SPECTRAL CONVERGENCE OF THE CONNECTION LAPLACIAN FROM RANDOM SAMPLES

AMIT SINGER AND HAU-TIENG WU

ABSTRACT. Spectral methods that are based on eigenvectors and eigenvalues of discrete graph Laplacians, such as Diffusion Maps and Laplacian Eigenmaps are extremely useful for manifold learning. It was previously shown by Belkin and Niyogi [4] that the eigenvectors and eigenvalues of the graph Laplacian converge to the eigenfunctions and eigenvalues of the Laplace-Beltrami operator of the manifold in the limit of infinitely many uniformly sampled data points. Recently, we introduced Vector Diffusion Maps and showed that the Connection Laplacian of the tangent bundle of the manifold can be approximated from random samples. In this paper, we present a unified framework for approximating other Connection Laplacians over the manifold by considering its principle bundle structure. We prove that the eigenvectors and eigenvalues of these Laplacians converge in the limit of infinitely many random samples. Our results for spectral convergence also hold in the case where the data points are sampled from a non-uniform distribution, and for manifolds with and without boundary.

1. INTRODUCTION

A recurring problem in fields such as neuroscience, computer graphics and image processing is that of classifying a set of 3-dim objects by pairwise comparisons. For example, the objects can be 3-dim brain functional magnetic resonance imaging (fMRI) images [15] that correspond to similar functional activity. In order to separate the actual sources of variability among the images from the nuisance parameters that correspond to different conditions of the acquisition process, the images are initially registered and aligned. Similarly, the shape space analysis problem in computer graphics [21] involves the organization of a collection of shapes. Also in this problem it is desired to factor out nuisance shape deformations, such as rigid transformations.

Once the nuisance parameters have been factored out, methods such as Diffusion Maps (DM) [10] or Laplacian Eigenmaps (LE) [2] can be used for non-linear dimensionality reduction, classification and clustering. In [26] we introduced Vector Diffusion Maps (VDM) as an algorithmic framework for organization of such data sets that also takes the nuisance parameters into account, using the eigenvectors and eigenvalues of the graph Connection Laplacian that encodes both data affinities and pairwise nuisance transformations. In [26] we also proved pointwise convergence of the graph Connection Laplacian to the Connection Laplacian of the tangent bundle of the data manifold in the limit of infinitely many sample points. The main contribution of the current paper is the spectral convergence of the graph Connection Laplacian to the Connection Laplacian operator over the vector bundle of the data manifold. In passing, we also provide a spectral convergence result for Diffusion Maps to the Laplace-Beltrami operator in the case of non-uniform sampling and for manifolds with non-empty boundary, thus strengthening a previous result of Belkin and Nyiogi [4].

At the center of LE, DM, and VDM is a weighted undirected graph, whose vertices correspond to the data objects and the weights quantify the affinities between them. A commonly used metric is the Euclidean distance, and the affinity can then be described using a kernel function of the distance. For example, if the data set $\{x_1, x_2, \ldots, x_n\}$ consist of n functions in $L_2(\mathbb{R}^3)$ then

$$d_E(x_i, x_j) := \|x_i - x_j\|_{L^2(\mathbb{R}^3)}, \tag{1}$$

and the weights can be defined using the Gaussian kernel with width $\sqrt{h}$ as

$$w_{ij} = e^{-\frac{d_E^2(x_i, x_j)}{2h}} \tag{2}$$

However, the Euclidean distance is sensitive to the nuisance parameters. In order to factor out the nuisance parameters, it is required to use a metric which is invariant to the group of transformations associated with those parameters, denoted by G. Let $\mathscr{X}$ be the total space from which data is sampled. The group G acts on $\mathscr{X}$ and instead of measuring distances between elements of $\mathscr{X}$, we want to measure distances between their orbits. The orbit of a point $x \in \mathscr{X}$ is the set of elements of $\mathscr{X}$ to which x can be mapped by the elements of G, denoted by

$$Gx = \{g \circ x \,|\, g \in G\}$$

The group action induces an equivalence relation on $\mathscr{X}$ and the orbits are the equivalence classes, such that the equivalence class $[x]$ of $x \in \mathscr{X}$ is Gx. The invariant metric is a metric on the orbit space $\mathscr{X}/G$ of equivalent classes.

