DIFFUSION MAPS FOR EMBEDDED MANIFOLDS WITH BOUNDARY WITH APPLICATIONS TO PDES

RYAN VAUGHN, TYRUS BERRY, AND HARBIR ANTIL

Abstract. Given only a collection of points sampled from a Riemannian manifold embedded in a Euclidean space, in this paper we propose a new method to solve elliptic and parabolic partial differential equations (PDEs) supplemented with boundary conditions. Notice that 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 such constructions. 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 weak (variational) sense. The latter reduces the smoothness requirements on the underlying functions which is crucial to approximating weak solutions to PDEs. As a by-product, we also provide a rigorous justification of the well-known relationship between diffusion maps and the Neumann eigenvalue problems. We then use a recently developed method of estimating the distance to boundary function (notice that the boundary location is assumed to be unknown and must be estimated from data) in order to correct the boundary error term in the diffusion maps construction. Finally, using this estimated distance, we illustrate how to impose Dirichlet and Neumann conditions for some common PDEs based on the Laplacian. Several numerical examples confirm 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 $\mathcal{M}$ embedded in an ambient Euclidean space via $\iota : \mathcal{M} \to \mathbb{R}^n$.

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 a collection of sample points, $\{x_i\}_{i=1}^N \subset \iota(\mathcal{M}) \subset \mathbb{R}^n$ 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

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

almost surely. In the statistical context the weight function $q \in C^3(\mathcal{M})$ is called the sampling density. Ultimately our method will be independent of q, meaning that we do not require a specific density or even a consistent quadrature rule. This is critical when the nodes are data points,but is also an advantage for synthetic data sets where creating uniform grids on manifolds can be challenging. There are several situations where this may arise:

Key words and phrases. mesh-free, partial differential equations, heat equation, Riemannian manifolds, boundary detection.

The first and second authors are partially supported by ..., the last author is partially supported by NSF grants DMS-1818772, DMS-1913004 and the Air Force Office of Scientific Research under Award NO: FA9550-19-1-0036.

- 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 Laplacian Eigenmaps [1] and the diffusion maps algorithm [7] which gives a method of approximating the intrinsic Laplacian operator on an unknown Manifold. In this manuscript we let Δ be the negative definite Laplacian, also know 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

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

by our quadrature formula

$$\mathcal{K}f(x) \equiv \frac{1}{N}\sum_{i=1}^{N} K(\epsilon, x, x_i) f(x_i) = \mathcal{I}f(x) + \text{Error}_{\text{Quad}}(N, f, q) \tag{1.2}$$

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 $\text{Error}_{\text{Quad}} = \mathcal{O}(N^{-1/2})$ with high probability [19, 2]. 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 a small ball on the manifold, which would clearly lead to a significant error in the quadrature formula. However, the probability of all such high-error events can be made arbitrarily small as $N^{-1/2}$ approaches zero [19, 2]. 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 $|x-y|$ 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 to converge to the semigroup $e^{\epsilon^2\Delta}$ associated to the intrinsic Laplacian Δ, so that

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

In fact, a more detailed asymptotic analysis [7, 11] reveals that for any kernel function of the form

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

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

$$\begin{aligned}\epsilon^{-m}\mathcal{K}f(x) = m_0 f(x)q(x) + \epsilon^2\frac{m_2}{2}\left(\omega(x)f(x)q(x) + \Delta(fq)(x)\right)\\ + \mathcal{O}(\epsilon^4) + \epsilon^{-m}\text{Error}_{\text{Quad}}(N, f, q).\end{aligned} \tag{1.3}$$

for all x with distance greater than ϵ from the boundary.

The expansion (1.3) is most commonly used for estimating the density function [15, 14, 12] by applying the operator $\mathcal{K}$ to the constant function $f \equiv 1$ to find,

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

Even for this simple problem the boundary creates a significant issue since the constant m_0 becomes a function $m_0^{\partial}(x)$ which depends on the distance to the boundary. This leads to the well known bias of Kernel Density Estimators (KDEs) near the boundary. In [4], the authors developed a method to estimate the distance to the boundary and then correct the bias of the KDE near the boundary. A significant advance of the method of [4] is that the location of the boundary is not needed and is effectively learned from the data.

A much more challenging and powerful use of (1.3) is for estimating the Laplace-Beltrami operator, Δ, and this expansion is the key component of justifying the diffusion maps algorithm [7]. 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 [7], this 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,

$$\int_{\mathcal{M}} \phi f \, d\text{vol} \,, \qquad \int_{\mathcal{M}} \nabla\phi \cdot \nabla f \, d\text{vol} \,, \qquad \int_{\partial\mathcal{M}} \phi f \, d\text{vol}_{\partial} \,. \tag{1.4}$$

where the first and the second terms correspond to mass and stiffness matrices. The final term arises in case of Neumann or Robin boundary conditions [17]. The first operator can be represented by a diagonal matrix with entries $D_{ii} = q(x_i)^{-1}$ and q can be estimated as in [4] (summarized below). In this paper, we show that the graph Laplacian (as constructed by the diffusion maps algorithm, described in Section 1.1) is a consistent estimator of the second operator. Finally, we introduce a novel consistent estimator of the third operator.

A key aspect of our method is that the location of the boundary is not known and must be estimated as in [4]. 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 we instead quantify the boundary by estimating the distance to the boundary for each data point. Taken together these results yield an algorithm that can 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 1.1 we briefly summarize the contributions of this paper. In Section 2 we review the method of [4] for estimating the distance to the boundary and the outward pointing normal vectors. In Section 3 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. Then we use this estimator in Section 5 to construct consistent estimators of the operators in (1.4). 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 [7]. Finally in Section 6 we show how to impose standard boundary conditions using these operator estimators. We first turn to a brief overview of the results to follow.

1.1. **Overview.** Using specially designed normalizations (see Section 5.3 below for more details) diffusion maps constructs a matrix $\mathbf{tK}_{ij}$ which represents the operator $t\mathcal{K}$ with expansion,

$$\frac{t\mathcal{K}f(x)}{t\mathcal{K}1(x)} = f(x) + \epsilon^2 \frac{m_2}{2m_0}\Delta f(x) + \mathcal{O}(\epsilon^4) + \epsilon^{-m}\text{Error}_{\text{Quad}}(N, f, q). \tag{1.5}$$

Based on the expansion (1.5) the normalized graph Laplacian

$$\mathbf{tL}_{\text{norm}} = \frac{2m_0}{m_2\epsilon^2}(\mathbf{I} - \mathbf{tD^{-1}tK})$$

(where $\mathbf{tD}_{ii} = (\mathbf{tK}\vec{1})_i$) converges to the Laplacian operator pointwise, meaning that $(\mathbf{L}\vec{f})_i \to \Delta f(x_i)$ in the limit of $N \to \infty$ and $\epsilon \to 0$.

In [7] the goal was to solve the associated eigenvalue problem $-\Delta\varphi_\ell = \lambda_\ell\varphi_\ell$ in $\mathcal{M}$. Notice that for bounded domains, Neumann boundary conditions must be assumed. More recently, the diffusion maps estimate of the Laplacian has been used for solving PDEs such as $-\Delta u = f$ [8, 10] where f is now the data and one is solving for u. We should note that in [8] a more general class of elliptic operators are considered using a more general class of kernel functions introduced in [3]. 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 [8].

We should also point out that [8] compared the diffusion maps approach to another popular meshless method based on radial-basis function (RBF) interpolation [16]. 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 [8], extending the RBF method to arbitrary manifolds would require extensive complex modifications, such as finding local coordinate systems and approximating the desired differential operators. The diffusion maps approach provides a large class of such operators directly with a global representation. Thus, when more information about the manifold structure is known, approaches such as [16] may have superior results (just as meshed results 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.

Returning to our discussion of (1.3), we note that the function $\omega(x)$ depends on the geometry of $\mathcal{M}$ and its embedding. Moreover, the constants

$$m_0 = \int_{z\in T_x\mathcal{M}} k(|z|^2)\,dz \qquad \text{and} \qquad m_2 = \int_{z\in T_x\mathcal{M}} z_i^2 k(|z|^2)\,dz \tag{1.6}$$

arise from the fact that integrating over a local region around x is asymptotically equivalent (due to the assumed exponential decay of K) to integrating over the tangent space $T_x\mathcal{M}$. Conspicuously absent in the above expansion are the first moment of K and any gradient terms. In fact (1.3) only holds in the interior of the manifold (far from the boundary). Near the boundary the m_0 and m_2 terms are not constant and depend on the distance to the boundary as was first noticed in [7] and further expanded on in [4].

In Section 5 we first extend the previous results by proving that the following expansion holds uniformly all the way to the boundary (where η_x the direction of the boundary),

$$\begin{aligned}\frac{t\mathcal{K}f(x)}{t\mathcal{K}1(x)} = f(x) + \epsilon\frac{m_1^\partial(x)}{m_0^\partial(x)}\frac{\partial f}{\partial \eta_x}(x) + \frac{\epsilon^2}{2m_0^\partial(x)}\left(m_2\Delta f(x) + \left(m_2^\partial(x) - m_2\right)\frac{\partial^2 f}{\partial \eta_x^2}(x)\right) \\ + \mathcal{O}(\epsilon^3 e^{-b_x^2/\epsilon^2}, \epsilon^4) + \epsilon^{-m}\text{Error}_{\text{Quad}}(N, f, q).\end{aligned} \tag{1.7}$$

In the above expansion the moments become functions of x such that when the distance from x to the boundary is greater than ϵ we have, $m_0^\partial(x) = m_0, m_2^\partial(x) = m_2, m_1^\partial(x) = 0$, which recovers

(1.3) in the interior of the manifold. In fact, for each x in the manifold, we can find ϵ sufficiently small that (1.3) will hold at x [7, 11, 20]. However, (1.3) clearly does not hold uniformly since the ϵ required will decrease to zero as x approaches the boundary. Moreover, the full expansion (4.2) has not been derived previously, although the order-ϵ term was known as early as [7].

A likely reason that the expansion was not continued is that the presence of the order-ϵ term already implies that the Laplacian estimator will blow-up pointwise at the boundary since,

$$\frac{1}{\epsilon^2}\left(f(x) - \frac{t\mathcal{K}f(x)}{t\mathcal{K}1(x)}\right) = \frac{1}{\epsilon}\frac{m_1^\partial(x)}{m_0^\partial(x)}\frac{\partial f}{\partial \eta_x}(x) + \mathcal{O}(1).$$

In order to avoid this issue, the diffusion maps paper [7] simply assumed that f satisfies the Neumann condition $\frac{\partial f}{\partial \eta_x} = 0$. This assumption is sometimes misunderstood as the reason that the diffusion maps algorithm is empirically observed to return the Neumann eigenfunctions of the Laplacian. In this paper we will provide the first rigorous explanation of this phenomenon. Moreover, even for Neumann functions we do not recover (1.3) near the boundary since $\frac{\partial^2 f}{\partial \eta_x^2}$ would also need to be zero. Nevertheless, in Section 5.1 below we will show that the unnormalized graph Laplacian,

$$\mathbf{tL} = \frac{2}{m_2\epsilon^2}(\mathbf{tD} - \mathbf{tK})$$

is a consistent estimator of the Dirichlet energy in the sense that

$$\vec{\phi}^\top \mathbf{tL}\vec{f} \to \int_{\mathcal{M}} \nabla\phi \cdot \nabla f \, dV$$

as $N \to \infty$ and $\epsilon \to 0$. Thus, despite the pointwise blow-up at the boundary, the graph Laplacian still converges, in the weak sense, to the appropriate symmetric operator associated to the Laplacian for a manifold with boundary.

