Manifold learning with arbitrary norms

Joe Kileel* · **Amit Moscovich** ·
Nathan Zelesko · **Amit Singer**

Abstract Manifold learning methods play a prominent role in nonlinear dimensionality reduction and other tasks involving high-dimensional data sets with low intrinsic dimensionality. Many of these methods are graph-based: they associate a vertex with each data point and a weighted edge with each pair. Existing theory shows that the Laplacian matrix of the graph converges to the Laplace-Beltrami operator of the data manifold, under the assumption that the pairwise affinities are based on the Euclidean norm. In this paper, we determine the limiting differential operator for graph Laplacians constructed using *any* norm. Our proof involves an interplay between the second fundamental form of the manifold and the convex geometry of the given norm's unit ball. To demonstrate the potential benefits of non-Euclidean norms in manifold learning, we consider the task of mapping the motion of large molecules with continuous variability. In a numerical simulation we show that a modified Laplacian eigenmaps algorithm, based on the Earthmover's distance, outperforms the classic Euclidean Laplacian eigenmaps, both in terms of computational cost and the sample size needed to recover the intrinsic geometry.

*Author to whom correspondence should be addressed.

Joe Kileel
Department of Mathematics and Oden Institute, University of Texas at Austin
E-mail: jkileel@math.utexas.edu

Amit Moscovich
Department of Statistics and Operations Research, Tel-Aviv University
E-mail: amit@moscovich.org

Nathan Zelesko
Department of Mathematics, Brown University
E-mail: nathan@zelesko.com

Amit Singer
Department of Mathematics and Program in Applied and Computational Mathematics, Princeton University
E-mail: amits@math.princeton.edu

1 Introduction

Manifold learning is broadly concerned with analyzing high-dimensional data sets that have a low intrinsic dimensionality. The standard assumption is that the input data set $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subseteq \mathbb{R}^D$ lies on or near a d-dimensional submanifold $\mathcal{M} \subseteq \mathbb{R}^D$ where $d \ll D$. The key tasks are dimensionality reduction [63,52,16,4,72,11,38,40,71], function representation and approximation [21, 10,35,59] and semi-supervised learning [5,24,45]. Most data analysis methods in this setting rely on pairwise Euclidean distances between the data points.

In this paper, we focus on manifold learning methods that use a graph Laplacian. These include the popular spectral embedding methods Laplacian eigenmaps [4,5] and diffusion maps [11]. Both methods map the input points to the eigenvectors of a graph Laplacian operator $\mathcal{L}_n$ (or weighted variant thereof). By definition, $\mathcal{L}_n$ acts on a function $f : \mathcal{X} \to \mathbb{R}$ via

$$\left(\mathcal{L}_n f\right)(\mathbf{x}_i) := \sum_{j=1}^{n} W_{ij} \left(f(\mathbf{x}_j) - f(\mathbf{x}_i)\right), \qquad W_{ij} := \exp\left(-\frac{\|\mathbf{x}_j - \mathbf{x}_i\|_2^2}{\sigma_n^2}\right). \quad (1)$$

Under suitable conditions, as $n \to \infty$ the discrete graph Laplacian operator $\mathcal{L}_n$ converges to the continuous Laplace-Beltrami operator $\Delta_{\mathcal{M}}$ on the manifold [7], and its eigenvectors converge to the Laplacian eigenfunctions [6]. While the convergence results can be extended to more general affinity kernels of the form $W_{ij} = K_{\sigma_n}(\|\mathbf{x}_i - \mathbf{x}_j\|_2)$, the role of the Euclidean norm here is essential. This poses a potential limitation for graph Laplacian methods since Euclidean metrics are not always the best choice for all application domains [8]. Furthermore, some non-Euclidean metrics use compressed representations, which can have practical benefits in terms of runtime and memory requirements. Following this line of reasoning leads to the following questions: can the machinery of discrete Laplacian operators be generalized to non-Euclidean metrics? Does doing so yield any practical benefits? If so, what is the underlying theory?

This paper is an initial step in answering these questions. The main contribution of the paper is the derivation of the continuum limit of discrete Laplacian operators similar to (1) but with an affinity kernel based on an arbitrary norm. Our key result (Theorem 6) is a proof that using *any norm*, graph Laplacians converge to an explicit second-order differential operator on $\mathcal{M}$. In contrast to the Euclidean case, in the general case the limiting operator is not intrinsic to the manifold, i.e., it depends on the embedding of $\mathcal{M}$ in $\mathbb{R}^D$. Furthermore, it has non-vanishing and possibly discontinuous first-order terms. The second-order coefficients of the limiting differential operator at a point $\mathbf{p} \in \mathcal{M}$ is given by the second moments of the intersection of the tangent plane to $\mathcal{M}$ at $\mathbf{p}$ and the given norm's unit ball. The first-order terms depend on the second fundamental form of $\mathcal{M}$ at $\mathbf{p}$ and the tangent cones to the norm's unit sphere, through a function we call tilt, defined in Section 3.3.

In a second contribution, which was the original motivation for this work, we present in Section 5 a variant of Laplacian eigenmaps that is based on a norm that approximates the Earthmover's distance (EMD), also known as

the Wasserstein-1 metric, for learning volumetric shape spaces. This is motivated by an important problem in structural biology: learning the conformation space of flexible proteins and other macromolecules with continuous variability from cryo-electron microscopy images. Empirically, we demonstrate that classical (Euclidean) Laplacian eigenmaps are at a disadvantage compared to Laplacian eigenmaps based on this approximate EMD, as it requires far fewer sample points to recover the intrinsic manifold of motion. Furthermore, as we show in Section 5.5, the proposed method can achieve faster runtime and a smaller memory footprint through an intermediate compressed representation. This demonstrates, at least for certain data sets, the use of non-Euclidean norms in Laplacian-based manifold learning is desirable from a practical view.

1.1 **Related work**

Our convergence proof builds on the well-known proof of Belkin and Niyogi's for the case of the Euclidean norm [7]. However, the argument for the Euclidean case is *not* directly adaptable to the case of other norms. It relies on a special property of the Euclidean norm: that Euclidean distances provide a second-order approximation to manifold geodesic distances (see [6, Figure 1]). This fails for general norms, which do not even give a first-order approximation to geodesic distances. This difference introduces a first-order derivative term in the limit in the general case. Another technical difference is that, in the standard case, the intersection of an embedded tangent space with the Euclidean unit ball is rotationally symmetric. This gives rise to the Laplace-Beltrami operator, which is the only second-order rotationally symmetric differential operator (up to scale). The property fails for general norms, thereby introducing "cross-terms" in the second-order term of the general limit.

In [64], a different extension of the convergence proof for graph Laplacian methods appeared. That work analyzed k-nearest neighbor graphs and other constructions, but based on a Euclidean norm. We do not pursue this direction.

To the best of our knowledge, most Laplacian-based manifold learning works employ the standard Euclidean norm. Two notable exceptions are the works of Mishne and collaborators [42, 41] where tree-based metrics [13] were used as a basis for diffusion maps. These metrics can be interpreted as hierarchical Earthmover's distances. However since the trees are data-dependent, our main theorem does not apply since we do require a data-independent norm.

The application section in this paper is an extension of [70], where we first proposed to use a variant of diffusion maps based on an approximate Earthmover's distance. In [50], the same approximate Earthmover's distance was used for the clustering of cryo-EM images. Lieu and Saito's work [36] is another that combines diffusion maps and the Earthmover's distance. However, the order of operations is different: they first use Euclidean diffusion maps and only then apply the Earthmover's distance to the resulting point clouds.

Symbol	Description
$\mathcal{M} \subseteq \mathbb{R}^D$	Compact embedded Riemannian submanifold
$d = \dim(\mathcal{M})$	Dimension of $\mathcal{M}$
$\mathbf{p} \in \mathcal{M}$	Point on $\mathcal{M}$
$T_{\mathbf{p}}\mathcal{M}$	Tangent space to $\mathcal{M}$ at $\mathbf{p}$
$\exp_{\mathbf{p}} : T_{\mathbf{p}}\mathcal{M} \to \mathcal{M}$	Exponential map for $\mathcal{M}$ at $\mathbf{p}$
$\mathbf{s} \in \mathbb{R}^d \cong T_{\mathbf{p}}\mathcal{M}$	Geodesic normal coordinates for $\mathcal{M}$ around $\mathbf{p}$
$f : \mathcal{M} \to \mathbb{R}$	Function on $\mathcal{M}$
$\widetilde{f} = f \circ \exp_{\mathbf{p}}$	Function pulled-back to tangent space
$\operatorname{grad} \widetilde{f} : \mathbb{R}^d \to \mathbb{R}^d$	Gradient of $\widetilde{f}$
$\operatorname{hess} \widetilde{f} : \mathbb{R}^d \to \mathbb{R}^{d \times d}$	Hessian of $\widetilde{f}$
$L_{\mathbf{p}} : T_{\mathbf{p}}\mathcal{M} \to \mathbb{R}^D$	Differential of exponential map at $\mathbf{p}$ (Eq. (6))
$Q_{\mathbf{p}} : T_{\mathbf{p}}\mathcal{M} \to \mathbb{R}^D$	Second fundamental form of $\mathcal{M}$ at $\mathbf{p}$ (Eq. (6))
$\mathcal{B} \subseteq \mathbb{R}^D$	Origin-symmetric convex body
$\|\cdot\|_{\mathcal{B}} : \mathbb{R}^D \to \mathbb{R}_{\geq 0}$	Norm with unit ball $\mathcal{B}$
$\|\cdot\|_2 : \mathbb{R}^D \to \mathbb{R}_{\geq 0}$	Euclidean norm
$\|\cdot\|_{\mathbf{w},1} : \mathbb{R}^D \to \mathbb{R}_{\geq 0}$	Weighted ℓ_1-norm
$K_{\sigma} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$	Affinity kernel with width parameter $\sigma > 0$
$\mathcal{L}_n$	Point-cloud Laplacian based on $\|\cdot\|_2$ (Eq. (5))
$\mathcal{L}_{n,\mathcal{B}}$	Point-cloud Laplacian based on $\|\cdot\|_{\mathcal{B}}$ (Definition 1)
$\Delta_{\mathcal{M}}$	Laplace-Beltrami operator on $\mathcal{M}$
$\Delta_{\mathcal{M},\mathcal{B}}$	Laplacian-like differential operator (Definition 2)
$\operatorname{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}}$	Tilt function at $\mathbf{p}$ (Proposition/Definition 4)
$\overline{S}, S^{\circ}, \partial S$	Closure, interior, boundary of a set
$TC_{\mathbf{y}}(\mathcal{Y}) \subseteq \mathbb{R}^D$	Tangent cone to $\mathcal{Y} \subseteq \mathbb{R}^D$ at $\mathbf{y} \in \mathcal{Y}$ (Eq. (12))
$\mathcal{W}$	Wavelet transform
$\langle \cdot, \cdot \rangle$	Inner product
$\mathbb{R}_{\geq 0}, \mathbb{R}_{>0}$	Non-negative/strictly positive real numbers

Table 1 List of notation.

2 Background: graph Laplacian methods

In this section, we review graph Laplacian methods in more detail than in the introduction. Given a subset $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subseteq \mathbb{R}^D$, and an affinity function $K_{\sigma_n} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, consider the symmetric matrix of pairwise affinities:

$$W_{ij} := K_{\sigma_n}(\|\mathbf{x}_i - \mathbf{x}_j\|_2). \tag{2}$$

The canonical choice for K_{σ_n} is the Gaussian kernel, $K_{\sigma_n}(t) = \exp(-t^2/\sigma_n^2)$ (up to normalization conventions) where the width parameters σ_n decay to zero at an appropriate rate. Another possibility is the 0/1 kernel, $K_{\sigma_n}(t) = \mathbb{1}(t \leq \sigma_n)$. The matrix W defines a weighted graph $G = (\mathcal{X}, E, W)$ where the set of edges E consists of all the pairs (i, j) for which $W_{ij} > 0$. Define the diagonal degree matrix by $D_{ij} = \delta_{ij} \sum_k W_{ik}$. The (unnormalized, negative semi-definite) Laplacian matrix of G, or the *graph Laplacian*, is defined to be

$$\mathcal{L}_G := W - D. \tag{3}$$

Remark 1 As a warning, several other authors use the *positive* semi-definite graph Laplacian convention, $\mathcal{L}_G^{\mathrm{psd}} := D - W$. In this paper, we chose the negative semi-definite conventions for both the discrete and continuous Laplacians.

The graph Laplacian acts on vectors $f \in \mathbb{R}^n$. We think of f as a real-valued function on the vertex set $\mathcal{X}$. Then the graph Laplacian averages, for each vertex, the differences between the function's value at the vertex and its neighbors:

$$(\mathcal{L}_G f)(\mathbf{x}_i) = \sum_{j=1}^{n} W_{ij}(f(\mathbf{x}_j) - f(\mathbf{x}_i)) = \sum_{j=1}^{n} K_{\sigma_n}(\|\mathbf{x}_j - \mathbf{x}_i\|_2)(f(\mathbf{x}_j) - f(\mathbf{x}_i)). \tag{4}$$

Note $\mathcal{L}_G$ is an $n \times n$ symmetric negative semi-definite matrix. We list its eigenvalues in descending order

$$0 = \lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{n-1},$$

and choose corresponding real orthonormal eigenvectors

$$\phi_0, \dots, \phi_{n-1} \in \mathbb{R}^n$$

where $\phi_0 = n^{-1/2} \mathbf{1}$. These eigenvectors give an orthonormal basis of functions on $\mathcal{X}$. Two common uses for the Laplacian eigenvectors are:

1. As a basis for function representation and approximation of real-valued functions g defined on $\mathcal{X}$ [5,12,30],

$$g(\mathbf{x}_i) = \sum_{j=0}^{n-1} \alpha_j \phi_j(\mathbf{x}_i).$$

2. As a method for dimensionality reduction of the input set $\mathcal{X}$ [4,11],

$$\mathbf{x}_i \mapsto (\phi_1(\mathbf{x}_i), \dots, \phi_m(\mathbf{x}_i)).$$

Here, each $\mathbf{x}_i$ is mapped into $\mathbb{R}^m$ via the i-th coordinates of the first m nontrivial Laplacian eigenvectors. This usage of eigenvectors is motivated by the fact that any closed connected Riemannian manifold is smoothly embedded into $\mathbb{R}^m$ by its first m Laplacian eigenfunctions for some m [3].

The Laplacian matrices $\mathcal{L}_G$ are often analyzed as points are added to $\mathcal{X}$ and the graph G grows. In this context, the *manifold assumption* is standard, namely that $\mathbf{x}_i$ are drawn i.i.d. from some embedded submanifold $\mathcal{M} \subseteq \mathbb{R}^D$ Other works, for example in the study of clustering, have analyzed the limit of graph Laplacians without making the manifold assumption (see [66,20]). It is convenient to work with an extension of the graph Laplacian that acts on any function f whose domain is a superset of $\mathcal{X}$. Specifically we define the (unnormalized, negative semi-definite) *point-cloud Laplacian* $\mathcal{L}_n$ computed using $\mathcal{X}$ as follows: for each $f : \mathcal{Y} \to \mathbb{R}$ where $\mathcal{X} \subseteq \mathcal{Y}$, define $\mathcal{L}_n f : \mathcal{Y} \to \mathbb{R}$ by

$$\mathcal{L}_n f(\mathbf{p}) := \frac{1}{n} \sum_{j=1}^{n} K_{\sigma_n}(\|\mathbf{x}_j - \mathbf{p}\|_2)(f(\mathbf{x}_j) - f(\mathbf{p})). \tag{5}$$

After rescaling by n the point-cloud Laplacian (5) extends the graph Laplacian (4), because for each $f : \mathcal{Y} \to \mathbb{R}$ where $\mathcal{X} \subseteq \mathcal{Y}$ it holds $(n\mathcal{L}_n f)|_{\mathcal{X}} = \mathcal{L}_G(f|_{\mathcal{X}})$.

2.1 Existing theory: graph Laplacians using the Euclidean norm

It is known that using Euclidean norms to compute affinities, the point-cloud Laplacian converges to the Laplace-Beltrami operator under the manifold assumption (see [25,57,7,22]). Here is a precise statement.

Theorem 1 ([7, Th. 3.1]: Convergence of the point-cloud Laplacian based on the Euclidean norm) *Let $\mathcal{M}$ be a compact d-dimensional embedded Riemannian submanifold of $\mathbb{R}^D$ with Laplace-Beltrami operator $\Delta_{\mathcal{M}}$. Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be i.i.d. draws from the uniform measure on $\mathcal{M}$. Fix any constant $\alpha > 0$, and set $\sigma_n = 2n^{-1/(2d+4+\alpha)}$ and $c_n = \frac{\pi^{d/2}}{4}\sigma_n^{d+2}$. Let $\mathcal{L}_n$ be the point-cloud Laplacian defined in Eq. (5) using the Gaussian affinity based on the Euclidean norm, $K_\sigma(\|\mathbf{x}_j - \mathbf{p}\|) = \exp(\|\mathbf{x}_j - \mathbf{p}\|_2^2/\sigma_n^2)$. Then given a three-times continuously differentiable function $f : \mathcal{M} \to \mathbb{R}$ and a point $\mathbf{p} \in \mathcal{M}$, we have the following convergence in probability:*

$$\frac{1}{c_n}\mathcal{L}_n f(\mathbf{p}) \xrightarrow{p} \frac{1}{\text{vol}(\mathcal{M})}\Delta_{\mathcal{M}} f(\mathbf{p}).$$

Our Theorem 6 extends Theorem 1 to the case of non-Euclidean norms.

Meanwhile, for a non-uniform sampling distribution, variants of the point-cloud Laplacian are known to converge pointwise to a weighted Laplacian

(or Fokker–Planck) operator, which has an additional density-dependent drift term [11, 46, 64]. Theorem 7 extends this to the case of non-Euclidean norms.

In addition to pointwise consistency, spectral consistency has been proven when the norm is Euclidean [6, 26, 37, 51, 20, 19, 68]. This is a stronger mode of convergence than pointwise convergence: the eigenvalues/eigenvectors of the graph Laplacians converge to the eigenvalues/eigenfunctions of the limiting operator. We leave such considerations for arbitrary norms to future work.

3 Ingredients, main theorem statement, first properties

In this section the primary goal is to formulate our main result, Theorem 6, in Section 3.4. Before this, we collect tools from differential geometry and convex geometry. Then in Section 3.3, we define a particular function that depends on the second-order geometry of a manifold and the unit ball of a given norm; this turns out to give the correct first-order derivative term in Theorem 6. After the main statement, we explain how it adapts to non-uniform sampling of the manifold (Theorem 7). Then we discuss first properties of the limiting differential operator and show that it reduces to the Laplace-Beltrami operator in the Euclidean case. As a non-Euclidean example, we calculate the limit explicitly for a circle in the plane where the ambient norm is weighted ℓ_1.

3.1 Preliminaries from Riemannian geometry

We start by reviewing some basics from Riemannian geometry. These notions are later used for our theorem statement and its proof. Textbook accounts of differential geometry are abundant; we particularly like Lee's books [32, 33].

Throughout the paper, $\mathcal{M} \subseteq \mathbb{R}^D$ denotes a d-dimensional compact embedded *Riemannian submanifold* of $\mathbb{R}^D$. We let $\mathbf{p} \in \mathcal{M}$ denote a point (typically fixed in our considerations). We write $T_{\mathbf{p}}\mathcal{M}$ for the abstract *tangent space* to $\mathcal{M}$ at $\mathbf{p}$ (defined in [33, Ch. 3]). In particular, $T_{\mathbf{p}}\mathcal{M}$ is a d-dimensional real vector space equipped with an inner product $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_{\mathbf{p}}$. Further, $0 \in T_{\mathbf{p}}\mathcal{M}$ and we do *not* consider $T_{\mathbf{p}}\mathcal{M}$ to be embedded in $\mathbb{R}^D$. The canonical mapping from the tangent space into the manifold is the *exponential map* at $\mathbf{p}$ [32, Ch. 5], denoted $\exp_{\mathbf{p}} : T_{\mathbf{p}}\mathcal{M} \to \mathcal{M}$.[1] This is a C^∞-map that carries straight lines on $T_{\mathbf{p}}\mathcal{M}$ through the origin to *geodesics* on $\mathcal{M}$ through the point $\mathbf{p}$. By the inverse function theorem, there exist open neighborhoods $\mathcal{U} \subseteq T_{\mathbf{p}}\mathcal{M}$ of 0 and $\mathcal{V} \subseteq \mathcal{M}$ of $\mathbf{p}$ (which we fix once and for all) such that the exponential map restricts to a *diffeomorphism* between these neighborhoods,

$$\exp_{\mathbf{p}} : U \xrightarrow{\sim} V.$$

Further, let us fix once and for all an orthonormal basis on $T_{\mathbf{p}}\mathcal{M}$ with respect to $\langle \cdot, \cdot \rangle_{\mathbf{p}}$, and write $\mathbf{s} = (s_1, \ldots, s_d)^\top$ for coordinates on U with respect to

---

[1] Compactness of $\mathcal{M}$ and the Hopf-Rinow theorem imply that $\exp_{\mathbf{p}}$ is defined on the entire tangent space $T_{\mathbf{p}}\mathcal{M}$.

this basis; these are *geodesic normal coordinates* for $\mathcal{M}$ around $\mathbf{p}$ with the chart given by the exponential map. Identifying $\exp_{\mathbf{p}}$ with $\iota \circ \exp_{\mathbf{p}}$, where $\iota : \mathcal{M} \hookrightarrow \mathbb{R}^D$ is inclusion, $\exp_{\mathbf{p}}$ is a smooth mapping from an open subset of Euclidean space $\mathbb{R}^d$ into Euclidean space $\mathbb{R}^D$, and thus it admits a Taylor expansion around $\mathbf{s} = 0$:

$$\exp_{\mathbf{p}}(\mathbf{s}) = \mathbf{p} + L_{\mathbf{p}}(\mathbf{s}) + \frac{1}{2}Q_{\mathbf{p}}(\mathbf{s}) + O(\|\mathbf{s}\|_2^3). \tag{6}$$

Equation (6) links *intrinsic coordinates* to *extrinsic coordinates* for $\mathcal{M}$. Here:

- $L_{\mathbf{p}} : T_{\mathbf{p}}\mathcal{M} \to \mathbb{R}^D$ is a homogeneous linear function, the *differential* of the exponential map at $\mathbf{p}$, namely $L_{\mathbf{p}} = D\exp_{\mathbf{p}}(0)$; and
- $Q_{\mathbf{p}} : T_{\mathbf{p}}\mathcal{M} \to \mathbb{R}^D$ is a homogeneous quadratic function (equivalently a linear function of $\mathbf{s}\mathbf{s}^\top$) called the *second fundamental form* of $\mathcal{M}$ at $\mathbf{p}$ [43].

Consider the image (respectively, translated image) of $L_{\mathbf{p}}$:

$$L_{\mathbf{p}}(T_{\mathbf{p}}\mathcal{M}) = \{L_{\mathbf{p}}(\mathbf{s}) : \mathbf{s} \in T_{\mathbf{p}}\mathcal{M}\} \subseteq \mathbb{R}^D, \text{ and} \tag{7}$$
$$\mathbf{p} + L_{\mathbf{p}}(T_{\mathbf{p}}\mathcal{M}) = \{\mathbf{p} + L_{\mathbf{p}}(\mathbf{s}) : \mathbf{s} \in T_{\mathbf{p}}\mathcal{M}\} \subseteq \mathbb{R}^D.$$

We call these the *linear (respectively, affine) embedded tangent space* of $\mathcal{M}$ at $\mathbf{p}$. It is well-known that $L_{\mathbf{p}}$ provides an *isometric embedding* of $T_{\mathbf{p}}\mathcal{M}$ into $\mathbb{R}^D$,

$$\|L_{\mathbf{p}}(\mathbf{s})\|_2 = \|\mathbf{s}\|_2 \text{ for all } \mathbf{s} \in T_{\mathbf{p}}\mathcal{M}. \tag{8}$$

Another important fact is that the second fundamental form $Q_{\mathbf{p}}$ takes values in the *normal* space to $\mathcal{M}$ at $\mathbf{p}$, that is $Q_{\mathbf{p}}(T_{\mathbf{p}}\mathcal{M}) \subseteq L_{\mathbf{p}}(T_{\mathbf{p}}\mathcal{M})^\perp \subseteq \mathbb{R}^D$, i.e.,

$$\langle L_{\mathbf{p}}(\mathbf{s}), Q_{\mathbf{p}}(\mathbf{s}') \rangle_{\mathbb{R}^D} = 0 \text{ for all } \mathbf{s}, \mathbf{s}' \in T_{\mathbf{p}}\mathcal{M}, \tag{9}$$

Finally, let μ denote the *density* on $\mathcal{M}$ uniquely determined by the Riemannian structure on $\mathcal{M}$ as in [33, Prop. 16.45]. The density determines a measure on $\mathcal{M}$, which we refer to as the *uniform measure*. The measure enables integration of measurable functions $f : \mathcal{M} \to \mathbb{R}$, which we write as $\int_{\mathcal{M}} f(\mathbf{x})d\mu(\mathbf{x})$. Then the Riemannian *volume* of $\mathcal{M}$ is $\text{vol}(\mathcal{M}) = \int_{\mathcal{M}} 1 d\mu(\mathbf{x})$.

3.2 Preliminaries from convex geometry

We next give a quick reminder on general norms in finite-dimensional vector spaces, and their equivalence with certain convex bodies. A few facts about tangent cones that we will need are also recorded. The only (possibly) novel content here is Proposition 3. A nice textbook on convex geometry is [27].

Let $\| \cdot \| : \mathbb{R}^D \to \mathbb{R}$ denote an arbitrary vector space *norm* on $\mathbb{R}^D$. This means:

- $\|\mathbf{v}\| \geq 0$ for all $\mathbf{v} \in \mathbb{R}^D$;
- $\|\lambda \mathbf{v}\| = |\lambda| \|\mathbf{v}\|$ for all $\lambda \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^D$;

$-\ \|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^D$.