One possible way of constructing the invariant metric d_G is through optimal alignment, given as

$$d_G([x_i],[x_j]) = \inf_{g_i, g_j \in G} d_E(g_i \circ x_i, g_j \circ x_j). \tag{3}$$

If the action of the group is an isometry, then

$$d_G([x_i],[x_j]) = \inf_{g \in G} d_E(x_i, g \circ x_j). \tag{4}$$

For example, if $\mathscr{X} = L^2(\mathbb{R}^3)$ and G is $O(3)$ (the group of 3×3 orthogonal matrices), then

$$(g \circ f)(x) = f(g^{-1}x) \tag{5}$$

and

$$d_G^2([f_i],[f_j]) = \min_{g \in O(3)} \int_{\mathbb{R}^3} |f_i(x) - f_j(g^{-1}x)|^2 \,\mathrm{d}x. \tag{6}$$

In this paper we only consider groups that are either orthogonal and unitary, for three reasons. First, this condition guarantees that the graph Connection Laplacian is symmetric (or Hermitian). Second, the action is an isometry and the invariant metric (4) is well defined. Third, it is a compact group and the minimizer of (6) is well defined.

The invariant metric d_G can be used to define weights between data samples, for example, the Gaussian kernel gives

$$w_{ij} = e^{-\frac{d_G^2([x_i],[x_j])}{2h}} \tag{7}$$

While LE and DM with weights given in (2) correspond to diffusion over the original space $\mathscr{X}$, LE and DM with weights given in (7) correspond to diffusion over the orbit space $\mathscr{X}/G$. In VDM, the weights (7) are also accompanied by the optimal transformations

$$g_{ij} = \operatorname*{argmin}_{g \in G} d_E(x_i, g \circ x_j). \tag{8}$$

VDM corresponds to diffusion over the vector bundle of the orbit space $\mathscr{X}/G$ associated with the group action. The following existing examples demonstrate the usefulness of such a diffusion process in data analysis:

- *Manifold learning:* Suppose we are given a point cloud randomly sampled from a d-dim smooth manifold M embedded in $\mathbb{R}^p$. Due to the smoothness of M, the embedded tangent bundle of M can be estimated by local principal component analysis [26]. All bases of an embedded tangent plane at x form a group isomorphic to $O(d)$. Since the bases of the embedded tangent planes form the frame bundle $O(\mathrm{M})$, from this point cloud we obtain a set of samples from the frame bundle which form the total space $\mathscr{X} = O(\mathrm{M})$. Since the set of all the bases of an embedded tangent plane is invariant under the action of $O(d)$, for the purpose of learning the manifold M, we take $O(d)$ as the nuisance group, and hence the orbit space is $\mathrm{M} = O(\mathrm{M})/O(d)$. As shown in [26], the generator of the diffusion process corresponding to VDM is the Connection Laplacian associated with the tangent bundle. With the eigenvalues and eigenvectors of the Connection Laplacian, we are able to embed the point cloud in an Euclidean space. We refer to the Euclidean distance in the embedded space as the *vector diffusion distance (VDD),* which provides a new metric for the point cloud. It is shown in [26] that VDD approximates the geodesic distance between nearby points on the manifold. Furthermore, by VDM, we extend the earlier spectral embedding theorem [5] by constructing a distance in a class of closed Riemannian manifolds with prescribed geometric conditions, which leads to a pre-compactness theorem on the class under consideration.

- *Orientability:* Suppose we are given a point cloud randomly sampled from a d-dim smooth manifold M and we want to learn its orientability. Since the frame bundle encodes whether or not the manifold is orientable, we take the nuisance group as $\mathbb{Z}_2$ defined as the determinant of the action $O(d)$ from the previous example. In other words, the orbit of each point on the manifold is $\mathbb{Z}_2$, the total space $\mathscr{X}$ is the $\mathbb{Z}_2$ bundle on M, and the orbit space is M. With the nuisance group $\mathbb{Z}_2$, VDM is equivalent to *Orientable Diffusion Map* (ODM) considered in [25] in order to estimate the orientability of M from a finite collection of random samples.
- *Cryo-EM:* The X-ray transform often serves as a mathematical model to many medical and biological imaging modalities, for example, in cryo-electron microscopy [12]. In cryo-electron microscopy, the 2-dim projection images of the 3-dim object are noisy and their projection directions are unknown. For the purpose of denoising, it is required to classify the images and average images with similar projection directions in order to increase the signal to noise ratio, a procedure known as class averaging. When the object of interest has no symmetry, the projection images have a one-to-one correspondence with a manifold diffeomorphic to $SO(3)$. Notice that $SO(3)$ can be viewed as the set of all right-handed bases of all tangent planes of S^2, and the set of all right-handed bases of a tangent plane is isomorphic to $SO(2)$. Since the projection directions are parameterized by S^2 and the set of images with the same projection direction is invariant under the $SO(2)$ action, we learn the projection direction by taking $SO(2)$ as the nuisance group and S^2 as the orbit space. The vector diffusion distance provides a metric for classification of the projection directions in S^2, and this metric has been shown to perform better compared to other classification methods [27, 23]