The key to making this novel connection is the realization that since the coefficient function $m_1^\partial(x)$ decays exponentially away from the boundary, the integral, $\int_M m_1^\partial(x)\, dV$ is localized in an ϵ neighborhood of the boundary. This implies that the integral is actually order-ϵ and moreover we show that,

$$\epsilon \int_{\mathcal{M}} m_1^\partial(x)\phi(x)\frac{\partial f}{\partial \eta_x}\, d\text{vol} = \epsilon^2 \frac{m_2}{2}\int_{\partial\mathcal{M}} \phi(x)\frac{\partial f}{\partial \eta_x}\, d\text{vol}_\partial + \mathcal{O}(\epsilon^3).$$

So the term that appeared to be order-ϵ (pointwise) is actually order-ϵ^2 in the weak sense. Similarly, the term $\epsilon^2\left(m_2^\partial(x) - m_2\right)\frac{\partial^2 f}{\partial \eta_x^2}$ is order-ϵ^3 in the weak sense. Finally, examining the Laplacian term in the weak sense we find (using integration by parts)

$$\epsilon^2\frac{m_2}{2}\int_{\mathcal{M}} \phi(x)\Delta f(x)\, d\text{vol} = -\epsilon^2\frac{m_2}{2}\int_{\partial\mathcal{M}} \phi(x)\frac{\partial f}{\partial \eta_x}\, d\text{vol}_\partial + \epsilon^2\frac{m_2}{2}\int_{\mathcal{M}} \nabla\phi(x)\cdot\nabla f(x)\, d\text{vol}.$$

Adding the previous two equations together, we see that the boundary integrals exactly cancel. Thus, the order-ϵ term (which blows-up pointwise) is required to obtain the desired symmetric form $\int_{\mathcal{M}} \nabla\phi(x)\cdot\nabla f(x)\, d\text{vol}$ in the weak sense. Finally, we note that such a cancellation should be expected since the symmetric matrix $\mathbf{tL}$ should represent a symmetric form.

The consistency of $\mathbf{tL}$, together with the novel boundary integral estimator introduced in Section 5.1 provides the required tools to set various boundary conditions as shown in Section 6. Next, we review some background information on estimating the distance to the boundary from [4]. Then in Section 3 we turn to the necessary geometry for rigorously establishing these results.

2. Estimating the normal vector field and distance to the boundary

In [4], following the results of [7], the authors extended the expansion (1.3) to manifolds with boundary as,

$$\epsilon^{-m}\mathcal{K}f(x) = m_0^{\partial}(x)q(x)f(x) + \mathcal{O}(\epsilon) + \epsilon^{-m}\text{Error}_{\text{Quad}}(N, f, q). \tag{2.1}$$

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 [4], the coefficients $m_\ell^{\partial}(x)$ appearing in (2.1) for manifolds with boundary are,

$$m_\ell^{\partial}(x) = \int_{\{z\in\mathbb{R}^m \,|\, z\cdot\eta_x < b_x/\epsilon\}} (z\cdot\eta_x)^\ell k(|z|^2)\, dz = \int_{\mathbb{R}^{m-1}} \int_{-\infty}^{b_x/\epsilon} z_m^\ell k\left(|z|^2\right)\, dz_m dz_1 \cdots dz_{m-1}\,. \tag{2.2}$$

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 will be discussed Section 3. 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 (1.6) up to higher order terms in ϵ.

For the exponential kernel,

$$k(z) = \exp(-z) \tag{2.3}$$

we can explicitly compute

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

Also, for the exponential kernel we can easily solve for higher moments, in terms of the first two, recursively using integration by parts

$$m_\ell^{\partial}(x) = \left(\frac{b_x}{\epsilon}\right)^{\ell-1} m_1^{\partial}(x) + \frac{\ell-1}{2} m_{\ell-2}^{\partial}(x). \tag{2.4}$$

For example we have

$$m_2^{\partial}(x) = \frac{\pi^{(m-1)/2}}{2}\left(-\frac{b_x}{\epsilon}e^{-b_x^2/\epsilon^2} + \frac{\sqrt{\pi}}{2}(1 + \text{erf}(b_x/\epsilon))\right). \tag{2.5}$$

These moments will be used in the Section 4 to extend the expansion (2.1) to higher order terms.

The motivation for the expansion (2.1) in [4] 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 (2.1) implies that

$$\mathbb{E}[q_{\epsilon,N}(x)] = m_0^{\partial}(x)q(x) + \mathcal{O}(\epsilon). \tag{2.6}$$

For manifolds without boundary, $q_{\epsilon,N}(x)$ can be made consistent after dividing by the normalization constant m_0 from (1.6). For manifolds with boundary, as a consequence of (2.1) we see that $q_{\epsilon,N}(x)$ is not consistent at the boundary. In [4] 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 the next section.

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 [4] 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 [4] 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 [4] by showing that

$$\mathbb{E}[\mu_{\epsilon,N}(x)] = \eta_x q(x) m_1^{\partial}(x) + \mathcal{O}(\epsilon\nabla q(x), \epsilon q(x)) \tag{2.7}$$

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, (2.7) implies that $\mu_{\epsilon,N}(x)$ points into the interior as expected from the above discussion.

In [4] 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 (2.7) by (2.6)),

$$\frac{\mathbb{E}\left[\mu_{\epsilon,N}(x)\right]}{\mathbb{E}\left[q_{\epsilon,N}(x)\right]} = \frac{\eta_x m_1^{\partial}(x) + \mathcal{O}(\epsilon)}{m_0^{\partial}(x) + \mathcal{O}(\epsilon)} = -\eta_x \frac{\pi^{-1/2} e^{-b_x^2/\epsilon^2}}{(1+\operatorname{erf}(b_x/\epsilon))} + O(\epsilon). \tag{2.8}$$

A significant feature of this approach is that (2.8) 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 (2.8) we have

$$\mathbb{E}\left[\left|\sqrt{\pi}\frac{\mu_{\epsilon,N}(x)}{q_{\epsilon,N}(x)}\right|\right] = \frac{e^{-b_x^2/\epsilon^2}}{(1+\operatorname{erf}(b_x/\epsilon))} + \mathcal{O}(\epsilon). \tag{2.9}$$

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 [4] used a Newton's method to solve (2.9) for b_x, we note that the right-hand-side is very well approximated by the following peicewise function

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

(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 (2.9) 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 (2.9).

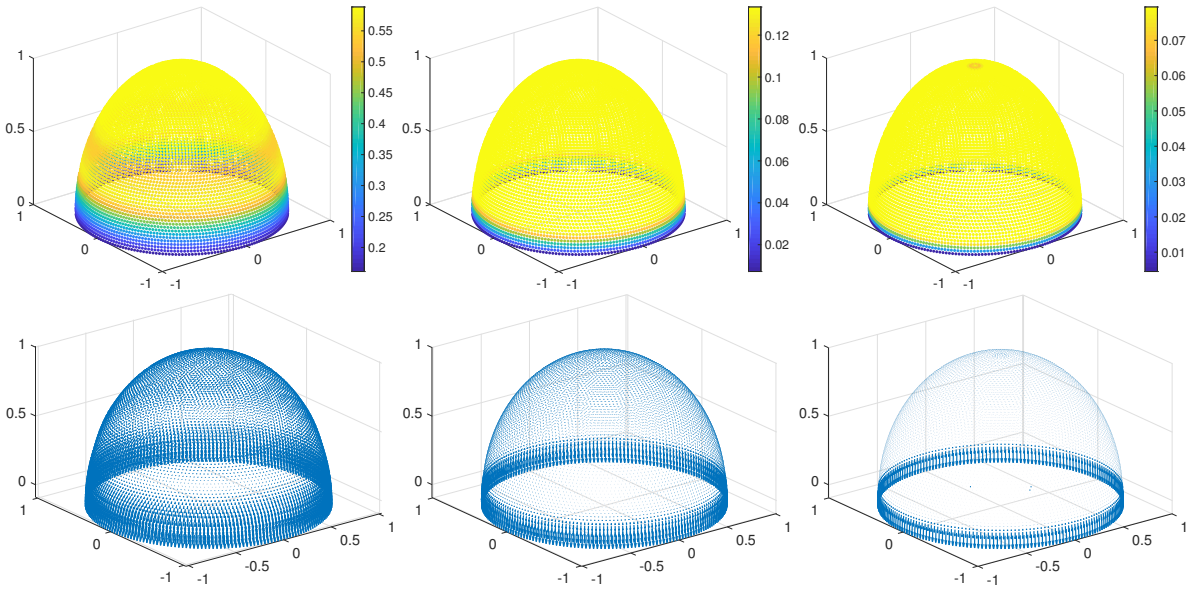

FIGURE 1. Estimating the distance to the boundary (top row) and the normal vector field (bottom) for bandwidth parameters $\epsilon \in \{0.5, 0.1, 0.05\}$ (left to right). The estimator is accurate up to a distance of approximately 1.5ϵ from the boundary.

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.

3. Geometric Preliminaries

Let $\mathcal{M}$ be a C^3 compact manifold with nonempty boundary smoothly embedded into $\mathbb{R}^d$ via the map $\iota : \mathcal{M} \to \mathbb{R}^d$. We endow $\mathcal{M}$ with the pullback metric $\iota^* g$ 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 probability density function that is absolutely continuous with respect to $d\text{vol}$.

Let $(\mathcal{M}, g)$ be a compact Riemannian manfiold. We let $\iota : \mathcal{M} \to \mathbb{R}^d$ be an isometric embedding, so that $\langle X, Y \rangle_g = \langle d\iota(X), d\iota(Y) \rangle_{\mathbb{R}^d}$ and in any local coordinates $(s^1, ..., s^m)$:

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

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 $\mathcal{M}$, where length is computed by:

$$L(\gamma(s)) = \int_0^s \langle \dot{\gamma}(t), \dot{\gamma}(t) dt \rangle_g .$$

If x and y are not in the same component, we define $d_g(x, y) = +\infty$. We also observe that the condition that ι is an isometry implies that the length of a curve γ in $\mathcal{M}$ is the same as the length of the curve $\iota \circ \gamma$ in $\mathbb{R}^d$.

3.1. **Normal Coordinates.** 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 $r_I(x) > 0$ such that $B_{r_I(x)}(0) \subseteq T_x\mathcal{M}$ is a diffeomorphism is called the injectivity radius of $\mathcal{M}$ at x. The infimum of all such $r_I(x)$ is called the r_I injectivity radius of $\mathcal{M}$ and it can be shown that for a compact manifold without boundary, this value is greater than zero.

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

Proposition 3.1. *Let $x \in \mathcal{M}$ and let $(s^1, .., s^m)$ denote a system of normal 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 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:

$$d_g(x, y) = |s|.$$

This distance comparison is the foundation of the diffusion maps expansion (1.3) in [7] and related papers. However, normal coordinates are not suitable for establishing uniform asymptotics near the boundary, so instead we turn to semigeodesic coordinates.

3.2. **The Normal Collar and Semigeodesic Coordinates.** Normal coordinates provide a particularly nice set of coordinates and are central to the proof of convergence of Diffusion maps in the case that $\mathcal{M}$ does not have a boundary. If $\mathcal{M}$ is a manifold with boundary, normal coordinates become less useful because the injectivity radius of $\mathcal{M}$ shrinks to zero as points approach the boundary. In addition, one is no longer is guaranteed to have a well-defined exponential map for points on the boundary. We instead make use of *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 [13, 18, 11].

Since $\mathcal{M}$ is compact, it admits a *normal collar* [13], 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 N. We will also define the set $N_{\frac{1}{2}} = \phi(\partial\mathcal{M} \times [0, \frac{1}{2}r_C])$. 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 $t \in [0, r_C)$. Since each $\partial\mathcal{M}^t$ are compact, each $\partial\mathcal{M}^t$ has a nonzero injectivity radius, and one can show that the infimum over all injectivity radii of all $\partial\mathcal{M}^t$ for $t \in [0, \frac{1}{2}r_C]$ is positive. We denote such an infimum as $r_B > 0$.

For clarity we use the letter, s, to denote normal coordinates (outside the collar) and u to denote semigeodesic coordinates (inside the collar). We use the term *geodesic chart* to refer to properties that hold in both normal coordinates and semigeodesic coordinates.

Inside of the normal collar, we can now construct semigeodesic coordinates centered at x. If $x \in \partial\mathcal{M}^t$ for some t, then we may construct normal coordinates in $\partial\mathcal{M}^t$ of dimension $m - 1$.

Proposition 3.2. *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 $\partial\mathcal{M}^t$, while the last coordinate entry parameterizes geodesic distance in $\mathcal{M}$ in the direction orthogonal to each of the $\partial\mathcal{M}^t$. Moreover, whereas the geodesic ball is a sphere, a semigeodesic ball is a cylinder. 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 $u^m \geq -b_x$. Finally, we note that following [13], 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.

