*, $\mu\mathcal{L}_p^p < \widehat{h}_-$ (which also guarantees $\mu\mathcal{L}_p^p < \widehat{h}_+$) and that $\widetilde{\ell}_p^p < h$. Since $\mu\mathcal{L}_p^p < \xi \leq \frac{1}{2}h = \frac{1}{2}\frac{\widehat{h}_-}{(1-\delta)} \leq \widehat{h}_-$, the $\mathcal{L}_p^p$ kernels evaluate to 1. To bound $\widetilde{\ell}_p^p$, we need to further consider the following two subcases:
Case 1a: $\mathcal{L}_p^p \leq n^{-\frac{1}{m}\left(1-\frac{2}{3p}\right)}$. In this case for n large enough the bound follows by considering the straight-line path, since

$$
\tilde{\ell}_p^p(x_i, x_j) \leq n^{\frac{p-1}{m}} \|x_i - x_j\|^p \lesssim n^{\frac{p-1}{m}} n^{-\frac{1}{m}(p-\frac{2}{3})} = n^{-\frac{1}{m}(1-\frac{2}{3})} = n^{-\frac{1}{3m}} \ll n^{-\frac{1}{m}(\frac{1}{3}-\epsilon)} \sim 2\xi \,,
$$

so that $\tilde{\ell}_p^p(x_i, x_j) < h$.
Case 1b: $n^{-\frac{1}{m}\left(1-\frac{2}{3p}\right)} \leq \mathcal{L}_p^p$. This case involves discrete-to-continuum control of metric

convergence at smaller scales than given in Theorem 3.2. However a repeat of the same arguments used to prove Theorem 3.2, but with choosing $q = 1+\frac{1}{4p}$ when applying Proposition 4.4 in the proof of Lemma E.4, gives that for $d(x_i, x_j) \geq 2(n\beta/2)^{-\frac{1}{m}\left(\frac{2p}{2p+1}-\epsilon\right)}$,

$$|\tilde{\ell}_p^p(x_i, x_j) - \mu\mathcal{L}_p^p(x_i, x_j)| = O((\mathcal{L}_p^p)^{(1+\frac{1}{4p})}) \tag{67}$$

with probability at least $1 - C_\epsilon n \exp\left(-c_\epsilon(\frac{n}{4\beta})^{\frac{\epsilon}{2p+1}\min\{\frac{1}{m},\frac{1}{p}\}}\right)$. We double check the lower bound in (23) is satisfied in Case 1b. Because $d_g(x_i, x_j) \geq \beta^{\frac{1-p}{m}}\mathcal{L}_p^p$ (see (2) and apply the upper bound on ρ) and $\mathcal{L}_p^p \geq n^{-\frac{1}{m}\left(1-\frac{2}{3p}\right)}$ by assumption, to apply (67) it is sufficient to check that

$$d(x_i, x_j) \geq \frac{1}{\beta^{\frac{p-1}{m}}} n^{-\frac{1}{m}\left(1-\frac{2}{3p}\right)} \geq 2(n\beta/2)^{-\frac{1}{m}\left(\frac{2p}{2p+1}-\epsilon\right)}, \tag{68}$$

i.e., we need $1 - \frac{2}{3p} < \frac{1}{1+(1/2p)} - \epsilon$. Since $\epsilon < \frac{1}{2(4p+3)} < \frac{1}{6p}$, we have

$$\frac{1}{1+(1/2p)} - \epsilon \geq \frac{1}{1+(1/2p)} - \frac{1}{6p} \geq 1 - \frac{1}{2p} - \frac{1}{6p} = 1 - \frac{2}{3p}$$

and indeed the lower bound in (68) holds for n large enough. Thus, w.h.p.

$$\tilde{\ell}_p^p(x_i, x_j) \leq \mu\mathcal{L}_p^p(x_i, x_j) + O(\xi^{\left(1+\frac{1}{4p}\right)}) = \mu\mathcal{L}_p^p + n^{-\frac{1}{m}\left(\frac{1}{3}-\epsilon\right)\left(1+\frac{1}{4p}\right)} < 2\mu\mathcal{L}_p^p < 2\xi$$

so that $\tilde{\ell}_p^p(x_i, x_j) < h$.