The main contribution of this paper is twofold. First, we use the mathematical framework of the *principal bundle* [7] in order to analyze the relationship between the nuisance group and the orbit space and how their combination can be used to learn the dataset. In this setup, the total space is the principal bundle, the orbit space is the base manifold, and the orbit is the fiber. This principal bundle framework unifies LE, DM, ODM, and VDM by providing a common mathematical language to all of them. Second, in addition to showing pointwise convergence of VDM in the general principal bundle setup, in Theorem 5.2 we prove that the algorithm converges in the spectral sense, that is, the eigenvalues and the eigenvectors computed by the algorithm converge to the eigenvalues and the eigen-vector-fields of the Connection Laplacian of the associated vector bundle, even when the sampling is nonuniform and the boundary is nonempty. We also show the spectral convergence of the VDM to the Connection Laplacian of the associated tangent bundle in Theorem 6.1 when the tangent bundle is estimated from the point cloud. The importance of these spectral convergence results stem from the fact that they provide a theoretical guarantee in the limit of infinite number of data samples for the above listed problems, namely, estimating vector diffusion distances, determining the orientability of a manifold from a point cloud, and classifying the projection directions of cryo-EM images.

The rest of the paper is organized as follows. In Section 2, we introduce background material and set up the notations. In Section 3, we review the VDM algorithm and clarify the relationship between the point cloud sampled from the manifold and the bundle structure of the manifold. In Section 4, we unify LE, DM, VDM, and ODM by taking the principal bundle structure of a manifold into account, and prove the first spectral convergence result that assumes knowledge of the bundle structure. The non-empty boundary and nonuniform sampling effects are simultaneously handled. In Section 6, we prove the second spectral convergence result when the bundle information is missing and needs to be estimated directly from the finite random point cloud.

2. Notations, Background and Assumptions

2.1. **Notations and Background of Differential Geometry.** Denote M to be a d-dim compact smooth manifold. If the boundary is non-empty, the boundary is smooth. Denote $\iota : \mathrm{M} \hookrightarrow \mathbb{R}^p$ to be a smooth embedding of M into $\mathbb{R}^p$ and we equip M with the metric g induced from the canonical metric on $\mathbb{R}^p$ via ι. With the metric g we have an induced measure, denoted as $\mathrm{d}V$, on M. Denote

$$\mathrm{M}_t = \{x \in \mathrm{M} : \min_{y \in \partial \mathrm{M}} d(x,y) \leq t\},$$

where $d(x,y)$ is the geodesic distance between x and y.

We refer the readers to Appendix A.1 for an introduction of the principal bundle. Denote $P(\mathrm{M},G)$ to be the principal bundle with a connection 1-form ω, where G is a Lie group right acting on $P(\mathrm{M},G)$ by $\circ$. Denote $\pi : P(\mathrm{M},G) \to \mathrm{M}$ to be the canonical projection. We call M the base space of $P(\mathrm{M},G)$ and G the structure group. From the view point of orbit space, $P(\mathrm{M},G)$ is the total space $\mathscr{X}$, G is the group acting on $P(\mathrm{M},G)$, and M is the orbit space of $P(\mathrm{M},G)$ under the action of G. In other words, G is the nuisance group, and the orbits are parametrize by the manifold M.

Denote ρ to be a representation of G into $O(m')$, where $m' > 0$. When there is no danger of confusion, we use the same symbol g to denote the Riemannian metric on M and an element of G. Denote $\mathscr{E}(P(\mathrm{M},G),\rho,\mathbb{R}^{m'})$, $m' \geq 1$, to be the associated vector bundle with the fiber diffeomorphic to $\mathbb{R}^{m'}$. We use $\mathscr{E}$ to simplify the notation. Given a fiber metric $g^{\mathscr{E}}$ in $\mathscr{E}$, which always exists since M is compact, the connection we consider is metric under which the parallel displacement of fiber of $\mathscr{E}$ is isometric with related to $g^{\mathscr{E}}$. The metric connection on $\mathscr{E}$ determined from ω is denoted as $\nabla^{\mathscr{E}}$. Note that by definition, each $u \in P(\mathrm{M},G)$ is a linear mapping from $\mathbb{R}^{m'}$ to E_x preserving the inner product structure, where $x = \pi(u)$, and satisfies