Recall that $\|\cdot\|$ is necessarily a continuous function on $\mathbb{R}^D$. Also standard is that all norms on $\mathbb{R}^D$ are *equivalent*, that is if $\|\cdot\|\!\|$ is another norm on $\mathbb{R}^D$, then there exist finite positive constants c, C (depending only on $\|\cdot\|, \|\cdot\|\!\|$) such that

$$c\|\mathbf{v}\|\!\| \leq \|\mathbf{v}\| \leq C\|\mathbf{v}\|\!\| \text{ for all } \mathbf{v} \in \mathbb{R}^D. \tag{10}$$

We write $\mathcal{B} \subseteq \mathbb{R}^D$ for the *unit ball* with respect to the norm $\|\cdot\|$,

$$\mathcal{B} = \{\mathbf{v} \in \mathbb{R}^D : \|\mathbf{v}\| \leq 1\}. \tag{11}$$

Then, $\mathcal{B}$ is a *convex body* in $\mathbb{R}^D$. This means: a compact convex subset of $\mathbb{R}^D$ with non-empty interior. Furthermore, the unit ball is *origin-symmetric*, i.e., $\mathbf{v} \in \mathcal{B}$ implies $-\mathbf{v} \in \mathcal{B}$ for all $\mathbf{v} \in \mathbb{R}^D$. Conversely, it is well-known that any origin-symmetric convex body in $\mathbb{R}^D$ occurs as the unit ball for some norm on $\mathbb{R}^D$. Thus, there is a one-to-one correspondence (see [27, Chapter 2]):

$$\{\text{norms } \|\cdot\| \text{ on } \mathbb{R}^D\} \quad \longleftrightarrow \quad \{\text{origin-symmetric convex bodies } \mathcal{B} \subseteq \mathbb{R}^D\}.$$

To emphasize this bijection, we shall let $\|\cdot\|_{\mathcal{B}}$ stand for the norm on $\mathbb{R}^D$ with unit ball $\mathcal{B}$ (except in the case of the ℓ_p-norm where we write $\|\cdot\|_p$), i.e.,

$$\|\cdot\|_{\mathcal{B}} \longleftrightarrow \mathcal{B}.$$

A few general topological remarks follow. Given any subset $\mathcal{Y} \subseteq \mathbb{R}^D$. The (relative) *topological boundary* of $\mathcal{Y}$ is the closure of $\mathcal{Y}$ minus the relative interior of $\mathcal{Y}$, written $\partial \mathcal{Y} := \overline{\mathcal{Y}} \setminus \mathrm{relint}(\mathcal{Y})$. In the case of the unit ball (11), the boundary is the *unit sphere*:

$$\partial \mathcal{B} = \{\mathbf{v} \in \mathbb{R}^D : \|\mathbf{v}\|_{\mathcal{B}} = 1\}.$$

Given any point $\mathbf{y} \in \mathcal{Y}$, the *tangent cone* to $\mathcal{Y}$ at $\mathbf{y}$ is defined to be

$$TC_{\mathbf{y}}(\mathcal{Y}) := \left\{\mathbf{d} \in \mathbb{R}^D : \exists (\mathbf{y}_k)_{k=1}^{\infty} \subseteq \mathcal{Y}, (\tau_k)_{k=1}^{\infty} \subseteq \mathbb{R}_{>0} \text{ s.t. } \tau_k \to 0, \frac{\mathbf{y}_k - \mathbf{y}}{\tau_k} \to \mathbf{d}\right\}. \tag{12}$$

Note that, unlike abstract tangent spaces to manifolds, tangent cones to sets reside in $\mathbb{R}^D$ by definition. We now give a few quick examples of tangent cones.

Example 1 (Tangent cones in familiar cases) For a submanifold $\mathcal{Y} \subseteq \mathbb{R}^D$, the tangent cone and embedded tangent space always agree: $TC_{\mathbf{y}}(\mathcal{Y}) = L_{\mathbf{p}}(T_{\mathbf{y}}(\mathcal{Y}))$. If $\mathcal{Y} = \{(x_1, x_2) : x_2^2 = x_1^3 + x_1^2\} \subseteq \mathbb{R}^2$ is the nodal cubic plane curve, and $\mathbf{y} = 0$ is the node, the tangent cone is the union of two lines: $TC_{\mathbf{y}}(\mathcal{Y}) = \mathbb{R}\begin{pmatrix}1\\1\end{pmatrix} \cup \mathbb{R}\begin{pmatrix}1\\-1\end{pmatrix}$. If $\mathcal{Y} = \{(x_1, x_2) : x_2^2 = x_1^3\} \subseteq \mathbb{R}^2$ is the cuspidal cubic plane curve, and $\mathbf{y} = 0$ is the cusp, the tangent cone is a half-line: $TC_{\mathbf{y}}(\mathcal{Y}) = \mathbb{R}_{\geq 0}\begin{pmatrix}1\\0\end{pmatrix}$. Finally if $\mathcal{Y} = \{\mathbf{x} : \sum_{i=1}^D |x_i| \leq 1\} \subseteq \mathbb{R}^D$ is the ℓ_1

unit ball, the tangent cone $TC_{\mathbf{y}}(\mathcal{Y})$ is all of $\mathbb{R}^D$, a half-space in $\mathbb{R}^D$ or a polyhedron in $\mathbb{R}^D$ depending on whether $\mathbf{y}$ lies in the interior of the unit ball, the relative interior of a facet of the unit sphere or elsewhere on the boundary.

We will use the following easy and well-known facts about tangent cones.

Lemma 2 *1. [53, Lem. 3.12] For all sets $\mathcal{Y} \subseteq \mathbb{R}^D$ and points $\mathbf{y} \in \mathcal{Y}$, we have $TC_{\mathbf{y}}(\mathcal{Y})$ is a closed cone.*
2. For all sets $\mathcal{Y} \subseteq \mathbb{R}^D$, points $\mathbf{y} \in \mathcal{Y}$ and linear subspaces $\mathcal{S} \subseteq \mathbb{R}^D$, we have $TC_{\mathbf{y}}(\mathcal{Y}) \cap \mathcal{S} = TC_{\mathbf{y}}(\mathcal{Y} \cap \mathcal{S})$.
3. [53, Lem. 3.13] For all convex sets $\mathcal{Y} \subseteq \mathbb{R}^D$ and points $\mathbf{y} \in \mathcal{Y}$, we have the explicit description (overline denotes closure in the Euclidean topology):

$$TC_{\mathbf{y}}(\mathcal{Y}) = \overline{\mathbb{R}_{\geq 0}(\mathcal{Y} - \mathbf{y})} := \overline{\{\beta(\widetilde{\mathbf{y}} - \mathbf{y}) \in \mathbb{R}^D : \beta \in \mathbb{R}_{>0}, \widetilde{\mathbf{y}} \in \mathcal{Y}\}}. \quad (13)$$

In particular, if $\mathcal{Y}$ is convex (respectively, convex with non-empty interior), then $TC_{\mathbf{y}}(\mathcal{Y})$ is convex (respectively, convex with non-empty interior).

In light of the third item, we know all the possibilities in the plane for the tangent cone to convex sets with non-empty interior.

Example 2 The closed convex cones in $\mathbb{R}^2$ with non-empty interior are precisely $\mathbb{R}^2$, closed half-spaces and the *conical hull* of two linearly independent vectors:

$$\mathrm{coni}\{\mathbf{d}_1, \mathbf{d}_2\} := \{\beta_1 \mathbf{d}_1 + \beta_2 \mathbf{d}_2 : \beta_1, \beta_2 \in \mathbb{R}_{\geq 0}\} \subseteq \mathbb{R}^2, \quad \mathbf{d}_1, \mathbf{d}_2 \in \mathbb{R}^2.$$

In the latter case, the pair $\mathbf{d}_1, \mathbf{d}_2$ is unique up to positive scales, and one says that they generate the cone's *extremal rays*, $\mathrm{coni}\{\mathbf{d}_1\}$ and $\mathrm{coni}\{\mathbf{d}_2\}$.

Finally, for technical purposes of the "tilt construction" developed in the next section, we need to observe that the topological boundary and tangent cone operations commute, at least in the case of our interest.

Lemma 3 *For $\mathcal{B} \subseteq \mathbb{R}^D$ the unit ball of a norm $\| \cdot \|_{\mathcal{B}}$ and a boundary point $\mathbf{y} \in \partial\mathcal{B}$, the boundary of the tangent cone is the tangent cone of the boundary:*

$$\partial\left(TC_{\mathbf{y}}(\mathcal{B})\right) = TC_{\mathbf{y}}(\partial\mathcal{B}). \quad (14)$$

We include a proof of Lemma 3 in Appendix A, since we could not readily find this statement in the literature.

3.3 Tilt construction

In this section, we present a construction that relates the second-order geometry of a submanifold $\mathcal{M} \subseteq \mathbb{R}^D$ around a point $\mathbf{p} \in \mathcal{M}$ to tangent cones to the unit sphere $\partial\mathcal{B} \subseteq \mathbb{R}^D$ of a norm $\| \cdot \|_{\mathcal{B}}$. We name this construction the *tilt function*, and denote it by $\mathrm{tilt}_{\mathcal{M}, \mathcal{B}, \mathbf{p}}$. Though not apparent initially, the relevance is that this function is required to define the limiting differential operator for point-cloud Laplacians formed by sampling $\mathcal{M}$ and computing affinities using $\| \cdot \|_{\mathcal{B}}$. Specifically, it appears in the first-order derivative term in Eq. (33).
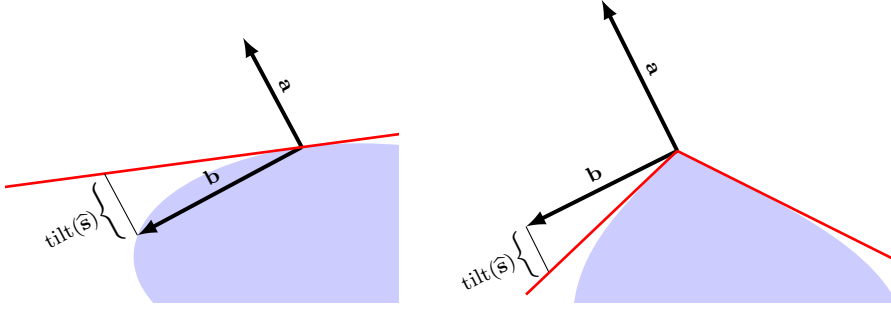
Fig. 1 *Tilt construction.* These diagrams take place in the 2D linear subspace $\mathcal{S} :=$ Span$\{\mathbf{a}, \mathbf{b}\} \subseteq \mathbb{R}^D$, where $\mathbf{a} := L_{\mathbf{p}}(\widehat{\mathbf{s}})$ and $\mathbf{b} := \frac{1}{2} Q_{\mathbf{p}}(\widehat{\mathbf{s}})$ are tangent and normal vectors to $\mathcal{M}$ at $\mathbf{p}$ respectively. Blue indicates $\widetilde{\mathcal{B}} = \mathcal{B} \cap \mathcal{S}$ (2D linear section of the unit ball $\mathcal{B}$). Red indicates the tangents to $\partial \widetilde{\mathcal{B}}$ at $\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}$. By definition, tilt$_{\mathcal{M}, \mathcal{B}, \mathbf{p}}(\widehat{\mathbf{s}})$ equals the signed ℓ_2-length of the braced line segment. (left) An example where $TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\partial \widetilde{\mathcal{B}})$ consists of one well-defined tangent line. Here tilt is positive; (right) An example where $TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\partial \widetilde{\mathcal{B}})$ consists of two tangent rays due to a singularity of $\partial \widetilde{\mathcal{B}}$ at $\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}$. Here tilt is negative.
.

Proposition/Definition 4 (Tilt function) *Let $\mathcal{M} \subseteq \mathbb{R}^D$ be a compact embedded Riemannian submanifold, let $\mathbf{p} \in \mathcal{M}$ be a point, and let $\widehat{\mathbf{s}} \in T_{\mathbf{p}}\mathcal{M}$ be a tangent vector to $\mathcal{M}$ at $\mathbf{p}$ with $\|\widehat{\mathbf{s}}\|_2 = 1$. Following Eq. (6), consider the differential of the exponential map at $\mathbf{p}$ and the second fundamental form at $\mathbf{p}$ both evaluated at $\widehat{\mathbf{s}}$, and write*

$$\mathbf{a} := L_{\mathbf{p}}(\widehat{\mathbf{s}}), \qquad \mathbf{b} := \frac{1}{2} Q_{\mathbf{p}}(\widehat{\mathbf{s}}).$$

Further, let $\|\cdot\|_{\mathcal{B}}$ denote a norm on $\mathbb{R}^D$ with unit ball $\mathcal{B} \subseteq \mathbb{R}^D$ and unit sphere $\partial \mathcal{B} \subseteq \mathbb{R}^D$. Also write TC to denote tangent cones as defined by Eq. (12).

Then, there exists a unique scalar $\eta \in \mathbb{R}$ such that

$$\frac{\mathbf{b}}{\|\mathbf{a}\|_{\mathcal{B}}^2} + \eta \mathbf{a} \in TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\partial \mathcal{B}). \tag{15}$$

We define the tilt function by

$$\text{tilt}_{\mathcal{M}, \mathcal{B}, \mathbf{p}}(\widehat{\mathbf{s}}) := \eta. \tag{16}$$

Hence tilt$_{\mathcal{M}, \mathcal{B}, \mathbf{p}}$ is a well-defined function from Euclidean-normalized tangent vectors to $\mathcal{M}$ at $\mathbf{p}$ into the real numbers.

Remark 2 In the course of proving Propostion/Definition 4 below, we shall show that the tilt function tilt$_{\mathcal{M}, \mathcal{B}, \mathbf{p}}(\widehat{\mathbf{s}})$ only depends on the norm $\|\cdot\|_{\mathcal{B}}$ through the following (typically) two-dimensional central slice of the unit ball:

$$\text{Span}\{L_{\mathbf{p}}(\widehat{\mathbf{s}}), Q_{\mathbf{p}}(\widehat{\mathbf{s}})\} \cap \mathcal{B}.$$

This is two-dimensional origin-symmetric convex body (unless $Q_{\mathbf{p}}(\widehat{\mathbf{s}}) = 0$, in which case it is an origin-symmetric line segment). We make two remarks.

First, as a consequence, we can visualize the tilt function using two-dimensional figures on the page (see Figure 1). Second, in general, the central planar sections of the unit ball of a norm can vary significantly, and indeed qualitatively, across different slices. For example, for the ℓ_1-ball in $\mathbb{R}^3$, there is not a unique combinatorial type for a central planar section: instead either a quadrilateral or a hexagon can occur depending on the specific slice.

Proof In the defining equation (15) for $\mathrm{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}}(\widehat{\mathbf{s}})$, note that it is equivalent to require the left-hand side lies in the linear slice of the tangent cone:

$$\mathcal{S} \cap TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\partial\mathcal{B}) \tag{17}$$
$$\text{where } \mathcal{S} := \mathrm{Span}\{L_{\mathbf{p}}(\widehat{\mathbf{s}}), Q_{\mathbf{p}}(\widehat{\mathbf{s}})\} \subseteq \mathbb{R}^D,$$

since membership in the linear space $\mathcal{S}$ is guaranteed by definition.

We shall rewrite the set (17) using basic properties relating tangent cones, boundaries, and intersection by linear spaces. Firstly, we have

$$\mathcal{S} \cap TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\partial\mathcal{B}) = TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\mathcal{S} \cap \partial\mathcal{B}) \tag{18}$$

by Lemma 2, item 2. Next, let

$$\widetilde{\mathcal{B}} := \mathcal{S} \cap \mathcal{B} \subseteq \mathcal{S},$$

and note this is the unit ball of the restriction of the norm $\|\cdot\|_{\mathcal{B}}$ to the subspace $\mathcal{S}$, which is a norm in its own right on $\mathcal{S}$ (in our notation, $\|\cdot\|_{\widetilde{\mathcal{B}}}$). Then,

$$\mathcal{S} \cap \partial\mathcal{B} = \{\mathbf{t} \in \mathcal{S} : \|\mathbf{t}\|_{\mathcal{B}} = 1\} = \{\mathbf{t} \in \mathcal{S} : \|\mathbf{t}\|_{\widetilde{\mathcal{B}}} = 1\} = \partial\widetilde{\mathcal{B}},$$

from which it follows

$$TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\mathcal{S} \cap \partial\mathcal{B}) = TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\partial\widetilde{\mathcal{B}}). \tag{19}$$

Now by Lemma 3,

$$TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\partial\widetilde{\mathcal{B}}) = \partial TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\widetilde{\mathcal{B}}). \tag{20}$$

Combining Eq. (18), (19) and (20), we get that

$$\mathcal{S} \cap TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\partial\mathcal{B}) = \partial TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\widetilde{\mathcal{B}}). \tag{21}$$

The upshot is that in the defining equation for $\mathrm{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}}$ it is equivalent to require membership in the right-hand side of (21).

We shall now obtain a more explicit description of the set (21). Firstly, note that $\mathbf{a} = L_{\mathbf{p}}(\widehat{\mathbf{s}}) \neq 0$, since $\|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_2 = \|\mathbf{s}\|_2$ (Eq. (8)). If $\mathbf{b} = \frac{1}{2}Q_{\mathbf{p}}(\widehat{\mathbf{s}}) = 0$, then the subspace $\mathcal{S}$ is one-dimensional. In this case, the existence and uniqueness of η in Eq. (15) is clear: $\widetilde{\mathcal{B}}$ is a line segment, $TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\widetilde{\mathcal{B}})$ is a ray (half-line), and its boundary $\partial TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\widetilde{\mathcal{B}})$ is the origin. Thus in the light of Eq. (21), we must take $\eta = 0$, so that $\mathrm{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}}(\widehat{\mathbf{s}}) = 0$ when the second fundamental form vanishes. Therefore, assume $\mathbf{b} \neq 0$. Since $\langle\mathbf{a}, \mathbf{b}\rangle = 0$ (Eq. (9)), it follows that $\mathcal{S} \cong \mathbb{R}^2$ is two-dimensional and $\widetilde{\mathcal{B}}$ is a convex body in $\mathbb{R}^2$. By Lemma 2, items

1 and 3, we know the tangent cone $TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}$ is a closed convex cone in $\mathbb{R}^2$ with non-empty interior. Then by Example 2, the tangent cone is either all of $\mathbb{R}^2$, a half-plane in $\mathbb{R}^2$ or it is conically spanned by two linearly independent vectors. We claim $TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\widetilde{\mathcal{B}}) \neq \mathbb{R}^2$. Indeed since $\widetilde{\mathcal{B}}$ is convex and $\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}$ lies in the boundary of $\widetilde{\mathcal{B}}$, the supporting hyperplane theorem implies:

$$\exists \mathbf{v} \in \mathcal{S} \setminus \{0\} \; \exists \gamma \in \mathbb{R} \;\; \text{s.t.} \; \langle \mathbf{v}, \mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}} \rangle = \gamma \; \wedge \; \forall \mathbf{u} \in \widetilde{\mathcal{B}}, \; \langle \mathbf{v}, \mathbf{u} \rangle \geq \gamma.$$

Combining this with Eq. (13), it follows

$$TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\widetilde{\mathcal{B}}) = \overline{\mathbb{R}_{\geq 0}(\widetilde{\mathcal{B}} - \mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}})} \subseteq \{\mathbf{u} \in \mathcal{S} : \langle \mathbf{v}, \mathbf{u} \rangle \geq 0\}.$$

In particular, $TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\widetilde{\mathcal{B}}) \neq \mathbb{R}^2$, so by Example 2 the tangent cone is either a half-plane or conically spanned by two extremal rays. For now, assume the latter case: there exist linearly independent vectors $\mathbf{d}_1, \mathbf{d}_2 \in \mathcal{S} \cong \mathbb{R}^2$ such that

$$TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\widetilde{\mathcal{B}}) = \text{coni}\{\mathbf{d}_1, \mathbf{d}_2\} \subseteq \mathbb{R}^2. \tag{22}$$

The set (20) is thus the union of two rays:

$$\partial TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\widetilde{\mathcal{B}}) = \mathbb{R}_{\geq 0}\mathbf{d}_1 + \mathbb{R}_{\geq 2}\mathbf{d}_2. \tag{23}$$

We shall now finish by proving the existence and uniqueness of $\eta \in \mathbb{R}$ such that

$$\frac{\mathbf{b}}{\|\mathbf{a}\|_{\mathcal{B}}^2} + \eta \mathbf{a} \; \in \; \mathbb{R}_{\geq 0}\mathbf{d}_1 \cup \mathbb{R}_{\geq 0}\mathbf{d}_2. \tag{24}$$

To this end, first note that

$$\begin{aligned}
-\mathbf{a} = 0 - \|\mathbf{a}\|_{\mathcal{B}}(\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}) \; &\in \; \text{relint}(\overline{\mathbb{R}_{\geq 0}(\widetilde{\mathcal{B}} - \mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}})}) \\
&= \; \text{relint}(TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\widetilde{\mathcal{B}})) \\
&= \; \mathbb{R}_{>0}\mathbf{d}_1 + \mathbb{R}_{>0}\mathbf{d}_2.
\end{aligned}$$

where the penultimate equality is again by Eq. (13) and the last equality is by Eq. (22). Thus there exist positive scalars $\beta_1, \beta_2 \in \mathbb{R}_{>0}$ such that $-\mathbf{a} = \beta_1 \mathbf{d}_1 + \beta_2 \mathbf{d}_2$. Substituting this into $\langle \mathbf{a}, \mathbf{b} \rangle = 0$ (Eq. (9)), we get

$$\beta_1 \langle \mathbf{d}_1, \mathbf{b} \rangle + \beta_2 \langle \mathbf{d}_2, \mathbf{b} \rangle = 0. \tag{25}$$

Since $\mathbf{d}_1, \mathbf{d}_2$ form a basis for $\mathcal{S}$ and $\mathbf{b} \in \mathcal{S}$ and we are presently assuming $\mathbf{b} \neq 0$, it cannot be that $\langle \mathbf{d}_1, \mathbf{b} \rangle = \langle \mathbf{d}_2, \mathbf{b} \rangle = 0$. Instead, Eq. (25) combined with $\beta_1, \beta_2 > 0$ imply that exactly one of the inner products $\langle \mathbf{d}_1, \mathbf{b} \rangle, \langle \mathbf{d}_2, \mathbf{b} \rangle$ is strictly positive while the other is strictly negative. Relabeling if necessary, we can assume that $\langle \mathbf{d}_1, \mathbf{b} \rangle > 0 > \langle \mathbf{d}_2, \mathbf{b} \rangle$. With this in hand, let us examine the membership (23).

Notice that for each $\eta \in \mathbb{R}$, it holds

$$\frac{\mathbf{b}}{\|\mathbf{a}\|_{\mathcal{B}}} + \eta \mathbf{a} \; \notin \; \mathbb{R}_{\geq 0}\mathbf{d}_2. \tag{26}$$

This is by taking inner products with $\mathbf{b}$: all vectors on the right-hand side of (26) have a non-positive inner product with $\mathbf{b}$ using $\langle \mathbf{d}_2, \mathbf{b} \rangle < 0$. Meanwhile on the left-hand side of (26), we have $\langle \mathbf{b}, \frac{\mathbf{b}}{\|\mathbf{a}\|_{\mathcal{B}}} + \eta \mathbf{a} \rangle = \frac{\|\mathbf{b}\|_2^2}{\|\mathbf{a}\|_{\mathcal{B}}} > 0$ (the equality is from $\langle \mathbf{a}, \mathbf{b} \rangle = 0$ and the strict inequality is from the assumption $\mathbf{b} \neq 0$).

On the other hand, there do exist scalars $\eta \in \mathbb{R}$ and $\beta \in \mathbb{R}_{\geq 0}$ satisfying

$$\frac{\mathbf{b}}{\|\mathbf{a}\|_{\mathcal{B}}^2} + \eta \mathbf{a} = \beta \mathbf{d}_1. \tag{27}$$

Indeed using that $\mathbf{a}, \mathbf{b}$ are an orthogonal basis for $\mathcal{S}$, $\|\mathbf{a}\|_2^2 = 1$ and $\mathbf{d}_1 \in \mathcal{S}$, note

$$\mathbf{d}_1 = \langle \mathbf{d}_1, \mathbf{a} \rangle \mathbf{a} + \frac{\langle \mathbf{d}_1, \mathbf{b} \rangle}{\|\mathbf{b}\|_2^2} \mathbf{b}. \tag{28}$$

Then substituting Eq. (28) into (27) and equating coefficients, we compute the following *unique* solution to Eq. (27):

$$\beta = \|\mathbf{b}\|_2^2 \Big/ \left(\|\mathbf{a}\|_{\mathcal{B}}^2 \langle \mathbf{d}_1, \mathbf{b} \rangle \right),$$

$$\eta = \|\mathbf{b}\|_2^2 \langle \mathbf{d}_1, \mathbf{a} \rangle \Big/ \left(\|\mathbf{a}\|_{\mathcal{B}}^2 \langle \mathbf{d}_1, \mathbf{b} \rangle \right). \tag{29}$$

This completes the case when $TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\widetilde{\mathcal{B}})$ is conically spanned by independent vectors. As for the third case afforded by Example 2, when the tangent cone is a half-plane, let the boundary of the half-plane be spanned by $\mathbf{d}_1 \in \mathcal{S} \cong \mathbb{R}^2$. Again we arrive at Eq. (27) but without the constraint that $\beta \geq 0$. Solving as before, η is uniquely determined and given by Eq. (29).

This completes the proof that η exists and is unique. In sum: if $Q_{\mathbf{p}}(\widehat{\mathbf{s}}) = 0$ then $\mathrm{tilt}_{\mathcal{M}, \mathcal{B}, \mathbf{p}}(\widehat{\mathbf{s}}) = 0$, and otherwise $\mathrm{tilt}_{\mathcal{M}, \mathcal{B}, \mathbf{p}}(\widehat{\mathbf{s}})$ is given by Eq. (29). □

In the next statement, we assume the local continuous differentiability of the norm to get a more explicit expression for the tilt function. The proof is in Appendix B.

Proposition 5 (Simplifications for tilt in the case of C^1-norm**)** *Regard the norm $\|\cdot\|_{\mathcal{B}}$ as a function from $\mathbb{R}^D$ to $\mathbb{R}$.*

1. *Let $\widehat{\mathbf{a}}$ be a point in $\mathbb{R}^D$ with $\|\widehat{\mathbf{a}}\|_{\mathcal{B}} = 1$. If $\|\cdot\|_{\mathcal{B}}$ is continuously differentiable in a neighborhood of $\widehat{\mathbf{a}}$, then the tangent cone to the $\|\cdot\|_{\mathcal{B}}$-unit sphere at $\widehat{\mathbf{a}}$ is the hyperplane:*

$$TC_{\widehat{\mathbf{a}}}(\partial \mathcal{B}) = \left\{ \mathbf{v} \in \mathbb{R}^D : \langle \mathbf{v}, \mathrm{grad} \, \|\cdot\|_{\mathcal{B}}(\widehat{\mathbf{a}}) \rangle = 0 \right\}. \tag{30}$$

2. *Assume the setup of Proposition/Definition 4, and further that $\|\cdot\|_{\mathcal{B}}$ is continuously differentiable in a neighborhood of the point $L_{\mathbf{p}}(\widehat{\mathbf{s}})$. Then, the tilt function equals*

$$\mathrm{tilt}_{\mathcal{M}, \mathcal{B}, \mathbf{p}}(\widehat{\mathbf{s}}) = \frac{-\left\langle \mathrm{grad} \, \|\cdot\|_{\mathcal{B}}(L_{\mathbf{p}}(\widehat{\mathbf{s}})), \frac{1}{2} Q_{\mathbf{p}}(\widehat{\mathbf{s}}) \right\rangle}{\left\langle \mathrm{grad} \, \|\cdot\|_{\mathcal{B}}(L_{\mathbf{p}}(\widehat{\mathbf{s}})), L_{\mathbf{p}}(\widehat{\mathbf{s}}) \right\rangle} \frac{1}{\|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}^2}. \tag{31}$$

3.4 **Laplacian-like operator and main theorem statement**

Here we state our main result in Theorem 6. We first need to define both sides of Eq. (34), in particular the differential operator $\Delta_{\mathcal{M},\mathcal{B}}$ (Definition 2). For simplicity, we consider only the standard Gaussian kernel with width σ:

$$K_\sigma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0} \quad \text{defined by} \quad K_\sigma(t) = \exp(-t^2/\sigma^2).$$

Definition 1 (Point-cloud Laplacian with respect to an arbitrary norm) Let $\| \cdot \|_{\mathcal{B}}$ be a norm on $\mathbb{R}^D$. Let $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subseteq \mathbb{R}^D$ be a set of points. Then we define the *point-cloud Laplacian* computed using the norm $\| \cdot \|_{\mathcal{B}}$ and the point set $\mathcal{X} \subseteq \mathbb{R}^D$ to act on functions whose domains contain $\mathcal{X}$ as follows: for each $f : \mathcal{Y} \to \mathbb{R}$ where $\mathcal{X} \subseteq \mathcal{Y}$, define $\mathcal{L}_{n,\mathcal{B}} f : \mathcal{Y} \to \mathbb{R}$ by

$$\mathcal{L}_{n,\mathcal{B}} f(\mathbf{p}) := \frac{1}{n} \sum_{i=1}^{n} K_{\sigma_n}(\|\mathbf{x}^{(i)} - \mathbf{p}\|_{\mathcal{B}})(f(\mathbf{x}^{(i)}) - f(\mathbf{p})). \tag{32}$$

Compare Eq. (5) with Eq. (32). In what follows, $d\mathbf{s}$ is the Lebesgue measure on $T_{\mathbf{p}}\mathcal{M}$, and $d\widehat{\mathbf{s}}$ is the uniform measure on the sphere $\{\widehat{\mathbf{s}} \in T_{\mathbf{p}}\mathcal{M} : \|\widehat{\mathbf{s}}\|_2 = 1\}$.