3.3. **Manifolds of Bounded Geometry.** Compact manifolds with boundary are an example of a wider class of manifolds called *manifolds with boundary and of bounded geometry.*

Definition 3.3. *Let $\mathcal{M}$ be a manifold with boundary $\partial\mathcal{M}$. We say that it is a manifold with bounded geometry if the following hold:*

(a) ***Normal Collar:*** *There exists a $R_C > 0$ such that the map $\phi : \partial\mathcal{M} \times [0, R_C) \to \mathcal{M}$ defined by:*

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

is a diffeomorphism from the set $\partial\mathcal{M} \times [0, R_C)$ onto its image, where $u^m = -\eta_x$ denotes the inward facing normal vector at $x \in \partial\mathcal{M}$.

(b) ***Positive injectivity radius of*** $\partial_{\mathcal{M}}$***:*** *There is a* $R_{\partial\mathcal{M}} > 0$ *such that for every point* $x \in \partial\mathcal{M}$*, the boundary exponential map* $\exp_x^{\partial\mathcal{M}}$ *is a diffeomorphism on a ball of radius* $R_{\partial\mathcal{M}}$ *in* $T_x\mathcal{M}$*.*

(c) ***Positive injectivity radius on the interior:*** *There is a* $R_I > 0$ *such that for every point* $x \in \mathcal{M} \setminus N_{\frac{1}{3}}$*, the exponential map* $\exp_x^{\mathcal{M}}$ *is a diffeomorphism on a ball of radius* R_I *in* $T_x\mathcal{M}$*.*

(d) ***Curvature bounds:*** *There exists a* $C > 0$ *such that for all* $k \geq 0$*, the Riemannian curvature tensor* R *and the second fundamental form of the boundary* Π *satisfy:*

$$|\nabla^k R| < C \text{ and } |\nabla^k \Pi| < C.$$

A central result about manifolds of bounded geometry is that they admit coverings by normal and semigeodesic coordinate charts. In [18], it is shown that the curvature bound requirement is equivalent to the following condition:

Theorem 3.4 (Schick, 2001)**.** *A manifold with boundary is of bounded geometry if and only if in addition to satisfying the first three axioms above, there exists* $0 < R_{\partial\mathcal{M}^t}$*,* $0 < R_C$ *and* $0 < R_I$ *and constants* $C_K > 0$ *(for each* $K \in \mathbb{N}$*) such that whenever we have semigeodesic coordinates of radius* r_1 *and height* r_2 *with* $r_1 \leq R_{\partial\mathcal{M}^t}$ *and* $r_2 \leq R_C$*, or normal coordinates of radius* $r_3 \leq R_I$ *then in these coordinates:*

$$|D^\alpha g_{ij}| \leq C_K \text{ and } |D^\alpha g^{ij}| \leq C_K$$

for all $|\alpha| \leq K$*.*

Using Theorem 3.4, we can now show that the norm in geodesic coordinates approximates the extrinsic distance induced by $\iota : \mathcal{M} \to \mathbb{R}^d$.

Proposition 3.5. *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. Since $\mathcal{M}$ has bounded geometry, there exists positive constants $\tilde{C}_0$ and $\tilde{C}_1$ such that

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

for any geodesic coordinate chart. Since the matrices with entrees g_{ij} and g^{ij} are symmetric and positive definite, this implies that there exists 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:

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

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

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

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

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

□

Now that we have the distance comparison, we must establish the existence of a global constant $C_{\mathcal{M}} > 0$ such that once ϵ is less than $C_{\mathcal{M}}$ we can always find a coordinate chart that contains the inverse image (under the embedding) of the ϵ-ball around any point on the manifold. This is the key to establishing uniform asymptotic expansions in the next section.

Proposition 3.6. *There exists a $C_{\mathcal{M}} > 0$ such that for any $0 < \epsilon < C_{\mathcal{M}}$, the set $\iota^{-1}(B^{\mathbb{R}^d}_\epsilon(\iota(x)))$ is completely contained inside of a geodesic coordinate chart in $\mathcal{M}$.*

Proof. For each $x \in \mathcal{M}$, let $\tilde{B}_{R_{\mathcal{M}}}(x)$ denote a geodesic ball or semigeodesic hypercylinder of radius $R_{\mathcal{M}}$. We then consider the union

$$\mathcal{B} = \bigcup_{x\in\mathcal{M}} \left((\mathcal{M}\setminus\{x\}) \times \tilde{B}_{R_{\mathcal{M}}}(x)\right).$$

We see that $\mathcal{B}$ is the set of pairs (x, y) such that y is contained in a geodesic chart at x. Since $\mathcal{B}$ is an open set, the set $\mathcal{M}\times\mathcal{M}\setminus\mathcal{B}$ is a compact set. The restriction of the function $d_{\mathbb{R}^d}$ to $\mathcal{B}$ is continuous, and nowhere zero. Since $\mathcal{B}$ is compact, $d_{\mathbb{R}^d}$ attains its minimum value $C_{\mathcal{M}}$. Therefore if $d_{\mathbb{R}^d}(\iota(x),\iota(y)) < C_{\mathcal{M}}$ then y is in the maximal geodesic coordinate chart centered at x. □

Now that we have established the distance comparison in Proposition 3.5 and an appropriate generalization of the injectivity radius to manifolds with boundary in Proposition 3.6 we will be able to establish uniform asymptotics for manifolds with boundary.

4. Uniform Asymptotic Expansion for Manifolds with Boundary

Our approach will be similar to a result given in [20], however, our result differs in that the convergence is uniform with respect to ϵ, and holds for boundary points. In addition, we get a more explicit relationship between first-order error and the curvature in $\mathcal{M}$. The proof breaks down into several steps. The first step is to localize the integral using the exponential decay of the kernel to restrict the integral to an ϵ-ball around x in geodesic coordinates. Now that we are localized near x, we compute the Taylor expansion of the kernel k, function f, and volume form $d\text{vol}$ in coordinates. Finally, we combine these Taylor expansions and use the symmetries of the domain to eliminate many of the terms. The result is the asymptotic expansion of the kernel operator.

Lemma 4.1 (Localization to a Geodesic Neighborhood)**.** *Let $0 \leq \gamma \leq \frac{1}{2}$. For any $\epsilon > 0$ such that $\epsilon^\gamma < \min\{\frac{r_{\mathcal{M}}}{C_1}, C_{\mathcal{M}}\}$,*

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

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 $\epsilon^\gamma < \frac{r_{\mathcal{M}}}{C_1}$, then $\tilde{B}^{\mathcal{M}}_{C_1\epsilon^\gamma}(x)$ 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 $L^2(\mathcal{M})$:

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

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 $z \geq 1$ by making the substitution $\delta = \epsilon^{-1}$ and applying Lhôpital's rule z-times. □

Now that we are localized we compute the required asymptotic expansions in semigeodesic coordinates. The next three lemma's derive the leading order error term between the ambient space distance $||\iota(u)||_{\mathbb{R}^n}$ and the intrinsic distance $||u||_{\mathcal{M}}$. This is the key to the asymptotic expansion of the kernel k. These lemmas are motivated by the work of [21] which derived similar formulas in standard geodesic coordinates, but these are novel results for semigeodesic coordinates.

We first connect the connection in coordinates to the derivatives of the embedding.

Lemma 4.2. *Let U be the vector field such that*

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

Then at the point p:

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

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 $u \in T_p\mathcal{M}$, 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:

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

as desired. □

Next we relate the connection to the second fundamental form $II_{\partial M^t}$ of the hypersurfaces $\partial\mathcal{M}^t$ as a submanifolds of $\mathcal{M}$.

Lemma 4.3. *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:*

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

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

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

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

$$\nabla_{u^j\partial_j}u^i\partial_i = u^i u^j\nabla_{\partial_j}\partial_i = u^i u^j\Gamma^k_{ij}\partial_k.$$

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

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

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 ∇ is a symmetric connection,

$$\nabla_{u^j\partial_j}u^i\partial_i = u^i u^j\Gamma^k_{ij}\partial_k = u^i u^j\Gamma^k_{ji}\partial_k = \nabla_{u^i\partial_i}u^j\partial_j$$

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:

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

□

The above lemmas show that the second derivatives appearing in the asymptotic expansion are connected to the second fundamental form of the boundary considered as a submanifold of $\mathcal{M}$. The next lemma shows that this is the leading order error in the distance comparison. This result is analogous to Proposition 6 of [21] except for semigeodesic coordinates instead of geodesic coordinates.

Lemma 4.4. *In semigeodesic coordinates we have the following distance comparison,*

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

Proof. The result follows from the straightforward computation,

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

where the final equations follow from the previous two lemmas. □

We note that in geodesic normal coordinates the error is order $||u||^4_{\mathcal{M}}$ instead of $||u||^3_{\mathcal{M}}$, and depends on the second fundamental form of the embedding of $\mathcal{M}$ into the ambient space. However, near the boundary we obtain a larger error as shown above.

Next, we expand the volume form $d\text{vol}$ in semigeodesic coordinates. First, we note that in normal coordinates the volume form has the expansion

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

where R_{ij} are the components of the Ricci curvature tensor [13]. However, in semigeodesic coordinates we have the following result.

Theorem 4.5 (First and Second Variation of Area, [6, 9])**.** *Let $\partial\mathcal{M}^t$ be a hypersurface in $\mathcal{M}$ with $x \in \partial\mathcal{M}$ and outward facing normal η_x. Let $\gamma(t) : (-\epsilon, \epsilon) \to \mathcal{M}$ be a geodesic with initial velocity $\dot\gamma(0) = -\eta_x$ Then for $y = \gamma(t)$ we have*

$$d\text{vol}(y) = 1 - (m-1)H(x)t + \omega_2(x)t^2 + \mathcal{O}(t^3)$$

where H is the mean curvature.

We note that [6] defines the mean curvature as simply the summation, whereas we follow [13] in including the factor $\frac{1}{m-1}$. For points y that are not along the geodesic γ 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,

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

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

$$\begin{aligned}\int_B k'(|u|^2)\left\langle \Pi_{\partial\mathcal{M}^t}(U^\top, U^\top), U^\perp \right\rangle_g du &= -\frac{(m-1)}{2}H(x)\int_B k(|u|^2)u^m\, du \\ &= \frac{(m-1)}{2}m_1^\partial(x)H(x) + \mathcal{O}(\epsilon^z) \end{aligned} \tag{4.1}$$

for any $z \geq 1$*, 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:

$$\left\langle \Pi_{\partial\mathcal{M}^t}(U^\top, U^\top), U^\perp \right\rangle_g = \langle \Pi_{\partial\mathcal{M}}(\partial_i, \partial_j), \partial_m \rangle_g \, u^i u^j u^m.$$

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

$$\int_B k'(|u|^2)\left\langle \Pi_{\partial\mathcal{M}^t}(U^\top, U^\top), U^\perp \right\rangle_g du = \langle \Pi_{\partial\mathcal{M}}(\partial_i, \partial_i), \partial_m \rangle_g \int_B k'(|u|^2)u^i u^i u^m du$$

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

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

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

$$\int_B k'(|u|^2)\left\langle \Pi_{\partial\mathcal{M}^t}(U^\top, U^\top), U^\perp \right\rangle_g du = -\frac{1}{2}\int_B k(|u|^2)u^m du \sum_{i=1}^{m-1} \langle \Pi_{\partial\mathcal{M}}(\partial_i, \partial_i), \partial_m \rangle_g$$

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

We are now ready to prove the uniform asymptotic expansion of the kernel operator on $\mathcal{M}$.