Case 2: Note $d(x_i, x_j) \geq \beta^{\frac{1-p}{m}}\mathcal{L}_p^p(x_i, x_j) \geq \beta^{\frac{1-p}{m}}\mu^{-1}\xi = 2(n\beta/2)^{-\frac{1}{m}(\frac{1}{3}-\epsilon)}$, and so by Theorem 3.2:

$$\mu\mathcal{L}_p^p(1-\delta) \leq \widetilde{\ell}_p^p \leq \mu\mathcal{L}_p^p(1+\delta)$$
$$\implies \frac{\mu\mathcal{L}_p^p}{\widehat{h}_+} = \frac{\mu\mathcal{L}_p^p}{h(1+2\delta)} \leq \frac{\widetilde{\ell}_p^p}{h} \leq \frac{\mu\mathcal{L}_p^p}{h(1-\delta)} = \frac{\mu\mathcal{L}_p^p}{\widehat{h}_-}$$

which establishes the corollary in this case.

Case 3: Now assume that $\mu\mathcal{L}_p^p > 2h$. We want to show that all kernels evaluate to 0, *i.e.,* we need to show that $\mathcal{L}_p^p > \widehat{h}_+$ (of course this guarantees $\mathcal{L}_p^p > \widehat{h}_-$) and also that $\widetilde{\ell}_p^p > h$. The first is true since $\mathcal{L}_p^p > 2h \geq h(1+2\delta) = \widehat{h}_+$. For $\widetilde{\ell}_p^p$, note that by Theorem 3.2

$$\widetilde{\ell}_p^p > \mu\mathcal{L}_p^p - C_1(\mathcal{L}_p^p)^2 - C_2(\mathcal{L}_p^p)^3 > \frac{\mu}{2}\mathcal{L}_p^p > h$$

whenever $\mathcal{L}_p^p \leq C_3$ for a fixed constant C_3. But when $\mathcal{L}_p^p > C_3$, it is clear that $\widetilde{\ell}_p^p > \frac{\mu}{2}\mathcal{L}_p^p$ from Theorem 1 in Hwang et al. (2016) (or from the proof of Proposition 4.4).

Finally, to conclude that (23) holds for all pairs of points we take a union bound, so that (23) holds with probability at least $1 - C_\epsilon n^3 \exp\left(-c_\epsilon(\frac{n}{4\beta})^{\frac{\epsilon}{2p+1}\min\{\frac{1}{m},\frac{1}{p}\}}\right)$ (note the probability from Case 1 dominates the probability from Case 2, since we applied Proposition 4.4 with $q = 1 + \frac{1}{4p}$ instead of $q = 2$).

■

Lemma 5.2 *For C_1, C_2 as in Theorem 3.2 and K_p as in Theorem B.2:*

$$\left|m_{\tilde{\ell}^p_p,h,i} - \rho_p(x_i)\right| \leq C\beta^p \left[\left(\frac{L_1}{\beta} + C_1 m\right) h + m(K_p + C_2\mu)h^2 + m\omega_m \frac{d^g_\infty(\nu,\nu_n)}{h}\right].$$

Proof From the proof of (García Trillos et al., 2019, Lemma 18), if a distance function d_0 approximates d_g by

$$\eta\left(\frac{d_g(x_i,x_j)}{h(1-\delta)}\right) \leq \eta\left(\frac{d_0(x_i,x_j)}{h}\right) \leq \eta\left(\frac{d_g(x_i,x_j)}{h(1+\delta)}\right)$$

for all $x_i, x_j \in \mathcal{X}_n$, then for $h \leq C_{\mathcal{M}}$ (see (García Trillos et al., 2019, Assumption 3))

$$|m_{d_0,h,i} - \rho(x_i)| \leq C\left(L_\rho h + \beta m K h^2 + \beta m\delta + \beta m\eta(0)\omega_m \frac{d^g_\infty(\nu,\nu_n)}{h}\right)$$