Definition 2 (Laplacian-like differential operator with respect to an arbitrary norm) The *Laplacian-like operator* on a submanifold $\mathcal{M} \subseteq \mathbb{R}^D$ with respect to a norm $\| \cdot \|_{\mathcal{B}}$ is defined to act on functions $f : \mathcal{M} \to \mathbb{R}$ according to

$$(\Delta_{\mathcal{M},\mathcal{B}} f)(\mathbf{p}) := \left\langle (\text{hess } \widetilde{f})(0), \frac{1}{2} \int_{\{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}} \leq 1\}} \mathbf{s}\mathbf{s}^\top d\mathbf{s} \right\rangle$$

$$+ \left\langle (\text{grad } \widetilde{f})(0), \int_{\{\widehat{\mathbf{s}} \in T_{\mathbf{p}}\mathcal{M} : \|\widehat{\mathbf{s}}\|_2 = 1\}} \widehat{\mathbf{s}} \|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}^{-d} \text{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}}(\widehat{\mathbf{s}}) \, d\widehat{\mathbf{s}} \right\rangle \tag{33}$$

where $\widetilde{f} = f \circ \exp_{\mathbf{p}} : T_{\mathbf{p}}\mathcal{M} \to \mathbb{R}$ and $L_{\mathbf{p}} = D\exp_{\mathbf{p}}(0) : T_{\mathbf{p}}\mathcal{M} \to \mathbb{R}^D$.

Remark 3 (Extrinsic interpretation of the integration domains and integrands in Eq. (33)) Both domains of integration in Definition 2 are subsets of the abstract tangent space $T_{\mathbf{p}}\mathcal{M}$, since we have written the integrals in parameterized form. Using the isometry $L_{\mathbf{p}}$, we can identify the first domain with the d-dimensional intersection of the embedded tangent space and the unit ball:

$$L_{\mathbf{p}}(T_{\mathbf{p}}\mathcal{M}) \cap \mathcal{B},$$

and the first integral in Eq. (33) with the second-moment of this convex body:

$$\int_{\mathbf{t} \in L_{\mathbf{p}}(T_{\mathbf{p}}\mathcal{M}) \cap \mathcal{B}} \mathbf{t}\mathbf{t}^\top d\mathbf{t}.$$

Meanwhile, under the mapping $s \mapsto L_{\mathbf{p}}(\mathbf{s})/\|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}}$, the second domain of integration in Definition 2 identifies with the $(d-1)$-dimensional intersection of the embedded tangent with the unit sphere:

$$L_{\mathbf{p}}(T_{\mathbf{p}}\mathcal{M}) \cap \partial\mathcal{B},$$

and the second integral in Eq. (33) is a weighted first-moment of this boundary.

Theorem 6 (Main result: Convergence of the point-cloud Laplacian based on an arbitrary norm) *Let $\| \cdot \|_{\mathcal{B}}$ be a norm on $\mathbb{R}^D$ with unit ball $\mathcal{B}$. Let $\mathcal{M}$ be a compact d-dimensional embedded Riemannian submanifold of $\mathbb{R}^D$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be i.i.d. draws from the uniform measure on $\mathcal{M}$. Fix any constant $\alpha > 0$, and set $\sigma_n = n^{-1/(2d+4+\alpha)}$ and $c_n := \Gamma(\frac{d+4}{2})\sigma_n^{d+2}$. Then given a three-times continuously differentiable function $f : \mathcal{M} \to \mathbb{R}$ and a point $\mathbf{p} \in \mathcal{M}$, we have the following almost sure convergence:*

$$\frac{1}{c_n}\mathcal{L}_{n,\mathcal{B}}f(\mathbf{p}) \xrightarrow{\text{a.s.}} \frac{1}{\text{vol}(\mathcal{M})}\Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p}). \tag{34}$$

Theorem 6 is proved in Section 4.

Remark 4 (Comparing $\Delta_{\mathcal{M},\mathcal{B}}$ and $\Delta_{\mathcal{M}}$) Note two key features that distinguish the continuum limit for general norms from the Laplace-Beltrami operator:

– The operator $\Delta_{\mathcal{M},\mathcal{B}}$ is typically **extrinsic**. By Remark 3, both terms in (33) vary with the orientation of the embedded tangent space $L_{\mathbf{p}}(T_{\mathbf{p}}\mathcal{M}) \subseteq \mathbb{R}^D$ in relation to the ball $\mathcal{B} \subseteq \mathbb{R}^D$. (For a concrete example, see Section 3.7.)

– The operator $\Delta_{\mathcal{M},\mathcal{B}}$ has a **first-order derivative term**.

The Euclidean norm is special on both counts. However despite the added complexity of general norms, there can be **practical advantages** to using $\Delta_{\mathcal{M},\mathcal{B}}$ over $\Delta_{\mathcal{M}}$. At least this can hold for a well-chosen norm, when reducing the dimension of certain data sets. We illustrate this numerically in Section 5.

We finish this section with an easy extension of Theorem 6 to the case where the sampling of $\mathcal{M}$ is non-uniform.

Theorem 7 (Convergence of the point-cloud Laplacian based on an arbitrary norm with non-uniform sampling) *Assume the same setup as Theorem 6 above, except $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are i.i.d. draws from a probability distribution on $\mathcal{M}$ described by a C^3 probability density function, $dP(\mathbf{x}) = P(\mathbf{x})d\mu(\mathbf{x})$. Then, the almost sure limit of the LHS of (34) exists and equals $\Delta_{\mathcal{M},\mathcal{B},P}f(\mathbf{p})$, where*

$$\Delta_{\mathcal{M},\mathcal{B},P} := \text{vol}(\mathcal{M})P\Delta_{\mathcal{M},\mathcal{B}} + \delta_{\mathcal{M},\mathcal{B},P}. \tag{35}$$

Here $\delta_{\mathcal{M},\mathcal{B},P}$ only modifies the first-order derivative term, and is defined by

$$(\delta_{\mathcal{M},\mathcal{B},P}f)(\mathbf{p}) := \left\langle (\text{grad }\widetilde{f}(0))(\text{grad }\widetilde{P}(0))^{\top}, \int_{\{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M}:\|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}}\leq 1\}} \mathbf{s}\mathbf{s}^{\top} d\mathbf{s} \right\rangle$$

$$= \int_{\{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M}:\|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}}\leq 1\}} \langle \text{grad }\widetilde{f}(0), \mathbf{s}\rangle\langle \text{grad }\widetilde{P}(0), \mathbf{s}\rangle d\mathbf{s}, \tag{36}$$

where $\widetilde{f} = f \circ \exp_{\mathbf{p}}$ and $\widetilde{P} = P \circ \exp_{\mathbf{p}}$.

Proof Using the reduction in [7, Sec. 5] and then Theorem 6, the LHS of (34) tends to $\Delta_{\mathcal{M},\mathcal{B}}h(\mathbf{p})$ where $h : \mathcal{M} \to \mathbb{R}$ is defined by $h(\mathbf{x}) := (f(\mathbf{x}) - f(\mathbf{p}))P(\mathbf{x})$. Let $\widetilde{h} = h \circ \exp_{\mathbf{p}}$, so $\widetilde{h} = \widetilde{f}\,\widetilde{P} - \widetilde{f}(0)\widetilde{P}$. Then $\text{grad }\widetilde{h}(0) = \widetilde{P}(0)\text{ grad }\widetilde{f}(0)$ and $\text{hess }\widetilde{h}(0) = \widetilde{P}(0)\text{ hess }\widetilde{f}(0) + (\text{grad }\widetilde{f}(0))(\text{grad }\widetilde{P}(0))^{\top} + (\text{grad }\widetilde{P}(0))(\text{grad }\widetilde{f}(0))^{\top}$. Inserting these formulas into Definition 2 and rearranging gives the result. □

3.5 First properties of $\Delta_{\mathcal{M},\mathcal{B}}$

We give a few basic properties of the limit in Theorem 6. Firstly, it is elliptic.

Lemma 8 ($\Delta_{\mathcal{M},\mathcal{B}}$ is elliptic) *For all compact embedded Riemannian submanifolds $\mathcal{M} \subseteq \mathbb{R}^D$ and all norms $\|\cdot\|_{\mathcal{B}}$ on $\mathbb{R}^D$, the Laplacian-like operator $\Delta_{\mathcal{M},\mathcal{B}}$ is a uniformly elliptic differential operator on $\mathcal{M}$.*

The proof of this lemma is in Appendix C.

Next, we investigate the regularity properties of the coefficients of $\Delta_{\mathcal{M},\mathcal{B}}$. Surprisingly, the first-order coefficient function need not even be continuous everywhere. Below, $T\mathcal{M} = \bigsqcup_{\mathbf{p}\in\mathcal{M}} T_{\mathbf{p}}\mathcal{M}$ is the tangent bundle of $\mathcal{M}$, and $\operatorname{Sym}^2(T\mathcal{M}) = \bigsqcup_{\mathbf{p}\in\mathcal{M}} \operatorname{Sym}^2(T_{\mathbf{p}}\mathcal{M})$ is its symmetric square bundle [33, Ch. 10].

Proposition 9 (Continuity properties of $\Delta_{\mathcal{M},\mathcal{B}}$) *For all compact embedded Riemannian submanifolds $\mathcal{M} \subseteq \mathbb{R}^D$ and all norms $\|\cdot\|_{\mathcal{B}}$ on $\mathbb{R}^D$, the Laplacian-like operator $\Delta_{\mathcal{M},\mathcal{B}}$ has the following continuity properties.*

1. As a section of $\operatorname{Sym}^2(T\mathcal{M})$, the coefficient of the second-order term,

$$\frac{1}{2} \int_{\mathbf{s}\in T_{\mathbf{p}}\mathcal{M}:\|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}}\leq 1} \mathbf{s}\mathbf{s}^{\top} d\mathbf{s}, \tag{37}$$

is continuous at all points $\mathbf{p} \in \mathcal{M}$.

2. As a section of $T\mathcal{M}$, the coefficient of the first-order term,

$$\int_{\widehat{\mathbf{s}}\in T_{\mathbf{p}}\mathcal{M}:\|\widehat{\mathbf{s}}\|_2=1} \widehat{\mathbf{s}}\|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}^{-d}\operatorname{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}}(\widehat{\mathbf{s}})d\widehat{\mathbf{s}}, \tag{38}$$

is continuous at all points $\mathbf{p} \in \mathcal{M}$ such that the norm $\|\cdot\|_{\mathcal{B}}$ is continuously differentiable in a neighborhood of $L_{\mathbf{p}}(T_{\mathbf{p}}\mathcal{M})\cap\mathbb{S}^{D-1}$. The first-order coefficient can be discontinuous at other points $\mathbf{p} \in \mathcal{M}$.

The proof of the second item relies on the expression for the tilt function in Proposition 5, where the norm is locally continuously differentiable. Details are in Appendix D.

3.6 Example: any manifold, Euclidean norm

Let us first check that Theorem 6 agrees with the standard Euclidean theory. Let $\mathcal{M} \subseteq \mathbb{R}^D$ be any d-dimensional compact smooth embedded Riemannian manifold, $\mathbf{p} \in \mathcal{M}$, and consider the Euclidean norm $\|\cdot\|_2$ with Euclidean unit ball $\mathcal{B} = \{\mathbf{x} : \|\mathbf{x}\|_2 \leq 1\} \subseteq \mathbb{R}^D$.

We first argue that $\operatorname{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}} \equiv 0$. Let $\widehat{\mathbf{s}} \in T_{\mathbf{p}}\mathcal{M}$ with $\|\widehat{\mathbf{s}}\|_2 = 1$. Set $\mathbf{a} = L_{\mathbf{p}}(\widehat{\mathbf{s}})$, $\mathbf{b} = \frac{1}{2}Q_{\mathbf{p}}(\widehat{\mathbf{s}})$ and $\mathcal{S} = \operatorname{Span}\{\mathbf{a},\mathbf{b}\}$. If $\mathbf{b} = 0$, then $\operatorname{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}}(\widehat{\mathbf{s}}) = 0$. Else, put $\widetilde{\mathcal{B}} := \mathcal{B} \cap \mathcal{S}$. By construction, $\operatorname{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}}(\widehat{\mathbf{s}}) = \eta$ for $\eta \in \mathbb{R}$ uniquely determined by

$$\mathbf{b} + \eta\mathbf{a} \in TC_{\mathbf{a}}(\partial\widetilde{\mathcal{B}}),$$

using $\|\mathbf{a}\|_2 = 1$ since $L_{\mathbf{p}}$ is an isometry (Eq. (8)). However, $\widetilde{\mathcal{B}}$ is a Euclidean unit disk in $\mathcal{S} \cong \mathbb{R}^2$, and $\partial\widetilde{\mathcal{B}}$ is a Euclidean unit circle in $\mathbb{R}^2$. So, $TC_{\mathbf{a}}(\partial\widetilde{\mathcal{B}})$ is the orthogonal complement of $\mathbb{R}\mathbf{a}$ inside $\mathcal{S}$. This gives $TC_{\mathbf{a}}(\partial\widetilde{\mathcal{B}}) = \mathbb{R}\mathbf{b}$ since $Q_{\mathbf{p}}$ takes values in the normal space (Eq. (9)). Clearly then, $\eta = 0$ and $\mathrm{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}} \equiv 0$. We have verified that the first-order term vanishes in $\Delta_{\mathcal{M},\mathcal{B}}$.

As for the second-order term, we compute the following second moment:

$$\int_{\{\mathbf{s}\in T_{\mathbf{p}}\mathcal{M}:\|L_{\mathbf{p}}(\mathbf{s})\|_2\leq 1\}} \mathbf{s}\mathbf{s}^\top d\mathbf{s} = \int_{\{\mathbf{s}\in T_{\mathbf{p}}\mathcal{M}:\|\mathbf{s}\|_2\leq 1\}} \mathbf{s}\mathbf{s}^\top d\mathbf{s} \qquad \text{[by Eq. (8)]}$$

$$= \left(\int_{\{\mathbf{s}\in T_{\mathbf{p}}\mathcal{M}:\|\mathbf{s}\|_2\leq 1\}} s_1^2 d\mathbf{s}\right) I_d \quad \text{[oddness, symmetry]}$$

$$\propto I_d. \tag{39}$$

Here the second equality used that the integration domain $\{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \|\mathbf{s}\|_2 \leq 1\}$ is preserved under sign flips of individual coordinates of $\mathbf{s}$, hence the off-diagonal terms $s_i s_j$ $(i \neq j)$ integrate to 0.

Plugging into Definition 2, we obtain

$$\Delta_{\mathcal{M},\mathcal{B}} f(\mathbf{p}) \propto \left\langle \mathrm{hess}\,\widetilde{f}(0), I_d \right\rangle$$

$$= \mathrm{trace}\left(\mathrm{hess}\widetilde{f}(0)\right)$$

$$= \Delta_{\mathcal{M}} f(\mathbf{p}).$$

This is the usual Laplace-Beltrami operator on $\mathcal{M}$ applied to f and evaluated at $\mathbf{p}$. Thus, we have checked that Theorem 6 indeed recovers Theorem 1.

Remark 5 Due to the lack of normalization in Definition 2, our Laplacian-like operator $\Delta_{\mathcal{M},\mathcal{B}}$ in the Euclidean case differs from the usual Laplace-Beltrami operator $\Delta_{\mathcal{M}}$ by a multiplicative constant. From Eq. (39), the constant is

$$\frac{1}{2}\int_{\{\mathbf{s}\in T_{\mathbf{p}}\mathcal{M}:\|\mathbf{s}\|_2\leq 1\}} s_1^2 d\mathbf{s} = \frac{1}{2d}\int_{\{\mathbf{s}\in T_{\mathbf{p}}\mathcal{M}:\|\mathbf{s}\|_2\leq 1\}} \|\mathbf{s}\|_2^2 d\mathbf{s}$$

$$= \frac{1}{2d}\int_0^1 r^{d+1}\,\mathrm{vol}(\mathbb{S}^{d-1})dr = \frac{1}{2d}\int_0^1 r^{d+1}\frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})}dr = \frac{\pi^{d/2}}{4\Gamma(\frac{d+4}{2})}.$$

This is simply the ratio of prefactors in the scales c_n in Theorems 1 and 6.

3.7 Example: circle in the plane, weighted ℓ_1-norm

Next, we look at a non-Euclidean example in full detail. Consider the (Euclidean) unit circle in $\mathbb{R}^2$ where the ambient norm is a weighted ℓ_1-norm. That is, let $\mathcal{M} = S^1 = \{\mathbf{x} = (x_1, x_2)^\top \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ and use the norm $\|\cdot\|_{\mathbf{w},1}$

defined by $\|\mathbf{x}\|_{\mathbf{w},1} = w_1|x_1| + w_2|x_2|$ where $\mathbf{w} = (w_1, w_2)^\top \in (\mathbb{R}_{>0})^2$. The unit ball of $\|\cdot\|_{\mathbf{w},1}$ is the region

$$\mathcal{B} = \left\{ \mathbf{x} = (x_1, x_2)^\top \in \mathbb{R}^2 : w_1|x_1| + w_2|x_2| \leq 1 \right\},$$

while the unit sphere of $\|\cdot\|_{\mathbf{w},1}$ is

$$\partial\mathcal{B} = \left\{ \mathbf{x} = (x_1, x_2)^\top \in \mathbb{R}^2 : w_1|x_1| + w_2|x_2| = 1 \right\},$$

a rhombus with vertices $(\pm(1/w_1), 0), (0, \pm(1/w_2))$. Let $\mathbf{p} = (\cos(\theta), \sin(\theta))^\top \in S^1$. Parameterize $T_\mathbf{p}S^1$ (with respect to a fixed unit basis vector) using $\psi \in \mathbb{R}$. The exponential map is

$$\exp_\mathbf{p} : T_\mathbf{p}S^1 \to S^1, \quad \psi \mapsto (\cos(\theta + \psi), \sin(\theta + \psi))^\top.$$

The differential of this is

$$L_\mathbf{p}(\psi) = \psi\,(-\sin(\theta), \cos(\theta))^\top.$$

The second fundamental form is

$$Q_\mathbf{p}(\psi) = -\psi^2\,(\cos(\theta), \sin(\theta))^\top.$$

We take on the terms in the limiting operator in Definition 2 in turn. For the second-order term, we need the second moment of a line segment:

$$\frac{1}{2}\int_{\{\psi : \|L_\mathbf{p}(\psi)\|_{\mathbf{w},1} \leq 1\}} \psi^2 d\psi = \frac{1}{2}\int_{|\psi| \leq \|(-\sin(\theta),\cos(\theta))^\top\|_{\mathbf{w},1}^{-1}} \psi^2 d\psi$$

$$= \frac{1}{3\,(w_1|\sin(\theta)| + w_2|\cos(\theta)|)^3}.$$

As for the first-order coefficient, this becomes a sum of over the two endpoints of the line segment. We shall show the first-order coefficient equals

$$\text{sign}(\cos(\theta)\sin(\theta))\frac{-w_1|\cos(\theta)| + w_2|\sin(\theta)|}{(w_1|\sin(\theta)| + w_2|\cos(\theta)|)^4}, \tag{40}$$

where $\text{sign} : \mathbb{R} \to \{-1, 0, 1\}$ is given by $\text{sign}(t) := 1$ if $t > 0$; $\text{sign}(t) := -1$ if $t < 0$; and $\text{sign}(0) := 0$. By the symmetry of the rhombus $\partial\mathcal{B}$ with respect to individual coordinate sign flips in $\mathbb{R}^2$, one easily sees formula (40) is correct for θ an integer multiple of $\frac{\pi}{2}$ (the first-order coefficient is zero). Otherwise, we may reduce to verifying correctness of the expression (40) when $\theta \in (0, \frac{\pi}{2})$. Then in this case, it is enough to show

$$\text{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}}(1) = \frac{1}{2}\frac{-w_1\cos(\theta) + w_2\sin(\theta)}{(w_1\sin(\theta) + w_2\cos(\theta))^3}. \tag{41}$$

To this end, let $\alpha \in (0, \frac{\pi}{2})$ be half the angle $\partial\mathcal{B}$ makes at $(\frac{1}{w_1}, 0)^\top$, so $\tan(\alpha) = w_1/w_2$. Let $\mathbf{a} = L_\mathbf{p}(1)$ and $\mathbf{b} = \frac{1}{2}Q_\mathbf{p}(1)$. Let ω be the signed angle at

$\mathbf{a}/\|\mathbf{a}\|_{\mathbf{w},1} \in \partial\mathcal{B}$ from $\mathbf{a}/\|\mathbf{a}\|_{\mathbf{w},1} + \mathbb{R}_{\geq 0}(\frac{-1}{w_1}, \frac{-1}{w_2})^\top$ to $\mathbf{a}/\|\mathbf{a}\|_{\mathbf{w},1} + \mathbb{R}_{\geq 0}\mathbf{b}$, where counterclockwise counts as positive. By elementary angle chasing,

$$\omega = \theta - \alpha.$$

Thus (see Figure 1),

$$\text{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}}(1) = \left\| \frac{\mathbf{b}}{\|\mathbf{a}\|_{\mathbf{w},1}^2} \right\|_2 \tan(\omega) = \frac{1}{2\|\mathbf{a}\|_{\mathbf{w},1}^2}\tan(\theta - \alpha)$$

$$= \frac{1}{2\left(w_1\sin(\theta) + w_2\cos(\theta)\right)^2} \frac{\tan(\theta) - (w_1/w_2)}{1 + \tan(\theta)(w_1/w_2)},$$

which indeed simplifies to Eq. (41).

Summarizing: for each angle $\theta \in [0, 2\pi]$, Theorem 6 implies the Laplacian-like differential operator $\Delta_{\mathcal{M},\mathcal{B}}$ is given by

$$\text{sign}(\cos\theta \, \sin\theta) \, \frac{-w_1|\cos\theta| + w_2|\sin\theta|}{\left(w_1|\sin\theta| + w_2|\cos\theta|\right)^4} \frac{d}{d\theta} \; + \; \frac{1}{3\left(w_1|\sin\theta| + w_2|\cos\theta|\right)^3} \frac{d^2}{d\theta^2}. \tag{42}$$

As an independent numerical verification of this formula, we performed the following experiment. We fixed a particular function $f : S^1 \to \mathbb{R}$ (namely a certain trigonometric polynomial). We drew n points uniformly i.i.d from the circle. We computed the empirical point-cloud Laplacian applied to f, using Eq. (32) and evaluating $\mathcal{L}_{n,\mathcal{B}}f$ along a dense regular grid. For comparison, we evaluated the Laplacian-like operator applied to f, using Eq. (42) and evaluating $\Delta_{\mathcal{M},\mathcal{B}}f$ along the same grid. Figure 2 shows a convincing match: as the sample size n grows, the empirical and theoretical plots match up increasingly well. Appendix F presents numerical results on the eigenfunctions of (42).

Remark 6 The coefficient of $\frac{d}{d\theta}$ in Eq. (42) is discontinuous at $\theta = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$. Thus, this example confirms the second sentence of Proposition 9, item 2.