$$(g \circ u)v = u(\rho(g)v) \in E_x,$$

where $u \in P(\mathrm{M},G)$, $g \in G$ and $v \in \mathbb{R}^{m'}$. We interpret the linear mapping u as finding the point $u(v) \in E_x$ possessing the coordinate $v \in \mathbb{R}^{m'}$. Intuitively, u is a "basis" of E_x. Note that $\mathscr{E}$ is the quotient space $P(\mathrm{M},G) \times \mathbb{R}^{m'} / \sim$, where the equivalence relationship $\sim$ is defined by the group action on $P(\mathrm{M},G) \times \mathbb{R}^{m'}$, that is, $g : (u,v) \to (g \circ u, \rho(g)^{-1}v)$, where $g \in G$, $u \in P(\mathrm{M},G)$ and $v \in \mathbb{R}^{m'}$. Denote E_x to be the fiber of $\mathscr{E}$ on $x \in \mathrm{M}$.[1] An important example is the frame bundle of the Riemannian manifold (M,g), denoted as $O(\mathrm{M}) = P(\mathrm{M},O(d))$, and the tangent bundle $T\mathrm{M}$, which is the associated vector bundle of the frame bundle $O(\mathrm{M})$ if we take $\rho = id$ and $m' = d$.

Denote $\Gamma(\mathscr{E})$ to be the set of sections, $C(\mathscr{E})$ to be the set of continuous sections, and $C^k(\mathscr{E})$ to be the set of k-th differentiable sections, where $k \geq 0$. With the above structure, we can rewrite a section $X \in \Gamma(\mathscr{E})$ as a $\mathbb{R}^{m'}$-valued function f_X defined on $P(\mathrm{M},G)$ by

$$f_X(u_x) := u_x^{-1}(X(x)),$$

where $u_x \in P(\mathrm{M},G)$ and $\pi(u_x) = x$. The covariant derivative of $X \in C^1(\mathscr{E})$ in the direction Y is defined as

$$\nabla^{\mathscr{E}}_{\dot{c}(0)}X = \lim_{h\to 0}\frac{1}{h}[\tau_0\tau_h^{-1}(X(c_h)) - X(c_0)], \tag{9}$$

where $c : [0,1] \to \mathrm{M}$ is the curve on M so that $c_0 = x$, $Y = \dot{c}(0)$ and τ_h is the horizontal lift of c_h to $P(\mathrm{M},G)$ so that $\pi(\tau_0) = x$. Let $/\!/_y^x$ denote the parallel displacement from y to x. When y is in the cut locus of x, we set $/\!/_y^x = 0$. We know $/\!/_{c_h}^{c_0} = \tau_0\tau_h^{-1}$ when h is small enough. We can better understand this definition in the frame bundle $O(\mathrm{M})$ and its associated tangent bundle. Take $X \in C^1(T\mathrm{M})$. First, we find the coordinate of $X(c_h)$ by $\tau_h^{-1}(X(c_h))$, and then we put this coordinate to x and map it back to the fiber $T_x\mathrm{M}$ by τ_0. In this way we can compare two different "abstract fibers" by comparing their coordinates.

Denote $L^p(\mathscr{E})$, $1 \leq p < \infty$ to be the set of L^p integrable sections, that is, $X \in L^p(\mathscr{E})$ iff $\int |g^{\mathscr{E}}(X,X)|^{p/2}\mathrm{d}V < \infty$. Denote ∇^2 the Connection Laplacian over M with respect to $\mathscr{E}$. Denote by $\mathscr{R}$, Ric, and s the Riemanian curvature tensor, the Ricci curvature, and the scalar curvature of (M,g), respectively. The second fundamental form of the embedding ι is denoted by II. Denote τ to be the largest positive number having the property: the open normal bundle about M of radius r is embedded in $\mathbb{R}^p$ for every $r < \tau$ [20]. $\tau > 0$ since M is compact. In all theorems, we assume that $\sqrt{h} < \tau$.