Note if $d_g = \mathcal{L}^p_p$ then data is sampled from density ρ^p, and L_ρ, β, K, δ all depend on p. Thus, for $d_g = \mathcal{L}^p_p$ and $d_0 = \widetilde{\ell}^p_p$ we obtain

$$\left|m_{\tilde{\ell}^p_p,h,i} - \rho^p(x_i)\right| \leq C\left(L_{\rho^p} h + \beta^p d K_p h^2 + \beta^p d\delta + \beta^p d\eta(0)\omega_d \frac{d^g_\infty(\nu,\nu_n)}{h}\right)$$

where K_p is computed in Theorem B.2, L_{ρ^p} is the Lipschitz constant of ρ^p, and we can choose $\delta = 4C_1 h + 8C_2\mu h^2$ by Corollary 5.1. We may take $L_{\rho^p} = C\beta^{p-1}L_1$ as for all $x, y \in \mathcal{M}$

$$|\rho^p(x) - \rho^p(y)| \leq |\rho(x) - \rho(y)| \sum_{j=0}^{p-1} |\rho^{p-j-1}(x)||\rho^j(y)| \leq (L_1)\left(C\beta^{p-1}\right).$$

■

Appendix G. De-Poissonization: Proof of Theorem 3.2

Proof To utilize the results of Theorem 4.3, we consider a coupling process of 2 random PPPs as follows. We have two infinite sequences with draws from ρ:

$$x_1, x_2, x_3, \ldots$$
$$y_1, y_2, y_3, \ldots$$

Now we let N_- be a Poisson random variable with mean $n_- = n - n^{\frac{1+\kappa}{2}}$, and N_{inc} a Poisson random variable with mean $2n^{\frac{1+\kappa}{2}}$; note $N_- + N_{\text{inc}}$ is Poisson with mean $n_+ = n + n^{\frac{1+\kappa}{2}}$. This process defines two PPPs:

$$H_{n_-\rho} := \{x_1, \ldots, x_{N_-}\},$$
$$H_{n_+\rho} := \{x_1, \ldots, x_{N_-}, y_1, \ldots, y_{N_{\text{inc}}}\}.$$

We define the following high probability events:

$$\Omega_1 := \{N_- \leq n\},$$
$$\Omega_2 := \{n \leq N_- + N_{\text{inc}}\}.$$

We now let X_n consist of the first $N_- \wedge n$ points from the x_i sequence, with an additional $n - N_-$ points taken from the y_i sequence if $n > N_-$; X_n consists of n iid samples from ρ. Note that on Ω_1, X_n gets to use all of the points available to $H_{n_-\rho}$, and possibly more; on Ω_2, $H_{n_+\rho}$ gets to use all of the points available to X_n, and possibly more. For notational brevity, we let $\zeta = C_1\mathcal{L}^2 + C_2\mathcal{L}^3$ where C_1, C_2 are the constants from Theorem 4.3 and define the following events:

$$A := \{n_-^{\frac{p-1}{m}} \ell_p^p(X_n) < \mu\mathcal{L} + \zeta\},$$
$$A' := \{n_-^{\frac{p-1}{m}} \ell_p^p(H_{n_-\rho}) < \mu\mathcal{L} + \zeta\},$$
$$B := \{n_+^{\frac{p-1}{m}} \ell_p^p(X_n) > \mu\mathcal{L} - \zeta\},$$
$$B' := \{n_+^{\frac{p-1}{m}} \ell_p^p(H_{n_+\rho}) > \mu\mathcal{L} - \zeta\}.$$

We then obtain:

$$P(A) \geq P(A \cap \Omega_1) \geq P(A' \cap \Omega_1) = P(A') - P(A' \cap \Omega_1^C) \geq P(A') - P(\Omega_1^C),$$
$$P(B) \geq P(B \cap \Omega_2) \geq P(B' \cap \Omega_2) = P(B') - P(B' \cap \Omega_2^C) \geq P(B') - P(\Omega_2^C)\,.$$

By Theorem 4.3

$$P(A') \geq 1 - Cn_- \exp\left(-c(n_-/2\beta)^\kappa\right) \geq 1 - Cn\exp\left(-c\,(n/4\beta)^\kappa\right)$$
$$P(B') \geq 1 - Cn_+ \exp\left(-c(n_+/2\beta)^\kappa\right) \geq 1 - Cn\exp\left(-c(n/2\beta)^\kappa\right)$$

since $n \geq 5$ and $2(n\beta/2)^{-\frac{1}{m}\left(\frac{1}{3}-\epsilon\right)} \leq d(x,y) \leq C_{\mathcal{M},\rho}$.

