Applied and Computational Harmonic Analysis 68 (2024) 101593

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

journal homepage: www.elsevier.com/locate/acha

Diffusion maps for embedded manifolds with boundary with applications to PDEs

Ryan Vaughn [a,*], Tyrus Berry [b], Harbir Antil [b]

[a] Courant Institute for the Mathematical Sciences, New York University, New York, NY 10012, USA
[b] Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030, USA

ARTICLE INFO

Communicated by Hau-Tieng Wu

Keywords:
Mesh-free
Partial differential equations
Heat equation
Riemannian manifolds
Boundary detection

ABSTRACT

Given only a finite collection of points sampled from a Riemannian manifold embedded in a Euclidean space, in this paper we propose a new method to numerically solve elliptic and parabolic partial differential equations (PDEs) supplemented with boundary conditions. Since the construction of triangulations on unknown manifolds can be both difficult and expensive, both in terms of computational and data requirements, our goal is to solve these problems without a triangulation. Instead, we rely only on using the sample points to define quadrature formulas on the unknown manifold. Our main tool is the diffusion maps algorithm. We re-analyze this well-known method in a variational sense for manifolds with boundary. Our main result is that the variational diffusion maps graph Laplacian is a consistent estimator of the Dirichlet energy on the manifold. This improves upon previous results and provides a rigorous justification of the well-known relationship between diffusion maps and the Neumann eigenvalue problem. Moreover, using semigeodesic coordinates we derive the first uniform asymptotic expansion of the diffusion maps kernel integral operator for manifolds with boundary. This expansion relies on a novel lemma which relates the extrinsic Euclidean distance to the coordinate norm in a normal collar of the boundary. We then use a recently developed method of estimating the distance to boundary function (notice that the boundary location is assumed to be unknown) to construct a consistent estimator for boundary integrals. Finally, by combining these various estimators, we illustrate how to impose Dirichlet and Neumann conditions for some common PDEs based on the Laplacian. Several numerical examples illustrate our theoretical findings.

1. Introduction

The goal of this paper is to analyze the diffusion maps algorithm in a weak (variational) form and to introduce a completely rigorous method to solve elliptic and parabolic partial differential equations (PDEs) with boundary conditions. These PDEs are posed on an m-dimensional Riemannian manifold $(\mathcal{M}, g)$ embedded in an ambient Euclidean space via $\iota : \mathcal{M} \to \mathbb{R}^d$. We assume that the Riemannian metric g on $\mathcal{M}$ is inherited from the ambient space $\mathbb{R}^d$ via the embedding.

Motivated by applications to machine learning or emergent structures in high-dimensional problems such as inertial manifolds, we will assume that we have no explicit description of the embedded Riemannian manifold. Instead, we assume only that we have

* Corresponding author.
E-mail addresses: rv2152@nyu.edu (R. Vaughn), tberry@gmu.edu (T. Berry), hantil@gmu.edu (H. Antil).

https://doi.org/10.1016/j.acha.2023.101593
Received 20 April 2022; Received in revised form 16 November 2022; Accepted 24 August 2023
Available online 9 September 2023
1063-5203/© 2023 Elsevier Inc. All rights reserved.

a collection of sample points, $\{x_i\}_{i=1}^N \subset \imath(\mathcal{M}) \subset \mathbb{R}^d$ which, together with equal weights, form a consistent weighted quadrature rule. Namely, for any square integrable function $f \in L^2(\mathcal{M}, g)$, we assume that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(x_i) = \int_{\mathcal{M}} f(x) q(x)\, d\text{vol} \tag{1.1}$$

almost surely. In the statistical context the weight function q is called the sampling density. Ultimately our method will be independent of q, meaning that we do not require any specific density and if a grid of samples is used this grid is not required to be uniformly spaced. This is critical when the nodes are data points (which are typically not sampled from the density that corresponds to the volume form), but is also an advantage for synthetic data sets where creating uniform grids on manifolds can be challenging (uniform in this context means that (1.1) holds with a constant q). There are several situations where this may arise:

- Random data on an unknown manifold, where q is the sampling density.
- Attractors and inertial manifolds for dynamical systems, where q is the invariant measure.
- Known but complex domains that are difficult to mesh, and difficult to sample uniformly.
- Known but moderate dimensional manifolds if one cannot afford a mesh.

There is a wide literature starting with the Laplacian eigenmaps [3] and the diffusion maps [10] algorithms which give a method of approximating the intrinsic Laplacian operator on an unknown manifold. In this manuscript we let Δ be the negative definite Laplacian, also known as the Laplace-Beltrami operator. The basic strategy for estimating the Laplacian starts with a kernel function $K(\epsilon, x, y)$ which approximates the heat kernel on a manifold, for example $K(\epsilon, x, y) = e^{-\frac{|x-y|^2}{4\epsilon^2}}$, where we choose ϵ^2 in the denominator so that ϵ has units of distance. Then for any f we can estimate the integral operator

$$\mathcal{I} f(x) := \int_{y\in\mathcal{M}} K(\epsilon, x, y) f(y) q(y)\, d\text{vol} \tag{1.2}$$

by our quadrature formula

$$\mathcal{K} f(x) \equiv \frac{1}{N} \sum_{i=1}^{N} K(\epsilon, x, x_i) f(x_i) = \mathcal{I} f(x) + \text{Error}_{\text{Quad}}(N, f, q) \tag{1.3}$$

where the $\text{Error}_{\text{Quad}}$ term is assumed to go to zero as $N \to \infty$. In fact, when the data, x_i, are independent identically distributed random variables it can be shown that $\text{Error}_{\text{Quad}} = O(N^{-1/2})$ with high probability [29,4]. The qualifier 'with high probability' is required because the data set is random and there is a finite (but extremely small) probability of all the data points landing in for instance, a small ball on the manifold. This would clearly lead to a significant error in the quadrature formula in most cases. However, the probability of all such high-error events can be made arbitrarily small as $N^{-1/2}$ approaches zero [4,29]. Finally, notice that if we represent f by a vector $\vec{f}_i = f(x_i)$ then we can represent $\mathcal{K}$ with the matrix with entries $\mathbf{K}_{ij} = K(\epsilon, x_i, x_j)$ so that $(\mathbf{K}\vec{f})_i = \mathcal{K} f(x_i)$.

The intuition behind the kernel function is that the exponential decay localizes the integral to an ϵ-ball around x, and in this neighborhood the Euclidean distance $d_{\mathbb{R}^d}(x, y) = \|x - y\|_{\mathbb{R}^d}$ is close to the geodesic distance $d_g(x, y)$ (under appropriate assumptions on the manifold and embedding). Thus, as $\epsilon \to 0$ the integral $\mathcal{I} f(x)$ can be shown [3] to converge to the semigroup $e^{\epsilon^2\Delta}$ associated to the intrinsic (negative definite) Laplace-Beltrami operator Δ, so that

$$\mathcal{K} f(x) = m_0 \epsilon^m e^{\epsilon^2\Delta}(fq)(x) + O(\epsilon^{m+2}) + \text{Error}_{\text{Quad}}(N, f, q).$$

In fact, a more detailed asymptotic analysis [10,16] reveals that for any kernel function of the form

$$K(\epsilon, x, y) = k\left(\frac{\|x - y\|^2_{\mathbb{R}^d}}{\epsilon^2}\right)$$

where $k : [0, \infty) \to [0, \infty)$ is sufficiently regular and has exponential decay ($k(z) \le a^{-bz}$ for some $a, b > 0$), we have

$$\begin{aligned} \epsilon^{-m} \mathcal{K} f(x) = {} & m_0 f(x) q(x) + \epsilon^2 \frac{m_2}{2} \left(\omega(x) f(x) q(x) + \Delta(fq)(x)\right) \\ & + O(\epsilon^4) + \epsilon^{-m} \text{Error}_{\text{Quad}}(N, f, q), \end{aligned} \tag{1.4}$$

for all x with distance greater than ϵ from the boundary. The constants m_0 and m_2 are the *zeroth* and *second moments* of the chosen kernel functions, namely,

$$m_0 := \int_{z\in\mathbb{R}^m} k(|z|^2)\, dz \qquad \text{and} \qquad m_2 := \int_{z\in\mathbb{R}^m} z_i^2 k(|z|^2)\, dz \tag{1.5}$$

where the domain of integration is determined by the intrinsic dimension of the manifold, m, rather than the extrinsic dimension of the embedding space, d.

The expansion (1.4) is commonly used for estimating the density function [20,23,24] by applying the operator $\mathcal{K}$ to the constant function $f \equiv 1$ to find,

$$\frac{\epsilon^{-m}}{m_0}\mathcal{K}1(x) = q(x) + O(\epsilon^2) + \frac{\epsilon^{-m}}{m_0}\text{Error}_{\text{Quad}}(N, f, q).$$

However, it is well-known that for manifolds with boundary this does not hold. This leads to the well known bias of Kernel Density Estimators (KDEs) near the boundary [8,17–19,22,25]. In [6], the authors developed a method to estimate the distance to the boundary of a manifold from data. They then used this estimate to correct the bias of the KDE near the boundary. A significant advance of the method of [6] is that the location of the boundary does not need to be known beforehand, and it is effectively learned from the data.

A much more challenging and powerful use of (1.4) is for estimating the Laplace-Beltrami operator, Δ, and this expansion is the key component of justifying the diffusion maps algorithm [10]. The Laplace-Beltrami operator is ubiquitous, especially when a physical process is modeled using PDEs. For such applications, it is critical that one be able to specify the appropriate boundary conditions. Moreover, while it is widely observed that the diffusion maps algorithm produces Neumann eigenfunctions [10], this empirical observation has not been adequately explained. In this paper we will show that the estimator defined by the diffusion maps algorithm is consistent in the weak sense even for manifolds with boundary, and that Neumann eigenfunctions are observed because of the naturality of the Neumann boundary conditions for the eigenproblem. Finally, allowing arbitrary boundary conditions to be specified requires us to introduce a new tool, namely a *boundary integral estimator*, which may have uses beyond these applications, and the consistency of this estimator is one of our key results.

In order to solve diffusion type PDEs in the weak-sense and specify boundary conditions we need consistent discrete estimators of the following bilinear forms,

$$\mathcal{G}(\phi, f) = \int_{\mathcal{M}} \phi f \, d\text{vol}\,, \qquad \mathcal{E}(\phi, f) = \int_{\mathcal{M}} \nabla\phi \cdot \nabla f \, d\text{vol}\,, \qquad \mathcal{B}(\phi, f) = \int_{\partial\mathcal{M}} \phi f \, d\text{vol}_{\partial}\,, \tag{1.6}$$

where $\mathcal{G}$ and $\mathcal{E}$ correspond to mass and stiffness matrices respectively. The L^2 inner product on the boundary, $\mathcal{B}$, arises in case of Neumann or Robin boundary conditions [27]. The first bilinear form, $\mathcal{G}$, is simply the L^2 inner product on the manifold, which can be represented by a diagonal matrix with entries $D_{ii} = (Nq(x_i))^{-1}$ since by (1.1) we have,

$$\lim_{N\to\infty} \vec{\phi}^{\top} D \vec{f} = \lim_{N\to\infty} \sum_{i=1}^{N} D_{ii}\vec{\phi}_i\vec{f}_i = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \frac{\phi(x_i)f(x_i)}{q(x_i)} = \int_{\mathcal{M}} \phi f \;\; d\text{vol} = \mathcal{G}(\phi, f)$$

where q can be estimated as in [6] (the method is summarized in Appendix B).

In this paper, we show that the graph Laplacian (as constructed by the diffusion maps algorithm [10]) is a consistent estimator of the Dirichlet energy, $\mathcal{E}$, even for manifolds with boundary (Theorem 6.2). Moreover, we introduce a novel consistent estimator for the boundary integral (Theorem 5.1). These results rely on the first uniform asymptotic expansion of kernel integral operators in a neighborhood of the boundary (Theorem 4.6). This in turn requires a subtle new distance comparison (Lemma 4.3) which expands the ambient space Euclidean distance between local points near the boundary of an embedded manifold with respect to semigeodesic coordinates. This new distance comparison is the direct analog of the expansion of the ambient space distance with respect to geodesic normal coordinates [31] which holds in the interior, but Lemma 4.3 carries out this expansion for points near the boundary using semigeodesic coordinates.

Recently, the diffusion maps estimate of the Laplacian has been used for solving PDEs such as $-\Delta u = f$ [13,15] where f is now the data and one is solving for u. We should note that in [13] a more general class of elliptic operators are considered using a more general class of kernel functions introduced in [5]. In this paper we restrict our attention to the Laplace-Beltrami operator in order to focus on the boundary conditions, however the theory and methods introduced here can also be used to impose new boundary conditions on the operators considered in [13]. We should also point out that [13] compared the diffusion maps approach to another popular meshless method based on radial-basis function (RBF) interpolation [26]. The RBF method outperforms the diffusion maps when an appropriate global coordinate system is available in which to form the basis functions. However, as pointed out in [13], extending the RBF method to arbitrary manifolds would require extensive modifications, such as finding local coordinate systems and approximating the desired differential operators. The diffusion maps approach provides a large class of operators directly with a global representation. Thus, when more information about the manifold structure is known, approaches such as [26] may have superior results (just as mesh-based methods may have better results when a mesh is available), so it should be emphasized that our focus is on mesh-free methods on an unknown manifold as motivated above.

A key aspect of our approach is that the location of the boundary is also unknown and must be estimated as in [6]. Since the boundary is a measure zero subset, we do not expect any data samples to lie exactly on the boundary, so rather than making a binary choice ('on' or 'off' the boundary) we instead locate the boundary implicitly by estimating the distance to the boundary for each data point. For completeness, we summarize the method of [6] for finding the distance to the boundary in Appendix B. This distance-to-the-boundary function turns out to be the key to building the boundary integral estimator that converges to $\mathcal{B}$. Together these results yield a collection of consistent estimators for $\mathcal{G}, \mathcal{E}$, and $\mathcal{B}$ that can be used to solve the heat equation with various boundary conditions given only a set of points lying on the unknown manifold.

The paper is organized as follows: In Section 2 we define the class of manifolds required for our theoretical results and establish bounds on the ratios of the intrinsic distance and the distances in coordinates near the boundary that will be required later. In Section 3 we use these bounds to show that kernel integral operators of the form (1.2) with kernels having fast decay can be localized to a small neighborhood of x (up to an error that is small with respect to the bandwidth, ϵ). In Section 4 we present the first asymptotic expansion of the kernel integral operator (1.2) that holds uniformly in a neighborhood of the boundary (meaning

ϵ is independent of x). In Section 5 we introduce the first boundary integral estimator (making use of the distance to the boundary function) which gives a consistent estimator of $\mathcal{B}$. In Section 6 we use the asymptotic expansion from Section 4 together with the boundary integral estimator in Section 5 to construct consistent estimators of the operators in (1.6). By analyzing the existing Laplacian estimators in light of our new results we show why these standard constructions result in Neumann boundary conditions, as observed empirically going back to [10]. Finally in Section 7 we show how to impose standard boundary conditions using these operator estimators.

2. Preliminaries

The contents of this section are devoted to establishing some coordinate computations which will be used several times in subsequent sections. We begin by setting some notation and recalling some fundamental properties from Riemannian geometry.

In what follows, we let $\mathcal{M}$ be a C^3 compact manifold with nonempty boundary smoothly and properly embedded into $\mathbb{R}^d$ via the map $\iota : \mathcal{M} \to \mathbb{R}^d$. We endow $\mathcal{M}$ with the pullback metric $g = \iota^* g^{\mathbb{R}^n}$ so that ι is an isometric embedding. We let $d\text{vol}$ be the Riemannian volume element defined by this metric and we let $q : \mathcal{M} \to \mathbb{R}$ denote a C^3 probability density function that is absolutely continuous with respect to $d\text{vol}$.

We also recall that in any local coordinates $(s^1, ..., s^m)$ on $\mathcal{M}$, the pullback metric evaluated on vector fields $X = x^i \partial_i, Y = y^j \partial_j$ is given by

$$\left\langle x^i \partial_i, y^j \partial_j \right\rangle_g = \delta_{\alpha\beta} \frac{\partial \iota^\alpha}{\partial s^i} \frac{\partial \iota^\beta}{\partial s^j} x^i y^j$$

where we are using Einstein notation so that indices appearing in both a superscript and subscript are implied to be summed over a common index. For convenience, we let Greek characters such as α, β range from 1 to d and Roman characters such as i, j range from 1 to $m = \dim(\mathcal{M})$.

We recall that the Riemannian metric on $\mathcal{M}$ induces a metric space structure on $\mathcal{M}$ with the metric by letting $d_g(x, y)$ denote the infimum of all piecewise smooth regular curves connecting x to y in $\mathcal{M}$. If x and y are not in the same component, we define $d_g(x, y) = +\infty$.

Recall that the Laplace-Beltrami operator Δ on $\mathcal{M}$ is defined by $\Delta f = \text{Div}(\text{grad } f)$ and has coordinate expression

$$\Delta f = \frac{1}{\sqrt{|\det g|}} \partial_i \left(\sqrt{|\det g|} g^{ij} \partial_j f \right)$$

in any coordinate system. In particular, when the metric is flat, $g_{ij} = \delta_{ij}$ this corresponds to the standard Euclidean Laplacian

$$\Delta_{\mathbb{R}^m} = \sum_{i=1}^{m} \frac{\partial^2}{\partial x^i \partial x^i}.$$

2.1. Kernel regularity

We let $k : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a C^2 real valued function. We furthermore assume for the remainder of the paper that $k, |k'|$, and $|k''|$ have exponential decay and $k(0) = 0$. We also assume that $q : \mathcal{M} \to \mathbb{R}$ is a strictly positive C^3 probability distribution function which is absolutely continuous with respect to the Riemannian volume element. These assumptions are identical to those found in [16].

We define a family of kernel averaging operators $\mathcal{I}_\epsilon$ indexed by parameter $\epsilon \in (0, \infty)$ by:

$$\mathcal{I}_\epsilon f = \int_{\mathcal{M}} k\left(\frac{\|x - y\|^2_{\mathbb{R}^n}}{\epsilon^2} \right) f(y) q(y) \, d\text{vol}.$$

We remark that this choice of kernel operator differs from the one originally used in [10] in that we choose to use squared distance over ϵ^2. This is simply to avoid issues of smoothness when $x = y$. For simplicity of exposition, we will state the results of the following section assuming that $q \equiv 1$. This presents no loss of generality, since one may make the substitution $f \mapsto fq$ and obtain the needed results.

2.2. Normal coordinates

Normal coordinates are a set of coordinates which are used extensively in the asymptotic analysis of manifold learning algorithms on Riemannian manifolds. In this section, we review several of the important properties used later in the paper.

Recall that the exponential map based at a point x is a mapping $\exp_x : U \subseteq T_x\mathcal{M} \to \mathcal{M}$ which maps a tangent vector v to the endpoint of the geodesic based at x with initial velocity v. On a small star-shaped neighborhood of $T_x\mathcal{M}$, $\exp_x$ is a diffeomorphism onto its image. The smallest value $\text{inj}(x) > 0$ such that $B_{\text{inj}(x)}(0) \subseteq T_x\mathcal{M}$ is a diffeomorphism is called the *injectivity radius* of $\mathcal{M}$ at x. The infimum of injectivity radii over all $x \in \mathcal{M}$ is called the *injectivity radius of* $\mathcal{M}$ and it can be shown that for a compact manifold without boundary, $\text{inj}(\mathcal{M})$ is positive.

By identifying $T_x\mathcal{M}$ with $\mathbb{R}^m$ using an orthonormal basis, one can use the exponential map to construct Riemannian normal coordinate charts centered at $x \in \mathcal{M}$, which are a system of coordinates $(s^1, ..., s^m)$ with the following nice properties:

Proposition 2.1. *Let $x \in \mathcal{M}$ and let $(s^1, .., s^m)$ denote a system of normal coordinates centered about x. Then*

(a) *The coordinates of x are $(0, .., 0)$.*
(b) *The components of the metric at x are $g_{ij}(x) = \delta_{ij}$.*
(c) *For every vector $v = v^i \partial_i$ at x, the radial geodesic with initial velocity v is represented in coordinates by:*

$$\gamma_v(t) = t(v^1, ..., v^m).$$

(d) *The Christoffel symbols and first partial derivatives of g_{ij} vanish at x.*

In particular, the geodesic distance between x and a point y in normal coordinates corresponds to the 2-norm of the coordinate representative s of y in normal coordinates:

$$d_g(x, y) = \|s\|_{\mathbb{R}^m}.$$

This expression for the intrinsic distance is the foundation of the asymptotic expansion (4.4) in [10] as well as many related papers [4,6,16,32,30,12].