Theorem 4.7. *Let* $\mathcal{M} \subset \mathbb{R}^n$ *be a compact* m*-dimensional* C^3 *Riemannian manifold either without boundary or with a* C^3 *boundary with a normal collar. Let* $k : \mathbb{R} \to \mathbb{R}$ *have exponential decay. Then for all* $x \in \mathcal{M}$ *and for* ϵ *sufficiently small we have*

$$\begin{aligned}\epsilon^{-m}\int_{y\in\mathcal{M}} k\left(\frac{|x-y|^2}{\epsilon^2}\right) f(y)\, dV = m_0^\partial(x)f(x) &+ \epsilon m_1^\partial(x)\left(\eta_x \cdot \nabla f(x) + \frac{m-1}{2}H(x)f(x)\right) \\ &+ \frac{\epsilon^2}{2}m_2\left(\tilde{\omega}(x)f(x) + \Delta f(x) + (m_2^\partial(x)/m_2 - 1)\frac{\partial^2}{\partial\eta_x^2}f(x)\right) \\ &+ \mathcal{O}(\epsilon^3 e^{-b_x^2/\epsilon^2}, \epsilon^4)\end{aligned} \tag{4.2}$$

where the moments $m_\ell^\partial(x)$ *are defined in* (2.2) *and* $\tilde{\omega}(x)$ *depends on the scalar curvature of* $\mathcal{M}$ *and second fundamental form of the embedding* $\mathcal{M} \subset \mathbb{R}^n$*, and* $H(x)$ *depends on the second fundamental form* $\partial\mathcal{M} \subset \mathcal{M}$*. Moreover, if* $\mathcal{M}$ *is compact this expansion holds uniformly on* $\mathcal{M}$*.*

Proof. We first localize the integral to a semigeodesic ϵ-ball, B by Lemma 4.1 making an error of higher order than $\mathcal{O}(\epsilon^4)$. Note that B is exactly the domain of the integral defining the coefficients $m_\ell^\partial(x)$ in (2.2). We then multiply three expansion. First, the kernel expansion,

$$k\left(\frac{||x-y||^2_{\mathbb{R}^d}}{\epsilon^2}\right) = k\left(\frac{||u||^2_{\mathcal{M}} - \left\langle \Pi_{\partial\mathcal{M}^t}(U^\top, U^\top), U^\perp\right\rangle_g + \omega_1(x,u) + \mathcal{O}(|u|^5)}{\epsilon^2}\right)$$
$$= k\left(\frac{||u||^2_{\mathcal{M}}}{\epsilon^2}\right) - k'\left(\frac{||u||^2_{\mathcal{M}}}{\epsilon^2}\right)\frac{1}{\epsilon^2}\left(\left\langle \Pi_{\partial\mathcal{M}^t}(U^\top, U^\top), U^\perp\right\rangle_g + \omega_1(x,u)\right) + \mathcal{O}(\epsilon^{-2}|u|^5)$$

which follows from Lemma 4.4 where $w_1(x,u)$ is a quartic polynomial in the components of u. Second, the Taylor expansion of f,

$$f(y) = f(x) + \frac{\partial f}{\partial u^i}u_i + \frac{1}{2}\frac{\partial^2 f}{\partial u^i \partial u^j}u^i u^j + \mathcal{O}(\epsilon^3)$$

and finally, by Theorem 4.5 we have the following expansion of the volume form

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

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 4.6. Finally, there are several order-ϵ^2 terms, first we consider,

$$\epsilon^2\int_B \frac{1}{2}k\left(|u|^2\right)\frac{\partial^2 f}{\partial u^i \partial u^j}u^i u^j\,du = \epsilon^2\sum_{i=1}^m\int_B \frac{1}{2}k\left(|u|^2\right)(u^i)^2\,du\frac{\partial^2 f}{\partial (u^i)^2}\,du$$
$$= \frac{m_2\epsilon^2}{2}\sum_{i=1}^{m-1}\frac{\partial^2 f}{\partial (u^i)^2} + \frac{m_2^\partial(x)\epsilon^2}{2}\frac{\partial^2 f}{\partial (u^m)^2}$$
$$= \frac{m_2\epsilon^2}{2}\Delta f(x) + \epsilon^2\frac{(m_2^\partial(x) - m_2)}{2}\frac{\partial^2 f}{\partial \eta_x^2}$$

where all the odd terms, $u^i u^j$ for $i \neq j$, cancel due to symmetry of the domain, and we add and subtract $m_2\frac{\partial^2 f}{\partial (u^m)^2}$ in order to identify the Laplace-Beltrami operator $\Delta f(x) = \sum_{i=1}^m \frac{\partial^2 f}{\partial (u^i)^2}$ since we are in semigeodesic coordinates. The next order-ϵ^2 term is

$$\epsilon^2 f(x)\int_B k\left(|u|^2\right)\omega_3(x)u^i u^j + k'\left(|u|^2\right)\left(\omega_1(x,u) + \frac{m-1}{2}\Pi(u,u)(u^m)^2 H(x)\right)\,du$$
$$= \epsilon^2 f(x)\tilde{\omega}(x)$$

which depends on the intrinsic and extrinsic curvatures of the manifold and the boundary. We should note that this term has been worked out for interior [11] but the Lemma 4.4 would need to

be extended to the quartic term to fully identify $\tilde{\omega}$. The final order-ϵ^2 term is,

$$\epsilon^2 \int_B -k\left(|u|^2\right) \frac{\partial f}{\partial u^i} u^i u^m (m-1) H(x) - k'\left(|u|^2\right) \frac{\partial f}{\partial u^i} u^i \Pi(\partial_j, \partial_k) u^j u^k u^m \, du$$

which follows from noting that each integral is zero unless $i = m$, and the second integral requires $j = k$ (recall that $1 \leq j, k \leq m-1$). Moreover, $k'\left(|u|^2\right) u^j u^j = \frac{1}{2}\left(\frac{\partial}{\partial u^j} k\left(|u|^2\right)\right) u^j$, and integrating by parts as in Lemma 4.6 yields

$$\epsilon^2 \int_B -k\left(|u|^2\right) (u^m)^2 (m-1) H(x) \frac{\partial f}{\partial u^m} + \frac{1}{2} \sum_{j=1}^{m-1} \Pi(\partial_j, \partial_j) \frac{\partial f}{\partial u^m} k\left(|u|^2\right) (u^m)^2 \, du$$

then substituting $\sum_{j=1}^{m-1} \Pi(\partial_j, \partial_j) = (m-1)H(x)$ and $\frac{\partial f}{\partial u^m} = -\frac{\partial f}{\partial \eta_x}$ and $u^m = -u \cdot \eta_x$ we have

$$-\frac{\epsilon^2}{2}(m-1)H(x)\frac{\partial f}{\partial u^m} \int_B k\left(|u|^2\right) (u \cdot \eta_x)^2 \, du = \frac{\epsilon^2}{2}(m-1)H(x)\frac{\partial f}{\partial \eta_x} m_2^\partial(x)$$

□

In order for the above asymptotic expansion to hold we require that ϵ is less than the width of the normal collar and also less than the infimum of the injectivity radius outside of the normal collar. Notice that for $b_x \gg \epsilon$ we have $m_0^\partial = m_0$ and $m_2^\partial = m_2$ and $m_1^\partial = 0$ up to higher order terms in ϵ, so this expansion agrees with (1.3) in the interior of the manifold.

4.1. **Examples.** We now turn to some simple examples to verify (4.2). 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. 2 we verify* (4.2) *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. 2. 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

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

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* $\mathcal{M} = \{(\tilde{x}, \tilde{y}) \,|\, \tilde{x}^2/a^2 + \tilde{y}^2/b^2 \leq 1\}$ *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*

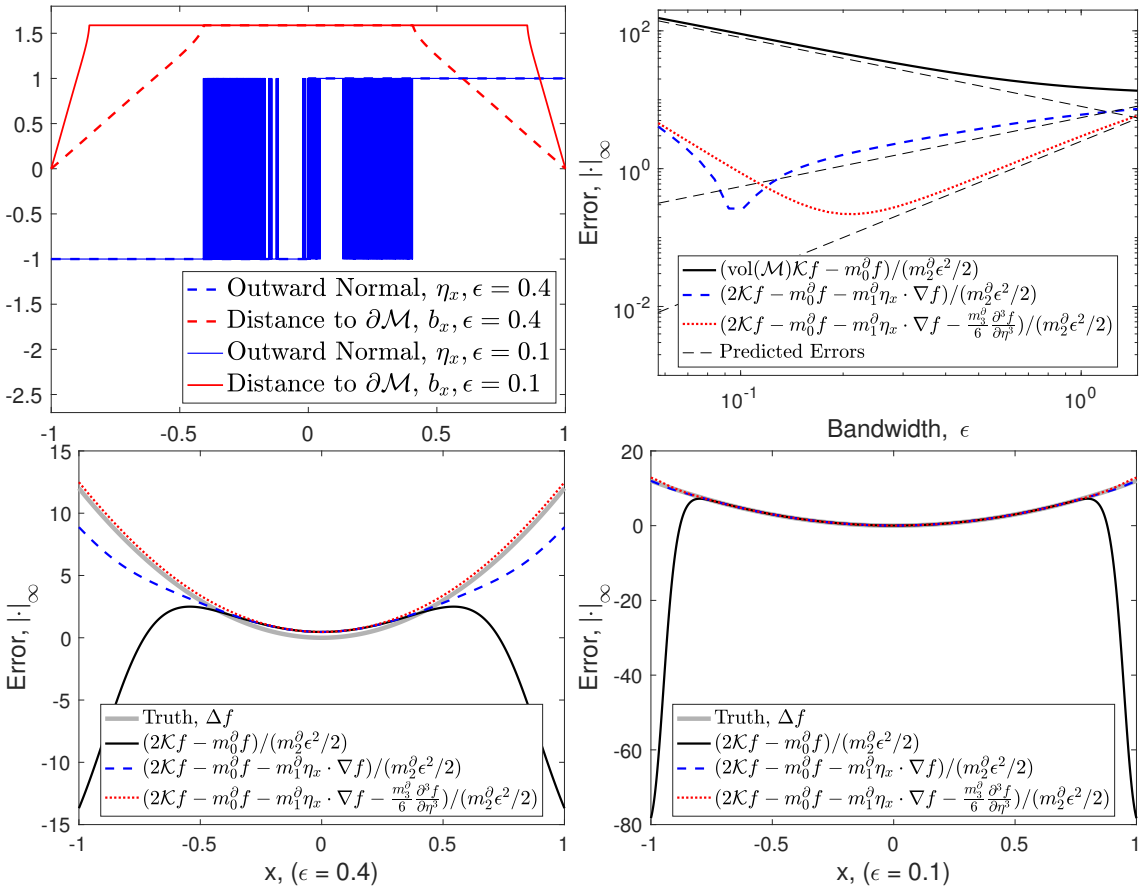

FIGURE 2. Verifying (4.2) by extracting the Laplacian on the interval $[-1, 1]$ applied to the function $f(x) = x^4$. Top, left: We show the estimate of b_x and η_x from the previous section, the b_x estimate saturates when $b_x \gg h$ and the η_x estimate is very noisy far from the boundary. Top, right: 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).

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 (4.2) becomes,

$$\epsilon^{-m} \int_{y \in \mathcal{M}} k\left(\frac{|x-y|^2}{\epsilon^2}\right) dV = m_0^\partial(x) + \epsilon m_1^\partial(x)\frac{m-1}{2}H(x) + \frac{\epsilon^2}{2}m_2\tilde{\omega}(x) + \mathcal{O}(\epsilon^3) \tag{4.3}$$

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. 3(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

$$(m-1)H(x) = \text{trace}\left(\left\langle \Pi(U^\top, U^\top), U^\perp \right\rangle\right) = \text{trace}\left(\left\langle \Pi\left(\frac{V^\top}{||V^\top||}, \frac{V^\top}{||V^\top||}\right), \frac{V^\perp}{||V^\perp||}\right\rangle\right)$$
$$= \frac{\text{trace}\left(\langle \Pi(V^\top, V^\top), V^\perp\rangle\right)}{||V^\top||^2||V^\perp||} = \frac{D^2\iota \cdot \frac{V^\perp}{||V^\perp||}}{||V^\top||^2} = \frac{ab}{(b^2\cos^2\theta + a^2\sin^2\theta)^{3/2}} \tag{4.4}$$

which is simply the standard (extrinsic) curvature of the parameterized curve. This function is shown as the solid grey curve in Fig. 3(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. 3(left) we only show points with distance to the boundary less than $\epsilon/4$.

Next we verify the derivative terms in (4.2) *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 (4.3) *by* $f(x)$ *matches many of the terms from* (4.2), *so subtracting this from* (4.2) *we isolate the terms*

$$\left(\mathbf{K}\vec{f}\right)_i - f(x_i)\left(\mathbf{K}\vec{1}\right)_i \to \epsilon^{-m}\int_{y\in\mathcal{M}} k\left(\frac{|x-y|^2}{\epsilon^2}\right) f(y)\, dV - f(x)\epsilon^{-m}\int_{y\in\mathcal{M}} k\left(\frac{|x-y|^2}{\epsilon^2}\right) dV$$
$$= \epsilon m_1^\partial(x)\nabla f(x)\cdot\eta_x + \epsilon^2\frac{m_2^\partial(x)}{2}\Delta f(x) \tag{4.5}$$

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. 3(right). This validates the derivative terms in* (4.2).