We now bound $P(\Omega_1^C), P(\Omega_2^C)$. Note that when $X = \text{Poisson}(\lambda)$, the following Poisson tail bounds can be derived from a Chernoff bound argument:

$$P(X \geq x) \leq \left(\frac{\lambda}{x}\right) e^{x-\lambda} \text{ for } x > \lambda \quad , \quad P(X \leq x) \leq \left(\frac{\lambda}{x}\right) e^{x-\lambda} \text{ for } x < \lambda\,.$$

Combining the first inequality with $\ln(1+x) \leq \frac{2x}{2+x}$ for $-1 < x \leq 0$ and $\frac{1}{1-x} \geq 1 + x$ for $|x| \leq 1$ gives $P(N_- \geq n) \leq e^{-\frac{1}{2}n^\kappa}$. Combining the second inequality with $\ln(1+x) \leq \frac{x}{\sqrt{1+x}}$ for $x \geq 0$ and $\frac{1}{\sqrt{1+x}} \leq 1 - \frac{x}{4}$ for $0 \leq x \leq 1$ gives $P(N_- + N_{\text{inc}} \leq n) \leq e^{-\frac{1}{4}n^\kappa}$. We thus obtain:

$$P(\Omega_1^C) \leq e^{-\frac{1}{2}n^\kappa} \quad , \quad P(\Omega_2^C) \leq e^{-\frac{1}{4}n^\kappa}$$

so that

$$P(A \cap B) \geq P(A) - P(B^C) \geq 1 - Cn\exp\left(-c\,(n/4\beta)^\kappa\right)\,.$$

Note that on $A \cap B$, we have:

$$n^{\frac{p-1}{m}} \ell_p^p(X_n) < \left(\frac{1}{1 - n^{\frac{\kappa-1}{2}}}\right)^{\frac{p-1}{m}} (\mu\mathcal{L} + \zeta) \leq \left(1 + 2n^{\frac{\kappa-1}{2}}\right)^{\frac{p-1}{m}} (\mu\mathcal{L} + \zeta)$$

since $n^{\frac{\kappa-1}{2}} \leq \frac{1}{2}$ (true since $\kappa \leq \frac{1}{21}, n \geq 5$). We utilize the inequality $(1+x)^q \leq 1 + x(1 \vee \lceil q \rceil^2)$ for $0 \leq x \leq \lceil q \rceil^{-1}$. If $q \leq 1$, then $(1+x)^q \leq 1 + x$ and the inequality clearly holds. If $q > 1$, then by the Binomial Theorem

$$(1+x)^{\lceil q \rceil} = \sum_{k=0}^{\lceil q \rceil} \binom{\lceil q \rceil}{k} x^k = 1 + \sum_{k=1}^{\lceil q \rceil} \binom{\lceil q \rceil}{k} x^k \leq 1 + \lceil q \rceil^2 x$$

as long as $\binom{\lceil q \rceil}{k} x^k \geq \binom{\lceil q \rceil}{k+1} x^{k+1}$ for $1 \leq k \leq \lceil q \rceil - 1$, which is guaranteed by $x \leq \lceil q \rceil^{-1}$. We thus obtain that for $2n^{\frac{\kappa-1}{2}} \leq \left\lceil \frac{p-1}{m} \right\rceil^{-1}$:

$$\begin{aligned} n^{\frac{p-1}{m}} \ell_p^p(X_n) &\leq \left[1 + 2n^{\frac{\kappa-1}{2}} \left(1 \vee \left\lceil \frac{p-1}{m} \right\rceil^2\right)\right] (\mu\mathcal{L} + \zeta) \\ &\leq \mu\mathcal{L} + \zeta + 2n^{\frac{\kappa-1}{2}} \left(1 \vee \left\lceil \frac{p-1}{m} \right\rceil^2\right) 3\mu\mathcal{L} \end{aligned}$$

since $\mu\mathcal{L} + \zeta \leq 3\mu\mathcal{L}$ for $d(x,y) \leq C_{\mathcal{M},\rho}$. A similar argument gives

$$n^{\frac{p-1}{m}} \ell_p^p(X_n) \geq \mu\mathcal{L} - \zeta - n^{\frac{\kappa-1}{2}} \left(1 \vee \left\lceil \frac{p-1}{m} \right\rceil^2\right) 3\mu\mathcal{L}$$

so that on $A \cap B$ we have

$$|n^{\frac{p-1}{m}} \ell_p^p(X_n) - \mu\mathcal{L}| \leq \zeta + 2n^{\frac{\kappa-1}{2}} \left(1 \vee \left\lceil \frac{p-1}{m} \right\rceil^2\right) 3\mu\mathcal{L}\,.$$