In the case of manifolds with boundary, we remark that the exponential map may be less well-behaved near the boundary than in the non boundary case. Since geodesics may intersect the boundary, there is no longer a nice one-one relationship between initial velocity vectors and geodesics through a given point.

As an illustrative example of this behavior, consider the closed two dimensional annulus $A \subseteq \mathbb{R}^2$ and let x be an interior point of A, y be a boundary point for which the straight line segment $\overline{xy}$ in $\mathbb{R}^2$ is not contained in A.

2.3. The normal collar and semigeodesic coordinates

We instead use *semigeodesic coordinates*, which will be more amenable to calculations for points near the boundary. We outline the needed results here, for more details on semigeodesic coordinates see [21]. Since $\mathcal{M}$ is compact, it admits a *normal collar* [21], which is a mapping $\phi : \partial\mathcal{M} \times [0, r_C) \to \mathcal{M}$ defined by:

$$\phi(x, t) = \exp_x(-t\eta_x)$$

where $r_C > 0$, and $-\eta_x$ is the inward-facing unit normal vector field at x. Such a mapping is a diffeomorphism onto its image, which we will denote as $\mathcal{N}$. For each $t \in [0, r_C)$, we note that the set $\partial\mathcal{M}_t := \phi^{-1}(\partial\mathcal{M} \times \{t\})$ is the hypersurface of points distance t from the boundary. Such $\partial\mathcal{M}_t$ are embedded submanifolds of $\mathcal{M}$ for each $t \in [0, r_C)$.

Inside of the normal collar, we can now construct semigeodesic coordinates centered at x. To do so, one first fixes a point x in the normal collar, and constructs normal coordinates $(u^1, ..., u^{m-1})$ in the $(m-1)$-dimensional hypersurface parallel to ∂M which intersects x. We shall refer to such a hypersurface as $\partial\mathcal{M}(x)$. One then uses the m-th coordinate u^m to parameterize the geodesic distance in $\mathcal{M}$ from $\partial\mathcal{M}(x)$. Thus semigeodesic coordinates are formed through a composition of the exponential map of $\partial\mathcal{M}$ and the inward-facing exponential map $\exp_x(-u^m\eta)$ of $\mathcal{M}$. We now list a few of their properties in contrast to the previous section.

Proposition 2.2. *Let $x \in \mathcal{M}$ and let $(u^1, .., u^m)$ denote a system of semigeodesic coordinates centered at x. Then*

(a) *The coordinates of x are $(0, .., 0)$.*
(b) *The components of the metric at x are $g_{ij}(x) = \delta_{ij}$.*
(c) *For every vector $v = v^i \partial_i$ at x, the radial geodesic in $\partial\mathcal{M}_t$ with initial velocity $\sum_{i=1}^{m-1} v^i$ is represented in coordinates by:*

$$\gamma(t) = t(v^1, ..., v^{m-1}, 0).$$

The geodesic starting at x which intersects each $\partial\mathcal{M}_t$ orthogonally is represented in coordinates as:

$$\gamma(t) = t(0, ..., 0, v^m).$$

We remark that in contrast to the case in normal coordinates, the norm in semigeodesic coordinates no longer measures a well-defined distance. The first $m-1$ coordinates parameterize geodesic distance in the parallel hypersurface, while the last coordinate entry parameterizes geodesic distance in $\mathcal{M}$ in the direction orthogonal to the hypersurface. Moreover, whereas the geodesic ball is a sphere, a semigeodesic chart may be viewed as a hypercylinder which intersects the boundary orthogonally. The cylinder is symmetric in the coordinates $u^1, ..., u^{m-1}$ with respect to any $(m-1)$-dimensional rotation but the symmetry does not extend the u^m which is the 'height' of the cylinder. Moreover, since u^m parametrizes the geodesic toward the boundary, the cylinder is truncated to $u^m \geq b_x = -d(x, \partial\mathcal{M})$. Finally, we note that following [21], u^m is oriented along an inward-facing pointing normal, so for any vector v we have $v^m = -v \cdot \eta_x$ since η_x is outward-pointing.

2.4. Coordinate estimates

Although the Euclidean norm in semigeodesic coordinates no longer corresponds to geodesic distance from x, the goal of this section is to relate the norm in such coordinates to the geodesic distance in $\mathcal{M}$.

We first note that compactness of $\mathcal{M}$ implies an upper and lower bound on the sectional curvature of the interior of $\mathcal{M}$ as well as bounds on the inward-facing sectional curvature of $\partial\mathcal{M}$. From the Gauss equation, it also follows that each parallel hypersurface $\partial\mathcal{M}_t$ for $t \in [0, \frac{1}{2}r_C]$ has inward-facing sectional curvature bounded above and below. We let K be an upper bound on the sectional curvature and inward facing sectional curvatures of such parallel hypersurfaces such that $-K$ is also a lower bound on such curvatures. By the results of [11], one also obtains a single lower bound $r_\partial > 0$ on the injectivity radius of each parallel hypersurface $\partial\mathcal{M}_t$ for all $t \in [0, \frac{1}{2}r_C]$.

The following lemma may be proven by an application of the Rauch comparison theorem (for normal coordinates) and Warner's generalization of the Rauch comparison theorem [33] (for semigeodesic coordinates. Such an argument was done in Proposition 2.6 of [28] and needs only slight modification for semigeodesic coordinates.

Lemma 2.3. *[28] Suppose $\mathcal{M}$ is a compact Riemannian manifold, then there exist constants $\tilde{C}_0 > 0, \tilde{C}_1 > 0$ such that in any geodesic chart in $\mathcal{M}$,*

$$|g_{ij}| \leq \tilde{C}_0 \text{ and } |g^{ij}| \leq \tilde{C}_1$$

Using Lemma 2.3, we can now show that the norm in geodesic coordinates approximates the extrinsic distance induced by $\iota : \mathcal{M} \to \mathbb{R}^d$.

Proposition 2.4. *There exists a $C_0, C_1 > 0$ such that in any geodesic chart centered at x, and any point y in that chart,*

$$C_0|u|^2 \leq d^2_{\mathbb{R}^d}(\iota(x), \iota(y)) \leq C_1|u|^2$$

where u is the coordinate representative of y in either normal or semigeodesic coordinates.

Proof. From Lemma 2.4, there exist positive constants $\tilde{C}_0$ and $\tilde{C}_1$ such that

$$|g_{ij}| \leq \tilde{C}_0 \text{ and } |g^{ij}| \leq \tilde{C}_1$$

for any geodesic coordinate chart. Since the matrices with entrees g_{ij} and g^{ij} are symmetric and positive definite, this implies that there exist positive bounds C_1 and C_0 on the largest eigenvalue of (g_{ij}) and (g^{ij}) in any geodesic coordinate chart.

We now let x be a point in $\mathcal{M}$, and choose either normal or semigeodesic coordinate charts for $\mathcal{M}$ centered at x, depending on whether x is in the normal collar. We then choose a normal coordinate chart for $\iota(x)$ in $\mathbb{R}^d$, which is simply centering $\iota(x)$ at zero. In these coordinates, we have that

$$d_{\mathbb{R}^d}(\iota(x), \iota(y)) = g^{\mathbb{R}^d}_{\alpha\beta}(0)\iota^\alpha(u)\iota^\beta(u),$$

where u is the coordinate representative of y in these coordinates. We then perform a Taylor expansion of $\iota(u)$, recalling that in these coordinates $\iota(0) = 0$. Therefore there exists a point $\tilde{u}$ in the domain such that:

$$\begin{aligned} g^{\mathbb{R}^d}_{\alpha\beta}(0)\iota^\alpha(u)\iota^\beta(u) &= g^{\mathbb{R}^d}_{\alpha\beta}(0)\left(\iota^\alpha(0) + \frac{\partial\iota^\alpha}{\partial u^i}\bigg|_{\tilde{s}} u^i\right)\left(\iota^\beta(0) + \frac{\partial\iota^\beta}{\partial u^j}\bigg|_{\tilde{u}} u^j\right) \\ &= g^{\mathbb{R}^d}_{\alpha\beta}(0)\frac{\partial\iota^\alpha}{\partial u^i}\bigg|_{\tilde{u}}\frac{\partial\iota^\beta}{\partial u^j}\bigg|_{\tilde{u}} u^i u^j. \end{aligned}$$

Since $g^{\mathbb{R}^d}_{\alpha\beta}(0) = g^{\mathbb{R}^d}_{\alpha\beta}(\iota(\tilde{u})) = \delta_{\alpha\beta}$, we have:

$$\begin{aligned} g^{\mathbb{R}^d}_{\alpha\beta}(0)\iota^\alpha(u)\iota^\beta(u) &= g^{\mathbb{R}^d}_{\alpha\beta}(\iota(\tilde{u}))\frac{\partial\iota^\alpha}{\partial u^i}\bigg|_{\tilde{u}}\frac{\partial\iota^\beta}{\partial u^j}\bigg|_{\tilde{u}} u^i u^j \\ &= g^{\mathcal{M}}_{ij}(\tilde{u})u^i u^j. \end{aligned}$$

We now note that the expression $g^{\mathcal{M}}_{ij}(\tilde{u})u^i u^j$ is maximized by the maximum eigenvalue of the matrix $g^{\mathcal{M}}_{ij}(\tilde{u})$ and minimized by the maximum eigenvalue of $(g^{\mathcal{M}})^{ij}(\tilde{u})$. Since we have previously show that these are bounded by C_1 and C_0 regardless of choice of geodesic chart, we have that

$$C_0|u|^2 \leq g^{\mathbb{R}^d}_{\alpha\beta}(0)\iota^\alpha(s)\iota^\beta(s) \leq C_1|u|^2$$

and therefore

$$C_0|u|^2 \leq d^2_{\mathbb{R}^d}(\iota(x), \iota(y)) \leq C_1|u|^2. \quad \square$$

Thus, we have established a relationship between the norm in semigeodesic coordinates and the extrinsic distance defined by the embedding $\iota : \mathcal{M} \to \mathbb{R}^d$.

3. Localization of asymptotic expansions for manifolds with boundary

We begin by rigorously defining two regions of $\mathcal{M}$ which are "close" and "far" from the boundary respectively. For each $\epsilon > 0$, we let $\mathcal{M}_\epsilon$ be the set of all points $x \in M$ such that $d(x, \partial\mathcal{M}) > \epsilon$. We then define $\mathcal{N}_\epsilon$ to be the points such that $d(x, \partial M) \leq \epsilon$. We refer to $\mathcal{M}_\epsilon$ as the *interior region* and $\mathcal{N}_\epsilon$ as the *closed collar region*. Due to the generalized Gauss lemma [21], one has that if ϵ is less than the normal collar width r_C, the topological boundary $\partial\mathcal{M}_\epsilon$ of $\mathcal{M}_\epsilon$ is a hypersurface in $\mathcal{M}$ parallel to $\partial\mathcal{M}$.

We first show that for each sufficiently small ϵ, the manifold $\mathcal{M}$ admits an atlas of charts which are each "large enough" to contain a metric ball of radius ϵ. This will be proven in Proposition 3.2, which we list here.

Proposition 3.2. *There exists a $C_{\mathcal{M}} > 0$ such that for all $0 < \epsilon < C_{\mathcal{M}}$, the preimage of an extrinsic ball $\iota^{-1}(B_\epsilon^{\mathbb{R}^d}(\iota(x)))$ centered at a point $x \in \mathcal{M}$ is contained in a normal coordinate chart if $x \in \mathcal{M}_\epsilon$ or in a semigeodesic coordinate chart if $x \in \mathbb{N}_\epsilon$.*

We let $\partial\mathcal{M}_t$ refer to the hypersurface of points distance t from the boundary. Such hypersurfaces are parallel to $\partial\mathcal{M}$ in the sense that geodesics with initial velocity normal to $\partial\mathcal{M}$ intersect the surfaces $\partial\mathcal{M}_t$ orthogonally. Using this fact, each tangent space $T_x\mathcal{N}$ for $x \in \mathcal{N}_\epsilon$ admits a decomposition:

$$T_x\mathcal{N}_\epsilon = T_x^\top\mathcal{N}_\epsilon \oplus T_x^\perp\mathcal{N}_\epsilon,$$

where $T_x^\top\mathcal{N} = T_x\partial\mathcal{M}_{d(x,\partial\mathcal{M})}$ is the tangent space of the parallel hypersurface intersecting x and $T_x^\perp\mathcal{N}_\epsilon$ is the space spanned by the unit vector normal to the hypersurface in $\mathcal{M}$.

For each $x \in \mathcal{M}$ and each $\epsilon > 0$, we define the *semigeodesic hypercylinder* $\tilde{B}_\epsilon(x)$ of radius and height ϵ as the set of all vectors $v \in T_x\mathcal{N}$ such that:

(a) $\|v^\top\|_g < \epsilon$
(b) $\|v^\perp\|_g < \epsilon$ if $v^\perp$ is inward-facing
(c) $\|v^\perp\|_g < d_g(x, \partial\mathcal{M})$ if $v^\perp$ is outward-facing.

By this construction, we have that if ϵ is less than the injectivity radius of the parallel hypersurface $\partial\mathcal{M}(x) = \partial\mathcal{M}_{d_g(x,\partial\mathcal{M})}$ through x, then $exp_x^{\partial\mathcal{M}(x)}(v^\top)$ is well-defined for all $v \in \tilde{B}_\epsilon(x)$. Similarly, if $\epsilon < \frac{r_C}{2}$ then $\exp_q(T_{pq}^{\partial\mathcal{M}(x)}v^\perp)$ is well-defined for any q in a normal neighborhood of x in $\partial\mathcal{M}(x)$, where here $T_{pq}^{\partial\mathcal{M}(x)}v^\perp$ denotes the parallel translate of $v^\perp$ in $\partial\mathcal{M}(x)$ to the point $q \in \partial\mathcal{M}(x)$ and r_C is the normal collar width. Thus we see that if ϵ is sufficiently small, the semigeodesic cylinder of radius and height ϵ may be identified with hypercylinder which is a submanifold of $\mathcal{M}$ through the exponential map of $\partial\mathcal{M}(x)$.

The next proposition shows that there is a constant $r_{\text{sem}} > 0$ such that if one chooses a radius smaller than some constant $C_{\mathcal{M}} > 0$, then a metric ball of radius $C_{\mathcal{M}}$ in $\mathcal{M}$ is small enough to fit inside any semigeodesic cylinder of radius and height r_{sem}. This is an essential step to show that one can uniformly localize the operator $\mathcal{I}_\epsilon$ to semigeodesic charts in the same way as normal coordinate charts.

Proposition 3.1. *Let $r_C > 0$ denote the normal collar width and let $r_\partial > 0$ be a lower bound on the injectivity radii of all parallel hypersurfaces $\partial\mathcal{M}_\epsilon$ with $\epsilon < \frac{1}{2}r_C$. Let $K \in \mathbb{R}$ be an upper bound on the sectional curvature of $\mathcal{M}$ and the inward-facing sectional curvature of $\partial\mathcal{M}$. Let $r_{\text{sem}} = \min\{\frac{1}{2}r_C, r_\partial, \frac{\pi}{2\sqrt{K}}\}$ where $\sqrt{K}^{-1}$ is defined to be infinite if $K \leq 0$ Then there exists a $\tilde{C}_{\mathcal{M}} > 0$ such that for all $x \in \mathcal{N}_{r_{\text{sem}}}$,*

$$B^{\mathcal{M}}_{\tilde{C}_{\mathcal{M}}}(x) \subseteq \tilde{B}_{r_{\text{sem}}}(x).$$

Proof. We first let X denote

$$X = \coprod_{x \in \mathcal{N}_{r_{\text{sem}}}} \partial\tilde{B}_{r_{\text{sem}}}(x).$$

In other words, consider the disjoint union of the boundary of all semigeodesic hypercylinders centered about all points in $x \in \mathcal{N}_{r_{\text{sem}}}$. It can be shown that such a set is a fiber bundle over $\mathcal{N}_{r_{\text{sem}}}$ with model fiber diffeomorphic to a hypercylinder with the interior of the "bottom" face removed. Such a model fiber is compact and since $\mathcal{N}_{r_{\text{sem}}}$ is also compact, it follows that the fiber bundle X is compact. Note that proof of existence of this fiber bundle follows in analogy to the construction of the unit tangent bundle on a Riemannian manifold without boundary.

We now consider the function $d_\partial : X \to \mathbb{R}$ which assigns to each (x, v) the distance from x to the geometric realization of v in $\mathcal{M}$. Such a map is clearly continuous on x, and thus obtains a minimum value $C_{\mathcal{M}}$ on X by compactness. It can be easily argued using properties of the exponential map that for each $x \in \mathcal{N}_{r_{\text{sem}}}$, we have that $x \notin \partial B_{r_{\text{sem}}}(x)$ and thus $d(x, v) > 0$ for all $(x, v) \in X$. Hence $\tilde{C}_{\mathcal{M}} > 0$.

Since we assume $r_{\text{sem}} < \frac{\pi}{4\sqrt{K}}$, we have that $\|v\|_g < \frac{\pi}{4\sqrt{K}}$ for each $(x,v) \in X$. It follows that the distance from x to the geometric realization of v is less than $\frac{\pi}{2\sqrt{K}}$. From Corollary 2 of [2], it follows that every pair of points (x,y) in a metric ball of radius $\frac{\pi}{2\sqrt{K}}$ has a length-minimizing curve γ connecting x and y in $\mathcal{M}$.

Now, fix $x \in \mathcal{N}_{r_{\text{sem}}}$ and suppose that $y \in B_{\tilde{C}_{\mathcal{M}}}(x)$ but $y \notin \tilde{B}_{r_{\text{sem}}}(x)$. Denote the length-minimizing curve connecting x and y by $\gamma_{xy}(t)$. A simple topological argument can be used to show that γ_{xy} eventually intersects the boundary of $\tilde{B}_{r_{\text{sem}}}(p)$. Namely, there exists a t_0 such that $\gamma_{xy}(t_0) \in \partial \tilde{B}_{r_{\text{sem}}}(p)$. It follows that $d_g(p, \gamma_{xy}(t_0)) \geq \tilde{C}_{\mathcal{M}}$ and hence the length of $\gamma_{xy} > C_{\mathcal{M}}$. This contradicts the fact that $y \in B_{\tilde{C}_{\mathcal{M}}}(x)$ and implies the result. □

Since in general the kernel function $k_\epsilon(x,y)$ is defined using extrinsic distance in $\mathbb{R}^d$ instead of distance in $\mathcal{M}$, we need to improve the result in Proposition 3.1 to account for extrinsic distance. We also wish to improve the result to hold for both semigeodesic and normal coordinates and hold for all ϵ sufficiently small. Most of this can be done by simply observing that the embedding map ι has a uniformly continuous inverse. The main remaining obstacle is that the region $\mathcal{M}_\epsilon$ grows as ϵ approaches zero. We therefore need to show that the injectivity radius of the region $\mathcal{M}_\epsilon$ does not shrink too fast as ϵ approaches zero.

Proposition 3.2. *There exists a $C_{\mathcal{M}} > 0$ such that for all $0 < \epsilon < C_{\mathcal{M}}$, the preimage of an extrinsic ball $\iota^{-1}(B^{\mathbb{R}^d}_\epsilon(\iota(x)))$ centered at a point $x \in \mathcal{M}$ is contained in a normal coordinate chart if $x \in \mathcal{M}_\epsilon$ or in a semigeodesic coordinate chart if $x \in \mathcal{N}_\epsilon$.*

Proof. First we consider the value

$$\text{inj}(\mathcal{M}_\epsilon) := \inf_{x \in \mathcal{M}_\epsilon} \text{inj}(x).$$

We first show that for sufficiently small ϵ, the value of $\text{inj}(\mathcal{M}_\epsilon) \geq \epsilon$. Consider the double $D(\mathcal{M})$ of $\mathcal{M}$ formed by gluing identical copies of $\mathcal{M}$ along the boundary of $\mathcal{M}$. It follows that $D(\mathcal{M})$ is compact and one may extend the metric on $\mathcal{M}$ arbitrarily to a metric on $D(\mathcal{M})$. In such a case, geodesic balls of radius ϵ or less on $\mathcal{M}_\epsilon$ coincide with geodesic balls of radius ϵ or less on $D(\mathcal{M})$. Since $D(\mathcal{M})$ is compact, it has positive injectivity radius and thus if $\epsilon < \text{inj}(D(\mathcal{M}))$, then exp_x is bijective on $B_\epsilon(x)$ in $\mathcal{M}_\epsilon$. Hence, if $\epsilon < \text{inj}(D(\mathcal{M}))$, $\mathcal{M}_\epsilon$ may be covered by normal coordinate charts which contain a metric ball of radius ϵ. By Proposition 3.2 if $\epsilon < \tilde{C}_{\mathcal{M}}$, then $\mathcal{N}_\epsilon$ may be covered in semigeodesic charts each of which contain a metric ball in $\mathcal{M}$ of radius ϵ.