2.2. **Notations and Background of Numerical Finite Samples.** To sample points from the manifold, we need a notion of probability density function (p.d.f.) [9]. Let the random vector $Y : (\Omega,\mathscr{F},\mathrm{d}P) \to \mathbb{R}^p$ be a measurable function defined on the probability space Ω. Suppose the range of X is supported on $\iota(\mathrm{M})$ and let $\tilde{\mathscr{B}}$ be the Borel sigma algebra of $\iota(\mathrm{M})$. Denote by $\mathrm{d}\tilde{P}_Y$ the probability measure of Y, defined on $\tilde{\mathscr{B}}$, induced from the probability measure $\mathrm{d}P$. Assume that $\mathrm{d}\tilde{P}_Y$ is absolutely continuous with respect to the volume measure on $\iota(\mathrm{M})$, that is, $\mathrm{d}\tilde{P}_Y(x) = p(\iota^{-1}(x))\iota_*\mathrm{d}V(x)$, where from now on we assume $p \in C^3(\mathrm{M})$. We call $p(x)$ the p.d.f. of the random vector Y whose range is supported

[1] We may also consider representing G into $U(m')$ if we take the fiber to be $\mathbb{C}^{m'}$. However, to simplify the discussion, we focus ourselves on $O(m')$ and the real vector space.

on M. Notice the difference between this definition and the traditional p.d.f. definition. With this definition, we have that for an integrable function $f:\iota(\mathrm{M})\to\mathbb{R}$:

$$\mathbb{E}f(Y)=\int_\Omega f(Y(\omega))\mathrm{d}P(\omega)=\int_{\iota(\mathrm{M})} f(x)\mathrm{d}\tilde{P}_Y(x)=\int_{\iota(\mathrm{M})} f(x)p(\iota^{-1}(x))\iota_*\mathrm{d}V(x)=\int_{\mathrm{M}} f(\iota(x))p(x)\mathrm{d}V(\iota(x)),$$

where the second equality follows from the fact that $\mathrm{d}\tilde{P}_Y$ is the induced probability measure, and the last one comes from the change of variable. When p is constant, we say the sampling is uniform; otherwise non-uniform. It is possible to discuss more general setups, but we restrict ourselves here to this definition to simplify the discussion. To simplify the notation, in the future we will not distinguish between x and $\iota(x)$ and M and $\iota(\mathrm{M})$ when there is not ambiguity.

Suppose the data points $\mathscr{X}:=\{x_1,x_2,\ldots,x_n\}\subset\mathbb{R}^p$ are identically and independently sampled from X. For each x_i we randomly pick $u_i\in P(\mathrm{M},G)$ so that $\pi(u_i)=x_i$. To simplify the notation, we denote $u_i:=u_{x_i}$ when $x_i\in\mathscr{X}$ and $/\!/^i_j:=/\!/^{x_i}_{x_j}$ when $x_i,x_j\in\mathscr{X}$. Denote nm' dimensional Euclidean vector spaces $V_{\mathscr{X}}:=\oplus_{x_i\in\mathscr{X}}\mathbb{R}^{m'}$ and $E_{\mathscr{X}}:=\oplus_{x_i\in\mathscr{X}}E_{x_i}$, which represents the discretized vector bundle. Note that $V_{\mathscr{X}}$ is isomorphic to $E_{\mathscr{X}}$ since E_{x_i} is isomorphic to $\mathbb{R}^{m'}$. Given a $\boldsymbol{v}\in V_{\mathscr{X}}=\mathbb{R}^{m'n}$, we denote $\boldsymbol{v}[l]\in\mathbb{R}^{m'}$ to be the l-th component in the direct sum by saying $\boldsymbol{v}[l]=[\boldsymbol{v}((l-1)m+1),\ldots,\boldsymbol{v}(lm)]^T\in\mathbb{R}^{m'}$ for all $l=1,\ldots,n$. Given a $\boldsymbol{w}\in E_{\mathscr{X}}$, we denote $\boldsymbol{w}=[\boldsymbol{w}[1],\ldots,\boldsymbol{w}[n]]$ and $\boldsymbol{w}[l]\in E_{x_l}$ to be the l-th component in the direct sum for all $l=1,\ldots,n$. Define $\boldsymbol{I}_m$ to be the $m\times m$ identity matrix.