In particular,

$$|n^{\frac{p-1}{m}} \ell_p^p(X_n) - \mu\mathcal{L}| \leq C_1\mathcal{L} + C_2(\mathcal{L}^3 + \mathcal{L}n^{\frac{\kappa-1}{2}})\,,$$

where we update $C_{p,d,\beta}$ in the definition of C_2 as needed.

Finally, we observe that the lower bound on $d(x,y)$ implies $\mathcal{L}^2 \geq C_{p,d,\beta} n^{-\frac{2}{m}\left(\frac{1}{3}-\epsilon\right)}$. Since $n^{-\frac{2}{m}\left(\frac{1}{3}-\epsilon\right)}$ dominates $n^{\frac{\kappa-1}{2}}$ whenever $3m - 4 + 10\epsilon \geq 0$ which is guaranteed by $m \geq 2$, we obtain the theorem statement. ∎

Appendix H. Additional Experimental Plots

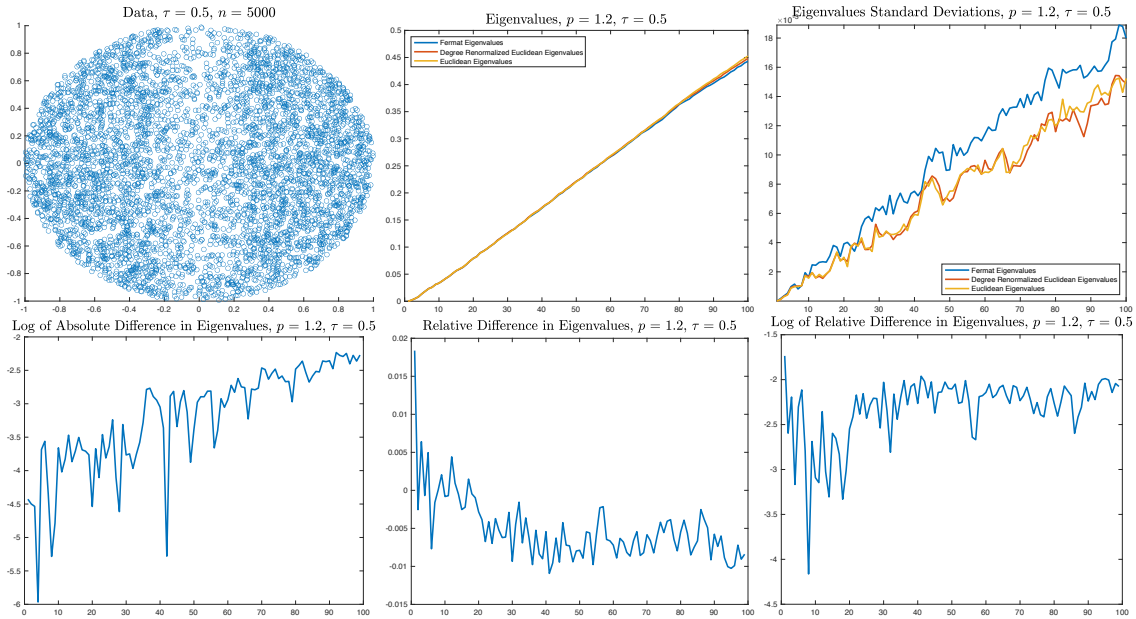


Figure 7: $p = 1.2$, $\tau = .5$. Runtime for Fermat Laplacian: 168.46 ± 10.70s. Runtime for Rescaled Euclidean Laplacian: 7.38 ± 1.10s.