We now let $C'_{\mathcal{M}} = \min\{\tilde{C}_{\mathcal{M}}, \text{inj}(D(\mathcal{M}))\}$. Since the embedding $\iota : \mathcal{M} \to \mathbb{R}^d$ is continuous on a compact set, its inverse ι^{-1} is uniformly continuous on its domain. Thus, there exists a $C_{\mathcal{M}} > 0$ which does not depend on $x \in \mathcal{M}$ for which $\iota^{-1}(B^{\mathbb{R}^d}_{C_{\mathcal{M}}}(\iota(x)) \subseteq B_{C'_{\mathcal{M}}}(x)$ for all $x \in \mathcal{M}$. Therefore for all $0 < \epsilon < C_{\mathcal{M}}$ and all $x \in \mathcal{M}$, we have $\iota^{-1}(B^{\mathbb{R}^d}_\epsilon(\iota(x))$ is contained in a normal coordinate chart if $x \in \mathcal{M}_\epsilon$ and a semigeodesic chart if $x \in \mathcal{N}_\epsilon$. □

We now put together Proposition 3.2 and Proposition 2.4 to show that for sufficiently small ϵ, one may localize the kernel integral operator to a geodesic coordinate chart up to order ϵ^z for arbitrarily large $z \in \mathbb{N}$.

Lemma 3.3 *(Localization to a Geodesic Neighborhood). Let $0 < \gamma < 1$. For any $\epsilon > 0$ such that $\epsilon^\gamma < \min\{\frac{r_{\mathcal{M}}}{C_1}, C_{\mathcal{M}}\}$,*

$$\left| \int_{\mathcal{M} \setminus \tilde{B}^{\mathcal{M}}_{\epsilon^\gamma}(x)} k\left(\frac{d^2_{\mathbb{R}^d}(\iota(x),\iota(y))}{\epsilon^2}\right) f(y) q(y)\, d\text{vol} \right| \in O(\epsilon^z)$$

where z may be chosen arbitrarily large in $\mathbb{N}$.

Proof. If $\epsilon^\gamma < C_{\mathcal{M}}$, we have that the preimage of an ϵ^γ ball in $\mathbb{R}^d$ centered about $\iota(x)$ is contained in a geodesic coordinate chart centered at x. If $\epsilon^\gamma < \frac{C'_{\mathcal{M}}}{C_1}$, then by Proposition 2.4, $\tilde{B}^{\mathcal{M}}_{C_1\epsilon^\gamma}(x)$ contains this preimage, and is also contained in the geodesic chart. Hence, any point in $\mathcal{M}$ outside of $\tilde{B}^{\mathcal{M}}_{C_1\epsilon^\gamma}(x)$ has extrinsic distance no less than ϵ^γ from x.

Using exponential decay of the kernel, this implies that

$$\begin{aligned}
\int_{\tilde{B}^{\mathcal{M}}_{C_1\epsilon^\gamma}(x)} k\left(\frac{d_{\mathbb{R}^d}(\iota(x),\iota(y))}{\epsilon^2}\right) q\, d\text{vol} &\leq \int_{\tilde{B}^{\mathcal{M}}_{C_1\epsilon^\gamma}(x)} \alpha e^{-\beta \frac{d^2_{\mathbb{R}^d}(\iota(x),\iota(y))}{\epsilon^2}} q\, d\text{vol} \\
&\leq \alpha e^{-\beta \frac{\epsilon^{2\gamma}}{\epsilon^2}} \\
&= \alpha e^{-\beta \epsilon^{2(\gamma-1)}}
\end{aligned}$$

We then apply Cauchy-Schwarz inequality in q-weighted $L^2(\mathcal{M})$:

$$\left\langle k\left(\frac{d_{\mathbb{R}^d}(\iota(x),\iota(y))}{\epsilon^2}\right), f\right\rangle^2 \leq \left\langle k\left(\frac{d_{\mathbb{R}^d}(\iota(x),\iota(y))}{\epsilon^2}\right), k\left(\frac{d_{\mathbb{R}^d}(\iota(x),\iota(y))}{\epsilon^2}\right)\right\rangle^2 \langle f, f\rangle$$
$$\leq \langle f, f\rangle \alpha e^{-2\beta\epsilon^{2(\gamma-1)}} \int_{\mathcal{M}\setminus \tilde{B}^{\mathcal{M}}_{C_1\epsilon^\gamma}(x)} q \, d\mathrm{vol}$$
$$\leq \langle f, f\rangle \alpha e^{-2\beta\epsilon^{2(\gamma-1)}}$$

We see that the term $\langle f, f\rangle \alpha e^{-2\beta\epsilon^{2(\gamma-1)}}$ is asymptotically bounded by any polynomial ϵ^z with $z \geq 1$. □

We have now shown that the value of $\mathcal{I}_\epsilon f(x)$ depends only on the behavior of f inside a single chart for small enough values of the parameter ϵ. If x is in the closed collar region $\mathcal{N}_\epsilon$ for small enough ϵ, this chart must be taken as a semigeodesic coordinate chart, while if x contained in the interior region $\mathcal{M}_\epsilon$ one may use a normal coordinate chart.

4. Uniform asymptotic expansion for manifolds with boundary

The results of the previous section show that the asymptotic analysis of $\mathcal{I}_\epsilon f(x)$ can be subdivided into two cases depending on whether x is in the interior region $\mathcal{M}_\epsilon$ or closed collar region $\mathcal{N}_\epsilon$. In this section, we derive new asymptotic expansions of $\mathcal{I}_\epsilon f$ in semigeodesic coordinates. When taken together with existing expansions in normal coordinates from [10,16], this yields an asymptotic expansion of $\mathcal{I}_\epsilon$ that is uniform in ϵ, meaning that for sufficiently small ϵ, the expansion holds for each $x \in M$. This uniformity is necessary for our later proof of convergence.

We begin by deriving asymptotic expansions of $\mathcal{I}_\epsilon f$ in semigeodesic coordinates. The following lemmas are used to show that the value of the constant in the leading order error term is related to the mean curvature of the boundary of $\mathcal{M}$. They are largely technical, but this specific value will give us some cancellation in the final expansion and is thus important.

We begin by observing coordinate expressions for the Levi-Civita connection on $\mathcal{M}$.

Lemma 4.1. *Let U be the vector field such that*

(a) *$U_x \in T_x\mathcal{M}$ maps to the point u in semigeodesic coordinates centered at x.*
(b) *The coordinate representation $U = u^i\partial_i$ has constant component functions u^i.*

Then at the point p:

$$2\langle \nabla_U U, U\rangle_g = \delta_{\alpha\beta}\left(\frac{\partial^2\iota^\alpha}{\partial u^a \partial u^c}\frac{\partial\iota^\beta}{\partial u^b} + \frac{\partial\iota^\beta}{\partial u^a}\frac{\partial^2\iota^\alpha}{\partial u^b\partial u^c}\right).$$

Proof. Since $\iota : \mathcal{M} \to \mathbb{R}^d$ is an isometric embedding, we may relate the components of the metric in $\mathcal{M}$ to those in $\mathbb{R}^d$ via:

$$g^{\mathcal{M}}_{ab}(u) = g^{\mathbb{R}^d}_{\alpha\beta}(\tilde{s})\frac{\partial\iota^\alpha}{\partial u^a}\frac{\partial\iota^\beta}{\partial u^b} = \delta_{\alpha\beta}\frac{\partial\iota^\alpha}{\partial u^a}\frac{\partial\iota^\beta}{\partial u^b}$$

We then take the partial derivative of both sides, noting that we are in normal coordinates in $\mathbb{R}^d$ and therefore all partial derivatives of the metric components vanish at 0. This yields:

$$\frac{\partial g^{\mathcal{M}}_{ab}(0)}{\partial u^c} = \frac{\partial g^{\mathbb{R}^d}_{\alpha\beta}(0)}{\partial u^c}\frac{\partial\iota^\alpha}{\partial u^\rho}\frac{\partial\iota^\rho}{\partial u^c}\frac{\partial\iota^\alpha}{\partial u^a}\frac{\partial\iota^\beta}{\partial u^b} + g^{\mathbb{R}^d}_{\alpha\beta}(0)\frac{\partial}{\partial u^c}\left(\frac{\partial\iota^\alpha}{\partial u^a}\frac{\partial\iota^\alpha}{\partial u^b}\right) = \delta_{\alpha\beta}\left(\frac{\partial^2\iota^\alpha}{\partial u^a\partial u^c}\frac{\partial\iota^\beta}{\partial u^b} + \frac{\partial\iota^\beta}{\partial u^a}\frac{\partial^2\iota^\alpha}{\partial u^b\partial u^c}\right).$$

Given any $u \in T_x\mathcal{M}$, we can extend u to the vector field $U = u^i\partial_i$ on the coordinate chart where u^i are constant functions and ∂_i are the coordinate vector fields. Using that the Levi-Civita connection is compatible with the metric, we obtain:

$$\frac{\partial g^{\mathcal{M}}_{ab}(0)}{\partial u^c}u^a u^b u^c = U\langle U, U\rangle_g = 2\langle \nabla_U U, U\rangle_g$$

as desired. □

Next we relate the Levi-Civita connection to the second fundamental form $\Pi_{\partial\mathcal{M}_t}$ of the hypersurfaces $\partial\mathcal{M}_t$ as a submanifolds of $\mathcal{M}$.

Lemma 4.2. *With u and U having the same conditions as above, decompose U into the vector field $U^\top = \sum_{i=1}^{m-1} u^i\partial_i$ tangential to the hypersurface $\partial\mathcal{M}_t$ and the normal vector field $U^\perp = u^m\partial_m = -u^m\eta_x$. Then we have:*

$$\langle \nabla_U U, U\rangle_g = -\left\langle \Pi_{\partial\mathcal{M}_t}(U^\top, U^\top), U^\perp\right\rangle_g.$$

Proof. We decompose $\langle \nabla_U, U\rangle_g$ into

$$\langle \nabla_U U, U\rangle_g = \left\langle \nabla_{(U^\top+U^\perp)}(U^\top + U^\perp), (U^\top + U^\perp)\right\rangle_g.$$

Since the connection is linear in both components over $\mathbb{R}$, and the component functions of U are constant, we may simply bilinearly expand the above term. We also note that

$$\nabla_{u^j\partial_j} u^i\partial_i = u^i u^j \nabla_{\partial_j}\partial_i = u^i u^j \Gamma_{ij}^k \partial_k.$$

Since many of the Christoffel symbols in semigeodesic coordinates are zero, we are left with:

$$\langle \nabla_U U, U\rangle_g = \left\langle \nabla_{U^\top} U^\top, U^\top \right\rangle_g + \left\langle \nabla_{U^\top} U^\top, U^\perp \right\rangle_g + \left\langle \nabla_{U^\top} U^\perp, U^\top \right\rangle_g + \left\langle \nabla_{U^\perp} U^\top, U^\top \right\rangle_g$$

Using the Gauss equation for the hypersurface embedded in $\mathcal{M}$, we get that $\nabla_{U^\top} U^\top = \Pi(U^\top, U^\top)$. This implies that the first term is zero and the second term is $\left\langle \Pi_{\partial\mathcal{M}_t}(U^\top, U^\top), U^\perp \right\rangle_g$.

For the next two terms, we first note that since $\bar{\nabla}$ is a symmetric connection,

$$\nabla_{u^j\partial_j} u^i\partial_i = u^i u^j \Gamma_{ij}^k \partial_k = u^i u^j \Gamma_{ji}^k \partial_k = \nabla_{u^i\partial_i} u^j \partial_j$$

and so ∇ is a symmetric tensor over $\mathbb{R}$. Thus, both of the remaining terms are equal. The Weingarten equation implies that:

$$\left\langle \nabla_{U^\top} U^\perp, U^\top \right\rangle_g = -\left\langle \Pi(U^\top, U^\top), U^\perp \right\rangle_g.$$

Putting this all together, we are left with:

$$\langle \nabla_U U, U\rangle_g = \left\langle \Pi_{\partial\mathcal{M}_t}(U^\top, U^\top), U^\perp \right\rangle_g - 2\left\langle \Pi(U^\top, U^\top), U^\perp \right\rangle_g = -\left\langle \Pi(U^\top, U^\top), U^\perp \right\rangle_g \quad \square$$

Putting the above three lemmas together, we now asymptotically compare the extrinsic distance in $\mathbb{R}^d$ of two nearby points $\iota(x), \iota(y)$ to the norm of the coordinate expression of y in semigeodesic coordinates centered at x. This result is analogous to Proposition 6 of [31] which makes a similar comparison, except in their case the comparison was between extrinsic and intrinsic distance instead of semigeodesic norm.

Lemma 4.3. *Let $x \in \mathcal{M}$ and let $y \in \mathcal{M}$ be such that $\|\iota(x) - \iota(y)\|_{\mathbb{R}^d} < C_{\mathcal{M}}$. Let u denote the coordinate representative of y in semigeodesic coordinates and let $\|u\|$ denote the norm of u in semigeodesic coordinates. Then*

$$\lim_{|u|^3\to 0} \frac{\|\iota(x) - \iota(y)\|_{\mathbb{R}^d}^2 - \|u\|^2}{\|u\|^3} = -\left\langle \Pi_{\partial\mathcal{M}_t}(U^\top, U^\top), U^\perp \right\rangle_g$$

Proof. Since $\|\iota(x) - \iota(y)\|_{\mathbb{R}^d} < C_{\mathcal{M}}$, it follows from 3.2 that y has a coordinate representative u in semigeodesic coordinates centered at x. We thus choose such a coordinate system and apply Taylor's theorem and apply Lemmas 4.1 and 4.2.

$$\begin{aligned}
\|\iota(u)\|_{\mathbb{R}^d}^2 &= g_{\alpha\beta}(0)\frac{\partial\iota^\alpha}{\partial u^a}\frac{\partial\iota^\beta}{\partial u^b}u^a u^b + \frac{1}{2}g_{\alpha\beta}(0)\frac{\partial^2\iota^\alpha}{\partial u^a u^c}\frac{\partial\iota^\beta}{\partial u^b}u^a u^b u^c + \frac{1}{2}g_{\alpha\beta}(0)\frac{\partial\iota^\beta}{\partial u^b}\frac{\partial^2\iota^\alpha}{\partial u^b u^c}u^a u^b u^c + O(|u|^4)\\
&= \|u\|_{\mathcal{M}}^2 + \frac{1}{2}\frac{\partial g_{ab}^{\mathcal{M}}(0)}{\partial u^c}u^a u^b u^c + O(\|u\|^4)\\
&= \|u\|_{\mathcal{M}}^2 + \langle \nabla_U U, U\rangle_g + O(\|u\|^4)\\
&= \|u\|_{\mathcal{M}}^2 - \left\langle \Pi_{\partial\mathcal{M}_t}(U^\top, U^\top), U^\perp \right\rangle_g + O(\|u\|^4)
\end{aligned}$$

where the final equations follow from the previous two lemmas. $\square$

We note that the analogous expansion done in Proposition 6 of [31] has order $d_{\mathcal{M}}(x,y)^4$ instead of $\|u\|^3$ and depends on the second fundamental form of the embedding of $\mathcal{M}$ into the ambient space. Thus, one of the tradeoffs for using semigeodesic coordinate charts is a lower order error term.

Next, we expand the volume form $d\mathrm{vol}$ in semigeodesic coordinates. In contrast, we first note that in normal coordinates the volume form has the expansion

$$d\mathrm{vol} = 1 + R_{ij}s^i s^j + O(|s|^3)$$

where R_{ij} are the components of the Ricci curvature tensor [21]. In semigeodesic coordinates, an analogous result follows from the first variation formula for the area of hypersurfaces.

Theorem 4.4 *(First variation of area, [9,14]). Let $\partial\mathcal{M}_t$ be a hypersurface in $\mathcal{M}$ with $x \in \partial\mathcal{M}$ and outward facing normal η_x. Let $\sigma(t) : (-\epsilon, \epsilon) \to \mathcal{M}$ be a geodesic with initial velocity $\dot\sigma(0) = -\eta_x$ Then for $y = \sigma(t)$ we have*

$$d\mathrm{vol}(y) = 1 - (m-1)H(x)t + O(t^2)$$

where H is the mean curvature of $\partial\mathcal{M}_t$.

We note that [9] defines the mean curvature as simply the summation, whereas we follow [21] in including the factor $\frac{1}{m-1}$. For points y that are not along the geodesic σ the first order term will be the same since semigeodesic coordinates are the same as normal coordinates on the submanifold $\partial\mathcal{M}^t$ which contains no first order term. Thus for y not along the geodesic we have,

$$d\text{vol}(y) = 1 - (m-1)H(x)u^m + \omega_3(x)_{ij}u^iu^j + O(|u|^3)$$

for some smooth tensor ω_3. The next result will be required to simplify some expressions in the theorem which involve the derivative of the kernel.

Lemma 4.5. *Integrating over a cylinder $B = \{u \mid \sum_{i=1}^{m-1}(u^i)^2 < \epsilon^2, u^m \in [-b_x/\epsilon, \epsilon]\}$ which is symmetric in coordinates u^i for $1 \le i \le m-1$ we have*

$$\begin{aligned}\int_B k'(|u|^2)\left\langle \Pi_{\partial\mathcal{M}_t}(U^\top, U^\top), U^\perp\right\rangle_g du &= -\frac{(m-1)}{2}H(x)\int_B k(|u|^2)u^m\,du \\ &= \frac{(m-1)}{2}m_1^\partial(x)H(x) + O(\epsilon^z)\end{aligned} \tag{4.1}$$

for any $z \ge 1$, where $H(x)$ is the mean curvature.

Proof. Linear expansion of $\left\langle \Pi_{\partial\mathcal{M}_t}(U^\top, U^\top), U^\perp\right\rangle_g$ in terms of the coordinate basis at x yields:

$$\left\langle \Pi_{\partial\mathcal{M}_t}(U^\top, U^\top), U^\perp\right\rangle_g = \left\langle \Pi_{\partial\mathcal{M}}(\partial_i, \partial_j), \partial_m\right\rangle_g u^iu^ju^m,$$

since the domain B is symmetric in the coordinates u^i for $1 \le i \le m-1$, all of the terms u^iu^j with $i \ne j$ will integrate to zero. Thus, we have

$$\int_B k'(|u|^2)\left\langle \Pi_{\partial\mathcal{M}_t}(U^\top, U^\top), U^\perp\right\rangle_g du = \left\langle \Pi_{\partial\mathcal{M}}(\partial_i, \partial_i), \partial_m\right\rangle_g \int_B k'(|u|^2)u^iu^iu^m du$$

and by the symmetry of the kernel, the integrals are equal for all $1 \le i \le m-1$, so we only need to compute

$$\int_B k'(|u|^2)u^1u^1u^m ds = \int_B \frac{1}{2}\left(\frac{\partial}{\partial u^1}k(|u|^2)\right)u^1u^m du^1du^2\cdots du^m = -\frac{1}{2}\int_B k(|u|^2)u^m du$$

where the last equality follows from integration by parts with respect to u^1. Finally, pulling the integral out of the sum, we have,

$$\int_B k'(|u|^2)\left\langle \Pi_{\partial\mathcal{M}_t}(U^\top, U^\top), U^\perp\right\rangle_g du = -\frac{1}{2}\int_B k(|u|^2)u^m du \sum_{i=1}^{m-1}\left\langle \Pi_{\partial\mathcal{M}}(\partial_i, \partial_i), \partial_m\right\rangle_g$$

and since the mean curvature is defined as $H(x) = \frac{1}{m-1}\sum_{i=1}^{m-1}\left\langle \Pi_{\partial\mathcal{M}}(\partial_i, \partial_i), \partial_m\right\rangle_g$ the first equality in (4.1) follows. Finally, substituting $u^m = -u \cdot \eta_x$ and extending the integral to all of $\{u \mid u^m > -b_x\} = \{u \mid u \cdot \eta_x < b_x\}$ by Lemma 3.3 we obtain the second equality of (4.1). □

We now compute the asymptotic expansion for points inside of $\mathcal{N}_\epsilon$. We introduce the following definition from [6] for the moments of a kernel function near the boundary,

$$m_\ell^\partial(x) = \int_{\{z\in\mathbb{R}^m \mid z\cdot\eta_x < b_x/\epsilon\}} (z\cdot\eta_x)^\ell k(|z|^2)\,dz = \int_{\mathbb{R}^{m-1}}\int_{-\infty}^{b_x/\epsilon} z_m^\ell k\left(|z|^2\right)\,dz_m dz_1\cdots dz_{m-1} \tag{4.2}$$

where η_x is a smooth extension of the boundary normal vector field into the normal collar of the boundary. For more information on these moments, see Appendix B. We can now state and prove the following theorem which includes a uniform asymptotic expansion for points near the boundary.