We need a parametrization to realize the isomorphism between $V_{\mathscr{X}}$ and $E_{\mathscr{X}}$. Define operators $B_{\mathscr{X}}:V_{\mathscr{X}}\to E_{\mathscr{X}}$ and $B^T_{\mathscr{X}}:E_{\mathscr{X}}\to V_{\mathscr{X}}$ by

$$\begin{aligned} B_{\mathscr{X}}\boldsymbol{v}&:=[u_1\boldsymbol{v}[1],\ldots u_n\boldsymbol{v}[n]]\in E_{\mathscr{X}},\\ B^T_{\mathscr{X}}\boldsymbol{w}&:=[(u_1^{-1}\boldsymbol{w}[1])^T,\ldots,(u_n^{-1}\boldsymbol{w}[n])^T]^T\in V_{\mathscr{X}}, \end{aligned}\tag{10}$$

where $\boldsymbol{w}\in E_{\mathscr{X}}$ and $\boldsymbol{v}\in V_{\mathscr{X}}$. Note that $B^T_{\mathscr{X}}B_{\mathscr{X}}\boldsymbol{v}=\boldsymbol{v}$ for all $\boldsymbol{v}\in V_{\mathscr{X}}$. And we define $\delta_{\mathscr{X}}:X\in C(\mathscr{E})\to E_{\mathscr{X}}$ by

$$\delta_{\mathscr{X}}X:=[X(x_1),\ldots X(x_n)]\in E_{\mathscr{X}}.$$

Here $\delta_{\mathscr{X}}$ is interpreted as the operator finitely sampling the section X and $B_{\mathscr{X}}$ the discretization of the action of a section from $\mathrm{M}\to P(\mathrm{M},G)$ on $\mathbb{R}^{m'}$. Note that $B_{\mathscr{X}}$ is aimed to recover the point on E_{x_i} from the coordinate $\boldsymbol{v}[i]$ with related to u_i, and the operator $B^T_{\mathscr{X}}$ is aimed to find the coordinates of $\boldsymbol{w}[i]$ associated with u_i. We can thus define

$$\boldsymbol{X}:=B^T_{\mathscr{X}}\delta_{\mathscr{X}}X\in V_{\mathscr{X}},\tag{11}$$

which is the coordinate of the discretized section X.

The following definitions are adapted from [29] to our setting in order to finish the spectral convergence proof. We refer the readers to [29] for the more general definition.

Definition 2.1. Take a probability space $(\Omega,\mathscr{F},P)$. For a pair of measurable functions $l:\Omega\to\mathbb{R}$ and $u:\Omega\to\mathbb{R}$, a bracket $[l,u]$ is the set of all measurable functions $f:\Omega\to\mathbb{R}$ with $l\le f\le u$. An ε-bracket in $L_1(P)$, where $\varepsilon>0$, is a bracket $[l,u]$ with $\int|u(y)-l(y)|\mathrm{d}P(y)\le\varepsilon$. Given a class of measurable function $\mathfrak{F}$. The bracketing number $N_{[]}(\varepsilon,\mathfrak{F},L_1(P))$ is the minimum number of ε-brackets needed to cover $\mathfrak{F}$.

We follow the standard notation defined in [29] to simplify the proof. Define the empirical measure as

$$\mathbb{P}_n:=\frac{1}{n}\sum_{i=1}^n\delta_{x_i}.$$

Thus, for a given measurable function $f:\mathrm{M}\to\mathbb{R}$ or a measurable vector-valued function $F:\mathrm{M}\to\mathbb{R}^m$ for $m>0$,

$$\mathbb{P}_nf:=\frac{1}{n}\sum_{i=1}^n f(x_i)\,,\ \mathbb{P}_nF:=\frac{1}{n}\sum_{i=1}^n F(x_i).$$

Then we define

$$\mathbb{P}f:=\int_{\mathrm{M}} f(x)p(x)\mathrm{d}V(x)\quad\text{and}\quad\mathbb{P}F:=\int_{\mathrm{M}} F(x)p(x)\mathrm{d}V(x).$$

Definition 2.2. Take a sequence of i.i.d. samples $\mathscr{X}:=\{x_1,\ldots,x_n\}\subset\mathrm{M}$ according to the p.d.f. p. We call a class $\mathfrak{F}$ of measurable functions a Glivenko-Cantelli class if

(1) $\mathbb{P}f$ exists for all $f \in \mathfrak{F}$
(2) $\sup_{f\in\mathfrak{F}} |\mathbb{P}_n f - \mathbb{P}f| \to 0$ almost surely when $n \to \infty$.