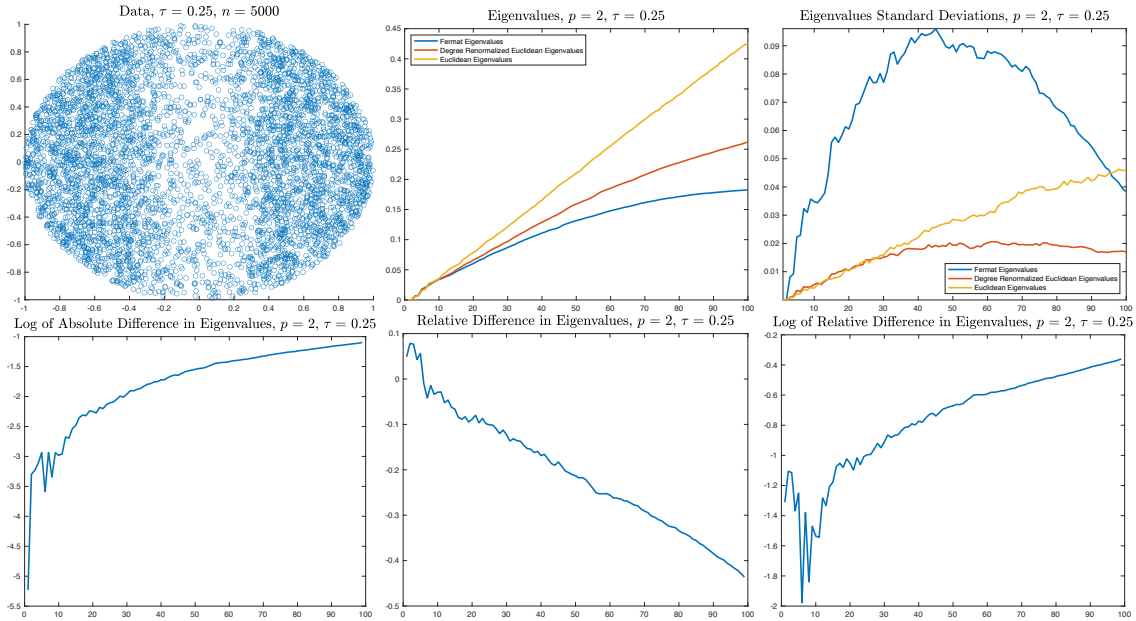


Figure 8: $p = 2$, $\tau = .1$. Runtime for Fermat Laplacian: 212.52 ± 70.49s. Runtime for Rescaled Euclidean Laplacian: $1.33 \pm .24$s.