Theorem 4.6 *(Expansion near the boundary). Let $\mathcal{M} \subset \mathbb{R}^n$ be a compact m-dimensional C^3 Riemannian manifold with a C^3 boundary. Let $k : \mathbb{R} \to \mathbb{R}$ have exponential decay. Suppose that for fixed $\gamma \in (0,1)$, that $\epsilon^\gamma < \min\{\frac{r_\mathcal{M}}{C_1}, C_\mathcal{M}\}$ Then for all $x \in \mathcal{N}_\epsilon := \{x \in \mathcal{M} : d(x, \partial\mathcal{M}) \le \epsilon\}$, we have*

$$\epsilon^{-m}\int_{y\in\mathcal{M}} k\left(\frac{|x-y|^2}{\epsilon^2}\right)f(y)\,d\text{vol} = m_0^\partial(x)f(x) + \epsilon m_1^\partial(x)\left(\eta_x\cdot\nabla f(x) + \frac{m-1}{2}H(x)f(x)\right) + O(\epsilon^2) \tag{4.3}$$

where the moments $m_\ell^\partial(x)$ are defined in (4.2) and $H(x)$ is the mean curvature of the hypersurface parallel to $\partial\mathcal{M}$ intersecting x (which depends on the second fundamental form of $\partial\mathcal{M} \subset \mathcal{M}$).

Proof. First, by Lemma 3.3 we localize the integral to a semigeodesic ϵ-ball, B making an error of higher order than $O(\epsilon^2)$. Note that B is exactly the domain of the integral defining the coefficients $m_\ell^\partial(x)$ in (4.2). We then multiply three expansion. First, the kernel expansion,

$$k\left(\frac{||x-y||^2_{\mathbb{R}^d}}{\epsilon^2}\right) = k\left(\frac{||u||^2_{\mathcal{M}} - \left\langle \Pi_{\partial\mathcal{M}_t}(U^\top, U^\top), U^\perp\right\rangle_g + \omega_1(x,u) + O(|u|^5)}{\epsilon^2}\right)$$
$$= k\left(\frac{||u||^2_{\mathcal{M}}}{\epsilon^2}\right) - k'\left(\frac{||u||^2_{\mathcal{M}}}{\epsilon^2}\right)\frac{1}{\epsilon^2}\left\langle \Pi_{\partial\mathcal{M}_t}(U^\top, U^\top), U^\perp\right\rangle_g + O(\epsilon^{-2}|u|^4)$$

which follows from Lemma 4.3. Second, the Taylor expansion of f,

$$f(y) = f(x) + \frac{\partial f}{\partial u^i}u_i + O(|u|^2)$$

and finally, by Theorem 4.4 we have the following expansion of the volume form

$$d\mathrm{vol}(y) = 1 - (m-1)H(x)u^m + O(|u|^2).$$

The product of these three terms appears inside the integral, so multiplying the three expansions and making the change of variables $u \mapsto \epsilon u$, we find the order-ϵ^0 term is $k\left(|u|^2\right) f(x)$ which integrates to $m_0^\partial(x)f(x)$. The order-ϵ^1 term is,

$$\epsilon\int_B k\left(|u|^2\right)\left(\frac{\partial f}{\partial u^i}u^i - f(x)(m-1)H(x)u^m\right) - k'\left(|u|^2\right)\left\langle \Pi_{\partial\mathcal{M}_t}(U^\top, U^\top), U^\perp\right\rangle_g f(x)\,du$$
$$= \epsilon m_1^\partial(x)\nabla f(x)\cdot\eta_x + \epsilon m_1^\partial(x)(m-1)H(x)f(x) - \epsilon m_1^\partial(x)\frac{m-1}{2}H(x)f(x)$$
$$= \epsilon m_1^\partial(x)\left(\nabla f(x)\cdot\eta_x + \frac{m-1}{2}H(x)f(x)\right)$$

where the first equality comes from noting that u^i integrates to zero by symmetry for $1 \le i \le m-1$ and then applying Lemma 4.5. □

Having proven an asymptotic expansion for the points in the boundary region $\mathcal{N}_\epsilon$, we now turn our attention to the interior region $\mathcal{M}_\epsilon$. The expansion in this region was computed previously in [10,16]. Combining these previous results together with the bounds on ϵ proven in Lemma 3.3 yields the following uniform expansion.

Theorem 4.7 *(Expansion in the interior due to [10,16]). Let $\mathcal{M} \subset \mathbb{R}^n$ be a compact m-dimensional C^3 Riemannian manifold with a C^3 boundary. Let $k : \mathbb{R} \to \mathbb{R}$ have exponential decay. Suppose that for fixed $\gamma \in (0,1)$, that $\epsilon^\gamma < \min\{\frac{r_{\mathcal{M}}}{C_1}, C_{\mathcal{M}}\}$ Then for all $x \in \mathcal{M}_\epsilon := \{x \in \mathcal{M} : d(x, \partial\mathcal{M}) > \epsilon$, we have*

$$\epsilon^{-m}\int_{y\in\mathcal{M}} k\left(\frac{|x-y|^2}{\epsilon^2}\right) f(y)\ d\mathrm{vol} = m_0 f(x) + \frac{m_2}{2}\epsilon^2\left(S(x)f(x) + \Delta f(x)\right) + O(\epsilon^3) \tag{4.4}$$

where $m_0 = \int_{\mathbb{R}^m} k(|u|)\,du$ and $m_2 = \int_{\mathbb{R}^m} u_1^2 k(|u|)\,du$ are the zeroth and second moments of the kernel and $S(x) = \frac{1}{2}(-R(x) + \frac{1}{2}||\sum_a \Pi(\partial_a, \partial_a)||^2)$ depends on the scalar curvature R and the second fundamental form Π at x.

The results of Theorems 4.6 and 4.7, when taken together, provide a uniform asymptotic treatment of the operator $\mathcal{I}_\epsilon$. That is, show that for ϵ sufficiently small, the asymptotic behavior of $\mathcal{I}_\epsilon f(x)$ can be computed for all points $x \in \mathcal{M}$. This subtle but important notion of uniformity comes from Lemma 3.3.

We remark that for small enough values of ϵ, the both expansions (4.4) and (4.3) hold in normal collar $\mathcal{N}_{r_C}$ but outside of $\mathcal{N}_\epsilon$. To reconcile these two expansions, notice that for $b_x \gg \epsilon$ we have $m_0^\partial = m_0$ and $m_1^\partial = 0$ up to higher order terms in ϵ. Thus, outside of the ϵ neighborhood of the boundary (4.3) reduces to $m_0 f(x) + O(\epsilon^2)$ which is consistent with (4.4).

4.1. Uniformity and compactness

Before presenting numerical experiments supporting the expansions of the previous section, we will briefly comment on the role of compactness of $\mathcal{M}$ and uniformity of the expansions in the variable ϵ. The authors of this paper speculate that the uniform expansions of this section should be obtainable in the noncompact case assuming some mild conditions such as sectional curvature bounds on $\mathcal{M}$, uniform normal collar, and positive injectivity radius lower bounds.

The expansion (4.4) in Theorem 4.7 was proven in [16] for noncompact manifolds with boundary assuming bounds on the sectional curvature. However, the expansions hold nonuniformly, that is, for every $x \in \mathcal{M} \setminus \partial\mathcal{M}$, there exists an $\epsilon_x > 0$ depending on x such that for all $\epsilon < \epsilon_x$ the expansion holds. The only other requirement in their proof is that the embedding map ι has uniformly continuous inverse. This condition is equivalent to assuming a lower bound on minimum radius of curvature, which is the condition that appears in [16].

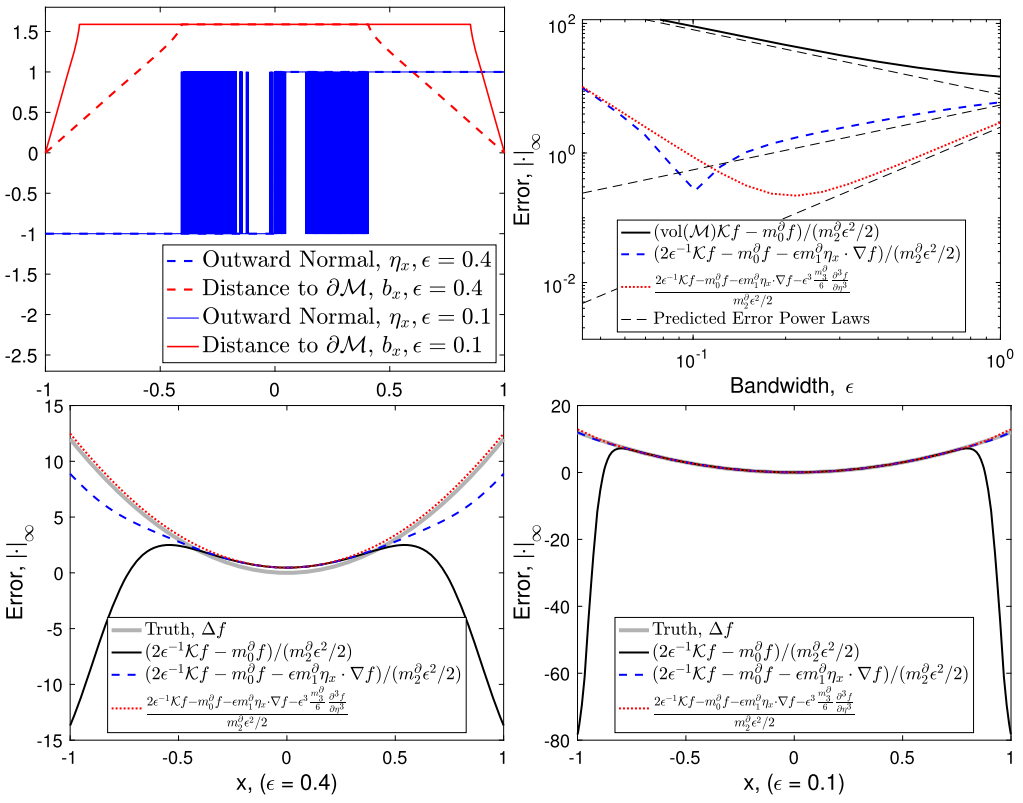

Fig. 1. Verifying (4.3) by extracting the Laplacian on the interval $[-1,1]$ applied to the function $f(x) = x^4$. Top, left: We show the estimate of b_x and η_x from the previous section, the b_x estimate saturates when $b_x \gg h$ and the η_x estimate is very noisy far from the boundary. Top, right: Error rates for various estimators of Δf extracted from $\mathcal{K}f$, for very small ϵ the quadrature error dominates. Bottom: True $\Delta f = 12x^2$ compared to various estimates for $\epsilon = 0.4$ (left) and $\epsilon = 0.1$ (right). (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

The fact that expansion (4.4) is uniform in the variable ϵ is a consequence of Lemma 3.3 in this paper, which requires compactness of $\mathcal{M}$ in two different parts, both of which could potentially be generalized to the noncompact case. First, one requires a universal lower bound on the injectivity radii of $\mathcal{M}_\epsilon$ for all ϵ as shown in the proof of Proposition 3.2. A sufficient condition for this to occur in the noncompact case is to assume that the double $D(\mathcal{M})$ admits a Riemannian metric which smoothly extends the metric on $\mathcal{M}$ and has a positive injectivity radius. This is manifestly true in the compact case.

Second, the argument in the proof of Proposition 3.1 computing a positive lower bound on the d_∂ function uses compactness of the closed collar region $\mathcal{N}_{r_{\text{sem}}}$. However, the authors conjecture that through using a triangle comparison argument using techniques from geometry of CAT(κ) spaces [2,1], one may be able to derive such a positive lower bound using only bounds on the curvature of $\mathcal{M}$ and inward-facing sectional curvature of $\partial\mathcal{M}$.

4.2. Examples

We now provide some simple numerical examples which verify the new boundary expansion in (4.3). We start with the interval, which is a flat manifold with a zero dimensional boundary, so the mean curvature $H(x) = 0$. We then consider a filled ellipse, so that the boundary has nontrivial curvature, but the manifold is still flat in the Riemannian sense.

Example 4.8 *(Interval).* In Fig. 1 we verify (4.3) using a uniform grid of $N = 5000$ data points on the interval $[-1,1]$ and the function $f(x) = x^4$. Since the grid is uniform, the density is $q(x) = 1/\mathrm{vol}(\mathcal{M}) = 1/2$ so in this simple example we can correct for the density by multiplying $\mathcal{K}$ by 2. After computing $2\mathcal{K}f$ we subtract the analytical value of $m_0^\partial(x)f(x)$ and divide by $\epsilon^2 m_2^\partial(x)/2$, which will agree with Δf in the interior of the manifold, but blows up like ϵ^{-1} near the boundary as shown by the solid black curves in Fig. 1. In order to obtain a consistent estimator we must also subtract the normal derivative term $m_1^\partial \eta_x \cdot \nabla f(x)$ as shown by the dashed blue curves.

In the next example, in order to eliminate the variance of a single random sample, we generated a uniform grid $\{x_i\}$ on $\mathcal{M}$ and then a very large set of uniform sampled random data points $\{y_j\}$ and computed the kernel $k\left(\frac{|x_i - y_j|^2}{\epsilon^2}\right)$ and the function $f(y_j)$ and then estimated the expected value by

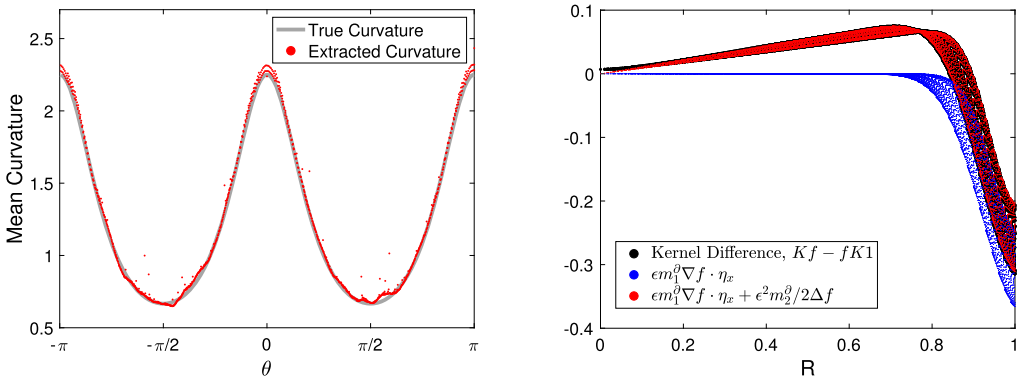

Fig. 2. Left: Verifying the mean curvature $H(x)$ in the order-ϵ term of (4.3) on the ellipse. **Right:** Verifying the derivative terms in the expansion (4.3).

$$\int_{\mathcal{M}} k\left(\frac{|x_i - y|^2}{\epsilon^2}\right) f(y)\, dV(y) = \mathbb{E}\left[k\left(\frac{|x_i - y|^2}{\epsilon^2}\right) f(y)\right] = \frac{1}{N}\sum_{j=1}^{N} k\left(\frac{|x_i - y_j|^2}{\epsilon^2}\right) f(y_j) + O(N^{-1/2}).$$

Since the average can be computed iteratively, this strategy allows us to compute the average over $N = 5 \times 10^7$ points and eliminate any variance (quadrature) error.

Example 4.9 *(Ellipse).* In this example we consider $\mathcal{M} = \{(\tilde{x}, \tilde{y}) \,|\, \tilde{x}^2/a^2 + \tilde{y}^2/b^2 \le 1\}$ with $a = 1, b = 2/3$ where we use $\tilde{x}, \tilde{y}$ to denote the coordinates since $x, y \in \mathbb{R}^d$ denote vectors. We note that this example is easy to sample uniformly by simply sampling points uniformly in $[0,1]^2$ and then selecting only the points that satisfy the inequality. We start by extracting the mean curvature term using the function $f \equiv 1$ so that $\nabla f \equiv 0 \equiv \Delta f$ and (4.3) becomes,

$$\epsilon^{-m} \int_{y \in \mathcal{M}} k\left(\frac{|x - y|^2}{\epsilon^2}\right) dV = m_0^\partial(x) + \epsilon m_1^\partial(x) \frac{m-1}{2} H(x) + \frac{\epsilon^2}{2} m_2 \tilde{\omega}(x) + O(\epsilon^3) \tag{4.5}$$

Using $\epsilon = 0.1$ (results were robust for $\epsilon \in [0.5, 0.15]$) we estimated the integral as described above and extracted the mean curvature term by subtracting m_0^∂ and dividing by $\epsilon m_1^\partial (m-1)/2$. In Fig. 2(left) we compare the extracted mean curvature with the following analytic derivation. Note that the boundary of the ellipse can be parameterized as $\iota(\theta) = (a\cos\theta, b\sin\theta)$ with first derivative (and tangent vector) $V^\top = D\iota(\theta) = (-a\sin\theta, b\cos\theta)$ so that the normal vector is $V^\perp = (b\cos\theta, a\sin\theta)$ and second derivative $D^2\iota(\theta) = (-a\cos\theta, -b\sin\theta)$. Thus, the projection of the second derivative onto the normal direction is $\iota''(\theta) \cdot \frac{V^\perp}{||V^\perp||} = \frac{-ab}{\sqrt{b^2\cos^2\theta + a^2\sin^2\theta}}$. However since we did not use a unit tangent vector we also need to divide by the norm-squared of the tangent vector which yields a mean curvature of

$$\begin{aligned}(m-1)H(x) &= \operatorname{trace}\left(\left\langle \Pi(U^\top, U^\top), U^\perp \right\rangle\right) = \operatorname{trace}\left(\left\langle \Pi\left(\frac{V^\top}{||V^\top||}, \frac{V^\top}{||V^\top||}\right), \frac{V^\perp}{||V^\perp||}\right\rangle\right) \\ &= \frac{\operatorname{trace}\left(\left\langle \Pi(V^\top, V^\top), V^\perp \right\rangle\right)}{||V^\top||^2 ||V^\perp||} = \frac{D^2\iota \cdot \frac{V^\perp}{||V^\perp||}}{||V^\top||^2} = \frac{ab}{(b^2\cos^2\theta + a^2\sin^2\theta)^{3/2}}\end{aligned} \tag{4.6}$$

which is simply the standard (extrinsic) curvature of the parameterized curve. This function is shown as the solid grey curve in Fig. 2(left) and compared to the empirically extracted curvature shown as red dots. This comparison is only valid for points near the boundary, and in Fig. 2(left) we only show points with distance to the boundary less than $\epsilon/4$.