To simplify the proof in Section 5.2, we introduce the following notations. Fix $h > 0$. Given a non-negative continuous kernel function K decaying fast enough, we denote

$$K_h(x,y) := K\Big(\frac{\|x-y\|_{\mathbb{R}^p}}{\sqrt{h}}\Big) \in C(\mathrm{M}\times\mathrm{M}). \tag{12}$$

More general kernels can be considered, but we focus on the above form to simplify the discussion. Then we define the following continuous functions:

$$d_h(x) := \int K_h(x,y)p(y)\mathrm{d}V(y) \in C(\mathrm{M}), \tag{13}$$

$$M_h(x,y) := \frac{K_h(x,y)}{d_h(x)} \in C(\mathrm{M}\times\mathrm{M}), \tag{14}$$

where $d_h > 0$ approximates the probability density function defined on the manifold and M_h can be viewed as a new kernel. Note that M indicates the "Markovian" behavior of the operator. When we have only finite samples, we estimate (13) and (14) by

$$\widehat{d}_h(x) := \frac{1}{n-1}\sum_{k=1,x_k\neq x}^{n} K_h(x,x_k) \in C(\mathrm{M}),$$
$$\widehat{M}_{h,n}(x,y) := \frac{K_h(x,y)}{\widehat{d}_h(x)} \in C(\mathrm{M}\times\mathrm{M}),$$

where $x,y \in \mathrm{M}$. Notice that the above kernels are not normalized, so the non-uniformality of the p.d.f. $p(x)$ cannot be ignored. In case we do not like the influence of the non-uniformality of $p(x)$, normalizing the kernel is needed. Notice that in the above definition we use $\frac{1}{n-1}\sum_{k=1,x_k\neq x}^{n}$ instead of $\frac{1}{n}\sum_{k=1}^{n}$ in order to avoid dependence. When $n \to \infty$, these estimators are not biased by using $n-1$. We define the following function to estimate the p.d.f.:

$$p_h(x) := \int K_h(x,y)p(y)\mathrm{d}V(y) \in C(\mathrm{M}) \tag{15}$$

where $p_h(x)$ is an estimation of the p.d.f. at x, which is simply by the approximation of identify. Note that although the appearance of (13) and (15) are the same, the roles they play are different. In practice when we have only finite samples, we approximate (15) by $\widehat{p}_{h,n}(x)$:

$$\widehat{p}_{h,n}(x) := \frac{1}{n-1}\sum_{k=1,x_k\neq x}^{n} K_h(x,x_k) \in C(\mathrm{M}). \tag{16}$$

For $0 \leq \alpha \leq 1$, we define the following p.d.f. adjusted kernel:

$$K_{h,\alpha}(x,y) := \frac{K_h(x,y)}{p_h^\alpha(x)p_h^\alpha(y)} \in C(\mathrm{M}\times\mathrm{M})$$
$$d_{h,\alpha}(x) := \int K_{h,\alpha}(x,y)p(y)\mathrm{d}V(y) \in C(\mathrm{M})$$
$$M_{h,\alpha}(x,y) := \frac{K_{h,\alpha}(x,y)}{d_{h,\alpha}(x)} \in C(\mathrm{M}\times\mathrm{M}),$$

where $K_{h,\alpha}(x,y)$ is understood as a new kernel function at (x,y) adjusted by the estimated p.d.f. at x and y, that is, the kernel is normalized to reduce the influence of the non-uniform p.d.f.. In practice when we have only finite samples, we approximate the above terms by the following estimators.

$$\widehat{K}_{h,\alpha,n}(x,y) := \frac{K_h(x,y)}{\widehat{p}^{\alpha}_{h,n}(x)\widehat{p}^{\alpha}_{h,n}(y)} \in C(\mathrm{M}\times\mathrm{M}) \tag{17}$$

$$\widehat{d}_{h,\alpha,n}(x) := \frac{1}{n-1}\sum_{k=1,x_k\neq x}^{n} \widehat{K}_{h,\alpha,n}(x,x_k) \in C(\mathrm{M}) \tag{18}$$

$$\widehat{M}_{h,\alpha,n}(x,y) := \frac{\widehat{K}_{h,\alpha,n}(x,y)}{\widehat{d}_{h,\alpha,n}(x)} \in C(\mathrm{M}\times\mathrm{M}). \tag{19}$$

Note that $\widehat{d}_{h,\alpha,n}$ is always positive by the assumption of K. We mention that the notation $\mathbb{P}_n\widehat{K}_{h,\alpha,n}(x,\cdot)$ should be carefully understood since $\widehat{K}_{h,\alpha,n}(x,\cdot)$ is itself a random variable. When we work with the proof, we need the following "intermittent terms", which play the role in estimating the error bounds:

$$\widehat{d}^{(p_h)}_{h,\alpha,n}(x) := \frac{1}{n-1}\sum_{k=1,x_k\neq x}^{n} K_{h,\alpha}(x,x_k) \in C(\mathrm{M}), \qquad \widehat{M}^{(d_{h,\alpha})}_{h,\alpha,n}(x,y) := \frac{K_{h,\alpha}(x,y)}{\widehat{d}_{h,\alpha,n}(x)} \in C(\mathrm{M}\times\mathrm{M}).$$