4 Proof of Theorem 6

To improve readability, we split the proof of Theorem 6 into several steps. First, we reduce to the population limit ($n = \infty$), replacing sums by integrals, via concentration of measure. The integrals are then parameterized by geodesic normal coordinates on the manifold. Both of these are standard steps in the analysis of empirical Laplacians based on the ℓ_2-norm. We then replace the Gaussian kernel by the 0/1 kernel, so all considerations are local. The domain of integration becomes the intersection of the manifold $\mathcal{M}$ with the convex body $\sigma\mathcal{B}$ (for $\sigma \to 0$). This being a potentially unwieldy domain, we substitute the Taylor series expansion of the exponential map to replace the manifold $\mathcal{M}$ by first- and second-order approximations around $\mathbf{p}$. For the term involving the second fundamental form, we switch to spherical coordinates. Then we study the radial domain of integration. We consider this step (Section 4.7) to be the proof's most technical. Following this analysis, the tilt function emerges (Proposition 16), and dominated convergence is used to finish.
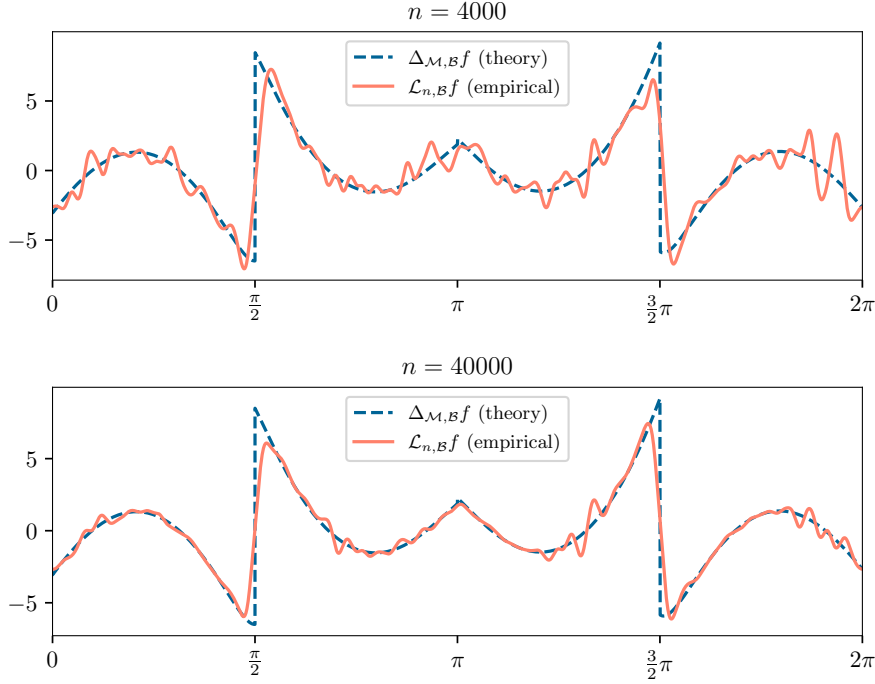
Fig. 2 *Empirical vs. theoretical weighted ℓ_1 Laplacian on the circle ($w_1 = 1, w_2 = 1.5$) applied to the function $f(\theta) = \sin(\theta) + \cos(2\theta) + \cos(5\theta)$. For the empirical Laplacian, the samples were drawn from the uniform distribution on the unit circle. (top panel) $n = 4,000$ samples; (bottom panel) $n = 40,000$ samples. Here $\mathcal{L}_{n,\mathcal{B}}f$ is scaled by $\mathrm{vol}(S^1)/(\Gamma(\frac{d+4}{2})\sigma_n^{d+2})$.*

4.1 Step 1: reduce to the population limit ($n \to \infty$)

This is a standard application of concentration of measure. Let

$$S_n^{(i)} := \mathrm{vol}(\mathcal{M})\frac{1}{nc_n}K_{\sigma_n}(\|\mathbf{x}_i - \mathbf{p}\|_{\mathcal{B}})(f(\mathbf{x}_i) - f(\mathbf{p})),$$

$$S_n := (\mathrm{vol}(\mathcal{M})/c_n)\mathcal{L}_{n,\mathcal{B}}f(\mathbf{p}) = \sum_{i=1}^{n} S_n^{(i)}.$$

For a fixed sample size n, the values $S_n^{(1)}, \ldots, S_n^{(1)}$ are i.i.d. random variables. By the continuity of f and compactness of $\mathcal{M}$, there is a constant $C_0 > 0$ such that $|f(\mathbf{x})| \leq c_0$ for all $\mathbf{x} \in \mathcal{M}$. Recalling $K_{\sigma_n} \leq 1$, it follows that

$$|S_n^{(i)}| \leq (2C_0 \,\mathrm{vol}(\mathcal{M}))/(nc_n). \tag{43}$$

Let $\epsilon > 0$. By inequality (43) together with Hoeffding's inequality,

$$\mathbb{P}\left(|S_n - \mathbb{E}[S_n]| \geq \frac{\epsilon}{2}\right) \leq 2\exp\left(\frac{-\epsilon^2 nc_n^2}{32C_0^2 \,\mathrm{vol}(\mathcal{M})^2}\right) = 2\exp\left(\frac{-\epsilon^2\Gamma(\frac{d+4}{2})n^{\frac{\alpha}{2d+4+\alpha}}}{32C_0^2 \,\mathrm{vol}(\mathcal{M})^2}\right) \tag{44}$$

where we substituted $c_n = \Gamma(\frac{d+4}{2})\sigma_n^{d+2}$ and $\sigma_n = n^{-1/(2d+4+\alpha)}$. Here

$$\mathbb{E}[S_n] \;=\; \frac{1}{\Gamma(\frac{d+4}{2})\sigma_n^{d+2}} \int_{\mathbf{x}\in\mathcal{M}} K_{\sigma_n}(\|\mathbf{x}-\mathbf{p}\|_{\mathcal{B}})\,(f(\mathbf{x})-f(\mathbf{p}))\,d\mu(\mathbf{x}),$$

where $d\mu$ is the Riemannian volume density on $\mathcal{M}$. Since $\sigma_n \to 0$ as $n \to \infty$, **assuming we proved**

$$\lim_{\sigma\to 0} \frac{1}{\Gamma(\frac{d+4}{2})\sigma^{d+2}} \int_{\mathbf{x}\in\mathcal{M}} K_{\sigma}(\|\mathbf{x}-\mathbf{p}\|_{\mathcal{B}})\,(f(\mathbf{x})-f(\mathbf{p}))\,d\mu(\mathbf{x}) \;=\; \Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p}),$$

$$(45)$$

then it would follow there exists $n_0 = n_0(\epsilon)$ such that for all $n > n_0$,

$$\big|\mathbb{E}[S_n] - \Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p})\big| \leq \frac{\epsilon}{2}. \tag{46}$$

Combining inequalities (44) and (46) gives, for all $n > n_0$,

$$\mathbb{P}\big(\big|S_n - \Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p})f(\mathbf{p})\big| \geq \epsilon\big) \leq \mathbb{P}\big(\big|S_n - \mathbb{E}[S_n]\big| \geq \tfrac{\epsilon}{2}\big) \tag{47}$$

$$\leq 2\exp\left(\frac{-\epsilon^2 \Gamma(\frac{d+4}{2})n^{\frac{\alpha}{2d+4+\alpha}}}{32C_0^2\,\mathrm{vol}(\mathcal{M})^2}\right). \tag{48}$$

By $\alpha > 0$, the RHS of (47) converges to 0 as $n \to \infty$. Since ϵ was arbitrary, this shows that S_n converges to $\Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p})$ in probability. Dividing by $\mathrm{vol}(\mathcal{M})$ gives $\mathcal{L}_{n,\mathcal{B}}f(\mathbf{p}) \xrightarrow{p} (1/\mathrm{vol}(\mathcal{M}))\Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p})$. We can upgrade this to almost sure convergence, simply by noting that the series

$$\sum_{n=1}^{\infty} 2\exp\left(\frac{-\epsilon^2 \Gamma(\frac{d+4}{2})n^{\frac{\alpha}{2d+4+\alpha}}}{32C_0^2\,\mathrm{vol}(\mathcal{M})^2}\right)$$

converges and citing the Borel-Cantelli lemma. It remains to actually prove (45).

4.2 Step 2: reduce to the indicator function kernel

In this step, we replace the Gaussian kernel K_{σ_n} by the indicator function kernel $\mathbb{1}_{\sigma_n}$, defined by

$$\mathbb{1}_{\sigma_n} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \text{ where } \mathbb{1}_{\sigma_n}(t) := 1 \text{ if } t \in [0,\sigma_n] \text{ and } \mathbb{1}_{\sigma_n}(t) := 0 \text{ if } t > \sigma_n.$$

Precisely, we show

$$\lim_{\sigma\to 0} \frac{1}{\sigma^{d+2}} \int_{\mathbf{x}\in\mathcal{M}} \mathbb{1}_{\sigma}(\|\mathbf{x}-\mathbf{p}\|_{\mathcal{B}})\,(f(\mathbf{x})-f(\mathbf{p}))\,d\mu(\mathbf{x}) \;=\; \Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p}) \tag{49}$$

implies the required formula (45), and thereby we will reduce to proving (49).

To achieve this reduction, we now assume (49). Write the Gaussian kernel K_σ as a sum of indicator functions:

$$K_\sigma(t) = \int_{s=0}^{\infty} \kappa_\sigma(s)\mathbb{1}_s(t)\,ds = \int_{s=t}^{\infty} \kappa_\sigma(s)\,ds, \tag{50}$$

where $\kappa_\sigma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is given by $\kappa_\sigma(s) := (2s/\sigma^2)\exp(-s^2/\sigma^2)$. Then,

$$\int_{\mathbf{x} \in \mathcal{M}} K_\sigma(\|\mathbf{x} - \mathbf{p}\|_\mathcal{B})(f(\mathbf{x}) - f(\mathbf{p}))d\mu(\mathbf{x})$$

$$= \int_{\mathbf{x} \in \mathcal{M}} \left(\int_{s=\|\mathbf{x}-\mathbf{p}\|_\mathcal{B}}^{\infty} \kappa_\sigma(s)ds \right) (f(\mathbf{x}) - f(\mathbf{p}))d\mu(\mathbf{x}) \qquad \text{[substituting (50)]}$$

$$= \int_{s=0}^{\infty} \kappa_\sigma(s) \left(\int_{\mathbf{x} \in \mathcal{M}:\|\mathbf{x}-\mathbf{p}\|_\mathcal{B} \leq s} f(\mathbf{x}) - f(\mathbf{p})d\mu(\mathbf{x}) \right) ds \qquad \text{[Fubini's theorem]}.$$
$$\tag{51}$$

Define

$$e(s) := \left(\int_{\mathbf{x} \in \mathcal{M}:\|\mathbf{x}-\mathbf{p}\|_\mathcal{B} \leq s} f(\mathbf{x}) - f(\mathbf{p})d\mu(\mathbf{x}) \right) - s^{d+2}\Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p}).$$

In light of (49), we have

$$e(s) = o(s^{d+2}) \quad \text{as } s \to 0. \tag{52}$$

Fix $\epsilon > 0$. By (52), we can fix $\delta > 0$ such that

$$0 \leq s \leq \delta \Rightarrow |e(s)| \leq \epsilon s^{d+2}. \tag{53}$$

Returning to Eq. (51), we may change the upper limit of integration with the following control on the approximation error:

$$\int_{s=0}^{\infty} \kappa_\sigma(s) \left(\int_{\mathbf{x} \in \mathcal{M}:\|\mathbf{x}-\mathbf{p}\|_\mathcal{B} \leq s} f(\mathbf{x}) - f(\mathbf{p})d\mu(\mathbf{x}) \right) ds \tag{54}$$

$$= \int_{s=0}^{\delta} \kappa_\sigma(s) \left(\int_{\mathbf{x} \in \mathcal{M}:\|\mathbf{x}-\mathbf{p}\|_\mathcal{B} \leq s} f(\mathbf{x}) - f(\mathbf{p})d\mu(\mathbf{x}) \right) ds \ + \ \exp(-\delta^2/\sigma^2)\text{poly}(\sigma).$$

To justify Eq. (54) holds, we note the parenthesized integral has absolute value bounded by $2c_0$ for all $s \in [0, \infty]$, by the compactness of $\mathcal{M}$. Thus, a tail bound for the Gaussian kernel implies (54) (set $k = 0$ in Eq. (103) in Appendix E). Now the main term in Eq. (54) is

$$\int_{s=0}^{\delta} \kappa_\sigma(s)\left(s^{d+2}\Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p}) \ + \ e(s)\right)ds.$$

From (53), this is bounded above by

$$\int_{s=0}^{\delta} \kappa_\sigma(s)s^{d+2}\left(\Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p}) + \epsilon\right)ds, \tag{55}$$

and below by

$$\int_{s=0}^{\delta} \kappa_\sigma(s)s^{d+2}\left(\Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p}) - \epsilon\right)ds. \tag{56}$$

By additional bounds for the Gaussian (Appendix E), the upper and lower bounds (55) and (56) are equal to

$$\int_{s=0}^{\infty} \kappa_{\sigma}(s)s^{d+2}\left(\Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p}) \pm \epsilon\right)ds \,+\, \exp(-\delta^2/\sigma^2)\mathrm{poly}(\sigma).$$

But, the main term is

$$\sigma^{d+2}\Gamma(\tfrac{d+4}{2})\left(\Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p}) \pm \epsilon\right),$$

by the formula for half the k-th absolute moment of κ_{σ} (Eq. (105), Appendix E). Using $\lim_{\sigma\to 0}\exp(-\delta^2/\sigma^2)\mathrm{poly}(\sigma) = 0$, and the fact that ϵ is arbitrary, we achieve what we wanted:

$$\int_{\mathbf{x}\in\mathcal{M}} \lim_{\sigma\to 0} \frac{1}{\sigma^{d+2}}K_{\sigma}(\|\mathbf{x}-\mathbf{p}\|_{\mathcal{B}})(f(\mathbf{x}) - f(\mathbf{p}))d\mu(\mathbf{x}) \;=\; \Gamma(\tfrac{d+4}{2})\Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p}).$$

To sum up, Eq. (49) implies Eq. (45). It remains to prove Eq. (49).

4.3 Step 3: use geodesic normal coordinates and Taylor expand

In this step, we express the integral in the LHS of (49) in normal coordinates,

$$\int_{\mathbf{x}\in\mathcal{M}:\|\mathbf{x}-\mathbf{p}\|_{\mathcal{B}}\leq\sigma} f(\mathbf{x}) - f(\mathbf{p})d\mu(\mathbf{x}). \tag{57}$$

We parameterize it using the exponential map (Section 3.1),

$$\exp_{\mathbf{p}} : U \xrightarrow{\sim} V,$$

where $U \subseteq T_{\mathbf{p}}\mathcal{M}$ and $V \subseteq \mathcal{M}$ are open neighborhoods of 0 and $\mathbf{p}$ respectively. Note that there exists some constant $\sigma_0 > 0$ such that for all $\sigma \leq \sigma_0$ the domain of integration in (57) is contained in V,

$$\{\mathbf{x} \in \mathcal{M} : \|\mathbf{x} - \mathbf{p}\|_{\mathcal{B}} \leq \sigma\} \subseteq V. \tag{58}$$

This follows from the fact that $\mathcal{M}$ is an embedded submanifold of $\mathbb{R}^D$, hence V can be written as an open set of $\mathbb{R}^D$ intersected with $\mathcal{M}$, and the fact $\|\cdot\|_{\mathcal{B}}$ is equivalent to the Euclidean norm on $\mathbb{R}^D$ and so induces the same open sets. Therefore, by a change of variables, for each $\sigma \leq \sigma_0$, the integral (57) equals

$$\int_{\mathbf{s}\in U:\|\exp_{\mathbf{p}}(\mathbf{s})-\mathbf{p}\|_{\mathcal{B}}\leq\sigma} \left(\widetilde{f}(\mathbf{s}) - \widetilde{f}(0)\right)\left|\det D\exp_{\mathbf{p}}(\mathbf{s})\right|d\mathbf{s}, \tag{59}$$

where $\mathbf{s} = (s_1,\ldots,s_d)^{\top}$ denotes coordinates for $T_{\mathbf{p}}\mathcal{M}$ with respect to an orthonormal basis and $d\mathbf{s}$ denotes the Lebesgue measure on $(T_{\mathbf{p}}\mathcal{M}, \langle\cdot,\cdot\rangle_{\mathbf{p}})$.

Our goal is to approximate the integral (59) up to order σ^{d+2}. To this end, we will consider three Taylor expansions:

$$\widetilde{f}(\mathbf{s}) \,=\, \widetilde{f}(0) + \mathrm{grad}\widetilde{f}(0)^{\top}\mathbf{s} + \tfrac{1}{2}\mathbf{s}^{\top}\mathrm{hess}\widetilde{f}(0)\mathbf{s} + O(\|\mathbf{s}\|_2^3), \tag{60}$$

$$\det D\exp_{\mathbf{p}}(\mathbf{s}) \,=\, 1 - \tfrac{1}{6}\mathbf{s}^{\top}\mathrm{Ric}(\mathbf{p})\mathbf{s} + O(\|\mathbf{s}\|_2^3) \,=\, 1 + O(\|\mathbf{s}\|_2^2), \tag{61}$$

$$\exp_{\mathbf{p}}(\mathbf{s}) \,=\, \mathbf{p} + L_{\mathbf{p}}(\mathbf{s}) + \tfrac{1}{2}Q_{\mathbf{p}}(\mathbf{s}) + O(\|\mathbf{s}\|_2^3). \tag{62}$$

Here $\mathrm{Ric}(\mathbf{p}) \in \mathbb{R}^{d \times d}$ stands for the Ricci curvature of $\mathcal{M}$ at $\mathbf{p}$ (see [32, Ch. 7]). Also, see Section 3.1 for discussion on $L_{\mathbf{p}}$ and $Q_{\mathbf{p}}$.

Substituting equations (60) and (61) into the integral (59) leads to

$$\int_{\mathbf{s}\in U:\|\exp_{\mathbf{p}}(\mathbf{s})-\mathbf{p}\|_{\mathcal{B}}\leq\sigma} \mathrm{grad}\widetilde{f}(0)^{\top}\mathbf{s} \,+\, \tfrac{1}{2}\mathbf{s}^{\top}\mathrm{hess}\widetilde{f}(0)\mathbf{s} \,+\, O(\|\mathbf{s}\|_2^3) \; d\mathbf{s}. \tag{63}$$

4.4 Step 4: approximate the domain of integration

In this step, we approximate the domain of integration in (63) using the Taylor expansion of the exponential map (Definition 3). Then we assess the quality of our approximations (Proposition 10).

Definition 3 For each $\sigma > 0$, we define three subsets of $T_{\mathbf{p}}\mathcal{M}$ as follows.

$$\mathrm{Exact}(\sigma) := \{\mathbf{s} \in U : \big\|\exp_{\mathbf{p}}(\mathbf{s}) - \mathbf{p}\big\|_{\mathcal{B}} \leq \sigma\},$$

$$\mathrm{Approx}^{(1)}(\sigma) := \{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}} \leq \sigma\},$$

$$\mathrm{Approx}^{(2)}(\sigma) := \{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \big\|L_{\mathbf{p}}(\mathbf{s}) + \tfrac{1}{2}Q_{\mathbf{p}}(\mathbf{s})\big\|_{\mathcal{B}} \leq \sigma\}.$$

The set $\mathrm{Exact}(\sigma) \subseteq T_{\mathbf{p}}\mathcal{M}$ is the exact domain of integration, parameterized on the tangent space. The sets $\mathrm{Approx}^{(1)}(\sigma)$ and $\mathrm{Approx}^{(2)}(\sigma)$ are approximations to $\mathrm{Exact}(\sigma)$, where the manifold around $\mathbf{p}$ is approximated to first and second order, respectively.

Proposition 10   *i. There exist constants $c_1, \sigma_1 > 0$ such that for all $\sigma \leq \sigma_1$,*

$$\mathrm{Exact}(\sigma) \subseteq \{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \|\mathbf{s}\|_2 \leq c_1\sigma\}. \tag{64}$$

ii. There exist constants $c_2, c_3, \sigma_2 > 0$ such that for all $\sigma \leq \sigma_2$,

$$(1 - c_2\sigma)\,\mathrm{Approx}^{(1)}(\sigma) \subseteq \mathrm{Exact}(\sigma) \subseteq (1 + c_3\sigma)\,\mathrm{Approx}^{(1)}(\sigma). \tag{65}$$

iii. There exist constants $c_4, c_5, \sigma_3 > 0$ such that for all $\sigma \leq \sigma_3$,

$$(1 - c_5\sigma^2)\,\mathrm{Approx}^{(2)}(\sigma) \subseteq \mathrm{Exact}(\sigma) \subseteq (1 + c_4\sigma^2)\,\mathrm{Approx}^{(2)}(\sigma). \tag{66}$$

Proof <u>part i</u>. First, note that for any $\epsilon > 0$, we can shrink U and V in (58) so as to guarantee that for all $\mathbf{s} \in U$ we have

$$\left\|\exp_{\mathbf{p}}(\mathbf{s}) - \mathbf{p} - L_{\mathbf{p}}(\mathbf{s})\right\|_{\mathcal{B}} \le \epsilon\|\mathbf{s}\|_2. \tag{67}$$

Let $\mathbf{s} \in \text{Exact}(\sigma)$. The result follows from:

$$
\begin{aligned}
\sigma &\ge \left\|\exp_{\mathbf{p}}(\mathbf{s}) - \mathbf{p}\right\|_{\mathcal{B}} && \text{[definition of Exact}(\sigma)\text{]} \\
&= \left\|L_{\mathbf{p}}(\mathbf{s}) - (L_{\mathbf{p}}(\mathbf{s}) - \exp_{\mathbf{p}}(\mathbf{s}) + \mathbf{p})\right\|_{B} \\
&\ge \|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}} - \|L_{\mathbf{p}}(\mathbf{s}) - \exp_{\mathbf{p}}(\mathbf{s}) + \mathbf{p}\|_{\mathcal{B}} && \text{[reverse triangle inequality]} \\
&\ge \|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}} - \epsilon\|\mathbf{s}\|_2 && \text{[using (67)]} \\
&\ge c\|L_{\mathbf{p}}(\mathbf{s})\|_2 - \epsilon\|\mathbf{s}\|_2 && \text{[norm equivalence, see (10)]} \\
&= (c - \epsilon)\|\mathbf{s}\|_2. && [L_{\mathbf{p}} \text{ is an isometry}]
\end{aligned}
$$

<u>part ii</u>. For the right inclusion, assume $\mathbf{s} \in \text{Exact}(\sigma)$. Then,

$$
\begin{aligned}
\sigma &\ge \left\|\exp_{\mathbf{p}}(\mathbf{s}) - \mathbf{p}\right\|_{\mathcal{B}} && \text{[definition of Exact}(\sigma)\text{]} \\
&= \left\|L_{\mathbf{p}}(\mathbf{s}) + O(\|\mathbf{s}\|^2)\right\|_{\mathcal{B}} && \text{[from (6)]} \\
&= \left\|L_{\mathbf{p}}(\mathbf{s}) + O(\sigma^2)\right\|_{\mathcal{B}} && \text{[shown in part i]} \\
&= \|L_{\mathbf{p}}(\mathbf{s})\|_{B} + O(\sigma^2). && \text{[triangle inequality]}
\end{aligned}
$$

Take c_3 to be the implicit constant inside the $O(\sigma^2)$ term, it follows that $\|L_{\mathbf{p}}(\mathbf{s})\|_B \le \sigma + c_3\sigma^2$ and therefore $\mathbf{s} \in (1 + c_3\sigma)\,\text{Approx}^{(1)}(\sigma)$.

For the left inclusion, let $\mathbf{s} \in (1 - c_2\sigma)\,\text{Approx}^{(1)}(\sigma)$. It follows by definition that $\|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}} \le \sigma - c_2\sigma^2$. We have just shown that

$$\left\|\exp_{\mathbf{p}}(\mathbf{s}) - \mathbf{p}\right\|_{\mathcal{B}} = \|L_{\mathbf{p}}(\mathbf{s})\|_B + O(\sigma^2)$$

Therefore,

$$\left\|\exp_{\mathbf{p}}(\mathbf{s}) - \mathbf{p}\right\|_{\mathcal{B}} \le \sigma - c_2\sigma^2 + O(\sigma^2)$$

Picking c_2 to be the implicit constant inside the $O(\sigma^2)$ term guarantees that $\left\|\exp_{\mathbf{p}}(\mathbf{s}) - \mathbf{p}\right\|_{\mathcal{B}} \le \sigma$ and hence that $\mathbf{s} \in \text{Exact}(\sigma)$.

<u>part iii</u>. From Eq. (6) we have that

$$\left\|\exp_{\mathbf{p}}(\mathbf{s}) - \mathbf{p}\right\|_{\mathcal{B}} = \left\|L_{\mathbf{p}}(\mathbf{s}) + \tfrac{1}{2}Q_{\mathbf{p}}(\mathbf{s}) + O(\|\mathbf{s}\|^3)\right\|_{\mathcal{B}} \tag{68}$$

$$= \left\|L_{\mathbf{p}}(\mathbf{s}) + \tfrac{1}{2}Q_{\mathbf{p}}(\mathbf{s})\right\|_{\mathcal{B}} + O(\|\mathbf{s}\|^3) \tag{69}$$

By part i, $\|\mathbf{s}\|_2 = O(\sigma)$. From (68) and the triangle inequality,

$$\left\|\exp_{\mathbf{p}}(\mathbf{s}) - \mathbf{p}\right\|_{\mathcal{B}} = \left\|L_{\mathbf{p}}(\mathbf{s}) + \tfrac{1}{2}Q_{\mathbf{p}}(\mathbf{s})\right\|_{\mathcal{B}} \le \sigma + c_4'\sigma^3, \tag{70}$$

for some constant $c_4' > 0$ and all sufficiently small σ. We want to find a constant $c_4 > 0$ such that $\mathbf{s}/(1 + c_4\sigma^2) \in \mathrm{Approx}^{(2)}(\sigma)$. To this end, compute

$$
\left\| \frac{1}{1 + c_4\sigma^2} L_{\mathbf{p}}(\mathbf{s}) + \frac{1}{(1 + c_4\sigma^2)^2} \frac{1}{2} Q_{\mathbf{p}}(\mathbf{s}) \right\|_{\mathcal{B}}
$$
$$
\leq \frac{1}{1 + c_4\sigma^2} \left\| L_{\mathbf{p}}(\mathbf{s}) + \tfrac{1}{2} Q_{\mathbf{p}}(\mathbf{s}) \right\|_{\mathcal{B}} + \left(\frac{1}{1 + c_4\sigma^2} - \frac{1}{(1 + c_4\sigma^2)^2} \right) \left\| \tfrac{1}{2} Q_{\mathbf{p}}(\mathbf{s}) \right\|_{\mathcal{B}}
$$
$$
\leq \frac{1}{1 + c_4\sigma^2} \left(\sigma + c_4'\sigma^3 \right) + \left(\frac{1}{1 + c_4\sigma^2} - \frac{1}{(1 + c_4\sigma^2)^2} \right) O(\sigma^2)
$$
$$
= \left(1 - c_4\sigma^2 + O(\sigma^4) \right) \left(\sigma + c_4'\sigma^3 \right) + \left(c_4\sigma^2 + O(\sigma^4) \right) O(\sigma^2)
$$
$$
= \sigma + (c_4' - c_4)\sigma^3 + O(\sigma^4). \tag{71}
$$

Here we used the triangle inequality, the bound (70), part i, and Taylor expansions in σ for $(1 + c_4\sigma^2)^{-1}$ and $(1 + c_4\sigma^2)^{-2}$. Thus, take $c_4 = 2c_4'$. For small enough σ, the RHS of Eq. (71) is at most σ, so $\mathbf{s}/(1 + c_4\sigma^2) \in \mathrm{Approx}^{(2)}(\sigma)$. Now consider the leftmost inclusion in part iii. Assume $\mathbf{s} \in \mathrm{Approx}^{(2)}(\sigma)$. We first prove that $\|\mathbf{s}\|_2 = O(\sigma)$. Indeed,

$$
\sigma \geq \left\| L_{\mathbf{p}}(\mathbf{s}) + \tfrac{1}{2} Q_{\mathbf{p}}(\mathbf{s}) \right\|_{\mathcal{B}} \gtrsim \left\| L_{\mathbf{p}}(\mathbf{s}) + \tfrac{1}{2} Q_{\mathbf{p}}(\mathbf{s}) \right\|_2
$$
$$
= \sqrt{\left\| L_{\mathbf{p}}(\mathbf{s}) \right\|_2^2 + \tfrac{1}{4} \left\| Q_{\mathbf{p}}(\mathbf{s}) \right\|_2^2} \geq \left\| L_{\mathbf{p}}(\mathbf{s}) \right\|_2 = \left\| \mathbf{s} \right\|_2. \tag{72}
$$

The first equality comes from orthogonality between the images of $Q_{\mathbf{p}}$ and $L_{\mathbf{p}}$ (9). Let the implicit constant in (72) be c_5'. Similarly to above, let us set $c_5 = 2c_5'$ and compute (for sufficiently small σ):

$$
\left\| \exp_{\mathbf{p}} \left((1 - c_5\sigma^2)\mathbf{s} \right) - \mathbf{p} \right\|_{\mathcal{B}} \leq \left\| (1 - c_5\sigma^2) L_{\mathbf{p}}(\mathbf{s}) + (1 - c_5\sigma^2)^2 \tfrac{1}{2} Q_{\mathbf{p}}(\mathbf{s}) \right\|_{\mathcal{B}} + c_5'\sigma^3
$$
$$
\leq (1 - c_5\sigma^2) \left\| L_{\mathbf{p}} + \tfrac{1}{2} Q_{\mathbf{p}}(\mathbf{s}) \right\|_{\mathcal{B}} + \left((1 - c_5\sigma^2) - (1 - c_5\sigma^2)^2 \right) \left\| \tfrac{1}{2} Q_{\mathbf{p}}(\mathbf{s}) \right\|_{\mathcal{B}} + c_5'\sigma^3
$$
$$
\leq (1 - c_5\sigma^2)\sigma + O(\sigma^4) + c_5'\sigma^3 = \sigma - c_5'\sigma^3 + O(\sigma^4) \leq \sigma.
$$

We used the Taylor expansion (62), the triangle inequality, the bound (72), and the triangle inequality again. Hence $(1 - c_5\sigma^2)\mathbf{s} \in \mathrm{Exact}(\sigma)$. $\qquad\square$

4.5 Step 5: drop $O(\|\mathbf{s}\|_2^3)$ and obtain the second-order term

In this step we prove that each of the terms in the integral (63) can be approximated up to an additive error of $O(\sigma^{d+3})$ by switching from the exact domain $\mathrm{Exact}(\sigma)$ to the approximate domains $\mathrm{Approx}^{(1)}(\sigma)$ and $\mathrm{Approx}^{(2)}(\sigma)$ defined in Definition 3.