Next we verify the derivative terms in (4.3) by defining a function on the ellipse by $f(\tilde{x}, \tilde{y}) = R^3$ where $R \equiv \sqrt{\tilde{x}^2/a^2 + \tilde{y}^2/b^2}$ so that $(\tilde{x}, \tilde{y}) = (aR\cos\theta, bR\sin\theta)$. The gradient $\nabla f = (\frac{\partial f}{\partial \tilde{x}}, \frac{\partial f}{\partial \tilde{y}})$ in the normal direction is

$$\nabla f \cdot \eta_x = 3R^2(\cos(\theta)/a, \sin(\theta)/b) \cdot \frac{V^\perp}{||V^\perp||} = 3R^2 \frac{(b/a)\cos^2\theta + (a/b)\sin^2\theta}{\sqrt{b^2\cos^2\theta + a^2\sin^2\theta}}$$

and the Laplacian is

$$\Delta f = \frac{\partial^2 f}{\partial \tilde{x}^2} + \frac{\partial^2 f}{\partial \tilde{y}^2} = 3R((\cos^2(\theta) + 1)/a^2 + (\sin^2(\theta) + 1)/b^2).$$

In order to eliminate the curvature terms, we note that multiplying (4.5) by $f(x)$ matches many of the terms from (4.3), so subtracting this from (4.3) we isolate the terms

$$\left(\mathbf{K}\vec{f}\right)_i - f(x_i)\left(\mathbf{K}\vec{1}\right)_i \to \epsilon^{-m}\int_{y\in\mathcal{M}} k\left(\frac{|x-y|^2}{\epsilon^2}\right) f(y)\,dV - f(x)\epsilon^{-m}\int_{y\in\mathcal{M}} k\left(\frac{|x-y|^2}{\epsilon^2}\right) dV$$

$$= \epsilon m_1^{\partial}(x)\nabla f(x)\cdot\eta_x + \epsilon^2\frac{m_2^{\partial}(x)}{2}\Delta f(x) \tag{4.7}$$

where the convergence is as the number of data points, $N\to\infty$. Using the averaging strategy described above to reduce variance, we estimate $\left(\mathbf{K}\vec{f}\right)_i - f(x_i)\left(\mathbf{K}\vec{1}\right)_i$ and compare the analytic expressions derived above in Fig. 2(right). This validates the derivative terms in (4.3).

5. Estimating boundary integrals

The purpose of this section is to use the distance to the boundary to construct a consistent estimator of the boundary integral $\mathcal{B}(\phi,f)=\int_{\partial\mathcal{M}}\phi(x)f(x)\,d\mathrm{vol}^{\partial}$ introduced in equation (1.6) as one of the key bilinear forms required for solving boundary value problems. It will turn out that the key result of this section will also be required in the next section to prove consistency of the graph Laplacian as an estimator of the Dirichlet energy, $\mathcal{E}$, in the weak sense.

We saw in the previous subsection that the standard graph Laplacian estimate of the Laplacian on a manifold is not consistent near the boundary. We now consider the weak form of the operators that the kernel matrix and graph Laplacian are estimating which requires a new result connecting the normal derivative term $m_1^{\partial}\eta_x\cdot\nabla f$ to a boundary integral. We first define a boundary integral estimator by

$$\mathcal{J}(f)=\frac{1}{\epsilon N}\sum_{i=1}^{N} k\left(\frac{b_{x_i}^2}{\epsilon^2}\right) f(x_i) \qquad\qquad \mathbb{E}[\mathcal{J}(f)]=\frac{1}{\epsilon}\int_{x\in\mathcal{M}} k\left(\frac{b_x^2}{\epsilon^2}\right) f(x)q(x)\,d\mathrm{vol}$$

where K is a kernel with exponential decay as above, for instance $K(z)=e^{-z^2}$ is the prototypical example. The expectation of the $\mathcal{J}$ functional is the integral over the entire manifold since we assume that the samples x_i yield a weighted quadrature on the manifold. However, the functional $\mathcal{J}$ uses the distance to the boundary b_x to weight the data points, so that only points near the boundary contribute significantly to the integral. In practice, in order to compute $\mathcal{J}$, we use the method described in Section Appendix B to estimate the distance to the boundary. We first show that $\mathcal{J}$ is a consistent estimator of a boundary integral.

For this result, it will now be convenient to use *boundary normal coordinates*, which are the special case of semigeodesic coordinates constructed on $\partial\mathcal{M}$. In this special case, we also will only need to parameterize the "height" of such charts, and so we will let ϵ parameterize only the n-th coordinate u^n in these charts.

Theorem 5.1. *In the same context as Theorem 4.6, let $d\mathrm{vol}_{\partial}$ be the natural volume element on the boundary inherited from $d\mathrm{vol}$, then we have*

$$\mathbb{E}[\mathcal{J}(f)]=\overline{m}_0\int_{y\in\partial\mathcal{M}} f(y)q(y)d\mathrm{vol}_{\partial}+\epsilon\overline{m}_1\int_{y\in\partial\mathcal{M}} f(y)q(y)H(y)-\eta_x\cdot\nabla(fq)(y)d\mathrm{vol}_{\partial}+O(\epsilon^2) \tag{5.1}$$

where $\overline{m}_0=\int_0^{\infty}k(u)\,du$ and $\overline{m}_1=\int_0^{\infty}uk(u)\,du$ and $H(y)$ is the mean curvature of $\partial\mathcal{M}$ at $y\in\partial\mathcal{M}$.

Proof. Let $0<\gamma<1$ and $\epsilon>0$ be such that ϵ^{γ} is less than the normal collar width R_C. In addition, let $\mathcal{N}_{\epsilon^{\gamma}}=\{y\in\mathcal{M}: b_y<\epsilon^{\gamma}\}$ denote the normal collar of width ϵ^{γ}. By using an identical argument as Lemma 3.3, one can localize the integral over $\mathcal{N}_{\epsilon^{\gamma}}$ so that

$$\frac{1}{\epsilon}\int_{\mathcal{M}} k\left(\frac{b_y^2}{\epsilon^2}\right) f(y)q(y)d\mathrm{vol}=\frac{1}{\epsilon}\int_{\mathcal{N}_{\epsilon}} k\left(\frac{b_y^2}{\epsilon^2}\right) f(y)q(y)d\mathrm{vol}+O(\epsilon^z)$$

for any choice of $z\in\mathbb{N}$.

Now let $\mathcal{U}=\{U_i\}_{i\in I}$ be a covering of $\partial\mathcal{M}$ in boundary normal coordinate charts. By taking intersections and complements, we can then generate a covering of $\mathcal{M}$ by measurable sets $\{V_j\}_{j\in J}$ of $\partial\mathcal{M}$ such that V_j are disjoint, and each contained in a single U_i. We then extend each V_j to a measurable subset of $\mathcal{M}$ by letting $\tilde{V}_j=\phi(V_j\times[0,\epsilon^{\gamma}))$. Therefore each V_j can be extended to a measurable set $\tilde{V}_j$ of $\mathcal{M}$ which is contained in a boundary normal coordinate chart map ϕ_{U_i}. The integral over the normal collar can then be parameterized as:

$$\int_{\mathcal{N}_{\epsilon^{\gamma}}} k\left(\frac{b_y^2}{\epsilon^2}\right) f(y)q(y)\,d\mathrm{vol}=\sum_{j\in J}\int_{\tilde{V}_j} k\left(\frac{b_y^2}{\epsilon^2}\right) f(y)q(y)\,d\mathrm{vol}. \tag{5.2}$$

In each of these charts, the coordinate representation of b_y is u^n. We then perform an order 1 Taylor expansion about $u^n=0$ of fq as well as a Taylor expansion of $d\mathrm{vol}$ about $u^n=0$ using the first variation of area:

$$\frac{1}{\epsilon}\int_{\tilde{V}_j} K\left(\frac{b_y^2}{\epsilon^2}\right) f(y)q(y)d\mathrm{vol}=\frac{1}{\epsilon}\int_{\phi_{U_i}(\tilde{V}_j)} k\left(\frac{(u^n)^2}{\epsilon^2}\right) f(u^{\top},u^n)q(u^{\top},u^n)\,d\mathrm{vol}(u^{\top},u^n)$$

$$
\begin{aligned}
&= \frac{1}{\epsilon}\Bigg(\int_{\phi_{U_i}(\tilde{V}_j)} k\left(\frac{(u^n)^2}{\epsilon^2}\right) f(u^\top,0)q(u^\top,0)\; d\mathrm{vol}(u^\top,0) \\
&+ \int_{\phi_{U_i}(\tilde{V}_j)} k\left(\frac{(u^n)^2}{\epsilon^2}\right) fq(u^\top,0)H(u^\top,0)u^n\; d\mathrm{vol}(u^\top,0) \\
&+ \int_{\phi_{U_i}(\tilde{V}_j)} k\left(\frac{(u^n)^2}{\epsilon^2}\right) \frac{\partial fq}{\partial u^n}(u^\top,0)u^n\; d\mathrm{vol}(u^\top,0) \\
&+ \int_{\phi_{U_i}(\tilde{V}_j)} k\left(\frac{(u^n)^2}{\epsilon^2}\right) \frac{\partial fq}{\partial u^n}(u^\top,0)H(u^\top,0)(u^n)^2\; d\mathrm{vol}(u^\top,0) \\
&+ \int_{\phi_{U_i}(\tilde{V}_j)} k\left(\frac{(u^n)^2}{\epsilon^2}\right) \omega(u^\top,\tilde{u}^n))(u^n)^2\; d\mathrm{vol}(u^\top,\tilde{u}^n)\Bigg)
\end{aligned}
$$

where $\omega(u^\top,\tilde{u}^n)$ is the sum of the second-order terms in both expansions with $0 \le \tilde{u}^n < \epsilon^\gamma$. Since $d\mathrm{vol} = \sqrt{|g|}$ in coordinates, and the n-th coordinate vector field $\partial_n = -\eta_y$ is orthogonal to each of the other coordinate vector fields, cofactor expansion of $\sqrt{|g|}$ implies that $d\mathrm{vol}(u^\top,0) = d\mathrm{vol}_\partial(u^\top)$. We can then separate terms involving u^n to obtain:

$$
\begin{aligned}
\frac{1}{\epsilon}\int_{\tilde{V}_j} K\left(\frac{b_y^2}{\epsilon^2}\right) f(y)q(y)d\mathrm{vol} &= \frac{1}{\epsilon}\int_{u^n=0}^{u^n=\epsilon^\gamma} k\left(\frac{u^n}{\epsilon}\right) du^n \int_{y\in V_j} f(y)q(y)\; d\mathrm{vol}_\partial \\
&+ \frac{1}{\epsilon}\int_{u^n=0}^{u^n=\epsilon^\gamma} k\left(\frac{u^n}{\epsilon}\right) u^n du^n \int_{V_j} f(y)q(y)H(y)\; d\mathrm{vol}_\partial \\
&+ \frac{1}{\epsilon}\int_{u^n=0}^{u^n=\epsilon^\gamma} k\left(\frac{u^n}{\epsilon}\right) u^n du^n \int_{V_j} -\eta_y\cdot\nabla f q(y)\; d\mathrm{vol}_\partial \\
&+ \frac{1}{\epsilon}\int_{u^n=0}^{u^n=\epsilon^\gamma} k\left(\frac{(u^n)^2}{\epsilon^2}\right) (u^n)^2 du^n \int_{y\in V_j} -\eta_y\cdot\nabla f q(y)H(y)\; d\mathrm{vol}_\partial \\
&+ \frac{1}{\epsilon}\int_{u^n=0}^{u^n=\epsilon^\gamma} k\left(\frac{(u^n)^2}{\epsilon^2}\right) (u^n)^2 du^n \int_{y\in W_j} \omega(u^\top,\tilde{u}^n)\; d\mathrm{vol}_{\partial\mathcal{M}^{\tilde{u}^n}}
\end{aligned}
$$

Where W_j is the coordinate image of $\partial\mathcal{M}^{\tilde{u}^n}$ in these coordinates (recall that $\partial\mathcal{M}_t$ indicates the hypersurface of points distance t away from $\partial\mathcal{M}$.)

Since the integral over V_j does not depend on ϵ, we may use exponential decay of the kernel to extend the integral over u^n to infinity. By then making a substitution $u^n \mapsto \epsilon u$, and letting $\overline{m}_0 = \int_0^\infty k(u)\,du$ and $\overline{m}_1 = \int_0^\infty k(u)u\,du$ we are left with:

$$
\begin{aligned}
\frac{1}{\epsilon}\int_{\tilde{V}_j} k\left(\frac{b_y}{\epsilon}\right) f(y)q(y)d\mathrm{vol} &= \overline{m}_0 \int_{\phi_{U_i}(V_j)} f(y)q(y)\; d\mathrm{vol}_\partial \\
&+ \epsilon\overline{m}_1 \int_{\phi_{U_i}(V_j)} f(y)q(y)H(y) - \eta_y\cdot\nabla f q(y)\; d\mathrm{vol}_\partial \\
&+ O(\epsilon^2)
\end{aligned}
$$

We remark that To compute the integral over the entire normal collar, we return to the parameterization of the integral in (5.2):

$$
\int_{\mathcal{N}_{\epsilon^\gamma}} k\left(\frac{b_y^2}{\epsilon^2}\right) f(y)q(y)\; d\mathrm{vol} = \sum_{j\in J}\int_{\tilde{V}_j} k\left(\frac{b_y^2}{\epsilon^2}\right) f(y)q(y)\; d\mathrm{vol}.
$$

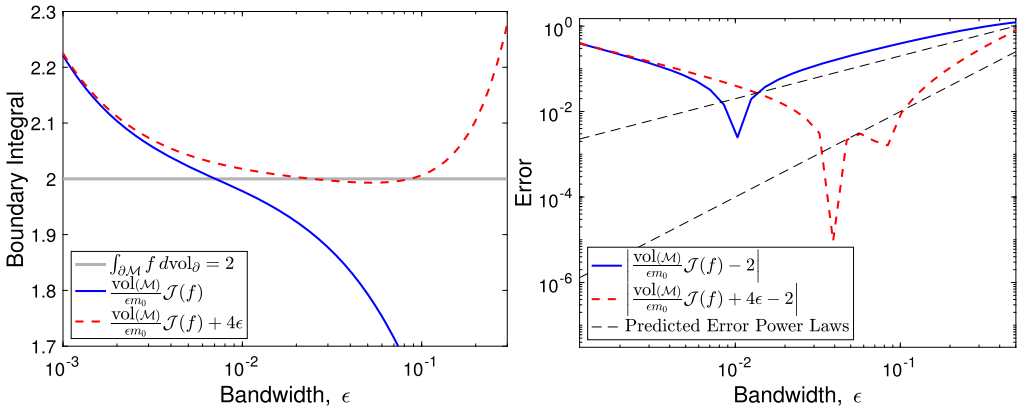

Fig. 3. Verifying (5.1) by extracting the boundary integral on the interval $[-1,1]$ applied to the function $f(x) = x^4$. Left: We show the estimate of the boundary integral compared to the true value, 2. Right: Error vs. bandwidth for various estimators of the boundary integral.

Summation over all $\tilde{V}_j$ in the manner above, we are left with:

$$\frac{1}{\epsilon}\int_{N_{\epsilon\mathcal{Y}}} k\left(\frac{b_y^2}{\epsilon^2}\right) f(y)q(y)d\text{vol} = \overline{m}_0 \int_{\partial\mathcal{M}} f(y)q(y)\, d\text{vol}_\partial + \epsilon\overline{m}_1 \int_{\partial\mathcal{M}} f(y)q(y)H(y) - \eta_y \cdot \nabla f q(y)\, d\text{vol}_\partial + O(\epsilon^2)$$

from which the result follows. □

Theorem 5.1 allows us to reinterpret integrals weighted by functions such as $m_1^\partial \propto e^{-b_x^2/\epsilon^2}$ (for which $\overline{m}_0 = \sqrt{\pi}/2$ and $\overline{m}_1 = 1/2$) as boundary integrals up to higher order terms. As in the previous section, this estimator is influenced by the sampling density q and we will correct this in the next section.

Returning to our example of the function $f(x) = x^4$ on the interval $[-1,1]$, in this case the boundary is the set $\{-1,1\}$ so the boundary integral is simply $\int_{\partial\mathcal{M}} x^4\, d\text{vol}_\partial = 1^4 + (-1)^4 = 2$. Using the estimator from Theorem 5.1 we can estimate this boundary integral as $\frac{\text{vol}(\mathcal{M})}{\epsilon m_0}\mathcal{J}(f)$ which will have error of order-ϵ as shown in Fig. 3. The next order term in the expansion (5.1) is $-\epsilon\overline{m}_1 \int_{x\in\partial\mathcal{M}} \eta_x \cdot \nabla(fq)(x) d\text{vol}_\partial = -4\epsilon$ so the error in $\frac{\text{vol}(\mathcal{M})}{\epsilon m_0}\mathcal{J}(f) + 4\epsilon$ should be order-ϵ^2 as shown in Fig. 3. Again, we emphasize that this example is purely for verification of Theorem 5.1, we will return to practical computation methods in Section 7.

6. Estimating the weak Laplacian

The standard method of estimating the Laplace operator is with a graph Laplacian,

$$\mathbf{L} = c(\mathbf{D} - \mathbf{K})$$

where $c = \epsilon^{-(m+2)}N^{-1}$ and $\mathbf{K}_{ij} = K(\epsilon, x_i, x_j)$ is the kernel matrix so that $(\mathbf{K}\vec{f})_i \propto \mathcal{K}f(x_i)$ where $\vec{f}_j = f(x_j)$. The matrix $\mathbf{D}$ is diagonal with $\mathbf{D}_{ii} = (\mathbf{K}\vec{1})_i \propto \mathcal{K}1(x_i)$. By Theorem 4.6, for x near the boundary we have

$$\epsilon^{-m}\mathbb{E}[\mathcal{K}(f)(x)] = \mathcal{I}(f)(x) = m_0^\partial(x)f(x)q(x) + m_1^\partial(x)\eta_x \cdot \nabla(fq)(x) + O(\epsilon^2)$$

and the expected value of $\mathbf{L}\vec{f}$ near the boundary is,

$$\begin{aligned}\mathbb{E}\left[(\mathbf{L}\vec{f})_i\right] &= \frac{\epsilon^{-m}\mathbb{E}[\mathbf{D}\vec{f}] - \epsilon^{-m}\mathbb{E}[\mathbf{K}\vec{f}]}{N\epsilon^2} \\ &= \frac{\epsilon^{-m}\mathbb{E}[f(x_i)\mathcal{K}(1)(x_i)] - \epsilon^{-m}\mathbb{E}[\mathcal{K}(f)(x_i)]}{N\epsilon^2} \\ &= -\epsilon^{-1}m_1^\partial(x_i)q(x_i)\eta_{x_i} \cdot \nabla f(x_i) + O(1).\end{aligned}$$

Notice that the estimator blows up like ϵ^{-1} near the boundary. This was first pointed out in [10] who derived the first two terms of (4.3). They argued that the graph Laplacian is a consistent estimator for Neumann functions (where $\eta_{x_i} \cdot \nabla f(x_i) = 0$), which is true but does not explain the empirically observed fact that the eigenvectors of $\mathbf{L}$ approximate the Neumann eigenfunctions of the Laplacian. While this pointwise blow-up seems discouraging, notice that it is blowing up with rate ϵ^{-1} in a neighborhood of the boundary that has volume of order ϵ. In this section we show that in fact, considered in the weak sense, the graph Laplacian $\mathbf{L}$ does have a well defined limit, namely the Dirichlet energy or weak Laplacian.

We now use (4.4) and (4.3) together with (5.1) to understand the weak form of the kernel operator and the graph Laplacian. In order to make this connection, we will require the following surprising result connecting the integral of the first moment near the boundary to the second moment in the interior.

Lemma 6.1. *For any $z \in \mathbb{N}$,*

$$\int_0^\infty m_1^\partial(b_x)\, db_x = \frac{m_2}{2} + O(\epsilon^z)$$

The proof of Lemma 6.1 is included in Appendix A.

When we apply the kernel operator to a function f and evaluate at all the data points we obtain a vector $\frac{\epsilon^{-m}}{N}(\mathbf{K}\vec{f})_i = \mathcal{K}f(x_i)$ which is simply the kernel matrix multiplied by the vector representation of the function $\vec{f}_i = f(x_i)$. If we then take another function $\vec{\phi}_i = \phi(x_i)$ and compute the inner product, the expectation will be a second integral,

$$\frac{\epsilon^{-m}}{N^2}\mathbb{E}\left[\vec{\phi}^\top \mathbf{K}\vec{f}\right] = \epsilon^{-m}\int_{x\in\mathcal{M}}\int_{y\in\mathcal{M}} \phi(x)q(x)K\left(\frac{|x-y|}{\epsilon}\right) f(y)q(y)\, d\text{vol}\, d\text{vol}. \tag{6.1}$$

By expanding the inner integral using (4.3) and applying Theorem 5.1 we derive the following result.