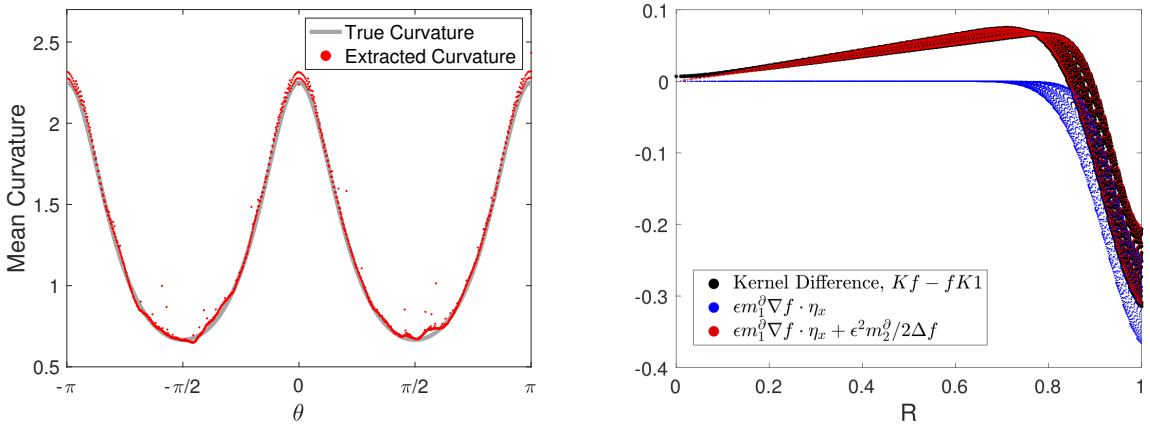

FIGURE 3. **Left:** Verifying the mean curvature $H(x)$ in the order-ϵ term of (4.2) on the ellipse. **Right:** Verifying the derivative terms in the expansion (4.2).

5. Estimating Boundary Integrals and the Weak Laplacian

The purpose of this section is to use the distance to the boundary to construct consistent estimators of the operators given in (1.4).

The standard method of extracting 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 $\mathbf{D}_{ii} = (\mathbf{K}\vec{1})_i \propto \mathcal{K}1(x_i)$. Since $\mathbb{E}[\mathcal{K}1] = m_0^\partial q + \epsilon^2 m_2 \tilde{\omega} q/2 + \mathcal{O}(\epsilon^3)$, the expected value of $\mathbf{L}\vec{f}$ is

$$\mathbb{E}[\mathbf{L}\vec{f}] = -\epsilon^{-1} m_1^\partial \eta_x \cdot \nabla(fq) - \Delta(fq) - \frac{1}{2}(m_2^\partial - m_2)\frac{\partial^2}{\partial \eta_x^2}(fq) + \mathcal{O}(\epsilon).$$

When q is not constant (meaning the sampling is nonuniform) several normalizations are required as will be discussed below in Section 5.3, but for a manifold with uniform sampling we have $q \equiv \mathrm{vol}(\mathcal{M})^{-1}$. Assuming uniform sampling, notice that we recover the Laplacian in the interior of the manifold, but that the estimator blows up like ϵ^{-1} near the boundary. This was first pointed out in [7] who derived the first two terms of (4.2). They argued that the graph Laplacian is a consistent estimator for Neumann functions, which is true but does not explain the empirically observed fact that the eigenvectors of $\mathbf{L}$ approximate the Neumann eigenfunctions of the Laplacian. In Section 5.2 we will finally be able to explain this long observed phenomenon. First, we show that integrals that are weighted by functions that decay away from the boundary approximate boundary integrals.

5.1. Estimating boundary integrals.

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

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

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 2 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 ∂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 4.7, let* $d\mathrm{vol}_\partial$ *be the natural volume element on the boundary inherited from* $d\mathrm{vol}$*, then we have*

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

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* ∂M *at* $y \in \partial M$.

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 $N_{\epsilon^\gamma} = \{y \in \mathcal{M} : b_y < \epsilon^\gamma\}$ denote the normal collar of width ϵ^γ. By using an identical argument as Lemma 4.1, one can localize the integral over N_{ϵ^γ} so that

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

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 coordinate charts. By taking intersections and complements, we can then generate a covering of M by measurable sets $\{V_j\}_{j\in J}$ of ∂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 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 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:

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

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 $d\text{vol}$ about $u^n = 0$ using the first variation of area:

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

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 $d\text{vol} = \sqrt{|g|}$ 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 $d\text{vol}(u^\top, 0) = d\text{vol}_\partial(u^\top)$.

We can then separate terms involving u^n to obtain:

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

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

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

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

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

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

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

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. 4. 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. 4. Again, we emphasize that this example is purely for verification of Theorem 5.1, we will return to practical computation methods in Section **??**.

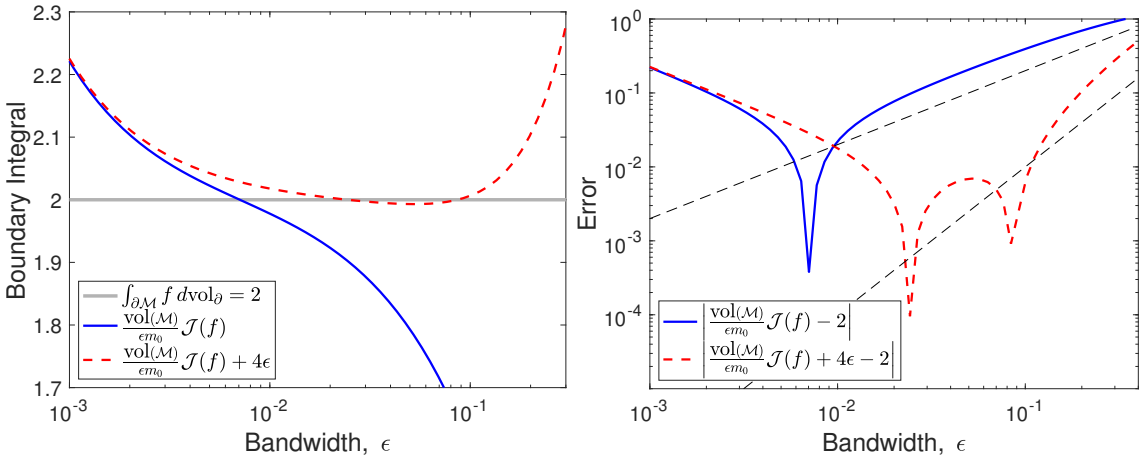

FIGURE 4. Verifying (5.1) by extracting the boundary integral on the interval $[-1, 1]$ applied to the function $f(x) = x^4$. Left: We show the estimate of the boundary integral compared to the true value, 2. Right: Error vs. bandwidth for various estimators of the boundary integral.

5.2. **Estimating the weak Laplacian.** We can now use (4.2) together with (5.1) to understand the weak form of the kernel operator and the graph Laplacian. 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,

$$\frac{\epsilon^{-m}}{N^2}\mathbb{E}\left[\vec{\phi}^\top \mathbf{K}\vec{f}\right] = \epsilon^{-m}\int_{x\in\mathcal{M}}\int_{y\in\mathcal{M}} \phi(x)q(x)K\left(\frac{|x-y|}{\epsilon}\right) f(y)q(y)\, d\text{vol}\, d\text{vol}. \tag{5.3}$$

By expanding the inner integral using (4.2) and applying Theorem 5.1 we derive the following result,

Theorem 5.2. *In the same context as Theorem 4.7 for all $\phi, f \in C^3(\mathcal{M})$ we have,*

$$\frac{2}{m_2 N}\mathbb{E}\left[\vec{\phi}^\top \mathbf{L}\vec{f}\right] = \int_{\mathcal{M}} (\nabla\phi \cdot \nabla f)\, q^2 d\text{vol} + \mathcal{O}(\epsilon). \tag{5.4}$$

Proof. For clarity we restrict to the case of uniform sampling, $q \equiv \text{vol}(\mathcal{M})^{-1}$ and $K(z) = e^{-z}$ since we have explicit formulas for the moments, for the full proof see Section A. Starting from

(5.3) and expanding the inner integral we have,

$$\frac{\epsilon^{-m}\mathrm{vol}(\mathcal{M})^2}{N^2}\mathbb{E}\left[\vec{\phi}^\top \mathbf{K}\vec{f}\right] = \int_{\mathcal{M}} m_0^\partial \phi f + \epsilon m_1^\partial \phi(\eta_x\cdot\nabla f - fH) + \frac{\epsilon^2}{2}m_2(\tilde{\omega}\phi f + \phi\Delta f) + \frac{\epsilon^2}{2}(m_2^\partial - m_2)\phi\frac{\partial^2}{\partial\eta_x^2} f\, d\mathrm{vol} + \mathcal{O}(\epsilon^3). \tag{5.5}$$

We can now interpret (5.5) in the light of (5.1) since the function $m_1^\partial(x) = -\pi^{(m-1)/2}\exp(-b_x^2/\epsilon^2)/2$ is concentrated at the boundary and has exponential decay away from the boundary. Since m_1^∂ satisfies the requirements of Theorem 5.1 we can rewrite the integral as a boundary integral up to higher order terms

$$\epsilon\int_{\mathcal{M}} m_1^\partial\phi\,(\eta_x\cdot\nabla f)\,d\mathrm{vol} = -\epsilon^2 m_0/4\int_{\partial\mathcal{M}}\phi\,(\eta_x\cdot\nabla f)\,d\mathrm{vol}_\partial + \mathcal{O}(\epsilon^3)$$

where notice that $-\pi^{(m-1)/2}\overline{m}_0/2 = -\pi^{m/2}/4 = -m_0/4$. Crucially, while this term is order-ϵ in the pointwise expansion, because it is concentrated in an ϵ-neighborhood of the boundary, it is order-ϵ^2 in the weak expansion. Similarly, since $m_2^\partial - m_2$ decays faster than any polynomial away from the boundary this term is actually order-ϵ^3 and can thus be dropped from (5.5). The only remaining term of order-ϵ is the mean curvature term, which will be cancelled by forming the graph Laplacian.

Next we recall that $\mathbf{D}_{ii} = \sum_{j=1}^N \mathbf{K}_{ij} = \sum_{j=1}^N \mathbf{K}_{ij}\vec{1}_j$, meaning that the entries of $\mathbf{D}$ estimate the kernel operator applied to constant function, thus

$$\frac{\epsilon^{-m}\mathrm{vol}(\mathcal{M})}{N}\mathbb{E}[\mathbf{D}_{ii}] = m_0^\partial(x_i) - \epsilon m_1^\partial(x_i)H(x_i) + \frac{\epsilon^2}{2}m_2\tilde{\omega}(x_i) + \mathcal{O}(\epsilon^3). \tag{5.6}$$

Since $\mathbf{D}$ is diagonal, the inner product $\vec{\phi}^\top\mathbf{D}\vec{f} = \sum_{i=1}^N \vec{\phi}_i D_{ii} f_i$ estimates the following integral

$$\frac{\epsilon^{-m}\mathrm{vol}(\mathcal{M})^2}{N^2}\mathbb{E}\left[\vec{\phi}^\top\mathbf{D}\vec{f}\right] = \int_{\mathcal{M}} m_0^\partial\phi f - \epsilon m_1^\partial\phi f H + \frac{\epsilon^2}{2}m_2\tilde{\omega}\phi f\, d\mathrm{vol} + \mathcal{O}(\epsilon^3). \tag{5.7}$$

Applying (5.5) and (5.7) to the graph Laplacian $\mathbf{L} = c(\mathbf{D}-\mathbf{K})$ we find that many terms cancel and we are left with,

$$\frac{\mathrm{vol}(\mathcal{M})^2}{N}\mathbb{E}\left[\vec{\phi}^\top\mathbf{L}\vec{f}\right] = \frac{m_0}{4}\int_{\partial\mathcal{M}}\phi\eta_x\cdot\nabla f\, d\mathrm{vol}_\partial - \frac{m_2}{2}\int_{\mathcal{M}}\phi\Delta f\, d\mathrm{vol} + \mathcal{O}(\epsilon). \tag{5.8}$$

so finally, we can apply integration by parts to recover a symmetric form

$$\begin{aligned}\frac{\mathrm{vol}(\mathcal{M})^2}{N}\mathbb{E}\left[\vec{\phi}^\top\mathbf{L}\vec{f}\right] &= \left(\frac{m_0}{4} - \frac{m_2}{2}\right)\int_{\partial\mathcal{M}}\phi\eta_x\cdot\nabla f\, d\mathrm{vol}_\partial + \frac{m_2}{2}\int_{\mathcal{M}}\nabla\phi\cdot\nabla f\, d\mathrm{vol} + \mathcal{O}(\epsilon)\\ &= \frac{m_2}{2}\int_{\mathcal{M}}\nabla\phi\cdot\nabla f\, d\mathrm{vol} + \mathcal{O}(\epsilon).\end{aligned} \tag{5.9}$$

where the first term cancels since $m_2 = m_0/2$. Finally, multiplying both sides by $2q^2/m_2 \equiv \frac{2}{m_2}\mathrm{vol}(\mathcal{M})^{-2}$ yields the result in uniform sampling case. □

The above proof can easily be generalized to all Gaussian kernels $K(z) = e^{-z/\sigma^2}$. In fact, the cancellation of the boundary integral term must hold for all kernels with exponential decay since $\mathbf{L}$ is a symmetric matrix and must represent a symmetric operator as is shown in Appendix A. Generalizing the non-uniform sampling is simply a more delicate computation and is left to Appendix A since the uniform sampling case more clearly illustrates the key ideas.