Proposition 11 *The following bounds hold:*

i. $\int_{\mathbf{s} \in \mathrm{Exact}(\sigma)} O(\|\mathbf{s}\|_2^3) d\mathbf{s} = O(\sigma^{d+3})$,

ii. $\int_{\mathbf{s} \in \mathrm{Exact}(\sigma)} \mathbf{s}\mathbf{s}^\top d\mathbf{s} = \int_{\mathbf{s} \in \mathrm{Approx}^{(1)}(\sigma)} \mathbf{s}\mathbf{s}^\top d\mathbf{s} + O(\sigma^{d+3})$,

iii. $\int_{\mathbf{s}\in\text{Exact}(\sigma)} \mathbf{s}d\mathbf{s} = \int_{\mathbf{s}\in\text{Approx}^{(2)}(\sigma)} \mathbf{s}d\mathbf{s} + O(\sigma^{d+3})$.

Proof Let $\ominus$ denote the symmetric difference of sets.
<u>part i.</u> Let $\sigma \leq \sigma_1$. By Proposition 10, part i, we have

$$\int_{\mathbf{s}\in\text{Exact}(\sigma)} O(\|\mathbf{s}\|_2^3)\, d\mathbf{s} \lesssim \int_{\|\mathbf{s}\|_2 \leq c_1\sigma} \|\mathbf{s}\|_2^3$$

$$\leq (c_1\sigma)^3 \text{vol}\{\mathbf{s}\in\mathbb{R}^d : \|s\|_2 \leq c_1\sigma\} = O(\sigma^{d+3}).$$

<u>part ii.</u> Let $\sigma \leq \sigma_2$. By Proposition 10, part ii, we see

$$\text{Exact}(\sigma) \ominus \text{Approx}^{(1)}(\sigma) \subseteq (1 + c_3\sigma)\,\text{Approx}^{(1)}(\sigma) \setminus (1 - c_2\sigma)\,\text{Approx}^{(1)}(\sigma).$$

Then we have

$$\left\| \int_{\mathbf{s}\in\text{Exact}(\sigma)} \mathbf{s}\mathbf{s}^\top d\mathbf{s} - \int_{\mathbf{s}\in\text{Approx}^{(1)}(\sigma)} \mathbf{s}\mathbf{s}^\top d\mathbf{s} \right\|_F$$

$$= \left\| \int_{\mathbf{s}\in\text{Exact}(\sigma)\setminus\text{Approx}^{(1)}(\sigma)} \mathbf{s}\mathbf{s}^\top d\mathbf{s} \right\|_F$$

$$\leq \int_{\mathbf{s}\in\text{Exact}(\sigma)\setminus\text{Approx}^{(1)}(\sigma)} \left\| \mathbf{s}\mathbf{s}^\top \right\|_F d\mathbf{s}$$

$$\leq \int_{\mathbf{s}\in\text{Exact}(\sigma)\ominus\text{Approx}^{(1)}(\sigma)} \left\| \mathbf{s}\mathbf{s}^\top \right\|_F d\mathbf{s}$$

$$\leq \int_{\mathbf{s}\in(1+c_3\sigma)\,\text{Approx}^{(1)}(\sigma)\setminus(1-c_2\sigma)\,\text{Approx}^{(1)}(\sigma)} \|\mathbf{s}\|_2^2 d\mathbf{s}$$

$$= \int_{\mathbf{s}\in(\sigma+c_3\sigma^2)\,\text{Approx}^{(1)}(1)\setminus(\sigma-c_2\sigma^2)\,\text{Approx}^{(1)}(1)} \|\mathbf{s}\|_2^2 d\mathbf{s}$$

$$= \left((\sigma + c_3\sigma^2)^{d+2} - (\sigma - c_2\sigma^2)^{d+2}\right) \int_{\|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}}\leq 1} \|\mathbf{s}\|_2^2 d\mathbf{s}.$$

This last quantity is $O(\sigma^{d+3})$, because $\mathcal{B}$ is bounded and $L_{\mathbf{p}}$ is an isometry.

<u>part iii.</u> Let $\sigma \leq \sigma_3$. By Proposition 10, part iii,

$$\text{Exact}(\sigma) \ominus \text{Approx}^{(2)}(\sigma)$$

$$\subseteq (1 + c_4\sigma^2)\,\text{Approx}^{(2)}(\sigma) \setminus (1 - c_5\sigma^2)\,\text{Approx}^{(2)}(\sigma).$$

Then,

$$\left\| \int_{\mathbf{s}\in\text{Exact}(\sigma)} \mathbf{s}d\mathbf{s} - \int_{\mathbf{s}\in\text{Approx}^{(2)}(\sigma)} \mathbf{s}d\mathbf{s} \right\|_2 = \left\| \int_{\mathbf{s}\in\text{Exact}(\sigma)\setminus\text{Approx}^{(2)}(\sigma)} \mathbf{s}d\mathbf{s} \right\|_2$$

$$\leq \int_{\mathbf{s}\in\text{Exact}(\sigma)\setminus\text{Approx}^{(2)}(\sigma)} \|\mathbf{s}\|_2\, d\mathbf{s} \leq \int_{\mathbf{s}\in\text{Exact}(\sigma)\ominus\text{Approx}^{(2)}(\sigma)} \|\mathbf{s}\|_2 d\mathbf{s}$$

$$\leq \int_{\mathbf{s}\in(1+c_4\sigma^2)\,\text{Approx}^{(2)}(\sigma)\setminus(1-c_5\sigma^2)\,\text{Approx}^{(2)}(\sigma)} \|\mathbf{s}\|_2 d\mathbf{s}.$$

The upper bound equals

$$\left((1 + c_4\sigma^2)^{d+1} - (1 - c_5\sigma^2)^{d+1}\right) \int_{\mathbf{s}\in\mathrm{Approx}^{(2)}(\sigma)} \|\mathbf{s}\|_2 d\mathbf{s}$$

$$= O(\sigma^2) \int_{\mathbf{s}'\in\mathrm{Approx}^{(2)}(1)} \|\sigma\mathbf{s}'\|_2 \sigma^d d\mathbf{s}' = O(\sigma^{d+3}) \int_{\mathbf{s}'\in\mathrm{Approx}^{(2)}(1)} \|\mathbf{s}'\|_2 d\mathbf{s}'$$

The last quantity is $O(\sigma^{d+3})$, because $\mathbf{s} \in \mathrm{Approx}^{(2)}(\sigma)$ implies $\|\mathbf{s}\|_2 = O(\sigma)$, as shown in the argument for Proposition 10, part iii. $\qquad\square$

Now, plug Proposition 11 into the integral (63):

$$\int_{\mathbf{s}\in\mathrm{Exact}(\sigma)} \mathrm{grad}\widetilde{f}(0)^\top\mathbf{s} + \tfrac{1}{2}\mathbf{s}^\top\mathrm{hess}\widetilde{f}(0)\mathbf{s} + O(\|\mathbf{s}\|_2^3) \; d\mathbf{s}$$

$$= \left\langle \mathrm{grad}\widetilde{f}(0), \int_{\mathbf{s}\in\mathrm{Approx}^{(2)}(\sigma)} \mathbf{s}d\mathbf{s} \right\rangle$$

$$+ \left\langle \mathrm{hess}\widetilde{f}(0), \tfrac{1}{2}\int_{\mathbf{s}\in\mathrm{Approx}^{(1)}(\sigma)} \mathbf{s}\mathbf{s}^\top d\mathbf{s} \right\rangle_F + O(\sigma^{d+3}),$$

where linearity of $L_{\mathbf{p}}$ gives

$$\int_{\mathbf{s}\in\mathrm{Approx}^{(1)}(\sigma)} \mathbf{s}\mathbf{s}^\top d\mathbf{s} = \sigma^{d+2}\int_{\mathbf{s}:\|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}}\leq 1} \mathbf{s}\mathbf{s}^\top d\mathbf{s}.$$

Thus, (63) divided by σ^{d+2} tends to $\Delta_{\mathcal{M},\mathcal{B}}f(\mathbf{p})$ as $\sigma \to 0$, as desired, **provided we can show**

$$\lim_{\sigma\to 0} \frac{1}{\sigma^{d+2}} \int_{\mathbf{s}\in\mathrm{Approx}^{(2)}(\sigma)} \mathbf{s}d\mathbf{s} = \int_{\|\widehat{\mathbf{s}}\|_2=1} \widehat{\mathbf{s}}\,\|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}^{-d}\mathrm{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}}(\widehat{\mathbf{s}})\,d\widehat{\mathbf{s}}. \qquad (73)$$

4.6 Step 6: use spherical coordinates

It remains to estimate

$$\int_{\mathbf{s}\in\mathrm{Approx}^{(2)}(\sigma)} \mathbf{s}d\mathbf{s}. \qquad (74)$$

First, we provide intuition for why this integral should scale like σ^{d+2}. Let $\mathbb{S}^{d-1} \subseteq \mathbb{R}^d \cong T_{\mathbf{p}}\mathcal{M}$ denote the ℓ_2-unit sphere, with density $d\widehat{\mathbf{s}}$, where $\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}$. Let $r \in \mathbb{R}_{\geq 0}$ be a radial variable with density dr. Consider the integral (74) in these spherical coordinates. Substituting $\mathbf{s} = r\widehat{\mathbf{s}}$ and $d\mathbf{s} = r^{d-1}drd\widehat{\mathbf{s}}$,

$$\int_{\mathbf{s}:\|L_{\mathbf{p}}(\mathbf{s})+\frac{1}{2}Q_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}}\leq\sigma} \mathbf{s}d\mathbf{s} = \int_{\widehat{\mathbf{s}}\in\mathbb{S}^{d-1}} \widehat{\mathbf{s}}\int_{r\in\mathrm{RadialDomain}(\widehat{\mathbf{s}},\sigma)} r^d drd\widehat{\mathbf{s}}, \qquad (75)$$

where we define

$$\mathrm{RadialDomain}(\widehat{\mathbf{s}},\sigma) := \left\{r \geq 0 : \left\|rL_{\mathbf{p}}(\widehat{\mathbf{s}}) + \tfrac{r^2}{2}Q_{\mathbf{p}}(\widehat{\mathbf{s}})\right\|_{\mathcal{B}} \leq \sigma\right\}. \qquad (76)$$

This is the second-order approximation set $\text{Approx}^{(2)}(\sigma)$ intersected with the ray in the direction of $\widehat{\mathbf{s}}$. Note that by definition,

$$\text{RadialDomain}(\widehat{\mathbf{s}}, \sigma) = \{r \geq 0 : r\widehat{\mathbf{s}} \in \text{Approx}^{(2)}(\sigma)\}. \tag{77}$$

Compare this domain of integration against the domain for $-\widehat{\mathbf{s}}$:

$$\text{RadialDomain}(-\widehat{\mathbf{s}}, \sigma) = \left\{ r \geq 0 : \left\| rL_{\mathbf{p}}(-\widehat{\mathbf{s}}) + \frac{r^2}{2} Q_{\mathbf{p}}(-\widehat{\mathbf{s}}) \right\|_{\mathcal{B}} \leq \sigma \right\}$$

$$= \left\{ r \geq 0 : \left\| rL_{\mathbf{p}}(\widehat{\mathbf{s}}) - \frac{r^2}{2} Q_{\mathbf{p}}(\widehat{\mathbf{s}}) \right\|_{\mathcal{B}} \leq \sigma \right\}. \tag{78}$$

Speaking roughly, the condition determining membership in (76) differs from that in (78) at the $O(r^2)$ term; the conditions would be the same without the $Q_{\mathbf{p}}$ term. On the other hand, the integrand in (75) is odd (that is, it flips sign upon inversion in the origin). Therefore, we should expect "near cancellation" between the inner integrals:

$$\widehat{\mathbf{s}} \int_{r \in \text{RadialDomain}(\widehat{\mathbf{s}}, \sigma)} r^d dr \;-\; \widehat{\mathbf{s}} \int_{r \in \text{RadialDomain}(-\widehat{\mathbf{s}}, \sigma)} r^d dr. \tag{79}$$

Supposing $r = O(\sigma)$ for r in each radial domain (justified in Lemma 12), then each of the two terms in (79) is $O(\sigma^{d+1})$. Thus after "near, but not complete, cancellation" the sum (79) is expected to be $O(\sigma^{d+2})$. Then, integrating over the unit sphere gives $O(\sigma^{d+2})$ in (75). This informal discussion explains why we expect the integral (74) to be $O(\sigma^{d+2})$, the mechanism being cancellation due to an approximate equality between $\text{RadialDomain}(\widehat{\mathbf{s}}, \sigma)$ and $\text{RadialDomain}(-\widehat{\mathbf{s}}, \sigma)$.

We shall now make this claim rigorous. Beyond proving that the integral (75) is $O(\sigma^{d+2})$, we will prove that dividing (75) by σ^{d+2} produces a well-defined limit as $\sigma \to 0$, namely the RHS of (73).

Remark 7 Steps 7-8 below are complicated by the fact that we do not assume smoothness of the norm $\| \cdot \|_{\mathcal{B}}$. As a consequence, *a priori* we cannot Taylor-expand the boundary points of the radial domain in the variable σ.

Lemma 12 *There exist constants $c_6, \sigma_4 > 0$ such that for all $\sigma \leq \sigma_4$ and all $\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}$, we have*

$$\text{RadialDomain}(\widehat{\mathbf{s}}, \sigma) \subseteq c_6[0, \sigma].$$

Proof Set $c_6 := 2c_1$ and $\sigma_4 := \min\left(\sigma_1, \sigma_3, \frac{1}{\sqrt{2c_5}}\right)$. For $\sigma \leq \sigma_4$, we have

$$\text{Approx}^{(2)}(\sigma) \subseteq \frac{1}{1 - c_5\sigma^2} \text{Exact}(\sigma) \qquad\qquad \text{[by (66)]}$$

$$\subseteq \frac{c_1}{1 - c_5\sigma^2} \{\mathbf{s} \in \mathbb{R}^d : \|\mathbf{s}\|_2 \leq \sigma\}. \qquad \text{[by (64)]}$$

Note that $\sigma \leq \frac{1}{\sqrt{2c_5}}$ implies $\frac{c_1}{1 - c_5\sigma^2} \leq 2c_1 = c_6$. Therefore,

$$\text{Approx}^{(2)}(\sigma) \subseteq c_6\{\mathbf{s} \in \mathbb{R}^d : \|\mathbf{s}\|_2 \leq \sigma\}.$$

By (77) it follows that $\text{RadialDomain}(\widehat{\mathbf{s}}, \sigma) \subseteq c_6[0, \sigma]$ for all $\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}$. $\qquad\square$

4.7 Step 7: study the boundary of RadialDomain$(\widehat{\mathbf{s}}, \sigma)$

In this step, we show that for small enough σ, the set RadialDomain$(\widehat{\mathbf{s}}, \sigma)$ is a single closed interval in $\mathbb{R}_{\geq 0}$. Then, we prove its nonzero boundary point is a continuous function of $(\widehat{\mathbf{s}}, \sigma)$ and we bound this to second-order in σ.

Let $G : \mathbb{S}^{d-1} \times \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}_{\geq 0};\ (\widehat{\mathbf{s}}, r) \longmapsto \left\| r L_{\mathbf{p}}(\widehat{\mathbf{s}}) + \frac{r^2}{2} Q_{\mathbf{p}}(\widehat{\mathbf{s}}) \right\|_{\mathcal{B}}.$

Lemma 13 *There exists a constant $c_7 > 0$ such that for all $\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}$, the function $r \mapsto G(\widehat{\mathbf{s}}, r)$ is strictly increasing in $r \in [0, c_7]$.*

Proof We will show that we can take

$$c_7 := \begin{cases} 1 & \text{if } Q_{\mathbf{p}} \equiv 0, \\ \min_{\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}} \|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}} / \max_{\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}} \|Q_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}} & \text{otherwise.} \end{cases}$$

Obviously $c_7 > 0$, because $\|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}} \gtrsim \|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_2 = \|\widehat{\mathbf{s}}\|_2 = 1$ by equivalence of norms and (8), and $Q_{\mathbf{p}}(\widehat{\mathbf{s}}) = O(1)$ by continuity and compactness.

Fix $\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}$. Set $g : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ by $g(r) = G(\widehat{\mathbf{s}}, r)$, $\mathbf{a} := L_{\mathbf{p}}(\widehat{\mathbf{s}})$ and $\mathbf{b} := \frac{1}{2} Q_{\mathbf{p}}(\widehat{\mathbf{s}})$. Since g is continuous, we may check g is strictly increasing on the half-open interval $[0, c_7)$.

Let $\lambda > 0$ satisfy $(1 + \lambda)r < c_7$. By the reverse triangle inequality and the definition of c_7, we have

$$\begin{aligned} g((1+\lambda)r) &= \left\| (1+\lambda)(r\mathbf{a} + r^2\mathbf{b}) + (\lambda + \lambda^2)r^2\mathbf{b} \right\|_{\mathcal{B}} \\ &\geq (1+\lambda)\|r\mathbf{a} + r^2\mathbf{b}\|_{\mathcal{B}} - (\lambda + \lambda^2)r^2\|\mathbf{b}\|_{\mathcal{B}} \\ &= \|r\mathbf{a} + r^2\mathbf{b}\|_{\mathcal{B}} + \lambda r(\|\mathbf{a} + r\mathbf{b}\|_{\mathcal{B}} - (1+\lambda)r\|\mathbf{b}\|_{\mathcal{B}}) \\ &\geq \|r\mathbf{a} + r^2\mathbf{b}\|_{\mathcal{B}} + \lambda r(\|\mathbf{a}\|_{\mathcal{B}} - r\|\mathbf{b}\|_{\mathcal{B}} - (1+\lambda)r\|\mathbf{b}\|_{\mathcal{B}}) \\ &> \|r\mathbf{a} + r^2\mathbf{b}\|_{\mathcal{B}} + \lambda r(\|\mathbf{a}\|_{\mathcal{B}} - \tfrac{1}{2}\|\mathbf{a}\|_{\mathcal{B}} - \tfrac{1}{2}\|\mathbf{a}\|_{\mathcal{B}}). \end{aligned}$$

The last quantity equals $g(r)$, and the lemma follows. $\qquad\square$

The following quantity is well-defined as a consequence of Lemma 13.

Corollary/Definition 4 There exists a constant $\sigma_5 > 0$ such that for all $\sigma \leq \sigma_5$ and all $\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}$, RadialDomain$(\widehat{\mathbf{s}}, \sigma)$ is a closed interval. Thus, there exists a function

$$r^* : \mathbb{S}^{d-1} \times [0, \sigma_5] \to \mathbb{R}_{\geq 0} \quad \text{such that } \text{RadialDomain}(\widehat{\mathbf{s}}, \sigma) = [0, r^*(\widehat{\mathbf{s}}, \sigma)]. \tag{80}$$

Lemma 14 *There exists a constant $\sigma_6 > 0$ such that the restriction of r^* to $\mathbb{S}^{d-1} \times [0, \sigma_6]$ is a continuous function.*

Proof Take $\sigma_6 = \sigma_5/2$. We shall verify continuity of r^* by bare hands. For notational convenience, within this proof, we denote the second argument of r^* by τ (subscripted and/or primed) rather than by σ. Fix $(\widehat{\mathbf{s}}_1, \tau_1) \in \mathbb{S}^{d-1} \times [0, \sigma_6]$ and let $\epsilon > 0$.

Lemma 13 says $r \mapsto G(\widehat{\mathbf{s}}_1, r)$ is continuous and strictly increasing around 0. By elementary facts, this has a well-defined continuous inverse function

around $G(0) = 0$. The inverse function is $\tau \mapsto r^*(\widehat{\mathbf{s}}_1, \tau)$ defined for $\tau \in [0, \sigma_5]$ (Lemma 4). So, we can take $\delta' \in (0, \sigma_6)$ such that for all $\tau'_2 \in [0, \sigma_5]$,

$$|\tau'_2 - \tau_1| < \delta' \implies |r^*(\widehat{\mathbf{s}}_1, \tau'_2) - r^*(\widehat{\mathbf{s}}_1, \tau_1)| < \epsilon. \tag{81}$$

Since $L_{\mathbf{p}}, Q_{\mathbf{p}}$ are continuous, there exist $\delta'', \delta''' > 0$ such that for all $\widehat{\mathbf{s}}_2 \in \mathbb{S}^{d-1}$,

$$\|\widehat{\mathbf{s}}_2 - \widehat{\mathbf{s}}_1\|_2 < \delta'' \implies \|L_{\mathbf{p}}(\widehat{\mathbf{s}}_2) - L_{\mathbf{p}}(\widehat{\mathbf{s}}_1)\|_{\mathcal{B}} < \frac{1}{c_7}\frac{\delta'}{3}, \tag{82}$$

$$\|\widehat{\mathbf{s}}_2 - \widehat{\mathbf{s}}_1\|_2 < \delta''' \implies \left\|\frac{1}{2}Q_{\mathbf{p}}(\widehat{\mathbf{s}}_2) - \frac{1}{2}Q_{\mathbf{p}}(\widehat{\mathbf{s}}_1)\right\|_{\mathcal{B}} < \frac{1}{c_7^2}\frac{\delta'}{3}. \tag{83}$$

Define $\delta := \min\left(\delta'/3, \delta'', \delta'''\right) > 0$. Let $(\widehat{\mathbf{s}}_2, \tau_2) \in \mathbb{S}^{d-1} \times [0, \sigma_6]$ satisfy $\|(\widehat{\mathbf{s}}_2, \tau_2) - (\widehat{\mathbf{s}}_1, \tau_1)\|_2 < \delta$. We shall verify $|r^*(\widehat{\mathbf{s}}_2, \tau_2) - r^*(\widehat{\mathbf{s}}_1, \tau_1)| < \epsilon$. Put $r_1 := r^*(\widehat{\mathbf{s}}_1, \tau_1), r_2 := r^*(\widehat{\mathbf{s}}_2, \tau_2)$ and $\tau'_2 := G(\widehat{\mathbf{s}}_1, r_2)$. By (81), it suffices to check $|\tau'_2 - \tau_1| < \delta'$, as then $\tau'_2 \in [0, \sigma_5]$ (because $\tau'_2 \leq \tau_1 + \delta' \leq 2\sigma_6 = \sigma_5$) and also $r^*(\widehat{\mathbf{s}}_1, \tau'_2) = r^*(\widehat{\mathbf{s}}_1, G(\widehat{\mathbf{s}}_1, r_2)) = r_2$. So, (81) gives $|r_2 - r_1| < \epsilon$.

To see that $|\tau'_2 - \tau_1| < \delta$ indeed holds, we write

$$\tau'_2 = \left\|r_2 L_{\mathbf{p}}(\widehat{\mathbf{s}}_1) + r_2^2\frac{1}{2}Q_{\mathbf{p}}(\widehat{\mathbf{s}}_1)\right\|_{\mathcal{B}} =$$

$$\left\|(r_2 L_{\mathbf{p}}(\widehat{\mathbf{s}}_2) + \frac{r_2^2}{2}Q_{\mathbf{p}}(\widehat{\mathbf{s}}_2)) + r_2(L_{\mathbf{p}}(\widehat{\mathbf{s}}_1) - L_{\mathbf{p}}(\widehat{\mathbf{s}}_2)) + r_2^2(\frac{1}{2}Q_{\mathbf{p}}(\widehat{\mathbf{s}}_2) - \frac{1}{2}Q_{\mathbf{p}}(\widehat{\mathbf{s}}_1))\right\|_{\mathcal{B}}.$$

Using the triangle inequality, (82), (83) and $r_2 \leq c_7$ (from the proof of Lemma 4),

$$|\tau'_2 - \tau_1| \leq |\tau_2 - \tau_1| + |\tau'_2 - \tau_2| < \frac{\delta'}{3} + c_7\frac{1}{c_7}\frac{\delta'}{3} + c_7^2\frac{1}{c_7^2}\frac{\delta'}{3} = \delta'.$$

This proves r^* is continuous on $\mathbb{S}^{d-1} \times [0, \sigma_6]$, when $\sigma_6 = \sigma_5/2$. $\qquad\square$

Lemma 15 *There exist constants $c_8 \geq 0$ and $\sigma_7 > 0$ such that for all $\sigma \leq \sigma_7$ and all $\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}$,*

$$\frac{1}{\|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}}\sigma - c_8\sigma^2 \leq r^*(\widehat{\mathbf{s}}, \sigma) \leq \frac{1}{\|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}}\sigma + c_8\sigma^2. \tag{84}$$

Proof We shall prove that we may take

$$c_8 = \frac{\max_{\widehat{\mathbf{s}}\in\mathbb{S}^{d-1}}\|Q_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}}{\min_{\widehat{\mathbf{s}}\in\mathbb{S}^{d-1}}\|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}^3}. \tag{85}$$

Given $\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}$. Write $\mathbf{a} = L_{\mathbf{p}}(\widehat{\mathbf{s}})$, $\mathbf{b} = \frac{1}{2}Q_{\mathbf{p}}(\widehat{\mathbf{s}})$, $a = \|\mathbf{a}\|_{\mathcal{B}}$ and $b = \|\mathbf{b}\|_{\mathcal{B}}$. If $\mathbf{b} = 0$, then $r^*(\widehat{\mathbf{s}}, \sigma) = (1/a)\sigma$ for all $\sigma \geq 0$, so (84) is obviously satisfied. Assume $\mathbf{b} \neq 0$. The triangle inequality gives

$$g_-(r) \leq g(r) \leq g_+(r) \quad \text{for all } r \in \mathbb{R}_{\geq 0}, \tag{86}$$

where $g(r) := G(\widehat{\mathbf{s}}, r) = \|r\mathbf{a} + r^2\mathbf{b}\|_{\mathcal{B}}$, $g_-(r) := ar - br^2$ and $g_+(r) := ar + br^2$. Note g_+ is strictly increasing over $r \in [0, \infty)$, while g_- is strictly increasing over $r \in [0, \frac{a}{2b}]$ and $g_-(\frac{a}{2b}) = \frac{a^2}{4b}$. Let

$$\sigma_6' := \min\left(\sigma_5, \frac{\min_{\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}} \|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}^2}{4 \max_{\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}} \|Q_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}}\right) > 0.$$

It follows from (86) and the intermediate value theorem that for all $\sigma \in [0, \sigma_7']$,

$$r_{+,+}^*(\sigma) \leq r^*(\sigma) \leq r_{-,-}^*(\sigma),$$

where $r_{+,+}^*(\sigma)$ denotes the greater of the two roots in r to the quadratic equation $g_+(r) = \sigma$ and where $r_{-,-}^*(\sigma)$ denotes the lesser of the two roots in r to $g_-(r) = \sigma$. Explicitly by the quadratic formula and Taylor series for the square root function, we have

$$r_{+,+}^*(\sigma) := \frac{-a + \sqrt{a^2 + 4b\sigma}}{2b} = \frac{1}{a}\sigma - \frac{b}{a^3}\sigma^2 + O(\sigma^3),$$

$$r_{-,-}^*(\sigma) := \frac{a - \sqrt{a^2 - 4b\sigma}}{2b} = \frac{1}{a}\sigma + \frac{b}{a^3}\sigma^2 + O(\sigma^3). \tag{87}$$

On the other hand, from the compactness of $\mathbb{S}^{d-1}$, one can check the implicit constants suppressed by the big O notation in (87) may all be taken independently of $\widehat{\mathbf{s}}$. At the same time, $\frac{2b}{a^3} \leq c_8$ for each $\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}$, by the definition (85). Taking $\sigma_7 > 0$ to be sufficiently smaller than σ_7' yields the lemma. $\qquad\square$

Definition 5 Set $\sigma_8 := \min(\sigma_6, \sigma_7) > 0$. Define $\eta^* : \mathbb{S}^{d-1} \times (0, \sigma_8] \to \mathbb{R}$ by

$$r^*(\widehat{\mathbf{s}}, \sigma) =: \frac{1}{\|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}}\sigma + \tfrac{1}{2}\eta^*(\widehat{\mathbf{s}}, \sigma)\sigma^2. \tag{88}$$

By Lemma 14, $\frac{1}{2}\eta^*$ is continuous. By Lemma 15, it is bounded uniformly in absolute value by $2c_8$.