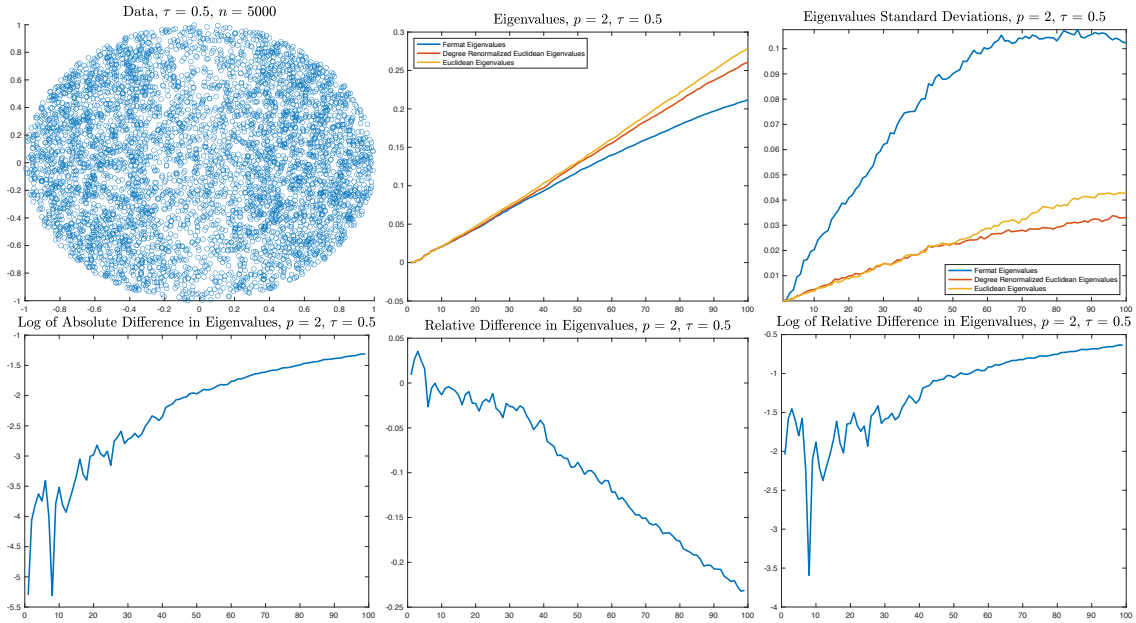


Figure 9: $p = 2$, $\tau = .5$. Runtime for Fermat Laplacian: 83.62 ± 10.50s. Runtime for Rescaled Euclidean Laplacian: $1.24 \pm .16$s.

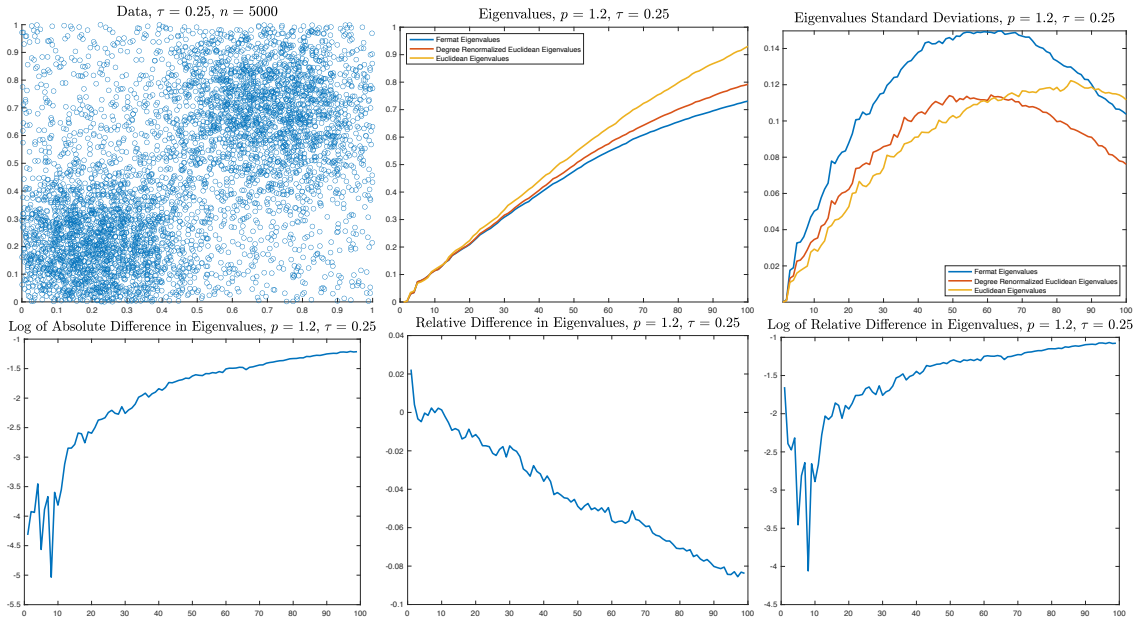


Figure 10: $p = 1.2$, $\tau = .25$. Runtime for Fermat Laplacian: 170.10 ± 58.73s. Runtime for Rescaled Euclidean Laplacian: $1.27 \pm .05$s.