Returning to our simple example of the uniform grid on the interval $[-1,1]$ with $f(x) = x^4$, we verify (5.4) by estimating $\int_{\mathcal{M}}|\nabla f|^2\, d\mathrm{vol} = 32/7$ 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 [5].

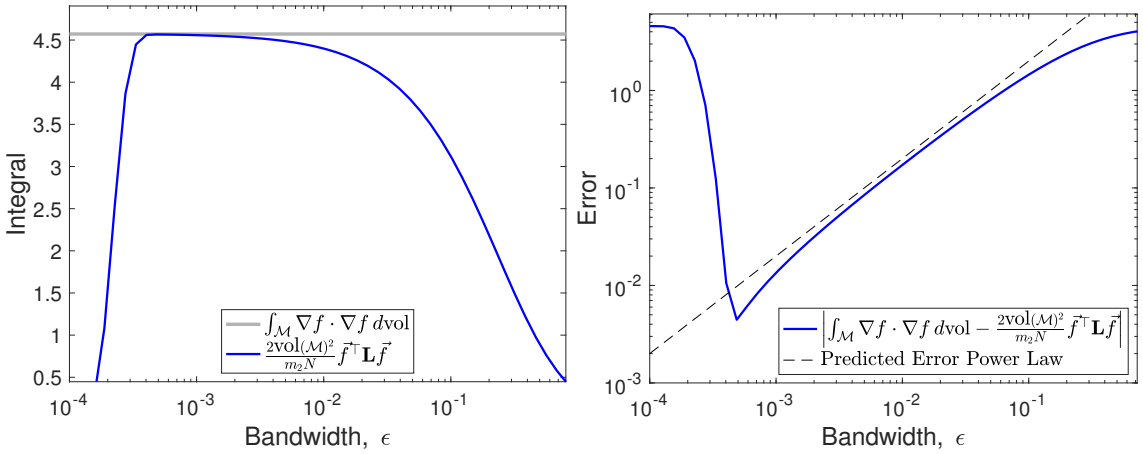

FIGURE 5. Verifying (5.9) computing the exact integral and the graph Laplacian estimator on the interval $[-1, 1]$ applied to the function $\phi(x) = f(x) = x^4$. Left: We compare the estimate to the true integral value as a function of the bandwidth parameter, ϵ. Right: Error vs. bandwidth for the graph 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 the bandwidth required for pointwise consistency will decrease to zero as you approach the boundary.

5.3. **Correcting for the sampling density.** In (4.2) all the terms are influenced by the density q, so to remove this influence we apply the 'right-normalization' introduced by [7]. The idea is to apply (4.2) 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. The one necessary change from the method of [7] is that we need to use a density estimator that is consistent all the way to the boundary which requires dividing by m_0^∂ as we will see. First, notice that

$$\begin{aligned} q_{\epsilon,N} = \epsilon^{-m} \mathbb{E}[\mathcal{K}1(x)] = m_0^\partial(x) q(x) + \epsilon m_1^\partial(x) \left(\eta_x \cdot \nabla q(x) - H(x) q(x)\right) \\ + \frac{\epsilon^2}{2} m_2 \left(\tilde{\omega}(x) q(x) + \Delta q(x) + (m_2^\partial(x)/m_2 - 1) \frac{\partial^2}{\partial \eta_x^2} q(x) \right) \\ + \mathcal{O}(\epsilon^3 e^{-b_x^2/\epsilon^2}, \epsilon^4) \end{aligned} \tag{5.10}$$

so as introduced in [4] we can form a consistent density estimator by dividing by $m_0^\partial(x)$ (estimated using the methods of Section 2). If we divide f by $q_{\epsilon,N}/m_0^\partial$ and then apply the kernel operator we

have,

$$\begin{aligned}\mathbb{E}\left[\mathcal{K}\left(\frac{m_0^\partial f}{\mathbb{E}[\mathcal{K}1]}\right)\right] &= \epsilon^{-m}\mathbb{E}\left[\mathcal{K}\left(\frac{f}{q}\left(1-\epsilon\omega_0+\epsilon^2\omega_1+\mathcal{O}(\epsilon^3)\right)\right)\right] \\ &= m_0^\partial f - \epsilon f\omega_0 + \epsilon m_1^\partial \eta_x\cdot\nabla f + \frac{\epsilon^2}{2}m_2(f\omega_2+\Delta f)+\frac{\epsilon^2}{2}(m_2^\partial - m_2)\frac{\partial^2 f}{\partial \eta_x^2}+\mathcal{O}(\epsilon^3)\end{aligned} \tag{5.11}$$

where $\omega_0 = m_1^\partial \eta_x\cdot\frac{\nabla q}{q}$ and $\omega_1 = \frac{m_1^\partial \eta_x\cdot\nabla q}{q} - m_2(\tilde{\omega}+(\Delta q)/q) - \frac{m_2^\partial - m_2}{q}\frac{\partial^2 q}{\partial \eta_x^2}$ and $\omega_2 = \tilde{\omega}+\omega_1$. Next we normalize on the left to form a symmetric matrix which will lead to a consistent weak-sense estimator.

Dividing f by the density estimator is incorporated into the kernel by multiplying on the right by the inverse of a diagonal matrix $\mathbf{D}_{ii} = \mathcal{K}1(x_i)/m_0^\partial(x_i)$. However, to remove the effect of the density on the weak-sense interpretation of the kernel matrix, we must also pre-divide ϕ by the density estimate, which corresponds to multiplying on the left by the $\mathbf{D}$.

$$\begin{aligned}\frac{\mathbb{E}\left[\mathcal{K}\left(\frac{m_0^\partial f}{\mathbb{E}[\mathcal{K}1]}\right)\right]}{\mathbb{E}\left[\frac{\mathcal{K}1}{m_0^\partial}\right]} &= \frac{m_0^\partial f - \epsilon f\omega_0 + \epsilon m_1^\partial\eta_x\cdot\nabla f + \frac{\epsilon^2}{2}m_2(f\omega_2+\Delta f)+\frac{\epsilon^2}{2}(m_2^\partial-m_2)\frac{\partial^2 f}{\partial\eta_x^2}}{q+\epsilon\frac{m_1^\partial}{m_0^\partial}(\eta_x\cdot\nabla q - Hq)+\frac{\epsilon^2}{2}\frac{m_2}{m_0^\partial}\left(\tilde{\omega}q+\Delta q+(m_2^\partial/m_2-1)\frac{\partial^2 q}{\partial\eta_x^2}\right)}+\mathcal{O}(\epsilon^3) \\ &= \frac{m_0^\partial f - \epsilon f\omega_0 + \epsilon m_1^\partial\eta_x\cdot\nabla f + \frac{\epsilon^2}{2}m_2(f\omega_2+\Delta f)+\frac{\epsilon^2}{2}(m_2^\partial-m_2)\frac{\partial^2 f}{\partial\eta_x^2}}{q\left(1+\epsilon\frac{m_1^\partial}{m_0^\partial}\left(\eta_x\cdot\frac{\nabla q}{q}-H\right)+\frac{\epsilon^2}{2}\frac{m_2}{m_0^\partial}\left(\tilde{\omega}+\frac{\Delta q}{q}+(m_2^\partial/m_2-1)q^{-1}\frac{\partial^2 q}{\partial\eta_x^2}\right)\right)}+\mathcal{O}(\epsilon^3) \\ &= \frac{1}{q}\left(m_0^\partial f - \epsilon f\omega_3 + \epsilon m_1^\partial\eta_x\cdot\nabla f\right) \\ &\quad + \frac{\epsilon^2}{2q}\left(f\omega_4 + m_2\Delta f + (m_2^\partial-m_2)\frac{\partial^2 f}{\partial\eta_x^2}+m_1^\partial\omega_5\eta_x\cdot\nabla f\right)+\mathcal{O}(\epsilon^3)\end{aligned} \tag{5.12}$$

Note that applying these kernel normalizations is a form of ratio estimator and computing the variance (equivalently quadrature error) is somewhat subtle [19, 2]. We have now formed a symmetrically normalized kernel matrix $\mathbf{tK} = \mathbf{D}^{-1}\mathbf{K}\mathbf{D}^{-1}$. Finally, we form the graph Laplacian for this normalized kernel matrix by defining the diagonal matrix $\mathbf{tD}_{ii} = \left(\mathbf{tK}\vec{1}\right)_i = \sum_{j=1}^N \mathbf{tK}_{ij}$ and the graph Laplacian matrix $\mathbf{tL} = c(\mathbf{tD}-\mathbf{tK})$. We note that in the infinite data limit the entries of $\mathbf{tD}$ converge to (setting $f\equiv 1$ in the above expression)

$$\frac{\mathbb{E}\left[\mathcal{K}\left(\frac{m_0^\partial 1}{\mathbb{E}[\mathcal{K}1]}\right)\right]}{\mathbb{E}\left[\frac{\mathcal{K}1}{m_0^\partial}\right]} = \frac{1}{q}\left(m_0^\partial - \epsilon\omega_3+\frac{\epsilon^2}{2}\omega_4\right)+\mathcal{O}(\epsilon^3). \tag{5.13}$$

Thus, when we analyze the graph Laplacian in the weak sense, $\vec{\phi}^\top\mathbf{tL}\vec{f}$ we see that the leading terms (order-ϵ^0) and the ω_3 and ω_4 terms from $\mathbf{tD}$ and $\mathbf{tK}$ exactly cancel. Moreover, the terms

$$\frac{\epsilon^2}{2}\left((m_2^\partial-m_2)\frac{\partial^2 f}{\partial\eta_x^2}+m_1^\partial\omega_5\eta_x\cdot\nabla f\right)$$

are concentrated near the boundary and thus are actually order-ϵ^3. The remaining terms are,

$$\vec{\phi}^\top\mathbf{tL}\vec{f}\to_{N\to\infty} c\int_{\mathcal{M}}\frac{1}{\epsilon}m_1^\partial\phi\eta_x\cdot\nabla f+\frac{m_2}{2}\phi\Delta f\,d\text{vol}+\mathcal{O}(\epsilon) = \int_{\mathcal{M}}\nabla\phi\cdot\nabla f\,d\text{vol}+\mathcal{O}(\epsilon) \tag{5.14}$$

where the final equality follows as in (5.9) and the convergence as $N\to\infty$ is in probability [19, 2].

To verify (5.14) 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. 6 we show that the normalized graph Laplacian $\mathbf{tL}$ constructed in this section recovers the weak-sense Laplacian for the same example as in the previous section on this nonuniform grid.