Theorem 6.2. *In the same context as Theorem 4.6 for all $\phi, f \in C^3(\mathcal{M})$ we have,*

$$\frac{2}{m_2 N}\mathbb{E}\left[\vec{\phi}^\top \mathbf{L}\vec{f}\right] = \int_{\mathcal{M}} (\nabla\phi\cdot\nabla f)\, q^2 d\text{vol} + O(\epsilon). \tag{6.2}$$

Proof. Recall that $\mathbf{L} = \epsilon^{-(m+2)}N^{-1}(\mathbf{D}-\mathbf{K})$ and $\mathbf{D}_{ii} = \sum_{j=1}^N \mathbf{K}_{ij}$ so that,

$$\begin{aligned}
\frac{1}{N}\mathbb{E}\left[\vec{\phi}^\top \mathbf{L}\vec{f}\right] &= \mathbb{E}\left[\frac{\epsilon^{-m-2}}{N^2}\sum_{i,j=1}^N \phi(x_i)K(\epsilon,x_i,x_j)(f(x_i)-f(x_j))\right] \\
&= \epsilon^{-m-2}\int_{x\in\mathcal{M}}\int_{y\in\mathcal{M}} q(x)\phi(x)K(\epsilon,x,y)(f(x)-f(y))q(y)\, d\text{vol}\, d\text{vol} \\
&= \epsilon^{-m-2}\int_{x\in\mathcal{M}} q(x)\phi(x)(f(x)\mathcal{K}(1)(x)-\mathcal{K}(f)(x))\, d\text{vol}
\end{aligned}$$

We now segment the manifold into the disjoint union $\mathcal{M} = M_{\epsilon^\gamma}\cup N_{\epsilon^\gamma}$ where $N_{\epsilon^\gamma} = B_{\epsilon^\gamma}(\partial\mathcal{M})$ is a neighborhood of the boundary and $M_{\epsilon^\gamma} = \mathcal{M}\backslash N_{\epsilon^\gamma}$ is the interior region. By Theorem 4.4, we see that the integral over the interior region is:

$$\begin{aligned}
&\epsilon^{-m-2}\int_{x\in M_{\epsilon^\gamma}} \phi(x)q(x)\left(f(x)\mathcal{I}_\epsilon(1)(x)-\mathcal{I}_\epsilon(f)(x)\right)\, d\text{vol} \\
&\quad= \epsilon^{-2}\int_{x\in M_{\epsilon^\gamma}} \phi(x)q(x)f(x)\left(m_0 q(x)+\frac{\epsilon^2 m_2}{2}(S(x)q(x)+\Delta q(x))+O(\epsilon^3)\right) \\
&\qquad -\phi(x)q(x)\left(m_0 f(x)q(x)+\frac{\epsilon^2 m_2}{2}\left(f(x)q(x)S(x)+\Delta(fq)(x)\right)+O(\epsilon^3)\right)\, d\text{vol} \\
&\quad= -\frac{m_2}{2}\int_{x\in M_{\epsilon^\gamma}} q(x)^2\phi(x)\Delta f(x)\, d\text{vol} - \frac{m_2}{2}\int_{x\in M_{\epsilon^\gamma}} 2q(x)\phi(x)\nabla f\cdot\nabla q\, d\text{vol} + O(\epsilon).
\end{aligned}$$

Moreover, applying the divergence theorem (for the negative definite Laplacian $\Delta = \text{div}\circ\nabla$, so that $\int_\partial h\eta\cdot\nabla f = \int \text{div}(h\nabla f) = \int h\Delta f + \nabla h\cdot\nabla f\, d\text{vol}$) we have,

$$\begin{aligned}
-\int_{M_{\epsilon^\gamma}} q^2\phi\Delta f\, d\text{vol} &= \int_{M_{\epsilon^\gamma}} \nabla(q^2\phi)\cdot\nabla f\, d\text{vol} - \int_{\partial M_{\epsilon^\gamma}} q^2\phi\eta\cdot\nabla f\, d\text{vol}_\partial \\
&= \int_{M_{\epsilon^\gamma}} q^2\nabla\phi\cdot\nabla f + 2q\phi\nabla q\cdot\nabla f\, d\text{vol} - \int_{\partial M_{\epsilon^\gamma}} q^2\phi\eta\cdot\nabla f\, d\text{vol}_\partial
\end{aligned}$$

and combining this with the previous equation we have,

$$\epsilon^{-m-2} \int_{x\in\mathcal{M}_{\epsilon^\gamma}} \phi(x)q(x)\left(f(x)\mathcal{I}_\epsilon(1)(x) - \mathcal{I}_\epsilon(f)(x)\right)\, d\text{vol}$$
$$= \frac{m_2}{2}\int_{\mathcal{M}_{\epsilon^\gamma}} q^2\nabla\phi\cdot\nabla f\, d\text{vol} - \frac{m_2}{2}\int_{\partial\mathcal{M}_{\epsilon^\gamma}} q^2\phi\eta\cdot\nabla f\, d\text{vol}_\partial + O(\epsilon) \tag{6.3}$$

By Theorem 4.3, the integral over the boundary region is:

$$\epsilon^{-m-2}\int_{x\in\mathcal{N}_{\epsilon^\gamma}} \phi(x)q(x)\left(f(x)\mathcal{I}_\epsilon(1)(x) - \mathcal{I}_\epsilon(f)(x)\right)\, d\text{vol}$$
$$= \epsilon^{-2}\int_{x\in\mathcal{N}_{\epsilon^\gamma}} q(x)\phi(x)\Big(m_0^\partial(x)q(x)f(x) + \epsilon m_1^\partial(x)f(x)\eta_x\cdot\nabla q(x) - \frac{m-1}{2}H(x)f(x)q(x)$$
$$- m_0^\partial(x)q(x)f(x) - \epsilon m_1^\partial(x)\eta_x\cdot\nabla(fq)(x) + \frac{m-1}{2}H(x)f(x)q(x) + O(\epsilon^2)\Big)\, d\text{vol}$$
$$= -\epsilon^{-1}\int_{\mathcal{N}_{\epsilon^\gamma}} m_1^\partial(x)q(x)^2\phi(x)\eta_x\cdot\nabla f(x) + O(\epsilon)\, d\text{vol}.$$

We now notice that the function $m_1^\partial(x)$ is purely a function of the distance to the boundary b_x and that it decays to zero exponentially as $b_x\to\infty$. Hence, we may apply the result of Theorem 5.1. We then have that,

$$\epsilon^{-m-2}\int_{x\in\mathcal{N}_{\epsilon^\gamma}} \phi(x)q(x)\left(f(x)\mathcal{I}_\epsilon(1)(x) - \mathcal{I}_\epsilon(f)(x)\right)\, d\text{vol}$$
$$= -\epsilon^{-1}\int_{\mathcal{N}_{\epsilon^\gamma}} m_1^\partial(x)q(x)^2\phi(x)\eta_x\cdot\nabla f(x) + O(\epsilon)\, d\text{vol}$$
$$= -\frac{m_2}{2}\int_{\partial\mathcal{M}} q(x)^2\phi(x)\eta_x\cdot\nabla f(x)\, d\text{vol}_\partial + O(\epsilon), \tag{6.4}$$

where the integral of $m_1^\partial(x)$ is exactly $\frac{m_2}{2}$ by Lemma 6.1. Adding (6.3) and (6.4) we have,

$$\frac{1}{N}\mathbb{E}\left[\vec{\phi}^\top\mathbf{L}\vec{f}\right] = \frac{m_2}{2}\int_{\mathcal{M}_{\epsilon^\gamma}} q^2\nabla\phi\cdot\nabla f\, d\text{vol} - \frac{m_2}{2}\int_{\partial\overline{\mathcal{N}}_{\epsilon^\gamma}} q^2\phi\eta\cdot\nabla f\, d\text{vol}_\partial + O(\epsilon) \tag{6.5}$$

where $\overline{\mathcal{N}}_{\epsilon^\gamma}$ is the closure of the boundary region, and its boundary is the union of the boundary of the interior region with the boundary of the manifold, $\partial\overline{\mathcal{N}}_{\epsilon^\gamma} = \partial\mathcal{M}_{\epsilon^\gamma}\cup\partial\mathcal{M}$. We then apply the Divergence theorem on the closure of the boundary region $\overline{\mathcal{N}}_{\epsilon^\gamma}$.

$$\int_{\partial\overline{\mathcal{N}}_{\epsilon^\gamma}} q^2\phi\eta\cdot\nabla f\, d\text{vol}_\partial = \int_{\overline{\mathcal{N}}_{\epsilon^\gamma}} \text{div}(q^2\phi\nabla f)\, d\text{vol}_\partial$$
$$= \int_{x\in\mathcal{M}} 1_{B_{\epsilon^\gamma}(\partial\mathcal{M})}(x)\text{div}(q^2\phi\nabla f)\, d\text{vol}_\partial$$
$$= O(\epsilon)$$

by Theorem 5.1, since the indicator function $1_{B_{\epsilon^\gamma}(\partial\mathcal{M})}(x)$ has exponential decay in the distance to the boundary as required by Theorem 5.1. Gathering the order-ϵ terms and dividing by $\frac{m_2}{2}$ we have,

$$\frac{2}{m_2 N}\mathbb{E}\left[\vec{\phi}^\top\mathbf{L}\vec{f}\right] = \int_{\mathcal{M}_{\epsilon^\gamma}} q^2\nabla\phi\cdot\nabla f\, d\text{vol} + O(\epsilon)$$

Similarly, we can add and subtract the integral of the same integrand over the boundary region (since this is order-ϵ) and we obtain,

$$\frac{2}{m_2 N}\mathbb{E}\left[\vec{\phi}^\top\mathbf{L}\vec{f}\right] = \int_{\mathcal{M}} q^2\nabla\phi\cdot\nabla f\, d\text{vol} + O(\epsilon)$$

as desired. □

Returning to our simple example of the uniform grid on the interval $[-1,1]$ with $f(x)=x^4$, we verify (6.2) by estimating $\int_{\mathcal{M}}|\nabla f|^2\, d\text{vol} = 32/7$ using the graph Laplacian $\mathbf{L}$. Notice that the optimal bandwidth for the weak sense estimator is much smaller than the optimal bandwidth for the pointwise estimator which results from the double summation being a lower variance estimator as shown in [7] (Fig. 4).

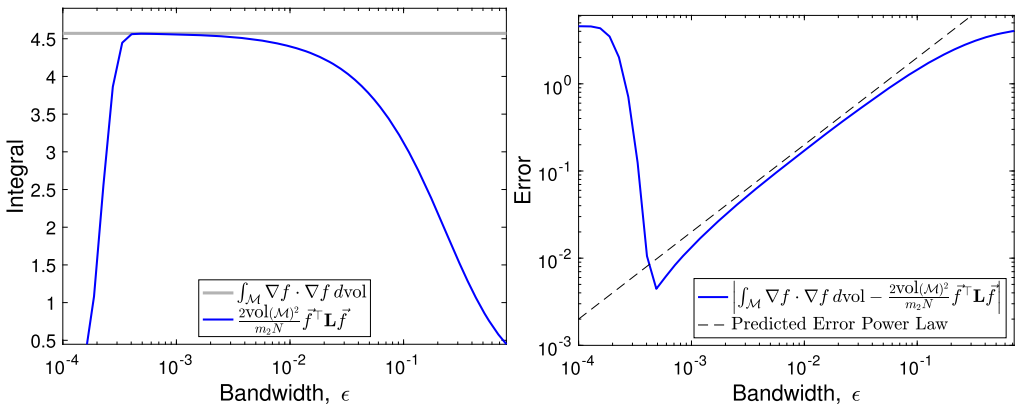

Fig. 4. Verifying (6.2) computing the exact integral and the graph Laplacian estimator on the interval $[-1, 1]$ applied to the function $\phi(x) = f(x) = x^4$. Left: We compare the estimate to the true integral value as a function of the bandwidth parameter, ϵ. Right: Error vs. bandwidth for the graph Laplacian as a weak-sense estimator.

The perhaps surprising conclusion of this section is that, even though the graph Laplacian is not a consistent pointwise estimator of the Laplacian for manifolds with boundary, it is a consistent weak-sense estimator. We should note that, if one simply removes the boundary and considers the interior, the graph Laplacian will be consistent pointwise at each point of the interior, however the rates of convergence will not be uniform, and the bandwidth required for pointwise consistency will decrease to zero as you approach the boundary.

6.1. Correcting for the sampling density

In (4.4), (4.3) and (6.2) all the terms are influenced by the density q, so to remove this influence we apply the 'right-normalization' introduced by [10]. The idea is to apply (4.3) to the function $f \equiv 1$ in order to extract a density estimate. Computationally this means multiplying the kernel matrix by a vector $\vec{1}$ of all ones, or equivalently summing the rows of the kernel matrix, which are the diagonal entries of the diagonal matrix $\mathbf{D}_{ii} = (\mathbf{K}\vec{1})_i$. We then normalize the kernel matrix $\mathbf{K}$ to form the matrices $\hat{\mathbf{K}}, \hat{\mathbf{D}}, \hat{\mathbf{L}}$ given by,

$$\hat{\mathbf{K}} = \mathbf{D}^{-1}\mathbf{K}\mathbf{D}^{-1} \qquad \text{and} \qquad \hat{\mathbf{D}}_{ii} = (\hat{\mathbf{K}}\vec{1})_i \qquad \text{and} \qquad \hat{\mathbf{L}} = \epsilon^{m-2} N^{-1} (\hat{\mathbf{D}} - \hat{\mathbf{K}}).$$

Notice that since $\mathbf{D}_{ii}$ is proportional to $q(x_i)$ we are essentially pre-dividing by a consistent estimator of $q(x)^2$. The next theorem shows that this normalization, introduced by [10], produces a consistent estimator of the Dirichlet energy that is independent of the sampling density of the points x_i.

Theorem 6.3. *In the same context as Theorem 4.6 for all $\phi, f \in C^3(\mathcal{M})$ we have,*

$$\frac{2m_0^2}{m_2 N} \mathbb{E}\left[\vec{\phi}^\top \hat{\mathbf{L}} \vec{f}\right] = \int_{\mathcal{M}} \nabla\phi \cdot \nabla f \, d\text{vol} + O(\epsilon). \tag{6.6}$$

Proof. Set $\hat{q}(x) = \epsilon^{-m}\mathcal{K}(1)(x) = \epsilon^{-m} \sum_{j=1}^N K(\epsilon, x, x_j)$ so that $\hat{q}(x_i) = \epsilon^{-m}\mathbf{D}_{ii}$ and we have

$$\begin{aligned}
\frac{1}{N}\mathbb{E}\left[\vec{\phi}^\top \hat{\mathbf{L}} \vec{f}\right] &= \mathbb{E}\left[\frac{\epsilon^{-m-2}}{N^2} \sum_{i,j=1}^N \phi(x_i) \frac{K(\epsilon, x_i, x_j)}{\hat{q}(x_i)\hat{q}(x_j)} (f(x_i) - f(x_j))\right] \\
&= \epsilon^{-m-2} \int_{x\in\mathcal{M}} \int_{y\in\mathcal{M}} \frac{q(x)}{\hat{q}(x)} \phi(x) K(\epsilon, x, y)(f(x) - f(y)) \frac{q(y)}{\hat{q}(y)} \, d\text{vol} \, d\text{vol} \\
&= \epsilon^{-m-2} \int_{x\in\mathcal{M}} \frac{q(x)}{\hat{q}(x)} \phi(x)(f(x)\mathcal{K}(1/\hat{q})(x) - \mathcal{K}(f/\hat{q})(x)) \, d\text{vol}
\end{aligned}$$

Note that by Theorems 4.4 and B.1, on the interior of the manifold we have $\hat{q}(x) = m_0 q(x) + O(\epsilon)$ and on the boundary we have $\hat{q}(x) = m_0^\partial(x) q(x) + O(\epsilon)$. So for x in the interior region, the above reduces to

$$\epsilon^{-m-2} \int_{x\in\mathcal{M}} \phi(x)(f(x)\mathcal{K}(1/\hat{q})(x) - \mathcal{K}(f/\hat{q})(x)) \, d\text{vol} = -\frac{m_2}{2m_0^2} \int_{x\in\mathcal{M}_{\epsilon\gamma}} \phi(x)\Delta f(x) \, d\text{vol}$$

and for x in the boundary region we have,

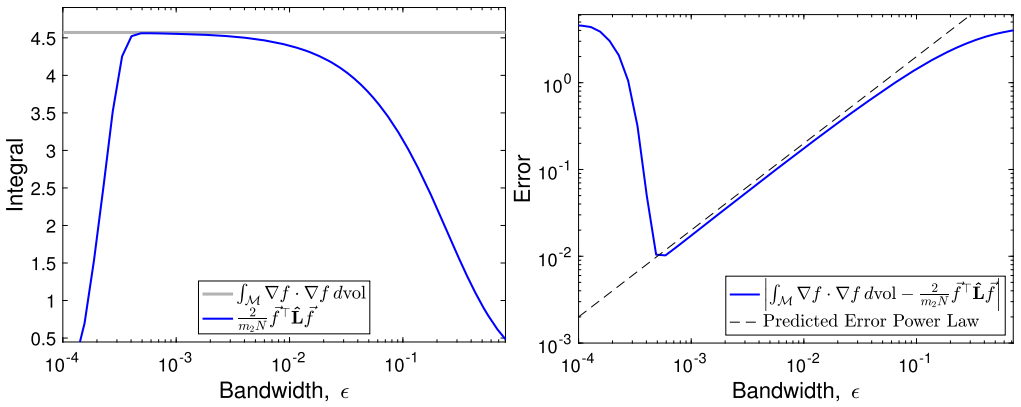

Fig. 5. Verifying (6.6) computing the exact integral and the graph Laplacian estimator on the interval $[-1,1]$ applied to the function $\phi(x) = f(x) = x^4$. Left: We compare the estimate to the true integral value as a function of the bandwidth parameter, ϵ. Right: Error vs. bandwidth for the graph Laplacian as a weak-sense estimator.

$$\frac{\epsilon^{-m-2}}{m_0^2} \int_{x \in \mathcal{M}} \phi(x)(f(x)\mathcal{K}(1/\hat{q})(x) - \mathcal{K}(f/\hat{q})(x)) \, d\text{vol} = -\epsilon^{-1} \int_{\mathcal{N}_{\epsilon^\gamma}} \frac{m_1^\partial(x)}{m_0^\partial(x)} \phi(x) \eta_x \cdot \nabla f(x) + O(\epsilon) \, d\text{vol}$$

and since $\frac{m_1^\partial}{m_0^\partial}$ has the necessary decay away from the boundary, it satisfies the hypotheses if Theorem 5.1 so that,

$$\int_{\mathcal{N}_{\epsilon^\gamma}} \frac{m_1^\partial(x)}{m_0^\partial(x)} \phi(x) \eta_x \cdot \nabla f(x) + O(\epsilon) \, d\text{vol} = \frac{m_2}{2m_0^2} \int_{\partial \mathcal{M}} \phi(x) \eta_x \cdot \nabla f(x) \, d\text{vol}_\partial$$

since the integral of m_1^∂ is $\frac{m_2}{2}$ by Lemma 6.1. Recombining the integral over the interior region and the integral of the boundary region we have,

$$\begin{aligned} \frac{2m_0^2}{N m_2} \mathbb{E}\left[\vec{\phi}^\top \hat{\mathbf{L}} \vec{f}\right] &= -\int_{x \in \mathcal{M}_{\epsilon^\gamma}} \phi(x) \Delta f(x) \, d\text{vol} + \int_{\partial \mathcal{M}} \phi(x) \eta_x \cdot \nabla f(x) \, d\text{vol}_\partial + O(\epsilon) \\ &= \int_{\mathcal{M}_{\epsilon^\gamma}} \nabla \phi \cdot \nabla f \, d\text{vol} + \int_{\partial \mathcal{M}_{\epsilon^\gamma} \cup \partial \mathcal{M}} \phi \eta_x \cdot \nabla f \, d\text{vol}_\partial + O(\epsilon) \end{aligned}$$

and as in the proof of Theorem 6.2, the boundary integral term above can be converted into an integral of the boundary region $\mathcal{N}_{\epsilon^\gamma}$ by the divergence theorem, and this integral is order-ϵ by Theorem 5.1. Similarly, extending the integral $\int_{\mathcal{M}_{\epsilon^\gamma}} \nabla \phi \cdot \nabla f \, d\text{vol}$ to all of $\mathcal{M}$ also creates an error of order-ϵ so we obtain (6.6). □

To verify (6.6) in our simple example on $[-1,1]$ we started with a uniform grid of $N = 5000$ points and then applied the nonlinear transformation $(x + .05)^{1.2}$ to each point and then shift and scale the resulting grid back to $[-1,1]$. The result is a nonuniform grid with higher density near -1 and lower density near 1. In Fig. 5 we show that the normalized graph Laplacian $\hat{\mathbf{L}}$ constructed in this section recovers the weak-sense Laplacian for the same example as in the previous section on this nonuniform grid.