4.8 Step 8: obtain $\mathrm{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}}$ and apply dominated convergence

It remains to establish Eq. (73): that is, to obtain the first-order term in the limiting differential operator. We do this using spherical coordinates (Section 4.6) and the results about the radial integration domain developed in Section 4.7. By swapping the order of limit and the integration (justified by dominated convergence), the tilt function emerges at last.

Proposition 16 *For each $\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}$,*

$$\lim_{\sigma \to 0} \tfrac{1}{2}\eta^*(\widehat{\mathbf{s}}, \sigma) = \mathrm{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}}(\widehat{\mathbf{s}}). \tag{89}$$

In particular, the limit on the LHS exists.

Proof By the bound (84) and compactness of $[-c_8, c_8]$, it suffices to show that every accumulation point of $\frac{1}{2}\eta^*(\widehat{\mathbf{s}}, \sigma)$ as $\sigma \to 0$ equals $\text{tilt}_{\mathcal{M}, \mathcal{B}, \mathbf{p}}(\widehat{\mathbf{s}})$. That is, assume $(\tau_k)_{k=1}^{\infty} \subseteq (0, \sigma_8]$ is such that $\tau_k \to 0$ and $\frac{1}{2}\eta^*(\widehat{\mathbf{s}}, \tau_k) \to \eta \in [-c_8, c_8]$ as $k \to \infty$; we will show $\eta = \text{tilt}_{\mathcal{M}, \mathcal{B}, \mathbf{p}}(\widehat{\mathbf{s}})$. Substituting (88) into (80), and putting $\mathbf{a} = L_{\mathbf{p}}(\widehat{\mathbf{s}})$, $\mathbf{b} = \frac{1}{2}Q_{\mathbf{p}}(\widehat{\mathbf{s}})$, $\eta_k = \eta(\widehat{\mathbf{s}}, \tau_k)$, gives

$$\tau_k = \left\| \left(\frac{\tau_k}{\|\mathbf{a}\|_{\mathcal{B}}} + \frac{1}{2}\eta_k\tau_k^2 \right)\mathbf{a} \; + \; \left(\frac{\tau_k}{\|\mathbf{a}\|_{\mathcal{B}}} + \frac{1}{2}\eta_k\tau_k^2 \right)^2 \mathbf{b} \right\|_{\mathcal{B}} .$$

Rearranging and dividing by τ_k, this reads

$$1 = \left\| \frac{\mathbf{a}}{\|\mathbf{a}\|_{\mathcal{B}}} \; + \; \left(\frac{1}{2}\eta_k\mathbf{a} + \frac{\mathbf{b}}{\|\mathbf{a}\|_{\mathcal{B}}^2} \right)\tau_k + \frac{\eta_k\mathbf{b}}{\|\mathbf{a}\|_{\mathcal{B}}}\tau_k^2 + \frac{\eta_k^2\mathbf{b}}{4}\tau_k^3 \right\|_{\mathcal{B}} . \tag{90}$$

By the definition of tangent cones, (90) witnesses that

$$\frac{\mathbf{b}}{\|\mathbf{a}\|_{\mathcal{B}}^2} + \eta\mathbf{a} \in TC_{\mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}}}(\partial\mathcal{B}) . \tag{91}$$

Indeed, in the definition (12), take $\mathcal{Y} = \partial\mathcal{B} \subseteq \mathbb{R}^D$; $\mathbf{y} = \mathbf{a}/\|\mathbf{a}\|_{\mathcal{B}} \in \partial\mathcal{B}$; $\mathbf{y}_k = \frac{\mathbf{a}}{\|\mathbf{a}\|_{\mathcal{B}}} + \left(\frac{1}{2}\eta_k\mathbf{a} + \frac{\mathbf{b}}{\|\mathbf{a}\|_{\mathcal{B}}^2} \right)\tau_k + \frac{\eta_k\mathbf{b}}{\|\mathbf{a}\|_{\mathcal{B}}}\tau_k^2 + \frac{\eta_k^2\mathbf{b}}{4}\tau_k^3 \in \partial\mathcal{B}$; the same τ_k; and $\mathbf{d} = \frac{\mathbf{b}}{\|\mathbf{a}\|_{\mathcal{B}}^2} + \eta\mathbf{a}$. Then, (91) follows because $(\mathbf{y}_k - \mathbf{y})/\tau_k = \left(\frac{1}{2}\eta_k\mathbf{a} + \frac{\mathbf{b}}{\|\mathbf{a}\|_{\mathcal{B}}^2} \right) + \frac{\eta_k\mathbf{b}}{\|\mathbf{a}\|_{\mathcal{B}}}\tau_k + \frac{\eta_k^2\mathbf{b}}{4}\tau_k^2 \to \mathbf{d}$ as $k \to \infty$, using $\frac{1}{2}\eta_k \to \eta$, $\tau_k \to 0$ and $\eta_k = O(1)$ (Lemma 15) as $k \to \infty$. From Proposition/Definition 4 and the membership (91), we obtain $\eta = \text{tilt}_{\mathcal{M}, \mathcal{B}, \mathbf{p}}(\widehat{\mathbf{s}})$. $\qquad\square$

Finishing the proof of Theorem 6: From the end of Section 4.5, it remains to establish (73). Let $\sigma \leq \min(\sigma_5, \sigma_6, \sigma_7, \sigma_8)$. Then using spherical coordinates:

$$\frac{1}{\sigma^{d+2}} \int_{\mathbf{s} \in \text{Approx}^{(2)}(\sigma)} \mathbf{s} d\mathbf{s}$$

$$= \frac{1}{\sigma^{d+2}} \int_{\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}} \widehat{\mathbf{s}} \int_{r \in \text{RadialDomain}(\widehat{\mathbf{s}}, \sigma)} r^d dr d\widehat{\mathbf{s}} \qquad\qquad [(75)]$$

$$= \frac{1}{\sigma^{d+2}} \int_{\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}} \widehat{\mathbf{s}} \int_{r=0}^{r^*(\widehat{\mathbf{s}}, \sigma)} r^d dr d\widehat{\mathbf{s}} \qquad\qquad [(80), \sigma \leq \sigma_5].$$

Substituting Eq. (88) for $r^*(\widehat{\mathbf{s}}, \sigma)$ and evaluating the inner integral, we obtain

$$\frac{1}{\sigma^{d+2}} \int_{\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}} \frac{\widehat{\mathbf{s}}}{d+1} \left(\frac{1}{\|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}} \sigma + \frac{1}{2} \eta^*(\widehat{\mathbf{s}}, \sigma) \sigma^2 \right)^{d+1} d\widehat{\mathbf{s}} \qquad [(88),\ \sigma \le \sigma_8]$$

$$= \frac{1}{(d+1)\sigma} \int_{\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}} \frac{\widehat{\mathbf{s}}}{\|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}^{d+1}} d\widehat{\mathbf{s}}$$

$$+ \int_{\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}} \widehat{\mathbf{s}} \left(\frac{\frac{1}{2} \eta^*(\widehat{\mathbf{s}}, \sigma)}{\|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}^d} + O(\sigma) \right) d\widehat{\mathbf{s}} \qquad [(84),\ \sigma \le \sigma_7,\ \|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}^{-1} = \omega(1)]$$

$$= \int_{\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}} \widehat{\mathbf{s}} \left(\frac{\frac{1}{2} \eta^*(\widehat{\mathbf{s}}, \sigma)}{\|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}^d} + O(\sigma) \right) d\widehat{\mathbf{s}}. \qquad [\text{oddness}]$$

By Eq. (89) and dominated convergence, as $\sigma \to 0$ this integral converges to

$$\int_{\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}} \widehat{\mathbf{s}} \, \|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}^{-d} \lim_{\sigma \to 0} \frac{1}{2} \eta^*(\widehat{\mathbf{s}}, \sigma) \, d\widehat{\mathbf{s}} = \int_{\widehat{\mathbf{s}} \in \mathbb{S}^{d-1}} \widehat{\mathbf{s}} \|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}^{-d} \mathrm{tilt}_{\mathcal{M}, \mathcal{B}, \mathbf{p}}(\widehat{\mathbf{s}}) \, d\widehat{\mathbf{s}}.$$

The use of dominated convergence is justified because $\widehat{\mathbf{s}} \mapsto \frac{1}{2}\eta^*(\widehat{\mathbf{s}}, \sigma)$ is continuous in $\widehat{\mathbf{s}}$ for each $\sigma \le \sigma_6$ (Lemma 14), it is uniformly bounded in absolute value by c_8 for each $\sigma \le \sigma_7$ (Lemma 15), and $\lim_{\sigma \to 0} \frac{1}{2}\eta^*(\widehat{\mathbf{s}}, \sigma)$ exists and is $\mathrm{tilt}_{\mathcal{M}, \mathcal{B}, \mathbf{p}}$ (Proposition 16). This completes the proof of Theorem 6. □

5 Application: mapping volumetric shape spaces

In this section, we demonstrate the use of non-Euclidean norms for embedding a set of 3D densities with continuous variability. The specific motivation comes from the field of single-particle cryo-electron microscopy (cryo-EM), an imaging technique for reconstructing the 3D structure of proteins, ribosomes, and other large molecules, using a large set of electron-microscope images. We give a description of cryo-EM and the continuous heterogeneity problem, which naturally lends itself to manifold learning. We then describe our method for mapping general volumetric shape spaces using non-Euclidean diffusion maps, and apply it to a simulated data set that satisfies the assumptions of Theorem 7.. For a broader general introduction to cryo-EM see Chapter 1 of [23] or the more mathematically-oriented reviews [58, 9]. Code for reproducing the numerical results is available at:

   http://github.com/mosco/manifold-learning-arbitrary-norms

5.1 Single-particle cryo-EM

The goal of single-particle cryo-EM is to obtain the 3D structure of a molecule of interest. This is done by obtaining a sample of the molecule and freezing it so that it forms a thin sheet of ice. This sheet typically contains hundreds of thousands of copies of the molecule, each suspended at a different orientation.
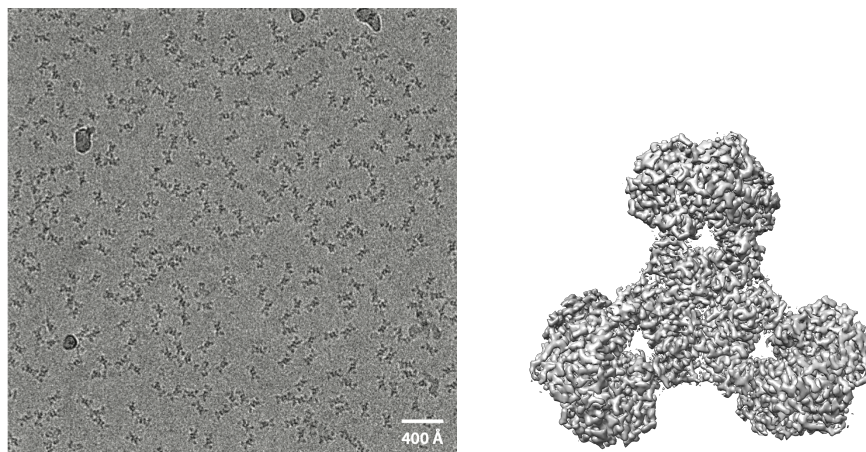
Fig. 3 (left) Cryo-EM image showing ~ 220 noisy tomographic projections of the PaaZ molecule and some contaminants (from [58]); (right) Surface plot of the reconstructed electrostatic density of the PaaZ molecule, based on 118,203 tomographic projections (from [54]).

The frozen sample is then imaged using a transmission electron microscope. This results in images that contain many noisy tomographic projections of the same molecule, viewed from different (unknown) directions (See Figure 3). The challenge is to compute a 3D reconstruction of the electrostatic density map. The field of cryo-EM has made such progress over the last decade that it is now common to see reconstructions of large rigid molecules, composed of tens of thousands of individual atoms, with resolutions finer than 3 ångströms, which allow for the accurate fitting of atomic models using specialized software. See the right panel of Figure 3 for an example experimental reconstruction.

The basic assumption behind most single-particle cryo-EM methods is that the molecule of interest is *rigid*. Hence, the different electron microscope images are tomographic projections of the same exact 3D volume from different angles (or the at least, there is only a finite set 3D volumes). However, this assumption does not always hold: some molecules have flexible components that can move independently. This fact, known as the *continuous heterogeneity problem* in cryo-EM, poses a difficulty for existing reconstruction methods. One of the key ongoing challenges in the field is the development of new methods that can map the entire space of molecular conformations [17, 28, 62, 18, 47, 14, 29, 73, 49]. See [60] for a survey. Several works have applied diffusion maps to this problem domain [15, 55, 14, 44]. In our conference paper [70], we applied diffusion maps with a particular non-Euclidean norm to a given set of 3D densities. Specifically, we used a fast wavelet-based approximation to the Earthmover's distance (WEMD). Those numerical results were the original motivation for the present paper, and the rest of Section 5 extends them.
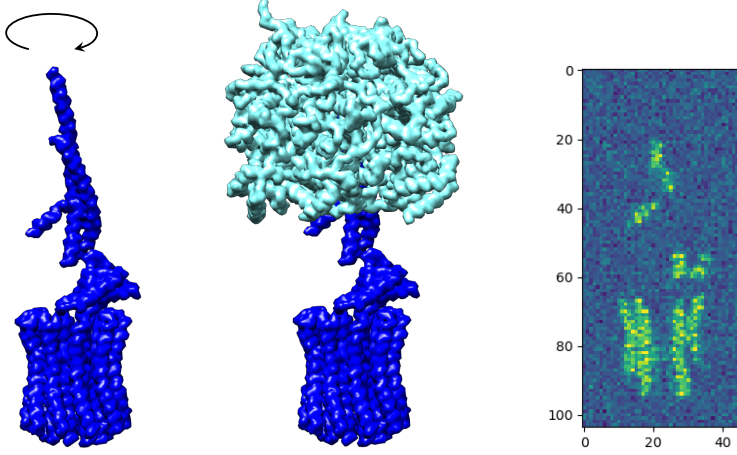
Fig. 4 *ATP synthase.* (left) F_0 and axle subunits. These jointly rotate in the presence of hydrogen ions, together forming a molecular electric motor; (middle) the F_1 subunit (in cyan) envelops the axle. As the axle rotates, the F_1 subunit assembles ATP; (right) representative 2D slice of the rotated F_0 and axle subunits with additive noise shown.

5.2 WEMD-based diffusion maps

Given a set of volumetric arrays $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{M} \subseteq \mathbb{R}^{N_x \times N_y \times N_z}$, we compute an approximate Earthmover's distance between all pairs of arrays [56]. This is done by first computing the discrete wavelet transform of each input and then using a weighted ℓ_1-norm on the pairwise differences of wavelet coefficients,

$$\|\mathbf{x}_i - \mathbf{x}_j\|_{\mathrm{WEMD}} := \sum_\lambda 2^{-5s/2} \left| \mathcal{W}\mathbf{x}_i(\lambda) - \mathcal{W}\mathbf{x}_j(\lambda) \right|. \tag{92}$$

Here, $\mathcal{W}\mathbf{x}$ denotes a 3D wavelet transform of $\mathbf{x}$ [39]. The index λ contains the wavelet shifts $(m_1, m_2, m_3) \in \mathbb{Z}^3$ and scale parameter $s \in \mathbb{Z}_{\geq 0}$. We then compute pairwise Gaussian affinities,

$$W_{ij} = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|_{\mathrm{WEMD}}^2 / \sigma^2\right), \tag{93}$$

proceed to construct a graph Laplacian, and perform the eigenvector-based embedding as described in Section 2. Since the construction uses a (fixed) norm, the theory described in Section 3.4 applies in this case. Hence, in the noiseless case, the graph Laplacian converges to an elliptic second-order differential operator on the relevant manifold $\mathcal{M}$ of ATP synthase conformations. Here, $\mathcal{M}$ is embedded in the Euclidean space of arrays of size $N_x \times N_y \times N_z$.

5.3 Simulation results

We tested our method on a synthetic volumetric data set that mimics the motion space of ATP synthase [69], see Figure 4. This enzyme is a stepper
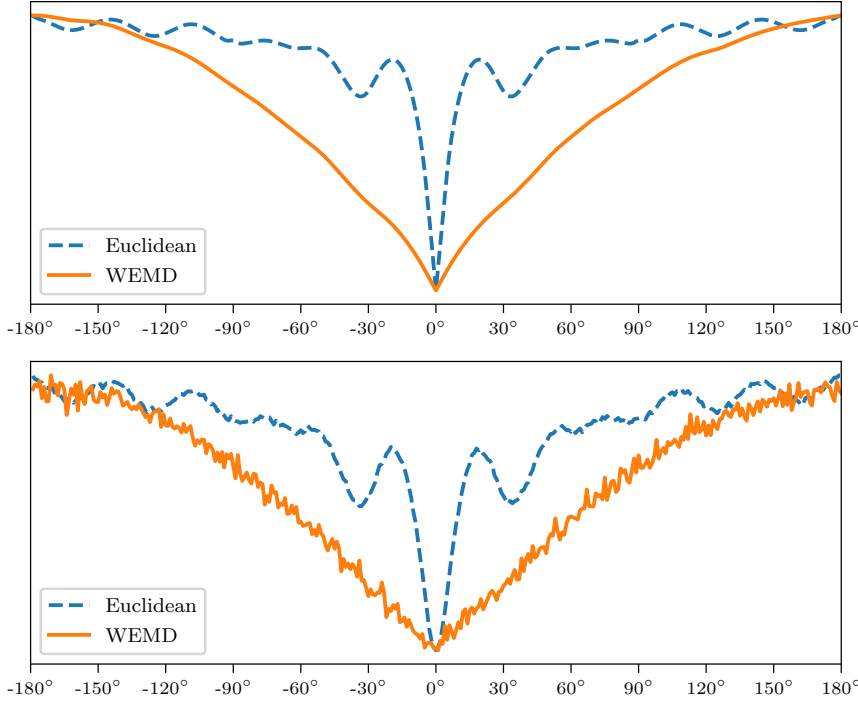
Fig. 5 *Euclidean distance vs. wavelet-based approximate Earthmover's distance* as functions of the angle between rotations of the ATP synthase rotor. (top) distances for rotated volumes without noise; (bottom) distances for the noisy data set. (Euclidean distances were scaled to be comparable to WEMDs.)

motor with a central asymmetric axle that rotates in 120° steps relative to the F_1 subunit, with short transient motions in-between the three dominant states. Our synthetic data was generated as follows: we produced 3D density maps of entry 1QO1 [61] from the Protein Data Bank [1] using the `molmap` command in UCSF Chimera [48]. These density maps have array dimensions $47 \times 47 \times 107$ and a resolution of 6Å per voxel. We then took random rotations of the F_0 and axle subunits, where the angles were drawn i.i.d. according to the following mixture distribution,

$$\tfrac{2}{5}U[0, 360] + \tfrac{1}{5}\mathcal{N}(0, 1) + \tfrac{1}{5}\mathcal{N}(120, 1) + \tfrac{1}{5}\mathcal{N}(240, 1).$$

The resulting density maps formed the clean dataset. The noisy dataset was generated in the same manner but also included additive i.i.d. Gaussian noise with mean zero and a standard deviation of 1/10 of the maximum voxel value.

The discrete wavelet transform of all the volumes in the dataset was computed using PyWavelets [31] with the `sym3` wavelet (symmetric Daubechies wavelets of order 3), though other wavelet choices also worked well [56, Sec 4.2]. The maximum scale level chosen was $s = 6$ to minimize the truncation in Eq. (92). The number of resulting wavelet coefficients was 40% larger than the number of voxels.

Table 2 *Running times [sec]* for computing the discrete wavelet transform (DWT), all pairwise wavelet-based Earthmover approximations (WEMD) not including the DWT, and all pairwise Euclidean (ℓ_2) distances.

n	DWT	WEMD	ℓ_2
25	0.3	0.13	0.09
50	0.61	0.49	0.38
100	1.2	1.93	1.5
200	2.4	7.6	5.5
400	4.9	31	22
800	11	126	86

Figure 5 compares the Euclidean norm to the WEMD norm for a range of angular differences for the noiseless and noisy datasets. Note that for the clean dataset, WEMD is monotonic in the absolute value of the X axis (equal to the angular difference between the ATP synthase rotors). This behavior also holds for the Euclidean norm, but only for small angular differences up to $\approx \pm 19°$. This suggests that an affinity graph built from this dataset using the Euclidean norm can capture the right geometry only when the dataset contains a dense sampling of the angles and when the kernel width is properly calibrated to nearly cut off connections at angles $> 19°$.

Figure 6 is the result of a two-dimensional Laplacian eigenmaps embedding, once with the Euclidean norm and once with WEMD norm (92). These embeddings use the unweighted graph Laplacian with a Gaussian kernel, which corresponds to the setting of Theorem 6. For similar results that use the density normalized diffusion maps of [11], see [70, Fig. 5]. We chose $\sigma = 30$ as the Gaussian kernel width in Eq. (93) for the WEMD embeddings, however the WEMD results were not very sensitive to the particular choice of σ. In contrast, the Euclidean embeddings required fine-tuning of σ to obtain the best results for each sample size. This makes sense given the results of Figure 5.

The key takeaway from Figure 6 is that for the standard Laplacian eigenmaps embedding based on the Euclidean norm, one needs > 400 samples to conclude that the intrinsic geometry is a circle. In contrast, for the embeddings based on WEMD, even small sample sizes give the right geometry.

5.4 Runtime

The running time of the WEMD-based diffusion maps is similar to that of the standard Euclidean diffusion maps. This follows from the fact that both algorithms need to compute $\binom{n}{2}$ pairwise ℓ_p-distances ($p \in \{1, 2\}$) for vectors of similar length. The cost of the discrete wavelet transform is negligible, since it is linear with respect to the input size. For our sample sizes, the time to form the Gaussian affinity matrix and compute its eigenvectors is also negligible. Table 2 lists single-core running times on an Intel Core i7-8569U CPU.
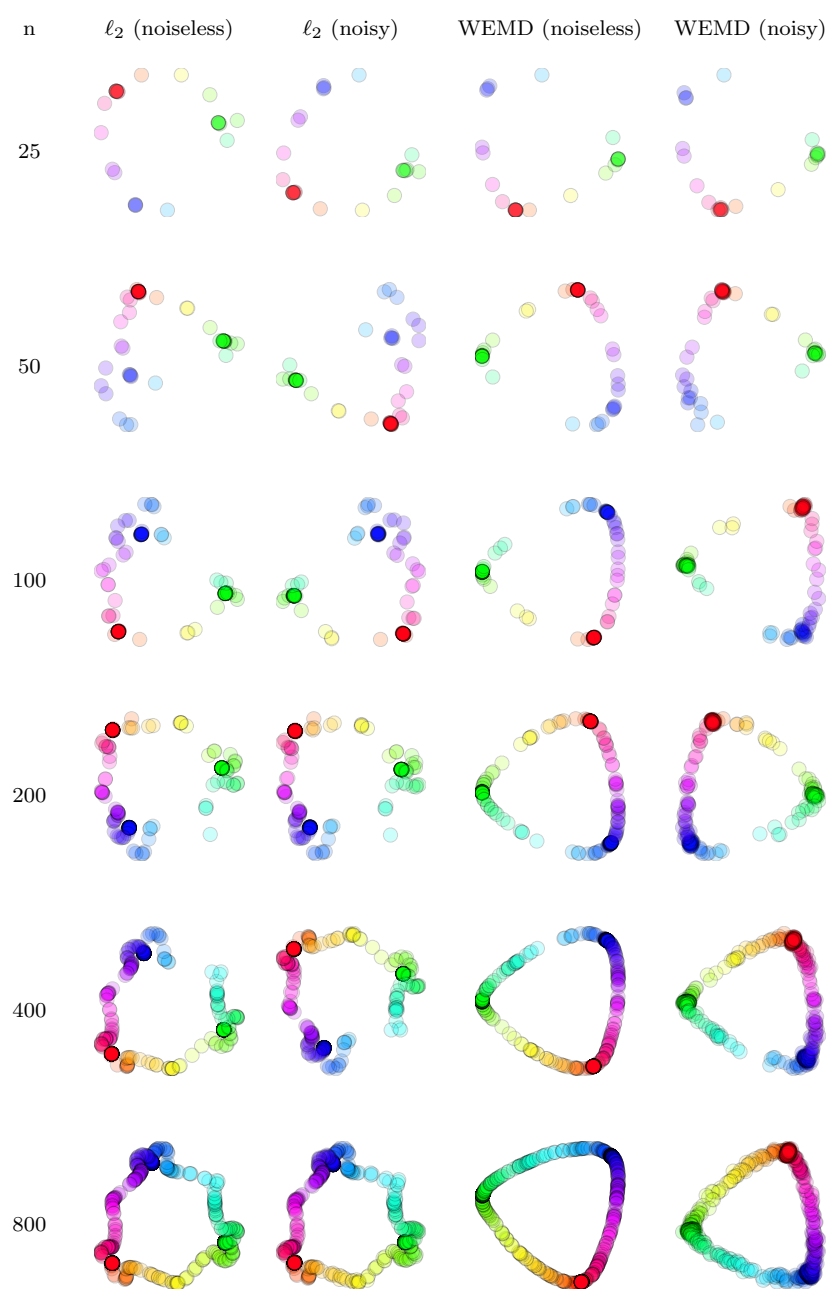
Fig. 6 *Simulation results.* Euclidean vs. WEMD-based Laplacian eigenmaps into $\mathbb{R}^2$ using the clean and noisy ATP synthase data sets. Sample sizes of $n = 25, 50, 100, 200, 400, 800$. The points are translucent to indicate density and the color is the groundtruth angle.