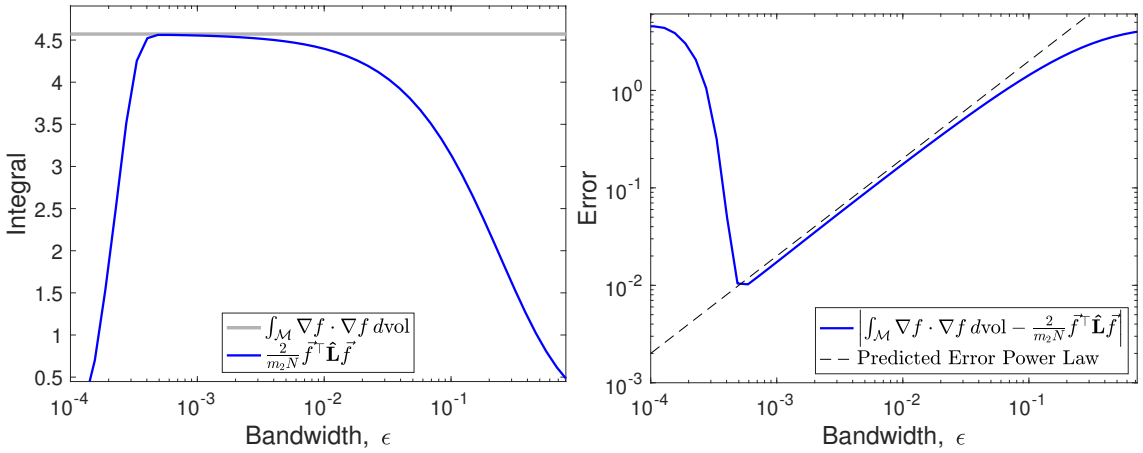

FIGURE 6. Verifying (5.14) computing the exact integral and the graph Laplacian estimator on the interval $[-1,1]$ applied to the function $\phi(x) = f(x) = x^4$. Left: We compare the estimate to the true integral value as a function of the bandwidth parameter, ϵ. Right: Error vs. bandwidth for the graph Laplacian as a weak-sense estimator.

6. Applications to Elliptic and Parabolic PDEs

Next, we shall adopt the tools created so far to solve elliptic and parabolic partial differential equations (PDEs) with various boundary conditions. For simplicity of presentation, we shall write our formulation for problems with Dirichlet and Neumann boundary conditions. However, the approach directly applies to problems with Robin or mixed boundary conditions.

6.1. **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

$$\begin{cases} -\Delta u &= f \quad \text{in } \Omega \\ u &= g \quad \text{on } \partial\Omega. \end{cases} \tag{6.1}$$

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 (6.1) in the following sense

$$\int_\Omega \nabla w \cdot \nabla v \, dx = \langle f, v \rangle - \int_\Omega \nabla \tilde{g} \cdot \nabla v \, dx \quad \forall v \in H_0^1(\Omega). \tag{6.2}$$

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

$$\begin{cases} -\Delta u + u &= f \quad \text{in } \Omega \\ \nabla u \cdot \eta &= g \quad \text{on } \partial\Omega \end{cases} \tag{6.3}$$

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 (6.3) in the following sense:

$$\int_\Omega \nabla u \cdot \nabla v + uv \, dx = \langle f, v \rangle + \int_{\partial\Omega} gv \, ds \quad \forall v \in H^1(\Omega). \tag{6.4}$$

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

$$\begin{cases} \partial_t u - \Delta u + u &= f \quad \text{in } \Omega \times (0, T) \\ u &= 0 \quad \text{on } \partial\Omega \times (0, T) \\ u &= u_0 \quad \text{in } \Omega \, . \end{cases} \tag{6.5}$$

The notion of weak solution to (6.5) is: find $u \in L^2(0, T; H_0^1(\Omega)) \cap H^1(0, T; H^{-1}(\Omega))$ solving

$$\langle \partial_t u, v \rangle + \int_\Omega \nabla u \cdot \nabla v \, dx = \langle f, v \rangle \quad \forall v \in H_0^1(\Omega) \tag{6.6}$$

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

6.2. **Discrete System.** Next we describe the linear algebraic systems we obtain after discretization of (6.2), (6.4), and (6.6). We recall that $\mathbf{tL}$ and $\mathbf{D}$ denote the discrete form of the Laplacian (in weak form) and the discretization of the integral over the entire manifold $\mathcal{M}$, respectively. Namely,

$$\int_{\mathcal{M}} \nabla u \cdot \nabla v \approx \mathbf{v}^\top \mathbf{tLu} \quad \text{and} \quad \int_{\mathcal{M}} fv \approx \mathbf{v}^\top \mathbf{Df}.$$

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

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

where $\mathbf{B}$ is a rectangular matrix with number of rows equal to the number of degrees of freedom in the interior and number of columns equal to the number of degrees of freedom on the boundary. The discrete form of (6.2) is given by

$$\mathbf{tL}(\mathrm{f_{dof}}, \mathrm{f_{dof}})\mathbf{w} = \mathbf{D}(\mathrm{f_{dof}}, :)\mathbf{f} - \mathbf{tL}(\mathrm{f_{dof}}, :)\tilde{\mathbf{g}}$$

where f_{dof} indicates the 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. Then the discrete solution on the interior nodes is $\mathbf{u}(\mathrm{f_{dof}}) = \mathbf{w}$.

The discrete form of the the system (6.4) is

$$(\mathbf{tL} + \mathbf{D})\mathbf{u} = \mathbf{Df} + \mathbf{Bg}.$$

Finally, we describe the discretization of (6.6). 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 $k = 1, \ldots, K$, we solve

$$\left(\mathbf{D}(\mathrm{f_{dof}}, \mathrm{f_{dof}}) + \tau \mathbf{tL}(\mathrm{f_{dof}}, \mathrm{f_{dof}})\right) \mathbf{u}^k = \tau \mathbf{D}(\mathrm{f_{dof}}, :)\mathbf{f}^k + \mathbf{D}(\mathrm{f_{dof}}, \mathrm{f_{dof}})\mathbf{u}^{k-1}.$$

6.3. **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 (6.2), (6.4), and (6.6). 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 6.1 (Elliptic homogeneous Dirichlet)**.** In (6.1) we set $g \equiv 0$, therefore $\tilde{g} \equiv 0$ (cf. (6.2)). 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: $\|u - u_h\|_{L^2(\Omega)} = 1.541989\text{e-}02$. Figure 7 shows a visual comparison between the computed and the exact solution.

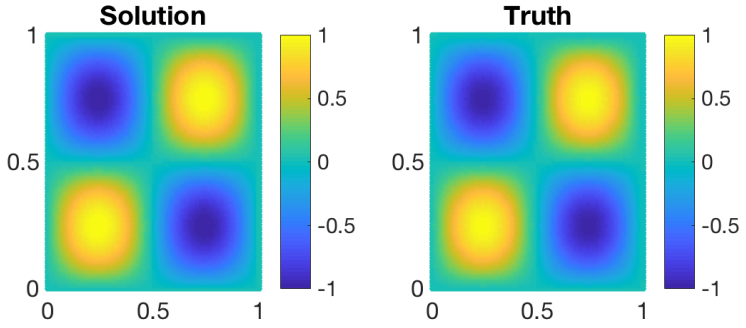

FIGURE 7. Left: Solution computed using our approach. Right: Exact solution.

Example 6.2 (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: $\|u - u_h\|_{L^2(\Omega)} = 6.378652\text{e-}03$. Figure 8 shows a visual comparison between the computed and the exact solution.

Example 6.3 (Elliptic homogeneous Neumann)**.** In (6.4) 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: $\|u - u_h\|_{L^2(\Omega)} = 2.125979\text{e-}02$. Figure 9 shows a visual comparison between the computed and the exact solution.

Example 6.4 (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: $\|u - u_h\|_{L^2(\Omega)} = 8.303406\text{e-}02$. Figure 10 shows a visual comparison between the computed and the exact solution.

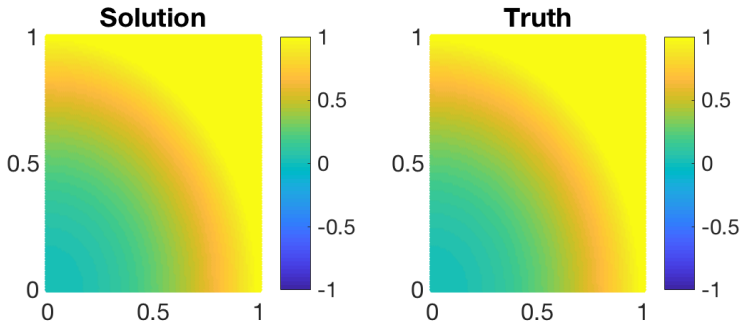

FIGURE 8. Left: Solution computed using our approach. Right: Exact solution.

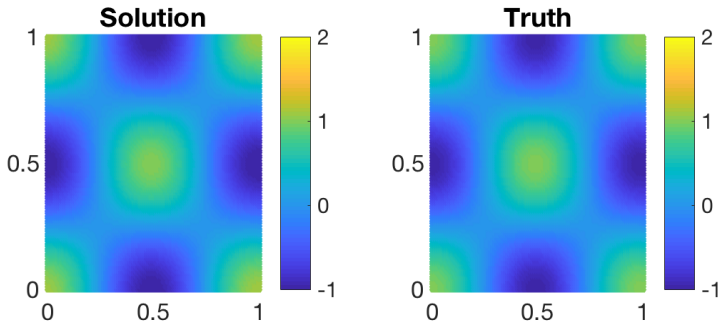

FIGURE 9. Left: Solution computed using our approach. Right: Exact solution.

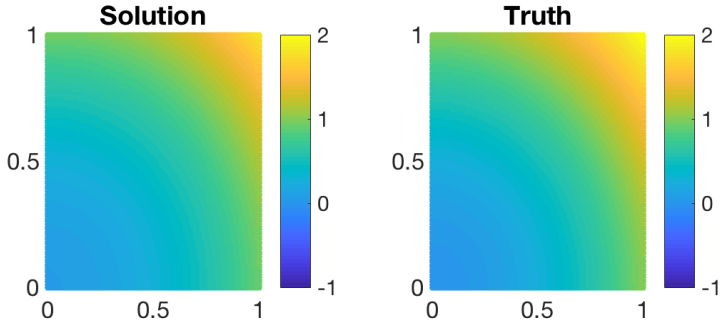

FIGURE 10. Left: Solution computed using our approach. Right: Exact solution.

Example 6.5 (Parabolic homogeneous Dirichlet)**.** In (6.6) 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 6.6 (Dirichlet on the hemisphere)**.** We now consider (6.1) 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

$$\Delta f = \frac{1}{\sqrt{|g|}} \begin{bmatrix} \frac{\partial}{\partial\theta} & \frac{\partial}{\partial\phi} \end{bmatrix} \left(\sqrt{|g|}\ g^{-1} \begin{bmatrix} \frac{\partial f}{\partial\theta} \\ \frac{\partial f}{\partial\phi} \end{bmatrix} \right) = \cot\theta\, \frac{\partial f}{\partial\theta} + \frac{\partial^2 f}{\partial\theta^2} + \csc^2\theta\, \frac{\partial^2 f}{\partial\phi^2}. \tag{6.7}$$

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

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

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 (6.1) using the estimator of the Laplacian **tL**. The resulting solution estimate u_h is compared in Fig. 11. The error between the exact solution u and it's approximation u_h using our proposed method in L^2-norm is: $\|u - u_h\|_{L^2(\mathcal{M})} = 5.067884\text{e-}03$.

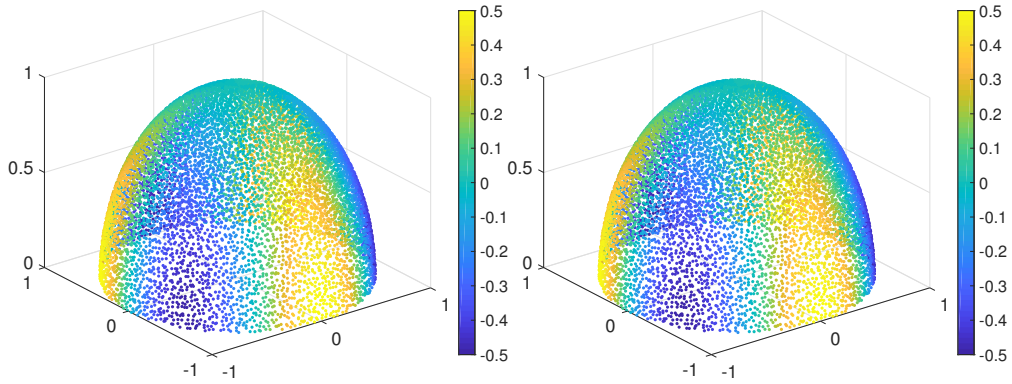

FIGURE 11. Left: Solution computed using our approach. Right: Exact solution.

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.

REFERENCES

[1] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[2] Tyrus Berry and John Harlim. Variable bandwidth diffusion kernels. *Applied and Computational Harmonic Analysis*, 40(1):68–96, 2016.