7. Applications to elliptic and parabolic PDEs

Now that we have shown consistency of the estimator of the weak Laplacian, we provide numerical experiments to demonstrate the validity of this method. For simplicity of presentation, we shall write our formulation for problems with Dirichlet and Neumann boundary conditions. However, the approach could be applied to problems with Robin or mixed boundary conditions.

It should be noted that while the previous results of this paper prove consistency of the method, they make no mention of explicit error bounds in terms of the number of data points used in the estimator. In order to provide such bounds, an analysis of the variance of the estimator. Similar variance analyses have been done in the non boundary case in for instance [16,29,30] as well as many others. Although such a variance analysis likely would rely on the results of this paper, it would require a different set of techniques and thus will be explored in a future work.

7.1. Continuous PDEs

Let $\Omega \subset \mathbb{R}^n$ be an open bounded set with Lipschitz boundary $\partial\Omega$. If $L^2(\Omega)$ denotes the set of square integrable functions then we define the Sobolev space $H^1(\Omega) := \{v \in L^2(\Omega) : \nabla v \in L^2(\Omega)\}$ where $\nabla v = (\partial_{x_i} v)_{i=1}^n$ denotes the weak gradient. We also define a closed subspace of $H^1(\Omega)$ as $H_0^1(\Omega)$ which is the set functions in $H^1(\Omega)$ which are zero on $\partial\Omega$ in the trace sense. We shall denote the dual

space of $H^1(\Omega)$ and $H_0^1(\Omega)$ by $H^1(\Omega)^*$ and $H^{-1}(\Omega)$, respectively and the duality pairing between these spaces as $\langle\cdot,\cdot\rangle$. Finally $H^{\frac{1}{2}}(\partial\Omega)$ denotes the standard fractional order space.

We start with the elliptic Dirichlet problem: Given $f \in H^{-1}(\Omega)$, $g \in H^{\frac{1}{2}}(\partial\Omega)$ we are interested in solving

$$\begin{cases} -\Delta u &= f \quad \text{in } \Omega \\ u &= g \quad \text{on } \partial\Omega. \end{cases} \tag{7.1}$$

We impose the nonzero boundary condition using the classical lifting argument: Let $\tilde{g} \in H^1(\Omega)$ denote an extension of g to Ω. Notice that due to trace theorem, such an extension exists. We then write $u = w + \tilde{g}$ where $w|_{\partial\Omega} = 0$ in the trace sense. Then by Lax-Milgram Theorem, under the stated assumptions on the data f, g, there exists a unique weak solution $w \in H_0^1(\Omega)$ to (7.1) in the following sense

$$\int_\Omega \nabla w \cdot \nabla v \, dx = \langle f, v\rangle - \int_\Omega \nabla\tilde{g} \cdot \nabla v \, dx \quad \forall v \in H_0^1(\Omega). \tag{7.2}$$

Next we turn to the Neumann boundary value problem. Given $f \in H^1(\Omega)^*$, $g \in L^2(\partial\Omega)$, we consider

$$\begin{cases} -\Delta u + u &= f \quad \text{in } \Omega \\ \nabla u \cdot \eta &= g \quad \text{on } \partial\Omega \end{cases} \tag{7.3}$$

where η denotes the outward unit normal to $\partial\Omega$. Again by Lax-Milgram Theorem it is not difficult to see that under the stated assumptions on the data f, g there exists a unique weak solution $u \in H^1(\Omega)$ to (7.3) in the following sense:

$$\int_\Omega \nabla u \cdot \nabla v + uv \, dx = \langle f, v\rangle + \int_{\partial\Omega} gv \, ds \quad \forall v \in H^1(\Omega). \tag{7.4}$$

We also state the parabolic homogeneous Dirichlet problem, the Neumann problem is similar and is omitted for brevity: Given $f \in L^2(0,T; H^{-1}(\Omega))$ and $u_0 \in L^2(\Omega)$, we consider

$$\begin{cases} \partial_t u - \Delta u + u &= f \quad \text{in } \Omega\times(0,T) \\ u &= 0 \quad \text{on } \partial\Omega\times(0,T) \\ u &= u_0 \quad \text{in } \Omega\,. \end{cases} \tag{7.5}$$

The notion of weak solution to (7.5) is: find $u \in L^2(0,T; H_0^1(\Omega)) \cap H^1(0,T; H^{-1}(\Omega))$ solving

$$\langle \partial_t u, v\rangle + \int_\Omega \nabla u \cdot \nabla v \, dx = \langle f, v\rangle \quad \forall v \in H_0^1(\Omega) \tag{7.6}$$

and almost every $t \in (0,T)$.

7.2. Discrete system

Next we describe the linear algebraic systems we obtain after discretization of (7.2), (7.4), and (7.6). Assume that we have N nodes sampled on the manifold, and recall that the N-by-N matrix $\hat{\mathbf{L}}$ and diagonal N-by-N matrix $\mathbf{D}$ denote the discrete form of the Laplacian (in weak form) and the discretization of the integral over the entire manifold $\mathcal{M}$, respectively. Namely,

$$\int_{\mathcal{M}} \nabla u \cdot \nabla v \approx \mathbf{v}^\top \hat{\mathbf{L}}\mathbf{u} \quad \text{and} \quad \int_{\mathcal{M}} fv \approx \mathbf{v}^\top \mathbf{D}\mathbf{f}.$$

We indicate the discrete boundary integral $\int_{\partial\mathcal{M}}$ as

$$\int_{\partial\mathcal{M}} gv = \mathbf{v}^\top \mathbf{B}\mathbf{g}$$

where $\mathbf{B}$ is a diagonal N-by-N matrix, whose diagonal entries are very close to zero for nodes that are far from the boundary, and $\mathbf{v}, \mathbf{g}$ specify the values of functions on all the nodes in the data set, but can be set to zero on interior nodes for the purposes of estimating the boundary integral since the corresponding diagonal entries of B will be very close to zero. Let N_{int} denote the number of interior degrees of freedom. We note that the interior degrees of freedom can be identified using the estimated distance to boundary function as the nodes, x_i, such that $b_{x_i} > \epsilon/2$. Choosing $\epsilon/2$ means that the boundary layer will be half the width of the ϵ tube around the boundary, which we found empirically to be an effective width. The discrete form of (7.2) is given by

$$\hat{\mathbf{L}}(\mathrm{f}_{\text{dof}}, \mathrm{f}_{\text{dof}})\mathbf{w} = \mathbf{D}(\mathrm{f}_{\text{dof}}, :)\mathbf{f} - \hat{\mathbf{L}}(\mathrm{f}_{\text{dof}}, :)\tilde{\mathbf{g}}$$

where f_{dof} indicates the interior degrees of freedom. Then the discrete solution on the interior nodes is $\mathbf{u}(\mathrm{f}_{\text{dof}}) = \mathbf{w}$. We note that f_{dof} refers to the indices of the interior nodes and the colon denotes the inclusion of all the nodes. So the $\hat{\mathbf{L}}(\mathrm{f}_{\text{dof}}, \mathrm{f}_{\text{dof}})$ matrix on the left

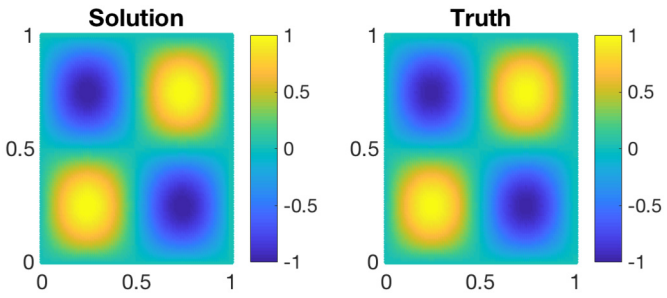

Fig. 6. Left: Solution computed using our approach. Right: Exact solution.

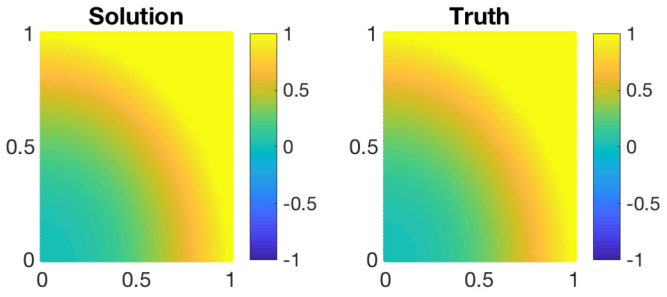

Fig. 7. Left: Solution computed using our approach. Right: Exact solution.

hand side of the above equation is N_{int}-by-N_{int}, whereas the matrices on the right hand side are N_{int}-by-N. Note that while these matrices can be quite large (depending on the number of nodes used to sample the domain) both $\mathbf{D}$ and $\mathbf{B}$ are diagonal, and $\hat{\mathbf{L}}$ is well approximated by a sparse matrix due to the fast decay of the kernel function used to construct it.

The discrete form of the system (7.4) is

$$(\hat{\mathbf{L}} + \mathbf{D})\mathbf{u} = \mathbf{Df} + \mathbf{Bg}$$

in this case, since we are solving the Neumann problem and $\hat{L}$ is consistent on the whole domain, we do not have to restrict to the interior nodes, so here all the matrices are N-by-N.

Finally, we describe the discretization of (7.6). In addition to the spatial discretization, we use Backward Euler to discretize in time. Let the number of time sub-intervals be K and the time step size is $\tau = T/K$. Then given $\mathbf{u}^0 = \mathbf{u}_0$, for $k = 1, \dots, K$, we solve

$$\left(\mathbf{D}(\mathrm{f}_{\text{dof}}, \mathrm{f}_{\text{dof}}) + \tau\hat{\mathbf{L}}(\mathrm{f}_{\text{dof}}, \mathrm{f}_{\text{dof}})\right)\mathbf{u}^k = \tau\mathbf{D}(\mathrm{f}_{\text{dof}}, :)\mathbf{f}^k + \mathbf{D}(\mathrm{f}_{\text{dof}}, \mathrm{f}_{\text{dof}})\mathbf{u}^{k-1}$$

where this last equation is again restricted to interior nodes so that all the matrices are N_{int}-by-N_{int} except for $\mathbf{D}(\mathrm{f}_{\text{dof}}, :)$ which is N_{int}-by-N.

7.3. Numerical examples

With the help of several examples, next we show that the approach introduced in this paper, can help solve the boundary value problems (7.2), (7.4), and (7.6). In the first 5 examples, we let $\Omega = (0,1)^2$. We first consider elliptic problems with both Dirichlet and Neumann boundary conditions. Afterwards, we illustrate the applicability of our approach on time-dependent PDE with Dirichlet boundary conditions. For numerical approximation in these 5 examples, we partition Ω into 100 uniform cells in each direction. Our final example is a semi-sphere with Dirichlet boundary conditions.

Example 7.1 *(Elliptic homogeneous Dirichlet).* In (7.1) we set $g \equiv 0$, therefore $\tilde{g} \equiv 0$ (cf. (7.2)). Consider the exact solution $u(x,y) = \sin(2\pi x)\sin(2\pi y)$, then $f(x,y) = 8\pi^2 \sin(2\pi x)\sin(2\pi y)$. The error between the exact solution u and it's approximation u_h using our proposed method in L^2-norm is: $\|u - u_h\|_{L^2(\Omega)} = 1.541989\text{e-}02$. Fig. 6 shows a visual comparison between the computed and the exact solution.

Example 7.2 *(Elliptic nonhomogeneous Dirichlet).* Let the exact solution $u = x^2 + y^2$. We set $g = u|_{\partial\Omega}$. The error between the exact solution u and it's approximation u_h using our proposed method in L^2-norm is: $\|u - u_h\|_{L^2(\Omega)} = 6.378652\text{e-}03$. Fig. 7 shows a visual comparison between the computed and the exact solution.

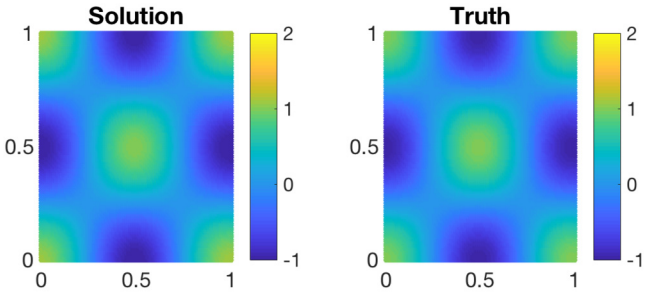

Fig. 8. Left: Solution computed using our approach. Right: Exact solution.

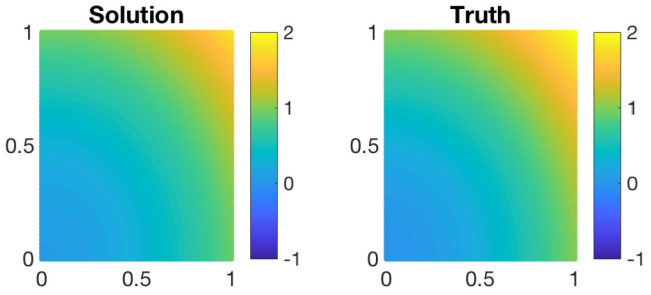

Fig. 9. Left: Solution computed using our approach. Right: Exact solution.

Example 7.3 *(Elliptic homogeneous Neumann).* In (7.4) we set $g \equiv 0$. Consider the exact solution $u(x,y) = \cos(2\pi x)\cos(2\pi y)$, then $f(x,y) = (8\pi^2+1)\cos(2\pi x)\cos(2\pi y)$. The error between the exact solution u and it's approximation u_h using our proposed method in L^2-norm is: $\|u-u_h\|_{L^2(\Omega)} = 2.125979\text{e-}02$. Fig. 8 shows a visual comparison between the computed and the exact solution.

Example 7.4 *(Elliptic nonhomogeneous Neumann).* Let the exact solution $u = x^2+y^2$. We set $g = \nabla u \cdot \eta$. The error between the exact solution u and it's approximation u_h using our proposed method in L^2-norm is: $\|u-u_h\|_{L^2(\Omega)} = 8.303406\text{e-}02$. Fig. 9 shows a visual comparison between the computed and the exact solution.

Example 7.5 *(Parabolic homogeneous Dirichlet).* In (7.6) we set $T = 1$. Consider the exact solution $u(x,y) = \sin(2\pi x)\sin(2\pi y)e^{-t}$, then $f(x,y) = (8\pi^2-1)u(x,y)$. We apply Backward-Euler scheme to do the time discretization with number of time steps equal to 50. The error between the exact solution u and it's approximation u_h using our proposed method in L^2-norm is: $\|u-u_h\|_{L^2(0,T;L^2(\Omega))} = 9.902258\text{e-}03$.

Example 7.6 *(Dirichlet on the hemisphere).* We now consider (7.1) on the hemisphere $\mathcal{M} = \{(x,y,z) \in \mathbb{R}^3 \,:\, x^2+y^2+z^2 = 1, z \geq 0\}$. This two dimensional manifold with boundary can be seen as the image of $(\theta,\phi) \in [0,\pi/2]\times[0,2\pi]$ under the embedding function, $\iota : [0,\pi]\times[0,2\pi] \to \mathbb{R}^3$ defined by

$$\iota(\theta,\phi) = \begin{bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{bmatrix}$$

where θ is the colatitude and ϕ is the azimuthal angle. The pullback metric in these coordinates is

$$g(x) = D\iota^T D\iota = \begin{bmatrix} \cos^2\theta\cos^2\phi + \cos^2\theta\sin^2\phi + \sin^2\theta & 0 \\ 0 & \sin^2\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{bmatrix}.$$

The Laplacian, Δ, on $\mathcal{M}$ in these coordinates is given by

$$\Delta f = \frac{1}{\sqrt{|g|}}\begin{bmatrix} \frac{\partial}{\partial\theta} & \frac{\partial}{\partial\phi} \end{bmatrix}\left(\sqrt{|g|}\; g^{-1}\begin{bmatrix} \frac{\partial f}{\partial\theta} \\ \frac{\partial f}{\partial\phi} \end{bmatrix}\right) = \cot\theta\,\frac{\partial f}{\partial\theta} + \frac{\partial^2 f}{\partial\theta^2} + \csc^2\theta\,\frac{\partial^2 f}{\partial\phi^2}.$$

We can avoid blowup at $\theta = 0$ by assuming a solution of the form $u(\theta,\phi) = \sin^2(\theta)\tilde{u}(\theta,\phi)$, and in this case we consider $u(\theta,\phi) = \sin^2(\theta)\sin(3\phi)/2$ which leads to

$$f = -\Delta u = (5/2 + 3\sin^2(\theta))\sin(3\phi).$$

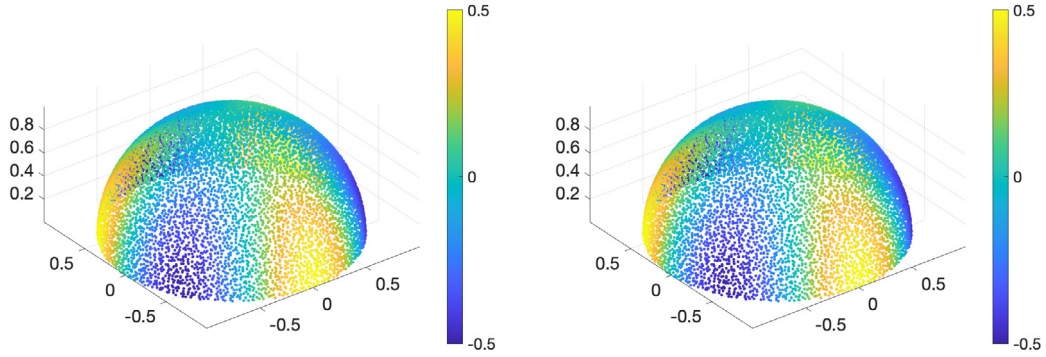

Fig. 10. Left: Solution computed using our approach. Right: Exact solution.

Using this f as the right-hand-side and using the true value of u on the boundary as a Dirichlet boundary condition, $g = u$ on $\partial\mathcal{M}$, we then solve (7.1) using the estimator of the Laplacian $\hat{\mathbf{L}}$. The resulting solution estimate u_h is compared in Fig. 10. The error between the exact solution u and it's approximation u_h using our proposed method in L^2-norm is: $\|u - u_h\|_{L^2(\mathcal{M})} = 5.067884\text{e-}03$.

In all the above examples, we observe that the solutions computed using our approach are highly accurate. We emphasize that the exact same code was used to solve the problem on the hemisphere as was used on the unit square. This is the advantage of the these diffusion maps based approaches, all that is needed is points sampled on the manifold.

8. Discussion

In this work, we have provided an analysis of the bias of the kernel averaging operator associated to the diffusion maps graph Laplacian. One of the main contributions of this work is the use of semigeodesic coordinates for making the bias analysis for points near the boundary. Using these coordinates, we are able to obtain new asymptotic estimates of the kernel averaging operator as well as prove consistency of a boundary integral estimator. It should be noted that our convergence results only address the bias of the estimator, that is, as $\epsilon \to 0$. In order to discuss a full treatment of error rates involving the number of data points, a treatment of variance must also be done.

Data availability

No data was used for the research described in the article.

Acknowledgments

The first author is partially supported by NSF DMS-2153561. The second author is partially supported by NSF DMS-1854204 and NSF DMS-2006808. The last author is partially supported by NSF grants DMS-1818772, DMS-1913004, DMS-2110263 and the Air Force Office of Scientific Research under Award NOs: FA9550-19-1-0036 and FA9550-22-1-0248.