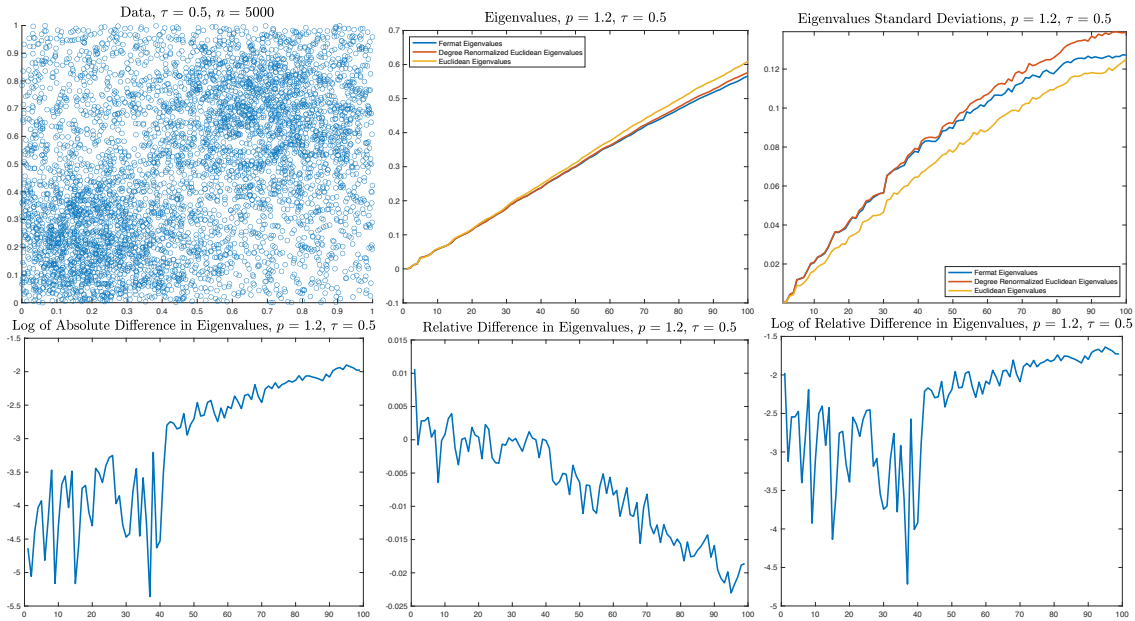


Figure 11: $p = 1.2$, $\tau = .5$. Runtime for Fermat Laplacian: 58.72 ± 1.51s. Runtime for Rescaled Euclidean Laplacian: $1.04 \pm .05$s.

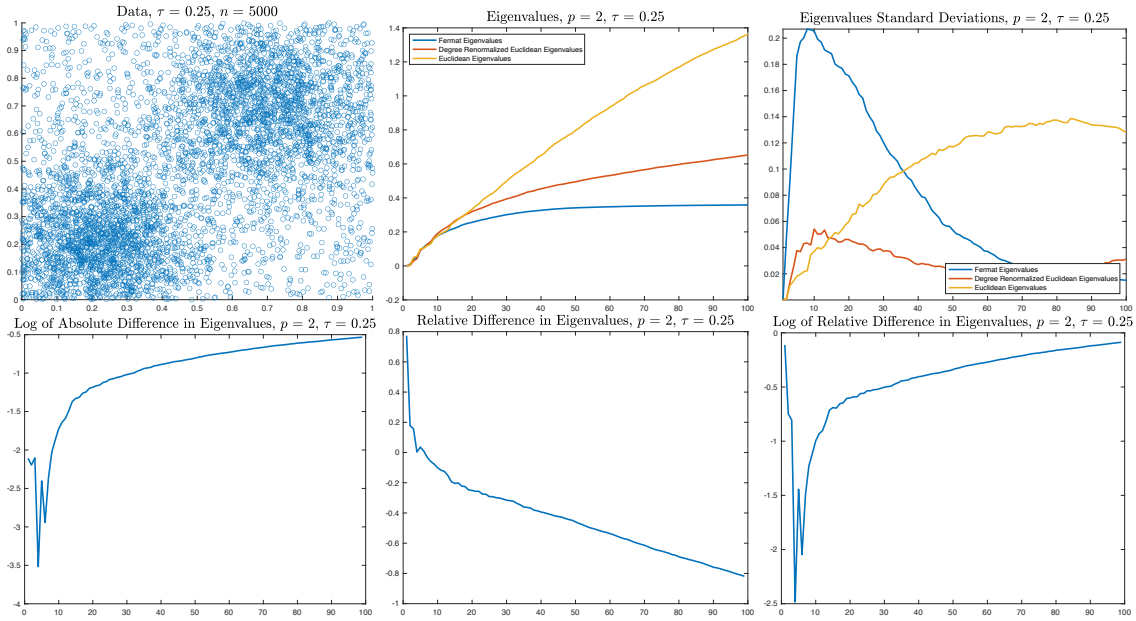


Figure 12: $p = 2$, $\tau = .25$. Runtime for Fermat Laplacian: 219.76 ± 19.41s. Runtime for Rescaled Euclidean Laplacian: $.60 \pm .28$s.