[3] Tyrus Berry and Timothy Sauer. Local kernels and the geometric structure of data. *Applied and Computational Harmonic Analysis*, 40(3):439 – 469, 2016.

[4] Tyrus Berry and Timothy Sauer. Density estimation on manifolds with boundary. *Computational Statistics and Data Analysis*, 107:1 – 17, 2017.

[5] Tyrus Berry and Timothy Sauer. Consistent manifold representation for topological data analysis. *Foundations of Data Science*, 1(1):1–38, 2019.
[6] Bennett Chow, Peng Lu, and Lei Ni. *Hamilton's Ricci flow*, volume 77. American Mathematical Soc., 2006.
[7] Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.
[8] Faheem Gilani and John Harlim. Approximating solutions of linear elliptic pde's on a smooth manifold using local kernel. *Journal of Computational Physics*, 395:563 – 582, 2019.
[9] Alfred Gray. *Tubes*, volume 221. Birkhäuser, 2012.
[10] John Harlim, Daniel Sanz-Alonso, and Ruiyi Yang. Kernel methods for bayesian elliptic inverse problems on manifolds. *arXiv preprint arXiv:1910.10669*, 2019.
[11] M Hein, JY Audibert, and U von Luxburg. Convergence of graph laplacians on random neighborhood graphs. *Journal of Machine Learning Research*, 8(1325-1370):6–19, 2007.
[12] Yoon Tae Kim and Hyun Suk Park. Geometric structures arising from kernel density estimation on Riemannian manifolds. *Journal of Multivariate Analysis*, 114:112–126, 2013.
[13] John Lee. *Introduction to Riemannian Manifolds*, volume 176. Springer International Publishing, 2018.
[14] Arkadas Ozakin and Alexander G Gray. Submanifold density estimation. In *Advances in Neural Information Processing Systems*, pages 1375–1382, 2009.
[15] Bruno Pelletier. Kernel density estimation on Riemannian manifolds. *Statistics & probability letters*, 73(3):297–304, 2005.
[16] S. A. Sarra. The matlab radial basis function toolbox. *Journal of Open Research Software*, 5, 2017.
[17] Francisco-Javier Sayas, Thomas S. Brown, and Matthew E. Hassell. *Variational techniques for elliptic partial differential equations*. CRC Press, Boca Raton, FL, 2019. Theoretical tools and advanced applications.
[18] Thomas Schick. Manifolds with boundary and of bounded geometry. *Mathematische Nachrichten*, 223(1):103–120, 2001.
[19] Amit Singer. From graph to manifold laplacian: The convergence rate. *Applied and Computational Harmonic Analysis*, 21(1):128–134, 2006.
[20] Amit Singer and Hau-Tieng Wu. Spectral convergence of the connection laplacian from random samples. *Information and Inference: A Journal of the IMA*, 6(1):58–123, 2016.
[21] Oleg G Smolyanov, Heinrich v Weizsäcker, and Olaf Wittich. Chernoff's theorem and discrete time approximations of brownian motion on manifolds. *Potential Analysis*, 26(1):1–29, 2007.

Appendix A. Proof of Theorem 5.2

In this section we present the full proof of Theorem 5.2 for non-uniform sampling.

Of Theorem 5.2. Starting from (5.3) and expanding the inner integral we have,

$$\frac{\epsilon^{-m}}{N^2}\mathbb{E}\left[\vec{\phi}^\top \mathbf{K}\vec{f}\right] = \int_{\mathcal{M}} m_0^\partial \phi f q^2 + \epsilon m_1^\partial \phi q(\eta_x\cdot\nabla(fq) - fqH) + \frac{\epsilon^2}{2}m_2(\tilde{\omega}\phi f q^2 + \phi q\Delta(fq)) + \frac{\epsilon^2}{2}(m_2^\partial - m_2)\phi q\frac{\partial^2}{\partial\eta_x^2}(fq)\,d\mathrm{vol} + \mathcal{O}(\epsilon^3). \tag{A.1}$$

We can now interpret (A.1) in the light of (5.1) since

$$\epsilon\int_{\mathcal{M}} m_1^\partial \phi q\eta_x\cdot\nabla(fq)\,d\mathrm{vol} = -\epsilon^2 m_0/4\int_{\partial\mathcal{M}}\phi q\eta_x\cdot\nabla(fq)\,d\mathrm{vol}_\partial + \mathcal{O}(\epsilon^3)$$

where $m_1^\partial = -\pi^{(m-1)/2}\exp(-b_x^2/\epsilon^2)/2$ and integrating the exponential we find $-\pi^{(m-1)/2}\overline{m}_0/2 = -\pi^{m/2}/4 = -m_0/4$. Similarly, since $m_2^\partial - m_2$ decays faster than any polynomial away from the boundary this term is actually order-ϵ^3 and can thus be dropped from (A.1). Next we recall that

$\mathbf{D}_{ii} = \sum_{j=1}^{N} \mathbf{K}_{ij}$ so that

$$\frac{\epsilon^{-m}}{N}\mathbb{E}[\mathbf{D}_{ii}] = m_0^{\partial}(x_i)q(x_i) + \epsilon m_1^{\partial}(x_i)(\eta_{x_i}\cdot\nabla q(x_i) - H(x_i)) + \frac{\epsilon^2}{2}m_2\left(\tilde{\omega}q(x_i) + \Delta q(x_i) + \left(\frac{m_2^{\partial}(x_i)}{m_2} - 1\right)\frac{\partial^2 q(x_i)}{\partial \eta_{x_i}^2}\right) + \mathcal{O}(\epsilon^3). \tag{A.2}$$

Since $\mathbf{D}$ is diagonal, the inner product $\vec{\phi}^{\top}\mathbf{D}\vec{f} = \sum_{i=1}^{N}\vec{\phi}_i D_{ii} f_i$ estimates the following integral

$$\frac{\epsilon^{-m}}{N^2}\mathbb{E}\left[\vec{\phi}^{\top}\mathbf{D}\vec{f}\right] = \int_{\mathcal{M}} m_0^{\partial}\phi f q^2 + \epsilon m_1^{\partial}\phi f q(\eta_x\cdot\nabla q - fqH) + \frac{\epsilon^2}{2}m_2(\tilde{\omega}\phi f q^2 + \phi f q\Delta q) + \frac{\epsilon^2}{2}(m_2^{\partial} - m_2)\phi f q\frac{\partial^2 q}{\partial \eta_x^2}\,d\mathrm{vol} + \mathcal{O}(\epsilon^3). \tag{A.3}$$

Applying (A.1) and (A.3) to the graph Laplacian $\mathbf{L} = c(\mathbf{D} - \mathbf{K})$ we find that many terms cancel and we are left with,

$$\begin{aligned}\frac{1}{N}\mathbb{E}\left[\vec{\phi}^{\top}\mathbf{L}\vec{f}\right] &= \frac{m_0}{4}\int_{\partial\mathcal{M}}\phi q\eta_x\cdot(\nabla(fq) - f\nabla q)\,d\mathrm{vol}_{\partial} - \frac{m_2}{2}\int_{\mathcal{M}}\phi q(\Delta(fq) - f\Delta q)\,d\mathrm{vol} + \mathcal{O}(\epsilon)\\ &= \frac{m_0}{4}\int_{\partial\mathcal{M}}\phi q^2\eta_x\cdot\nabla f\,d\mathrm{vol}_{\partial} - \frac{m_2}{2}\int_{\mathcal{M}}\phi q(q\Delta f + 2\nabla q\cdot\nabla f)\,d\mathrm{vol} + \mathcal{O}(\epsilon)\end{aligned} \tag{A.4}$$

where we apply the product rules for the gradient and Laplacian. Finally, we can apply integration by parts to recover a symmetric form

$$\begin{aligned}\frac{1}{N}\mathbb{E}\left[\vec{\phi}^{\top}\mathbf{L}\vec{f}\right] &= \left(\frac{m_0}{4} - \frac{m_2}{2}\right)\int_{\partial\mathcal{M}}\phi q^2\eta_x\cdot\nabla f\,d\mathrm{vol}_{\partial} + \frac{m_2}{2}\int_{\mathcal{M}}\nabla(\phi q^2)\cdot\nabla f - 2\phi q\nabla q\cdot\nabla f\,d\mathrm{vol} + \mathcal{O}(\epsilon)\\ &= \frac{m_2}{2}\int_{\mathcal{M}} q^2\nabla\phi\cdot\nabla f + \phi\nabla q^2\cdot\nabla f - 2\phi q\nabla q\cdot\nabla f\,d\mathrm{vol} + \mathcal{O}(\epsilon)\\ &= \frac{m_2}{2}\int_{\mathcal{M}}(\nabla\phi\cdot\nabla f)\,q^2 d\mathrm{vol} + \mathcal{O}(\epsilon).\end{aligned} \tag{A.5}$$

where the first equality follows from $m_2 = m_0/2$ and the second since $\nabla q^2 = 2q\nabla q$. Finally we show that,

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

for any natural number ℓ and for all kernel functions that have exponential decay. Notice that we typically consider $m_1^{\partial}(x)$ depending on the base point x on the manifold, however, the constant only depends on the distance to the boundary as seen in the definition (2.2) so here we consider $m_1^{\partial}(\cdot)$ as a function only of the distance to the boundary. We first recall that due to the exponential decay of the kernel we can localize the integrals to a cube where each coordinate is in $[-\epsilon^{\gamma}, \epsilon^{\gamma}]$ for any $\gamma < 1$. Thus, we find that

$$\int_0^{\infty} m_1^{\partial}(\epsilon b_x)\,db_x = \int_0^{\epsilon^{\gamma}}\int_{\mathbb{R}^{m-1}}\int_{-\epsilon^{\gamma}}^{b_x} z_{\parallel}K\left(\sqrt{|z_{\perp}|^2 + z_{\parallel}^2}\right)\,dz_{\parallel}dz_{\perp}db_x$$

We can now move the integral with respect to x inside the other integrals by noticing that the domain of integration for the variables $(x, z_{\parallel})$ is the region

$$[0,\epsilon^{\gamma}]\times[-\epsilon^{\gamma},0]\cup\{(x,z_{\parallel})\,:\,x\in[0,\epsilon^{\gamma}], 0 < z_{\parallel} < b_x\}$$

which is a rectangle below the x-axis and a triangle above the x-axis. Since the integrand is odd in $z_\parallel$, the triangle cancels out the reflected triangle below the x-axis, which is half of the rectangle. The remaining domain of integration is a triangle below the x-axis given by $\{(x, z_\parallel) : x \in [0, \epsilon^\gamma], -\epsilon^\gamma < z_\parallel < -b_x\}$ which we can rewrite as $\{(x, z_\parallel) : z \in [-\epsilon^\gamma, 0], 0 < b_x < z_\parallel\}$. By rewriting the domain of integration we have exchanged the integrals so that

$$\begin{aligned}
\int_0^\infty m_1^\partial(\epsilon b_x)\, db_x &= \int_{\mathbb{R}^{m-1}} \int_{-\epsilon^\gamma}^0 \int_0^{z_\parallel} z_\parallel K\left(\sqrt{|z_\perp|^2 + z_\parallel^2}\right) db_x dz_\parallel dz_\perp \\
&= \int_{\mathbb{R}^{m-1}} \int_{-\epsilon^\gamma}^0 z_\parallel^2 K\left(\sqrt{|z_\perp|^2 + z_\parallel^2}\right) dz_\parallel dz_\perp \\
&= \frac{1}{2} \int_{\mathbb{R}^{m-1}} \int_{-\epsilon^\gamma}^{\epsilon^\gamma} z_\parallel^2 K\left(\sqrt{|z_\perp|^2 + z_\parallel^2}\right) dz_\parallel dz_\perp \\
&= \frac{1}{2} \int_{\mathbb{R}^m} (z \cdot \eta_x)^2 K(|z|)\, dz + \mathcal{O}(\epsilon^\ell) = \frac{m_2}{2} + \mathcal{O}(\epsilon^\ell)
\end{aligned} \tag{A.6}$$

where the second to last equality follows from the exponential decay of the kernel and the last equality follows by the radial symmetry of the kernel. □

Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030, USA.
E-mail address: tberry@gmu.edu

Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030, USA.
E-mail address: tberry@gmu.edu

Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030, USA.
E-mail address: hantil@gmu.edu