Appendix A. Proof of Lemma 6.1

In this section we prove Lemma 6.1 connecting the integral of the first moment of a kernel near the boundary to the second moment of the kernel in the interior.

Proof of Lemma 6.1. Since m_1^{∂} has fast decay, we can localize the integral to an ϵ^{γ} region for any $\gamma \in (0,1)$ by the same argument as in the proof of Lemma 3.3. Thus, we have

$$\begin{aligned}\int_0^{\infty} m_1^{\partial}(b_x)\, db_x &= \int_0^{\epsilon^{\gamma}} m_1^{\partial}(b_x) db_x + O(\epsilon^z) \\ &= \int_0^{\epsilon^{\gamma}} \int_{-b_x}^{\epsilon^{\gamma}} \int_{\mathbb{R}^{m-1}} u^m k(\|u\|^2_{\text{sem}}) du^1 \ldots du^{m-1} du^m db_x\end{aligned}$$

We notice that there is considerable symmetry in the domain of integration between the b_x and u^m variables. Since $u^m k(\|u\|^2_{\text{sem}})$ is an odd function with respect to u^m, we obtain cancellation of the domain in the area indicated in red in Fig. A.11. This leads to the following simplification:

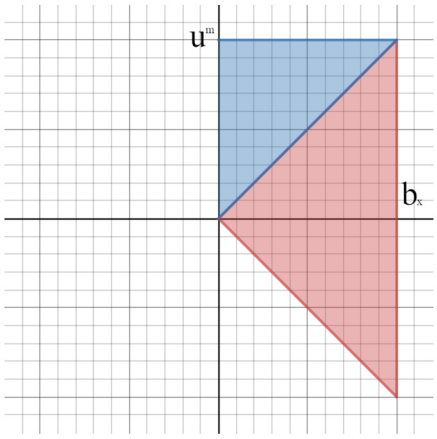

Fig. A.11. The domain of integration for b_x and u^m. The red region indicates the domain which cancels due to symmetry. The remaining blue region is evaluated.

$$\int_0^{\epsilon^\gamma}\int_{-b_x}^{\epsilon^\gamma}\int_{\mathbb{R}^{m-1}} u^m k(\|u\|_{\text{sem}}^2) du^1 \dots du^{m-1} du^m db_x = \int_0^{\epsilon^\gamma}\int_{b_x}^{\epsilon^\gamma}\int_{\mathbb{R}^{m-1}} u^m k(\|u\|_{\text{sem}}^2) du^1 \dots du^{m-1} du^m db_x$$

We then apply Fubini's theorem to obtain,

$$\begin{aligned}
&= \int_0^{\epsilon^\gamma}\int_0^{u^m}\int_{\mathbb{R}^{m-1}} u^m k(\|u\|_{\text{sem}}^2) du^1 \dots du^{m-1} db_x du^m \\
&= \int_0^{\epsilon^\gamma}\int_{\mathbb{R}^{m-1}}\int_0^{u^m} u^m k(\|u\|_{\text{sem}}^2) db_x du^1 \dots du^{m-1} du^m \\
&= \int_0^{\epsilon^\gamma}\int_{\mathbb{R}^{m-1}} (u^m - 0) u^m k(\|u\|_{\text{sem}}^2) du^1 \dots du^{m-1} du^m \\
&= \int_0^{\epsilon^\gamma}\int_{\mathbb{R}^{m-1}} (u^m)^2 k(\|u\|_{\text{sem}}^2) du^1 \dots du^m \\
&= \frac{1}{2}\int_{-\epsilon^\gamma}^{\epsilon^\gamma}\int_{\mathbb{R}^{m-1}} (u^m)^2 k(\|u\|_{\text{sem}}^2) du^1 \dots du^m + O(\epsilon^z) \\
&= \frac{m_2}{2} + O(\epsilon^z),
\end{aligned}$$

where in the last line we expand the integral to all of $\mathbb{R}^m$ and recover the second moment $m_2 = \int_{\mathbb{R}^m} (u^i)^2 k(||u||^2)\, du$ which is the same for any choice of i due to the radial symmetry of the kernel. □

Appendix B. Estimating the normal vector field and distance to the boundary

In [6], following the results of [10], the authors extended the expansion (4.4) to manifolds with boundary as,

$$\epsilon^{-m}\mathcal{K} f(x) = m_0^\partial(x) q(x) f(x) + O(\epsilon) + \epsilon^{-m}\text{Error}_{\text{Quad}}(N, f, q), \tag{B.1}$$

where the coefficient m_0^∂ is no longer constant but depends on the distance b_x from x to the boundary (defined as the infimum over smooth curves). As shown in [6], the coefficients $m_\ell^\partial(x)$ appearing in (B.1) for manifolds with boundary are given in (4.2) (repeated here for clarity)

$$m_\ell^\partial(x) = \int_{\{z\in\mathbb{R}^m \,|\, z\cdot\eta_x < b_x/\epsilon\}} (z\cdot\eta_x)^\ell k(|z|^2)\, dz = \int_{\mathbb{R}^{m-1}}\int_{-\infty}^{b_x/\epsilon} z_m^\ell k\left(|z|^2\right)\, dz_m dz_1 \cdots dz_{m-1}\,.$$

The vector field η_x is equal to the outward pointing normal when $x \in \partial\mathcal{M}$. We can smoothly extend the vector field η_x to a tubular neighborhood of the boundary called a *normal collar* as discussed in Section 2. It can easily be seen that when $b_x > \epsilon$ (meaning that x is further from the boundary than ϵ) these reduce to the formulas (1.5) up to higher order terms in ϵ.

For the exponential kernel,

$$k(z) = \exp(-z) \tag{B.2}$$

we can explicitly compute

$$m_0^\partial(x) = \frac{\pi^{m/2}}{2}(1 + \operatorname{erf}(b_x/\epsilon)), \qquad m_1^\partial(x) = -\frac{\pi^{(m-1)/2}}{2}\exp\left(-\frac{b_x^2}{\epsilon^2}\right).$$

These moments are used in the Section 4 to extend the expansion (B.1) to higher order terms.

The motivation for the expansion (B.1) in [6] was to analyze the standard Kernel Density Estimator (KDE) for manifolds with boundary. The standard KDE is (up to a constant) given by,

$$q_{\epsilon,N}(x) \equiv \epsilon^{-m}\mathcal{K}1(x) = \frac{1}{N\epsilon^m}\sum_{i=1}^{N} k\left(\frac{|x - X_i|^2}{\epsilon^2}\right)$$

and (B.1) implies that

$$\mathbb{E}[q_{\epsilon,N}(x)] = m_0^\partial(x)q(x) + O(\epsilon). \tag{B.3}$$

For manifolds without boundary, $q_{\epsilon,N}(x)$ can be made consistent after dividing by the normalization constant m_0 from (1.5). For manifolds with boundary, as a consequence of (B.1) we see that $q_{\epsilon,N}(x)$ is not consistent at the boundary. In [6] it has been shown how to fix this estimator by estimating the distance to the boundary b_x. We briefly summarize this method since it will be a key tool in constructing boundary integral estimators as defined in Section 5.

Since the standard estimator mixes information about the density and distance to the boundary, additional information is needed in order to estimate the density. Thus, in [6] the Boundary Direction Estimator (BDE) was introduced, which is defined as,

$$\mu_{\epsilon,N}(x) \equiv \frac{1}{N\epsilon^m}\sum_{i=1}^{N} k\left(\frac{|x - X_i|^2}{\epsilon^2}\right)\frac{(X_i - x)}{\epsilon}.$$

Notice that the BDE is a kernel weighted average of the vectors pointing from the specified point x to all the other data points $\{X_i\}$. The kernel weighting ensures that only the nearest neighbors of the point x contribute significantly to the summation. Moreover, for data points in the interior of the manifold, and for sufficiently small bandwidth parameter ϵ, we expect the nearest neighbors to be evenly distributed in all the directions tangent to the manifold. The resulting cancellations imply that the summation should result in a relatively small value (it is shown to be order-ϵ in [6] for points further than ϵ from the boundary). On the other hand, when x is on the boundary, if we look in the direction normal to the boundary we expect all the data points to be on one side of x, and thus the BDE will have a significant component in exactly the normal direction (inward pointing since we are averaging vectors pointing into the manifold). For points near the boundary (relative to the size of the bandwidth ϵ) this effect will be diminished smoothly until the distance to the boundary becomes greater than the bandwidth and we return to the case of an interior point. This intuitive description of the behavior of the BDE was made rigorous in [6] by showing that

$$\mathbb{E}[\mu_{\epsilon,N}(x)] = \eta_x q(x) m_1^\partial(x) + O(\epsilon\nabla q(x), \epsilon q(x)) \tag{B.4}$$

where $\eta_x \in T_x\mathcal{M}$ is a unit vector pointing towards the closest boundary point (for $x \in \partial\mathcal{M}$, η_x is the outward pointing normal). Moreover, $\mathbb{E}[\cdot]$ denotes the expected value. Notice that since $m_1^\partial(x) < 0$ and η_x is outward pointing, (B.4) implies that $\mu_{\epsilon,N}(x)$ points into the interior as expected from the above discussion.

In [6] the authors combined the BDE with the classical density estimator. Indeed by dividing $\mu_{\epsilon,N}$ by $q_{\epsilon,N}$, the dependence on the true density $q(x)$ cancels and the result depends only on the distance to the boundary b_x, namely (dividing (B.4) by (B.3)),

$$\frac{\mathbb{E}\left[\mu_{\epsilon,N}(x)\right]}{\mathbb{E}\left[q_{\epsilon,N}(x)\right]} = \frac{\eta_x m_1^\partial(x) + O(\epsilon)}{m_0^\partial(x) + O(\epsilon)} = -\eta_x \frac{\pi^{-1/2} e^{-b_x^2/\epsilon^2}}{\left(1 + \operatorname{erf}\left(b_x/\epsilon\right)\right)} + O(\epsilon). \tag{B.5}$$

A significant feature of this approach is that (B.5) can be easily estimated without any explicit dependence on the dimension m of the manifold. By combining the definitions of $\mu_{\epsilon,N}$ and $q_{\epsilon,N}$ above we find,

$$\frac{\mu_{\epsilon,N}(x)}{q_{\epsilon,N}(x)} = \frac{\sum_{i=1}^{N} k\left(\frac{|x-X_i|^2}{\epsilon^2}\right)\frac{(X_i - x)}{\epsilon}}{\sum_{i=1}^{N} k\left(\frac{|x-X_i|^2}{\epsilon^2}\right)}.$$

In order to compute the distance to the boundary, we compute the norm of the vector of the previous equation, and applying (B.5) we have

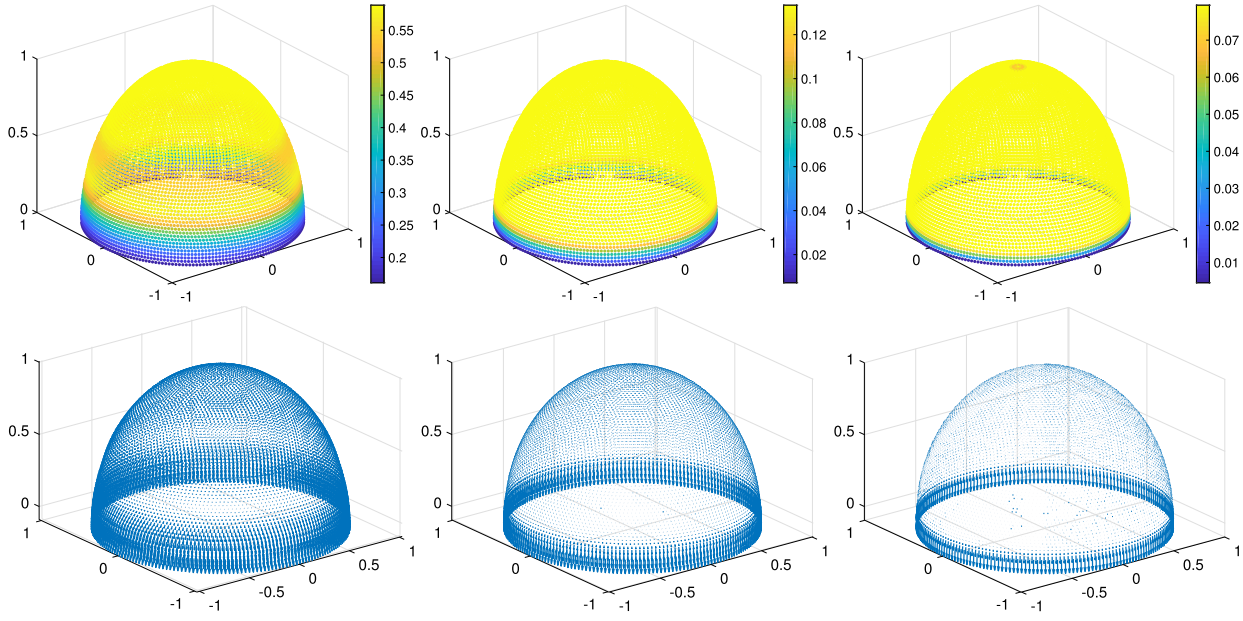

Fig. B.12. Estimating the distance to the boundary (top row) and the normal vector field (bottom) for bandwidth parameters $\epsilon \in \{0.5, 0.1, 0.05\}$ (left to right). The estimator is accurate up to a distance of approximately 1.5ϵ from the boundary.

$$\mathbb{E}\left[\left|\sqrt{\pi}\frac{\mu_{\epsilon,N}(x)}{q_{\epsilon,N}(x)}\right|\right] = \frac{e^{-b_x^2/\epsilon^2}}{\left(1+\operatorname{erf}\left(b_x/\epsilon\right)\right)} + O(\epsilon). \tag{B.6}$$

This is now a scalar equation with a known quantity on the left-hand-side, so it remains only to invert the function on the right-hand-side in order to estimate the distance to the boundary b_x. We note that this computation must be performed at each data point since we require an estimate of the distance to the boundary for each of our data points.

While [6] used a Newton's method to solve (B.6) for b_x, we note that the right-hand-side is very well approximated by the following piecewise function

$$\frac{e^{-b_x^2/\epsilon^2}}{\left(1+\operatorname{erf}\left(b_x/\epsilon\right)\right)} \approx \begin{cases} 1 - 1.15\frac{b_x}{\epsilon} + 0.35\left(\frac{b_x}{\epsilon}\right)^2 & b_x < 1.4\epsilon \\ \frac{1}{2}\exp\left(\left(\frac{b_x}{\epsilon}\right)^2\right) & b_x \geq 1.4\epsilon \end{cases}$$

(The above quadratic approximation was derived by interpolating the function at $\frac{b_x}{\epsilon} \in \{0, 1/2, 1\}$ and for $b_x \geq 1.4\epsilon$ the denominator of (B.6) is very close to 2.) Since the quadratic and the exponential are both explicitly invertible, this approximation avoids requiring a numerical inversion of the right-hand-side of (B.6) (Fig. B.12).

We now have consistent estimators for both the direction of the boundary, η_x, and the distance to the boundary, b_x, and these will be essential in imposing the desired boundary conditions for our grid free solvers.

Appendix C. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.acha.2023.101593.

References

[1] Stephanie Alexander, Vitali Kapovitch, Anton Petrunin, Alexandrov Geometry: Preliminary Version No. 1, 2019.
[2] Stephanie B. Alexander, I. David Berg, Richard L. Bishop, Geometric curvature bounds in Riemannian manifolds with boundary, Trans. Am. Math. Soc. 339 (2) (1993) 703–716.
[3] Mikhail Belkin, Partha Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.
[4] Tyrus Berry, John Harlim, Variable bandwidth diffusion kernels, Appl. Comput. Harmon. Anal. 40 (1) (2016) 68–96.
[5] Berry Tyrus, Timothy Sauer, Local kernels and the geometric structure of data, Appl. Comput. Harmon. Anal. 40 (3) (2016) 439–469.
[6] Berry Tyrus, Timothy Sauer, Density estimation on manifolds with boundary, Comput. Stat. Data Anal. 107 (2017) 1–17.
[7] Berry Tyrus, Timothy Sauer, Consistent manifold representation for topological data analysis, Found. Data Sci. 1 (1) (2019) 1–38.
[8] SongXi Chen, Probability density function estimation using gamma kernels, Ann. Inst. Stat. Math. 52 (3) (2000) 471–480.
[9] Bennett Chow, Peng Lu, Lei Ni, Hamilton's Ricci Flow, vol. 77, American Mathematical Soc., 2006.
[10] Ronald R. Coifman, Stéphane Lafon, Diffusion maps, Appl. Comput. Harmon. Anal. 21 (1) (2006) 5–30.
[11] Paul E. Ehrlich, Continuity properties of the injectivity radius function, Compos. Math. 29 (2) (1974) 151–178.
[12] Tingran Gao, The diffusion geometry of fibre bundles: horizontal diffusion maps, Appl. Comput. Harmon. Anal. 50 (2021) 147–215.
[13] Faheem Gilani, John Harlim, Approximating solutions of linear elliptic pde's on a smooth manifold using local kernel, J. Comput. Phys. 395 (2019) 563–582.

[14] Alfred Gray, Tubes, vol. 221, Birkhäuser, 2012.
[15] John Harlim, Daniel Sanz-Alonso, Ruiyi Yang, Kernel methods for Bayesian elliptic inverse problems on manifolds, arXiv preprint arXiv:1910.10669, 2019.
[16] M. Hein, J.Y. Audibert, U. von Luxburg, Convergence of graph Laplacians on random neighborhood graphs, J. Mach. Learn. Res. 8 (1325–1370) (2007) 6–19.
[17] M.C. Jones, Simple boundary correction for kernel density estimation, Stat. Comput. 3 (3) (1993) 135–146.
[18] M.C. Jones, P.J. Foster, A simple nonnegative boundary correction method for kernel density estimation, Stat. Sin. 6 (4) (1996) 1005–1013.
[19] Rhoana J. Karunamuni, Tom Alberts, On boundary correction in kernel density estimation, Stat. Methodol. 2 (3) (2005) 191–212.
[20] Yoon Tae Kim, Hyun Suk Park, Geometric structures arising from kernel density estimation on Riemannian manifolds, J. Multivar. Anal. 114 (2013) 112–126.
[21] John Lee, Introduction to Riemannian Manifolds, vol. 176, Springer International Publishing, 2018.
[22] Peter Malec, Melanie Schienle, Nonparametric kernel density estimation near the boundary, Comput. Stat. Data Anal. 72 (2014) 57–76.
[23] Arkadas Ozakin, Alexander G. Gray, Submanifold density estimation, in: Advances in Neural Information Processing Systems, 2009, pp. 1375–1382.
[24] Bruno Pelletier, Kernel density estimation on Riemannian manifolds, Stat. Probab. Lett. 73 (3) (2005) 297–304.
[25] M.C. Jones, S. Zhang, R.J. Karunamuni, An improved estimator of the density function at the boundary, J. Am. Stat. Assoc. 94 (448) (1999) 1231–1241.
[26] S.A. Sarra, The Matlab radial basis function toolbox, J. Open Res. Softw. 5 (2017).
[27] Francisco-Javier Sayas, Thomas S. Brown, Matthew E. Hassell, Variational techniques for elliptic partial differential equations, in: Theoretical Tools and Advanced Applications, CRC Press, Boca Raton, FL, 2019.
[28] Thomas Schick, Manifolds with boundary and of bounded geometry, Math. Nachr. 223 (1) (2001) 103–120.
[29] Amit Singer, From graph to manifold Laplacian: the convergence rate, Appl. Comput. Harmon. Anal. 21 (1) (2006) 128–134.
[30] Amit Singer, H-T. Wu, Vector diffusion maps and the connection Laplacian, Commun. Pure Appl. Math. 65 (8) (2012) 1067–1144.
[31] Oleg G. Smolyanov, Heinrich v. Weizsäcker, Olaf Wittich, Chernoff's theorem and discrete time approximations of brownian motion on manifolds, Potential Anal. 26 (1) (2007) 1–29.
[32] Hau Tieng Wu, Nan Wu, When Locally Linear Embedding Hits Boundary, 2018.
[33] F.W. Warner, Extension of the Rauch comparison theorem to submanifolds, Trans. Am. Math. Soc. 122 (2) (1966) 341–356.