Table 3 *Running times [sec]* for computing all pairs of sparsified wavelet-based Earthmover's distances (sparse-WEMD), as compared to the dense computation.

n	Sparse runtime (noiseless data)	Sparse runtime (noisy data)	Dense runtime
25	0.01	0.037	0.13
50	0.013	0.1	0.49
100	0.026	0.39	1.93
200	0.046	1.5	7.6
400	0.16	6.2	31
800	0.6	25	126

5.5 Using wavelet sparsity

To compute the approximate Earthmover's distance between all pairs of volumes, we first compute a weighted discrete wavelet transform of each volume in the data set. For smooth signals, this results in sparse vectors of wavelet coefficients [39]. We can use this property by thresholding the vectors of weighted wavelet coefficients, and then storing them in a sparse matrix. This is beneficial because computing the ℓ_1-distance between two sparse vectors has a runtime that is linear in the number of their non-zero elements. Since the computation of all pairwise ℓ_1 differences is the slowest part of our procedure, this approach can reduce the running time significantly. To test this, we used the ATP synthase data described in the previous section. First, we subtracted the mean volume from all volumes in the data set. This mean-centering does not change the pairwise WEMD distances but makes the resulting vectors more sparse.

We used the hard-thresholding function h_t defined as follows:

$$h_t(x) := \begin{cases} 0 & \text{for } |x| \leq t, \\ x & \text{for } |x| > t. \end{cases}$$

We found a threshold t for the wavelet coefficients such that the ℓ_1-norm of the post-thresholding weighted wavelet coefficients are $> 90\%$ of the ℓ_1-norm of the dataset prior to thresholding. This threshold was computed on the smallest simulation of size $n = 25$ and then applied to the rest of the runs. Figure 7 shows the results of the WEMD embedding following this sparsification step. Table 3 shows the running times for the sparsified WEMD. Note that the running times are different for the noiseless and noisy data, since the noisy data is less sparse. However, in both cases, there are significant gains to the running times, with few visually-noticeable changes to the embedding results.
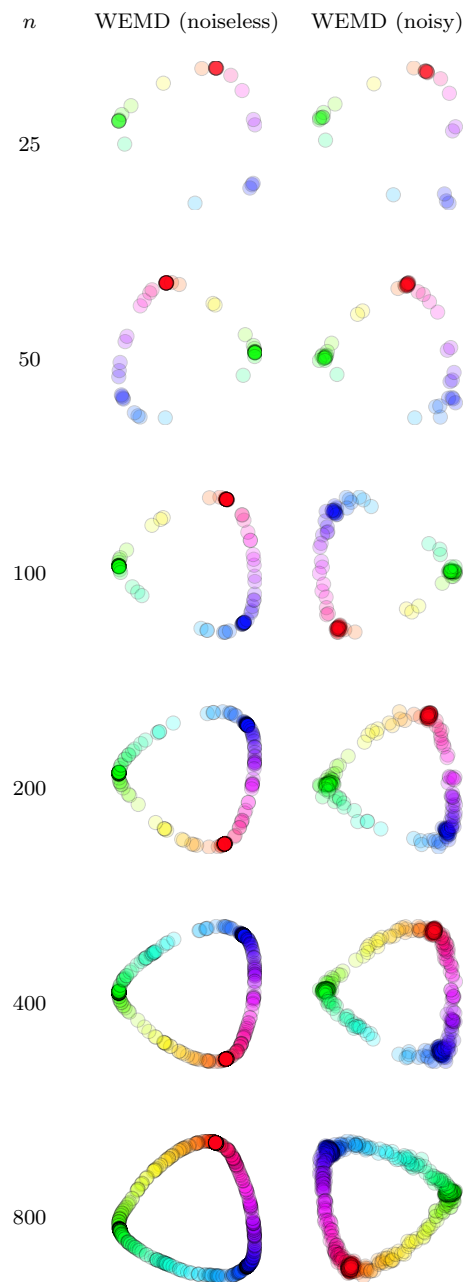
Fig. 7 *Sparsified results.* Wavelet-EMD-based Laplacian eigenmaps of the clean and noisy ATP synthase data sets, after applying hard-thresholding to obtain sparse coefficient vectors.

6 Conclusion

In this paper, we placed Laplacian-based manifold learning methods that use non-Euclidean norms on a firmer theoretical footing. We proved the pointwise convergence of graph Laplacians computed using general norms to elliptic second-order differential operators. In particular, our proof involved a novel second-order interaction between the manifold $\mathcal{M}$ and the unit ball $\mathcal{B}$, encoded by the function $\text{tilt}_{\mathcal{M},\mathcal{B},\mathbf{p}}$. We showed that some properties of the usual Laplace-Beltrami operator are lost in the general case. The limiting operator $\Delta_{\mathcal{M},\mathcal{B}}$ changes with the embedding of $\mathcal{M}$. Further, the limit $\Delta_{\mathcal{M},\mathcal{B}}$ carries a first-order term that can exhibit discontinuities at certain points of $\mathcal{M}$.

In addition, this paper demonstrated practical advantages for using non-Euclidean norms in manifold learning. We considered the task of learning molecular shape spaces. Here data points are conformations represented by 3D arrays, and we want to capture the range of motion. A simulation found that using Laplacian eigenmaps with the wavelet Earthmover's distance (a weighted-ℓ_1 norm in wavelet coefficients) resulted in a qualitative improvement of sample complexity compared to the Euclidean norm. Thresholding the wavelet coefficients before computing norms reduced the computational cost.

This work suggests several directions worthy of future study:

- **Convergence rates.** With what rate does the convergence in Theorem 6 occur? How does this depend on the choice of norm $\|\cdot\|_{\mathcal{B}}$?

- **Eigenfunctions.** What can be said about the eigenfunctions of $\Delta_{\mathcal{M},\mathcal{B}}$? How do the discontinuities of the first-order term in $\Delta_{\mathcal{M},\mathcal{B}}$ impact them? Due to space limitations, we only gave numerical examples in Appendix F.

- **Spectral convergence.** For general norms, do the eigenvectors of the graph Laplacian converge to the eigenfunctions of the operator $\Delta_{\mathcal{M},\mathcal{B}}$?

- **Concentration.** The operator $\Delta_{\mathcal{M},\mathcal{B}}$ depends on d- and 2-dimensional linear sections of the convex body $\mathcal{B} \subseteq \mathbb{R}^D$. When $D \gg d$, is there a sense in which these slices look "increasingly Euclidean"? Does $\Delta_{\mathcal{M},\mathcal{B}}$ concentrate?

- **Data-dependent norms.** If the norm chosen is some fixed function of the data set, does a well-defined limit of the graph Laplacians still exist?

- **Applications.** Are some applied domains better-suited for non-Euclidean norms than others? How should a practitioner decide which norm to use?

Appendices

A Proof of Lemma 3

Step 1: LHS $\subseteq$ RHS. By the identity (13) for tangent cones of convex sets, we have

$$TC_{\mathbf{y}}(\mathcal{B}) = \overline{\mathbb{R}_{>0}(\mathcal{B} - \mathbf{y})}.$$

By definition of ∂ and of tangent cones (13), the LHS of Eq. (14) is

$$\partial\left(TC_{\mathbf{y}}(\mathcal{B})\right) = \overline{\mathbb{R}_{>0}(\mathcal{B} - \mathbf{y})} \setminus (\mathbb{R}_{>0}(\mathcal{B} - \mathbf{y}))^{\circ}. \tag{94}$$

Let $\mathbf{d} \in \partial\left(TC_{\mathbf{y}}(\mathcal{B})\right)$. By Eq. (94), $\mathbf{d} = \lim_{k \to \infty} \beta_k(\widetilde{\mathbf{y}}_k - \mathbf{y})$ for some $\beta_k \in \mathbb{R}_{>0}$ and $\widetilde{\mathbf{y}}_k \in \mathcal{B}$. Without loss of generality, we assume $\widetilde{\mathbf{y}}_k \in \partial\mathcal{B}$ for each k. We break into cases.

- *Case A: $\widetilde{\mathbf{y}} = \mathbf{y}$.*
  Either $\mathbf{d} = 0 \in T_{\mathbf{y}}(\partial\mathcal{B})$, or $\tau_k := 1/\beta_k \to \infty$ as $k \to \infty$. If the latter, the sequences $(\widetilde{\mathbf{y}}_k)_{k=1}^{\infty} \subseteq \partial\mathcal{B}$ and $(\tau_k)_{k=1}^{\infty} \subseteq \mathbb{R}_{>0}$ witness $\mathbf{d} \in TC_{\mathbf{y}}(\partial\mathcal{B})$.

- *Case B: $\widetilde{\mathbf{y}} \neq \mathbf{y}$.*
  Here, $\lim_{k \to \infty} \beta_k =: \beta \in \mathbb{R}_{\geq 0}$ exists, and $\mathbf{d} = \beta(\widetilde{\mathbf{y}} - \mathbf{y})$. If $\beta = 0$, then $\mathbf{d} = 0 \in T_{\mathbf{y}}(\partial\mathcal{B})$. Suppose $\beta \neq 0$. Let the line segment joining $\widetilde{\mathbf{y}}$ and $\mathbf{y}$ be

$$\mathtt{conv}\{\widetilde{\mathbf{y}}, \mathbf{y}\} := \{\alpha\widetilde{\mathbf{y}} + (1 - \alpha)\mathbf{y} \in \mathbb{R}^D : \alpha \in [0, 1]\}.$$

So, $\mathtt{conv}\{\widetilde{\mathbf{y}}, \mathbf{y}\} \subseteq \mathcal{B}$. We claim $\mathtt{conv}\{\widetilde{\mathbf{y}}, \mathbf{y}\} \subseteq \partial\mathcal{B}$. Assume not. That is,

$$\exists \alpha \in (0, 1) \text{ such that } \mathbf{z} := \alpha\widetilde{\mathbf{y}} + (1 - \alpha)\mathbf{y} \in \mathcal{B}^{\circ}.$$

But then,

$$\mathbf{d} = \beta(\widetilde{\mathbf{y}} - \mathbf{y}) = (\beta/\alpha)(\mathbf{z} - \mathbf{y}) \in \mathbb{R}_{>0}(\mathcal{B}^{\circ} - \mathbf{y}) \subseteq (\mathbb{R}_{>0}(\mathcal{B} - \mathbf{y}))^{\circ}.$$

This contradicts $\mathbf{d} \in \partial\left(TC_{\mathbf{y}}(\mathcal{B})\right)$ (see Eq. (94)). So, indeed $\mathtt{conv}\{\widetilde{\mathbf{y}}, \mathbf{y}\} \subseteq \partial\mathcal{B}$. Now, define

$$\widehat{\mathbf{y}}_k := \frac{1}{k}\widetilde{\mathbf{y}} + (1 - \frac{1}{k})\mathbf{y} \in \partial\mathcal{B} \quad \text{and} \quad \tau_k := \frac{1}{k} \in \mathbb{R}_{>0}.$$

Then, $\frac{\widehat{\mathbf{y}}_k - \mathbf{y}}{\tau_k} = \mathbf{d}$ for each k, and $(\widehat{\mathbf{y}}_k)_{k=1}^{\infty}$ and $(\tau_k)_{k=1}^{\infty}$ witness $\mathbf{d} \in TC_{\mathbf{y}}(\partial\mathcal{B})$.

In all cases, we have verified $\mathbf{d} \in TC_{\mathbf{y}}(\partial\mathcal{B})$. This gives LHS $\subseteq$ RHS in (14).

Step 2: LHS $\supseteq$ RHS. Let $\mathbf{d} \in TC_{\mathbf{y}}(\partial\mathcal{B})$. By the definition of tangent cones (12), $\mathbf{d} = \lim_{k \to \infty} \tau_k^{-1}(\widetilde{\mathbf{y}}_k - \mathbf{y})$ for some $\tau_k \in \mathbb{R}_{>0}$ and $\widetilde{\mathbf{y}}_k \in \partial\mathcal{B}$ with $\tau_k \to 0$ and $\widetilde{\mathbf{y}}_k \to \mathbf{y}$ as $k \to \infty$. By (94), we need to show $\mathbf{d} \notin (\mathbb{R}_{>0}(\mathcal{B} - \mathbf{y}))^{\circ}$.

First, we will prove $\mathtt{conv}\{\mathbf{d} + \mathbf{y}, \mathbf{y}\} \cap \mathcal{B}^{\circ} = \emptyset$. Assume not, i.e.,

$$\exists \alpha \in (0, 1) \text{ such that } \mathbf{z} := \alpha(\mathbf{d} + \mathbf{y}) + (1 - \alpha)\mathbf{y} = \alpha\mathbf{d} + \mathbf{y} \in \mathcal{B}^{\circ}.$$

Let $\widehat{\tau}_k = \tau_k/\alpha \in \mathbb{R}_{>0}$, so that

$$\alpha\mathbf{d} = \lim_{k \to \infty} \widehat{\tau}_k^{-1}(\widetilde{\mathbf{y}}_k - \mathbf{y}). \tag{95}$$

Since $\mathcal{B}^{\circ}$ is open, there exists $\delta > 0$ with

$$\mathcal{N} := \{\mathbf{w} \in \mathbb{R}^D : \|\mathbf{w} - \mathbf{z}\|_2 \leq \delta\} \subseteq \mathcal{B}^{\circ}.$$

By Eq. (95), there exists K such that for all $k \geq K$,

$$\widehat{\tau}_k^{-1}(\widetilde{\mathbf{y}}_k - \mathbf{y}) + \mathbf{y} \in \mathcal{N}.$$

On the other hand, it is easy to see for each $\mathbf{w} \in \mathcal{B}^\circ$,

$$\left(\mathbf{y} + \mathbb{R}_{\geq 0}(\mathbf{w} - \mathbf{y})\right) \cap \mathcal{B} = \mathtt{conv}\{\mathbf{y}, \mathbf{w}'\} \tag{96}$$

for some $\mathbf{w}' \in \partial\mathcal{B}$, using convexity and compactness of $\mathcal{B}$. In addition,

$$\left(\mathbf{y} + \mathbb{R}_{\geq 0}(\mathbf{w} - \mathbf{y})\right) \cap \partial\mathcal{B} = \{\mathbf{y}, \mathbf{w}'\}, \tag{97}$$

using $\mathbf{w} \in \mathtt{conv}\{\mathbf{y}, \mathbf{w}'\}$, $\|\mathbf{w}\|_\mathcal{B} < 1$, and the triangle inequality for $\|\cdot\|_\mathcal{B}$. Clearly,

$$\|\mathbf{w}' - \mathbf{y}\|_2 > \|\mathbf{w} - \mathbf{y}\|_2. \tag{98}$$

Now, let $\epsilon := \min_{\mathbf{w} \in \mathcal{N}} \|\mathbf{w} - \mathbf{y}\|_2$. Note $\epsilon > 0$. For each $k \leq K$, we apply (96), (97) to $\mathbf{w} = \widehat{\tau}_k^{-1}(\widetilde{\mathbf{y}}_k - \mathbf{y}) + \mathbf{y} \in \mathcal{N}$. Then, $\mathbf{w}' = \widetilde{\mathbf{y}}_k$. By (98),

$$\|\widetilde{\mathbf{y}}_k - \mathbf{y}\|_2 > \|\mathbf{w} - \mathbf{y}\|_2 \geq \epsilon \quad \text{for all } k \geq K. \tag{99}$$

But (99) contradicts $\widetilde{\mathbf{y}}_k \to \mathbf{y}$ as $k \to \infty$. Therefore, $\mathtt{conv}\{\mathbf{d} + \mathbf{y}, \mathbf{y}\} \cap \mathcal{B}^\circ = \emptyset$.

Translating by $-\mathbf{y}$, $\mathtt{conv}\{\mathbf{d}, 0\} \cap (\mathcal{B} - \mathbf{y})^\circ = \emptyset$. By this and convexity, it follows there exists a properly separating hyperplane:

$$\exists\, \mathbf{v} \in \mathbb{R}^D \setminus \{0\}, \exists\, \gamma \in \mathbb{R} \text{ such that } \forall\, \mathbf{u}_1 \in \mathtt{conv}\{\mathbf{d}, 0\}, \forall\, \mathbf{u}_2 \in \mathcal{B} - \mathbf{y}$$
$$\langle \mathbf{v}, \mathbf{u}_1 \rangle \geq \gamma, \langle \mathbf{v}, \mathbf{u}_2 \rangle \leq \gamma \text{ and } \exists\, \widetilde{\mathbf{u}}_2 \in \mathcal{B} - \mathbf{y} \text{ such that } \langle \mathbf{v}, \widetilde{\mathbf{u}}_2 \rangle < \gamma.$$

In particular,

$$\mathbb{R}_{>0}(\mathcal{B} - \mathbf{y}) \subseteq \{\mathbf{u} \in \mathbb{R}^D : \langle \mathbf{v}, \mathbf{u} \rangle \leq \gamma\}.$$

Also, for any open neighborhood $\mathcal{D} \subseteq \mathbb{R}^D$ with $\mathbf{d} \in \mathcal{D}$,

$$\exists\, \widetilde{\mathbf{d}} \in \mathcal{D} \text{ such that } \langle \mathbf{v}, \widetilde{\mathbf{d}} \rangle > \langle \mathbf{v}, \mathbf{d} \rangle \geq \gamma.$$

We conclude $\mathbf{d} \notin (\mathbb{R}_{>0}(\mathcal{B} - \mathbf{y}))^\circ$, as desired. This gives $\mathbf{d} \in \partial(TC_\mathbf{y}(\mathcal{B}))$, and LHS $\supseteq$ RHS in Eq. (14), completing the proof of the lemma. $\qquad\square$

B Proof of Proposition 5

For item 1, we first note that $\mathrm{grad}\,\|\cdot\|_\mathcal{B}(\widehat{\mathbf{a}})$ is nonzero, since the directional derivative of the norm function at $\widehat{\mathbf{a}}$ in the direction of $\widehat{\mathbf{a}}$ is nonzero. Indeed the function $\mathbb{R} \to \mathbb{R}; \lambda \mapsto \|\widehat{\mathbf{a}} + \lambda\widehat{\mathbf{a}}\|_\mathcal{B}$ has derivative $\|\widehat{\mathbf{a}}\|_\mathcal{B} = 1$ at $\lambda = 0$, using homogeneity of $\|\cdot\|_\mathcal{B}$ under positive scaling. Item 1 now follows immediately from [53, Thm. 3.15] and the preceding paragraph in that reference that metric regularity is implied by the linear independence of the gradients.

For item 2, we note that due to homogeneity of the norm, since $\|\cdot\|_\mathcal{B}$ is C^1 around $L_\mathbf{p}(\widehat{\mathbf{s}})$, it is also C^1 around $L_\mathbf{p}(\widehat{\mathbf{s}})/\|L_\mathbf{p}(\widehat{\mathbf{s}})\|_\mathcal{B}$ and it holds

$$\mathrm{grad}\,\|\cdot\|_\mathcal{B}(L_\mathbf{p}(\widehat{\mathbf{s}})/\|L_\mathbf{p}(\widehat{\mathbf{s}})\|_\mathcal{B}) = (1/\|L_\mathbf{p}(\widehat{\mathbf{s}})\|_\mathcal{B})\,\mathrm{grad}\,\|\cdot\|_\mathcal{B}(L_\mathbf{p}(\widehat{\mathbf{s}})).$$

Thus, item 1 applies and implies the tangent cone in right-hand side of Eq. (15) is the hyperplane normal to $L_\mathbf{p}(\widehat{\mathbf{s}})$. Now we finish by equating the inner product of $\mathrm{grad}\,\|\cdot\|_\mathcal{B}(L_\mathbf{p}(\widehat{\mathbf{s}}))$ and the LHS of Eq. (15) with 0, and solving for η. $\qquad\square$

C Proof of Lemma 8

Given $\mathcal{M}$ and $\mathcal{B}$, we need to show that there exists a positive constant C (independent of $\mathbf{p}, \xi$) such that for all $\mathbf{p} \in \mathcal{M}$ and all vectors $\xi \in T_{\mathbf{p}}\mathcal{M}$ we have

$$\left\langle \xi\xi^{\top}, \tfrac{1}{2} \int_{\{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}} \leq 1\}} \mathbf{s}\mathbf{s}^{\top} d\mathbf{s} \right\rangle \ \geq \ C\, \|\xi\|_2^2. \tag{100}$$

To this end, use linearity of the integral to rewrite the left-hand side of (100) as

$$\frac{1}{2} \int_{\{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}} \leq 1\}} \langle \xi\xi^{\top}, \mathbf{s}\mathbf{s}^{\top} \rangle \, d\mathbf{s} \ = \ \frac{1}{2} \int_{\{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}} \leq 1\}} \langle \xi, \mathbf{s} \rangle^2 \, d\mathbf{s}. \tag{101}$$

By the equivalence of norms on $\mathbb{R}^D$, there exists a positive constant c such that for all $\mathbf{v} \in \mathbb{R}^D$ we have that $\|\mathbf{v}\|_2 \leq c$ implies $\|\mathbf{v}\|_{\mathcal{B}} \leq 1$. In particular, the domain of integration in Eq. (101) is inner-approximated by

$$\{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \|L_{\mathbf{p}}(\mathbf{s})\|_2 \leq c\} \ \subseteq \ \{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}} \leq 1\}.$$

Since the integrand in (101) is non-negative, it follows

$$\frac{1}{2} \int_{\{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}} \leq 1\}} \langle \xi, \mathbf{s} \rangle^2 \, d\mathbf{s} \ \geq \ \frac{1}{2} \int_{\{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \|L_{\mathbf{p}}(\mathbf{s})\|_2 \leq c\}} \langle \xi, \mathbf{s} \rangle^2 \, d\mathbf{s}.$$

Since $L_{\mathbf{p}}$ is an isometry, the right-hand side is

$$\frac{1}{2} \int_{\{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \|\mathbf{s}\|_2 \leq c\}} \langle \xi, \mathbf{s} \rangle^2 \, d\mathbf{s}.$$

Using rotational symmetry of Euclidean balls, this equals

$$\left(\frac{1}{2} \int_{\{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \|\mathbf{s}\|_2 \leq c\}} s_1^2 \, d\mathbf{s} \right) \|\xi\|_2^2, \tag{102}$$

where s_1 denotes the first coordinate of $\mathbf{s}$ with respect to the fixed orthonormal basis on $T_{\mathbf{p}}\mathcal{M}$ (Section 3.1). Now note the parenthesized quantity in (102) is a positive constant C depending only on c and the manifold dimension d. By what we have said, it satisfies the bound (100) as desired. □

D Proof of Proposition 9

1. Denote the function (37) by $F : \mathcal{M} \to \mathrm{Sym}^2(T\mathcal{M})$. Let $(\mathbf{p}_k)_{k=1}^{\infty} \subseteq \mathcal{M}$ be a sequence converging to $\mathbf{p} \in \mathcal{M}$. To move to one fixed space we identify tangent spaces using the Levi-Civita connection on $\mathcal{M}$. After choosing a smooth path $\gamma : [0,1] \to \mathcal{M}$ such that $\gamma(\frac{1}{k}) = \mathbf{p}_k$ for each $k \geq 1$ and $\gamma(0) = 1$, the Levi-Civita connection gives isometries $\tau_k : T_{\mathbf{p}}\mathcal{M} \to T_{\mathbf{p}_k}\mathcal{M}$. Furthermore, τ_k converges to the identity map on $T_{\mathbf{p}}\mathcal{M}$ as elements of $(T\mathcal{M})^* \otimes T\mathcal{M}$ as $k \to \infty$.

   We want to show $F(\mathbf{p}_k) \to F(\mathbf{p})$ in $\mathrm{Sym}^2(T\mathcal{M})$. It suffices to show $(\tau_k^{-1} \otimes \tau_k^{-1})(F(\mathbf{p}_k)) \to F(\mathbf{p})$ in $\mathrm{Sym}^2(T_{\mathbf{p}}\mathcal{M})$ (last sentence of the previous paragraph). Changing variables $\mathbf{s} \leftarrow \tau_k^{-1}(\mathbf{s})$ and using that τ_k is an isometry, we have

$$(\tau_k^{-1} \otimes \tau_k^{-1})(F(\mathbf{p}_k)) = \frac{1}{2} \int_{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M} : \|L_{\mathbf{p}_k}(\tau_k(\mathbf{s}))\|_{\mathcal{B}} \leq 1} \mathbf{s}\mathbf{s}^{\top} d\mathbf{s}.$$

Write this as

$$\int_{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M}} \mathbb{1}(\|L_{\mathbf{p}_k}(\tau_k(\mathbf{s}))\|_{\mathcal{B}} < 1)\, \mathbf{s}\mathbf{s}^{\top} d\mathbf{s} \in \mathrm{Sym}^2(T_{\mathbf{p}}\mathcal{M}).$$

Compare this to

$$F(\mathbf{p}) = \int_{\mathbf{s} \in T_{\mathbf{p}}\mathcal{M}} \mathbb{1}(\|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}} < 1)\mathbf{s} \in \mathrm{Sym}^2(T_{\mathbf{p}}\mathcal{M}).\mathbf{s}^{\top} d\mathbf{s}$$

Since $L_{\mathbf{p}_k} \to L_{\mathbf{p}}$, $\tau_k \to \mathrm{Id}_{T_{\mathbf{p}}\mathcal{M}}$ and $\|\cdot\|_{\mathcal{B}}$ is continuous on $\mathbb{R}^D$, for each $\mathbf{s} \in T_{\mathbf{p}}\mathcal{M}$ there is the pointwise convergence:

$$\mathbb{1}(\|L_{\mathbf{p}_k}(\tau_k(\mathbf{s}))\|_{\mathcal{B}} < 1)\, \mathbf{s}\mathbf{s}^{\top} \longrightarrow \mathbb{1}(\|L_{\mathbf{p}}(\mathbf{s})\|_{\mathcal{B}} < 1)\mathbf{s}\mathbf{s}^{\top}$$

Also, letting $c \in \mathbb{R}_{>0}$ be a constant such that $\|\mathbf{u}\|_2 \le c\|\mathbf{u}\|_{\mathcal{B}}$ for all $\mathbf{u} \in \mathbb{R}^D$, we have the uniform bound:

$$\|\mathbb{1}(\|L_{\mathbf{p}_k}(\tau_k(\mathbf{s}))\|_{\mathcal{B}} < 1)\, \mathbf{s}\mathbf{s}^{\top}\|_F \le c^2 \quad \text{for all } \mathbf{s} \in T_{\mathbf{p}}\mathcal{M} \text{ and } k \ge 1,$$

since $L_{\mathbf{p}}$ and τ_k are both isometries. Hence, the bounded convergence theorem is applicable, and implies $(\tau_k \otimes \tau_k)(F(\mathbf{p}_k)) \to F(\mathbf{p})$.

2. Denote the function (38) by $G : \mathcal{M} \to \mathrm{Sym}(T\mathcal{M})$. Let $\mathbf{p}$ be a point satisfying the stated assumption. There exists an open neighborhood $\mathcal{U}$ of $\mathbf{p}$ in $\mathcal{M}$ such that for each $\mathbf{p}_* \in \mathcal{U}$ the norm $\|\cdot\|_{\mathcal{B}}$ is C^1 in some neighborhood of $L_{\mathbf{p}_*}(T_{\mathbf{p}_*}\mathcal{M}) \cap \mathbb{S}^{D-1}$. Let $(\mathbf{p}_k)_{k=1}^{\infty} \subseteq \mathcal{U}$ be a sequence converging to $\mathbf{p}$. Identifying tangent spaces as above, it suffices to show $\tau_k^{-1}(F(\mathbf{p}_k)) \to F(\mathbf{p})$.
By the local C^1 assumption, Proposition 5, item 2 applies and gives

$$F(\mathbf{p}) = \int_{\widehat{\mathbf{s}} \in T_{\mathbf{p}}\mathcal{M}:\|\widehat{\mathbf{s}}\|_2=1} -\widehat{\mathbf{s}}\|L_{\mathbf{p}}(\widehat{\mathbf{s}})\|_{\mathcal{B}}^{-d-2} \frac{\left\langle \mathrm{grad}\, \|\cdot\|_{\mathcal{B}}(L_{\mathbf{p}}(\widehat{\mathbf{s}})),\, \frac{1}{2}Q_{\mathbf{p}}(\widehat{\mathbf{s}}) \right\rangle}{\left\langle \mathrm{grad}\, \|\cdot\|_{\mathcal{B}}(L_{\mathbf{p}}(\widehat{\mathbf{s}})),\, L_{\mathbf{p}}(\widehat{\mathbf{s}}) \right\rangle}\, d\widehat{\mathbf{s}}.$$

Likewise, by a change of variables using that τ_k^{-1} preserves unit spheres:

$$\tau_k^{-1}(F(\mathbf{p}_k)) =$$

$$\int_{\widehat{\mathbf{s}} \in T_{\mathbf{p}}\mathcal{M}:\|\widehat{\mathbf{s}}\|_2=1} -\widehat{\mathbf{s}}\|L_{\mathbf{p}_k}(\tau_k(\widehat{\mathbf{s}}))\|_{\mathcal{B}}^{-d-2} \frac{\left\langle \mathrm{grad}\, \|\cdot\|_{\mathcal{B}}(L_{\mathbf{p}_k}(\tau_k(\widehat{\mathbf{s}}))),\, \frac{1}{2}Q_{\mathbf{p}_k}(\tau_k(\widehat{\mathbf{s}})) \right\rangle}{\left\langle \mathrm{grad}\, \|\cdot\|_{\mathcal{B}}(L_{\mathbf{p}_k}(\tau_k(\widehat{\mathbf{s}}))),\, L_{\mathbf{p}_k}(\tau_k(\widehat{\mathbf{s}})) \right\rangle}\, d\widehat{\mathbf{s}}.$$

Boundedness and pointwise convergence hold since $\mathrm{grad}\, \|\cdot\|_{\mathcal{B}}$ is locally continuous. So the bounded convergence theorem implies the first sentence in the statement. The second sentence follows from the example in Section 3.7. □

E Tail bounds and absolute moments of the Gaussian

We recall some basic properties of the Gaussian. As in Section 4.2, $\kappa_{\sigma}(s) := \frac{2s}{\sigma^2}e^{-s^2/\sigma^2}$.

− For each even $k \ge 0$ and $\delta \ge 0$, by substitution and then integration by parts $k/2$ times,

$$\int_{s=\delta}^{\infty} s^k \kappa_{\sigma}(s)ds$$

$$= \sigma^k e^{-\delta^2/\sigma^2}\left(\left(\frac{\delta^2}{\sigma^2}\right)^{\frac{k}{2}} + \frac{k}{2}\left(\frac{\delta^2}{\sigma^2}\right)^{\frac{k}{2}-1} + \frac{k}{2}\left(\frac{k}{2}-1\right)\left(\frac{\delta^2}{\sigma^2}\right)^{\frac{k}{2}-2} + \ldots + \left(\frac{k}{2}\right)!\right)$$

$$= e^{-\delta^2/\sigma^2}\mathrm{poly}(\sigma, \delta). \tag{103}$$

– For each odd $k \geq 0$ and $\delta > 0$, using $s/\delta \geq 1$ for $s \in [\delta, \infty]$ and Eq. (103),

$$\int_{s=\delta}^{\infty} s^k \kappa_\sigma(s) ds \leq (1/\delta) \int_{s=\delta}^{\infty} s^{k+1} \kappa_\sigma(s) ds = e^{-\delta^2/\sigma^2} (1/\delta) \operatorname{poly}(\sigma, \delta). \qquad (104)$$

– For each $k \geq 0$, from [67, Equation 18],

$$\int_{s=0}^{\infty} s^k \kappa_\sigma(s) ds = \sigma^k \Gamma \left(\frac{k+2}{2} \right). \qquad (105)$$

F Numerical estimation of one-dimensional eigenfunctions

The eigenfunctions of the limiting operator are of key interest for manifold learning methods in general. For the case of the circle example (Section 3.7), these are the functions $\varphi : [0, 2\pi] \to \mathbb{R}$ that solve the following generalized Helmholtz boundary value problem:

$$\Delta_{\mathcal{M},\mathcal{B}} \varphi + \lambda \varphi = 0, \qquad (106)$$

where $\Delta_{\mathcal{M},\mathcal{B}}$ is the limiting Laplacian-like differential operator in Eq. (42), subject to the periodic boundary conditions:

$$\varphi(\theta + 2\pi) = \varphi(\theta),$$
$$\varphi'(\theta + 2\pi) = \varphi'(\theta).$$

Figure 8 shows numerically computed solutions of Eq. (106) for different choices of w_1, w_2. Notice the eigenfunctions are oscillatory, as dictated by Sturm-Liouville theory [2].

We describe the numerical computation of these limiting eigenfunctions. We used a standard finite-difference scheme where the first derivative was replaced by a symmetrized difference

$$\frac{df}{d\theta} \to \frac{f(\theta + \Delta\theta) - f(\theta - \Delta\theta)}{2\Delta\theta}, \qquad (107)$$

and the second derivative by

$$\frac{d^2 f}{d\theta^2} \to \frac{f(\theta + \Delta\theta) - 2f(\theta) + f(\theta - \Delta\theta)}{(\Delta\theta)^2}. \qquad (108)$$

In this equation, f is taken to be a cyclic function defined over the discrete range

$$\left\{ 0, \frac{2\pi}{n}, \ldots, \frac{2\pi(n-1)}{n} \right\}.$$

To compute the solutions we formed a sparse $n \times n$ matrix L that corresponds to the finite-difference operator formed by substituting (107) and (108) into the first and second derivative terms in Eq. (106). The eigenvalues and eigenvectors of L were found using the function `scipy.sparse.linalg.eigs()` from the `SciPy` package [65]. It is a wrapper of the ARPACK library for large-scale eigenvalue problems [34]. Recall that in our problem, all eigenvalues are non-positive. To obtain the smallest (in magnitude) eigenvalues and their corresponding eigenvectors, we used the `eigs()` function in shift-invert mode with $\sigma = 1$. The particular choice of σ did not seem to matter much when $\sigma > 0$, however choosing $\sigma = 0$ resulted in instabilities and convergence errors. This is due to the fact that shift-invert mode attempts to find solutions to $(L - \sigma I)^{-1} \mathbf{x} = \lambda \mathbf{x}$, and since zero is an eigenvalue of L, the choice $\sigma = 0$ results in the inversion of an ill-conditioned matrix. The use of sparse matrices allows one to take large values of n, since applying the finite-difference operator defined above costs only $O(n)$.
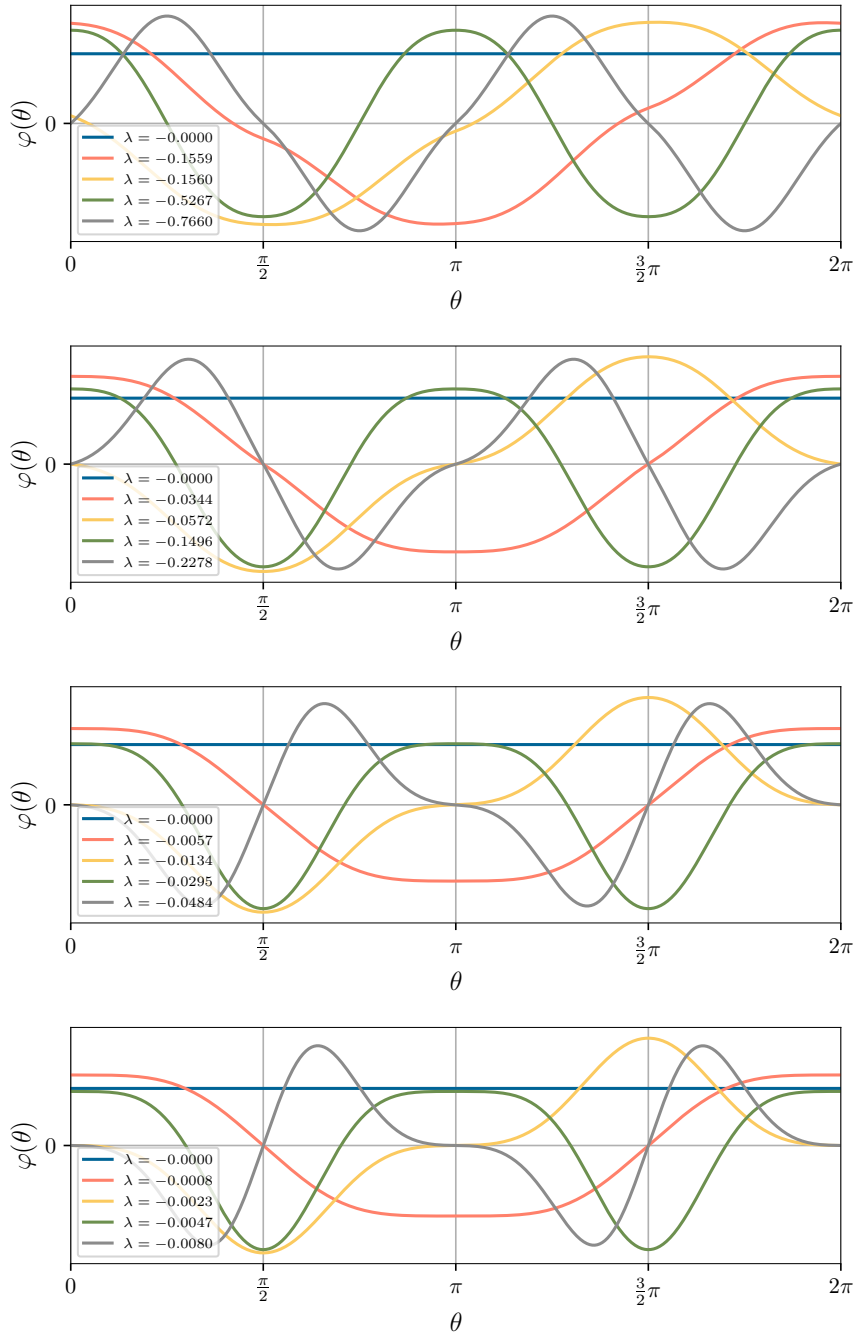
Fig. 8 *Eigenfunctions.* The five eigenfunctions with eigenvalues smallest in magnitude for the weighted ℓ_1 Laplacian on the unit circle (42). These were computed numerically. In these plots, $w_2 = 1$. All the choices $w_1 \in \{1, 2, 4, 8\}$ are displayed from top to bottom.

References

1. Rose et al., P.: The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic acids research* **45**(D1), D271–D281 (2017). doi:10.1093/nar/gkw1000

2. Al-Gwaiz, M.: Sturm-Liouville theory and its applications. Springer London, London (2008). doi:10.1007/978-1-84628-972-9

3. Bates, J.: The embedding dimension of Laplacian eigenfunction maps. *Applied and Computational Harmonic Analysis* **37**(3), 516–530 (2014). doi:10.1016/j.acha.2014.03.002

4. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* **15**(6), 1373–1396 (2003). doi:10.1162/089976603321780317

5. Belkin, M., Niyogi, P.: Semi-supervised learning on Riemannian manifolds. *Machine Learning* **56**(1-3), 209–239 (2004). doi:10.1023/B:MACH.0000033120.25363.1e

6. Belkin, M., Niyogi, P.: Convergence of Laplacian Eigenmaps. *Neural Information Processing Systems (NIPS)* (2007). doi:10.7551/mitpress/7503.003.0021

7. Belkin, M., Niyogi, P.: Towards a theoretical foundation for Laplacian-based manifold methods. *Journal of Computer and System Sciences* **74**(8), 1289–1308 (2008). doi:10.1016/j.jcss.2007.08.006

8. Bellet, A., Habrard, A., Sebban, M.: Metric Learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* **9**(1), 1–151 (2015). doi:10.2200/S00626ED1V01Y201501AIM030

9. Bendory, T., Bartesaghi, A., Singer, A.: Single-Particle Cryo-Electron Microscopy: Mathematical Theory, Computational Challenges, and Opportunities. *IEEE Signal Processing Magazine* **37**(2), 58–76 (2020). doi:10.1109/MSP.2019.2957822

10. Cheng, M.Y., Wu, H.T.: Local linear regression on manifolds and its geometric interpretation. *Journal of the American Statistical Association* **108**(504), 1421–1434 (2013). doi:10.1080/01621459.2013.827984

11. Coifman, R.R., Lafon, S.: Diffusion maps. *Applied and Computational Harmonic Analysis* **21**(1), 5–30 (2006). doi:10.1016/j.acha.2006.04.006

12. Coifman, R.R., Lafon, S., Lee, A.B., Maggioni, M., Nadler, B., Warner, F., Zucker, S.W.: 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). doi:10.1073/pnas.0500334102

13. Coifman, R.R., Leeb, W.: Earth mover's distance and equivalent metrics for spaces with hierarchical partition trees. Tech. rep., Yale University (2013)

14. Dashti, A., et al.: Retrieving functional pathways of biomolecules from single-particle snapshots. *Nature Communications* **11**(1), 4734 (2020). doi:10.1038/s41467-020-18403-x

15. Dashti, A., et al.: Trajectories of the ribosome as a Brownian nanomachine. *Proceedings of the National Academy of Sciences* **111**(49), 17492–17497 (2014). doi:10.1073/pnas.1419276111

16. Donoho, D.L., Grimes, C.: Hessian eigenmaps: locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences* **100**(10), 5591–5596 (2003). doi:10.1073/pnas.1031596100

17. Frank, J.: New Opportunities Created by Single-Particle Cryo-EM: The Mapping of Conformational Space. *Biochemistry* **57**(6), 888–888 (2018). doi:10.1021/acs.biochem.8b00064

18. Frank, J., Ourmazd, A.: Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods* **100**, 61–67 (2016). doi:10.1016/j.ymeth.2016.02.007

19. García Trillos, N., Gerlach, M., Hein, M., Slepcev, D.: Error Estimates for Spectral Convergence of the Graph Laplacian on Random Geometric Graphs Toward the Laplace–Beltrami Operator. *Foundations of Computational Mathematics* **20**(4), 827–887 (2020). doi:10.1007/s10208-019-09436-w

20. García Trillos, N., Slepcev, D.: A variational approach to the consistency of spectral clustering. *Applied and Computational Harmonic Analysis* **45**(2), 239–281 (2018). doi:10.1016/j.acha.2016.09.003

21. Gavish, M., Nadler, B., Coifman, R.R.: Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning. *International Conference on Machine Learning (ICML)* (2010)
22. Giné, E., Koltchinskii, V.: Empirical graph Laplacian approximation of Laplace–Beltrami operators: Large sample results. *High Dimensional Probability, vol. 51*, pp. 238–259. Institute of Mathematical Statistics, Beachwood, Ohio, USA (2006). doi:10.1214/074921706000000888
23. Glaeser, R.M., Nogales, E., Chiu, W. (eds.): Single-particle Cryo-EM of Biological Macromolecules. IOP Publishing (2021). doi:10.1088/978-0-7503-3039-8
24. Goldberg, A.B., Zhu, X., Singh, A., Xu, Z., Nowak, R.: Multi-Manifold Semi-Supervised Learning. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 169–176 (2009)
25. Hein, M., Audibert, J.Y., von Luxburg, U.: From graphs to manifolds – weak and strong pointwise consistency of graph Laplacians. *International Conference on Computational Learning Theory (COLT)*, pp. 470–485 (2005). doi:10.1007/11503415_32
26. Hein, M., Audibert, J.Y., von Luxburg, U.: Graph Laplacians and their Convergence on Random Neighborhood Graphs. *Journal of Machine Learning Research* **8**, 1325—-1368 (2007)
27. Hug, D., Weil, W.: Lectures on convex geometry, *Graduate Texts in Mathematics*, vol. 286. Springer International Publishing, Cham (2020). doi:10.1007/978-3-030-50180-8
28. Jin, Q., et al.: Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes. *Structure* **22**(3), 496–506 (2014). doi:10.1016/j.str.2014.01.004
29. Lederman, R.R., Andén, J., Singer, A.: Hyper-molecules: on the representation and recovery of dynamical structures for applications in flexible macro-molecules in cryo-EM. *Inverse Problems* **36**(4), 044005 (2020). doi:10.1088/1361-6420/ab5ede
30. Lee, A.B., Izbicki, R.: A spectral series approach to high-dimensional nonparametric regression. *Electronic Journal of Statistics* **10**(1), 423–463 (2016). doi:10.1214/16-EJS1112
31. Lee, G., Gommers, R., Waselewski, F., Wohlfahrt, K., O'Leary, A.: PyWavelets: a Python package for wavelet analysis. *Journal of Open Source Software* **4**(36), 1237 (2019). doi:10.21105/joss.01237
32. Lee, J.M.: Riemannian manifolds, *Graduate Texts in Mathematics*, vol. 176. Springer New York, New York, NY (1997). doi:10.1007/b98852
33. Lee, J.M.: Introduction to smooth manifolds, *Graduate Texts in Mathematics*, vol. 218. Springer New York, New York, NY (2012). doi:10.1007/978-1-4419-9982-5
34. Lehoucq, R.B., Sorensen, D.C., Yang, C.: ARPACK users' guide. Society for Industrial and Applied Mathematics (1998). doi:10.1137/1.9780898719628
35. Liao, W., Maggioni, M., Vigogna, S.: Learning adaptive multiscale approximations to data and functions near low-dimensional sets. *2016 IEEE Information Theory Workshop (ITW)*, pp. 226–230. IEEE (2016). doi:10.1109/ITW.2016.7606829
36. Lieu, L., Saito, N.: Signal ensemble classification using low-dimensional embeddings and earth mover's distance. *Wavelets and Multiscale Analysis*, 9780817680947, pp. 227–256. Birkhäuser Boston, Boston (2011). doi:10.1007/978-0-8176-8095-4_11
37. von Luxburg, U., Belkin, M., Bousquet, O.: Consistency of spectral clustering. *The Annals of Statistics* **36**(2), 555–586 (2008). doi:10.1214/009053607000000640
38. van der Maaten, L., Hinton, G.: Visualizing Data using t-SNE Laurens. *Journal of Machine Learning Research* **9**(86), 2579–2605 (2008)
39. Mallat, S.: A wavelet tour of signal processing, 3rd edn. Elsevier (2009). doi:10.1016/B978-0-12-374370-1.X0001-8
40. McInnes, L., Healy, J., Saul, N., Großberger, L.: UMAP: uniform manifold approximation and projection. *Journal of Open Source Software* **3**(29), 861 (2018). doi:10.21105/joss.00861
41. Mishne, G., Talmon, R., Cohen, I., Coifman, R.R., Kluger, Y.: Data-driven tree transforms and metrics. *IEEE Transactions on Signal and Information Processing over Networks* **4**(3), 451–466 (2018). doi:10.1109/TSIPN.2017.2743561
42. Mishne, G., Talmon, R., Meir, R., Schiller, J., Lavzin, M., Dubin, U., Coifman, R.R.: Hierarchical coupled-geometry analysis for neuronal structure and activity pattern discovery. *IEEE Journal of Selected Topics in Signal Processing* **10**(7), 1238–1253 (2016). doi:10.1109/JSTSP.2016.2602061

43. Monera, M.G., Montesinos-Amilibia, A., Sanabria-Codesal, E.: The Taylor expansion of the exponential map and geometric applications. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales - Serie A: Matematicas* **108**(2), 881–906 (2014). doi:10.1007/s13398-013-0149-z

44. Moscovich, A., Halevi, A., Andén, J., Singer, A.: Cryo-EM reconstruction of continuous heterogeneity by Laplacian spectral volumes. *Inverse Problems* **36**(2), 024003 (2020). doi:10.1088/1361-6420/ab4f55

45. Moscovich, A., Jaffe, A., Nadler, B.: Minimax-optimal semi-supervised regression on unknown manifolds. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 933—-942. PMLR (2017)

46. Nadler, B., Lafon, S., Coifman, R.R., Kevrekidis, I.G.: Diffusion maps, spectral clustering and eigenfunctions of Fokker–Planck operators. *Neural Information Processing Systems (NIPS)*, pp. 955–962 (2005)

47. Nakane, T., Kimanius, D., Lindahl, E., Scheres, S.H.: Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, 1–18 (2018). doi:10.7554/eLife.36861

48. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E.: UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**(13), 1605–1612 (2004). doi:10.1002/jcc.20084

49. Punjani, A., Fleet, D.J.: 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of Structural Biology* **213**(2), 107702 (2021). doi:10.1016/j.jsb.2021.107702

50. Rao, R., Moscovich, A., Singer, A.: Wasserstein K-Means for Clustering Tomographic Projections. *Machine Learning for Structural Biology Workshop, NeurIPS* (2020)

51. Rosasco, L., Belkin, M., De Vito, E.: On learning with integral operators. *Journal of Machine Learning Research* **11**, 905–934 (2010)

52. Roweis, S.T., Saul, L.K.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* **290**(5500), 2323–2326 (2000). doi:10.1126/science.290.5500.2323

53. Ruszczyński, A.: Nonlinear optimization. Princeton University Press (2011). doi:10.2307/j.ctvcm4hcj

54. Sathyanarayanan, N., Cannone, G., Gakhar, L., Katagihallimath, N., Sowdhamini, R., Ramaswamy, S., Vinothkumar, K.R.: Molecular basis for metabolite channeling in a ring opening enzyme of the phenylacetate degradation pathway. *Nature Communications* **10**(1), 4127 (2019). doi:10.1038/s41467-019-11931-1

55. Schwander, P., Fung, R., Ourmazd, A.: Conformations of macromolecules and their complexes from heterogeneous datasets. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**(1647), 1–8 (2014). doi:10.1098/rstb.2013.0567

56. Shirdhonkar, S., Jacobs, D.W.: Approximate earth mover's distance in linear time. *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8. IEEE (2008). doi:10.1109/CVPR.2008.4587662

57. Singer, A.: From graph to manifold Laplacian: the convergence rate. *Applied and Computational Harmonic Analysis* **21**(1), 128–134 (2006). doi:10.1016/j.acha.2006.03.004

58. Singer, A., Sigworth, F.J.: Computational methods for single-particle electron cryomicroscopy. *Annual Review of Biomedical Data Science* **3**(1), 163–190 (2020). doi:10.1146/annurev-biodatasci-021020-093826

59. Sober, B., Aizenbud, Y., Levin, D.: Approximation of functions over manifolds: A Moving Least-Squares approach. *Journal of Computational and Applied Mathematics* **383**, 113140 (2021). doi:10.1016/j.cam.2020.113140

60. Sorzano, C.O.S., et al.: Survey of the analysis of continuous conformational variability of biological macromolecules by electron microscopy. *Acta Crystallographica Section F Structural Biology Communications* **75**(1), 19–32 (2019). doi:10.1107/S2053230X18015108

61. Stock, D., Leslie, A., Walker, J.: Molecular architecture of the rotary motor in ATP synthase. *Science* **286**(5445), 1700–1705 (1999). doi:10.1126/science.286.5445.1700

62. Tagare, H.D., Kucukelbir, A., Sigworth, F.J., Wang, H., Rao, M.: Directly reconstructing principal components of heterogeneous particles from cryo-EM images. *Journal of Structural Biology* **191**(2), 245–262 (2015). doi:10.1016/j.jsb.2015.05.007

63. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* **290**(5500), 2319–2323 (2000). doi:10.1126/science.290.5500.2319

64. Ting, D., Huang, L., Jordan, M.: An analysis of the convergence of graph Laplacians. *International Conference on Machine Learning (ICML)* (2010)

65. Virtanen, P., et al.: SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**(3), 261–272 (2020). doi:10.1038/s41592-019-0686-2

66. Von Luxburg, U.: A tutorial on spectral clustering. *Statistics and Computing* **17**(4), 395–416 (2007). doi:10.1007/s11222-007-9033-z

67. Winkelbauer, A.: Moments and absolute moments of the normal distribution. *arXiv preprint arXiv:1209.4340v2* pp. 1–4 (2012)

68. Wormell, C.L., Reich, S.: Spectral Convergence of Diffusion Maps: Improved Error Bounds and an Alternative Normalization. *SIAM Journal on Numerical Analysis* **59**(3), 1687–1734 (2021). doi:10.1137/20M1344093

69. Yoshida, M., Muneyuki, E., Hisabori, T.: ATP synthase – a marvellous rotary engine of the cell. *Nature Reviews Molecular Cell Biology* **2**(9), 669–677 (2001). doi:10.1038/35089509

70. Zelesko, N., Moscovich, A., Kileel, J., Singer, A.: Earthmover-based manifold learning for analyzing molecular conformation spaces. *IEEE International Symposium on Biomedical Imaging (ISBI)* (2020). doi:10.1109/ISBI45749.2020.9098723

71. Zhang, S., Moscovich, A., Singer, A.: Product Manifold Learning. *International Conference on Artificial Intelligence and Statistics (AISTATS)* (2021)

72. Zhang, Z., Zha, H.: Principal Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment. *SIAM Journal on Scientific Computing* **26**(1), 313–338 (2004). doi:10.1137/S1064827502419154

73. Zhong, E.D., Bepler, T., Berger, B., Davis, J.H.: CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods* **18**(2), 176–185 (2021). doi:10.1038/s41592-020-01049-4