3. Vector Diffusion Map Algorithm

Take an undirected affinity graph $\mathbb{G}=(\mathbb{V},E)$, where $\mathbb{V}=\{x_i\}_{i=1}^n$ and fix a $m'\in\mathbb{N}$. Suppose we have assigned the edge $(i,j)\in E$ a scalar $w_{ij}>0$ and $g_{ij}\in O(m')$. We call w_{ij} the *affinity* between x_i and x_j and call g_{ij} the *translation group* between *vector status* of x_i and x_j. Construct the following $n\times n$ block matrix $\boldsymbol{S}_n$ with $m'\times m'$ entries:

$$\boldsymbol{S}_n(i,j) = \begin{cases} w_{ij}g_{ij} & (i,j)\in E, \\ 0_{d\times d} & (i,j)\notin E. \end{cases} \tag{20}$$

Notice that the square matrix $\boldsymbol{S}_n$ is symmetric since $w_{ij}=w_{ji}$ and $g_{ij}^T=g_{ji}$. Then define a $n\times n$ diagonal block matrix $\boldsymbol{D}_n$ with $m'\times m'$ entries, where the diagonal blocks are scalar multiples of the identity matrices given by

$$\boldsymbol{D}_n(i,i) = \sum_{j:(i,j)\in E} w_{ij}\boldsymbol{I}_{m'}. \tag{21}$$

The *graph Connection Laplacian* and the *normalized graph Connection Laplacian* are defined by [26, 1]

$$\boldsymbol{D}_n - \boldsymbol{S}_n, \quad \boldsymbol{I}_n - \boldsymbol{D}_n^{-1}\boldsymbol{S}_n$$

respectively. Note that the matrix $\boldsymbol{D}_n^{-1}\boldsymbol{S}_n$ is an operator acting on $\boldsymbol{v}\in\mathbb{R}^{nm'}$:

$$(\boldsymbol{D}_n^{-1}\boldsymbol{S}_n\boldsymbol{v})[i] = \frac{\sum_{j:(i,j)\in\mathbb{E}} w_{ij}g_{ij}v_j}{\sum_{j:(i,j)\in\mathbb{E}} w_{ij}}, \tag{22}$$

which suggests the interpretation of $\boldsymbol{D}_n^{-1}\boldsymbol{S}_n$ as a generalized Markov chain with the random walker (e.g., diffusive particle) characterized by several statuses. In other words, a particle at i is endowed with a vector status represented by a m'-dim vector, and at each time step it can move from i to j with probability $\frac{w_{ij}}{\sum_{j:(i,j)\in E} w_{ij}}$. If the translation group is trivial, these statuses are separately viewed as m' functions defined on $\mathbb{G}$. Notice that the graph Laplacian is a special case in the way that $m'=1$ and $g_{ij}=1$. However, if $m'>1$ and we have a non-trivial relationship between the vector statuses of particles on i and j characterized by the translation group, we may want to take it into account – when a particle moves, its status must change according to the translation group g_{ij}. Thus, if a particle with a vector status $v_i\in\mathbb{R}^{m'}$ moves along a path of length t from i to j containing vertices $i,j_1,\ldots,j_{t-1},j$ so that $(i,j_1)\in E$, $(j_{t-1},j)\in E$ and $(j_l,j_{l+1})\in E$ for $l=1,\ldots,t-2$, when the particle arrives j, its vector status is influenced by a series of rotational groups along the path from i to j and becomes $g_{j,j_{t-1}}\cdots g_{j_2,j_1}g_{j_1,i}v_i$. In case there are more than two paths from i to j, we may get cancelation while adding transformations of different paths. Intuitively, "the closer two points are" or "the less variance of the translational group on the paths is", the more consistent the vector statuses are between i and j. We can thus define a new affinity between i and j by the consistency between the vector statuses. Notice that the matrix $(\boldsymbol{D}_n^{-1}\boldsymbol{S}_n)^{2t}(i,j)$, where $t>0$, contains the average of the rotational information on all paths of length $2t$ from i to j. Thus, the Hilbert-Schmidt norm of $(\boldsymbol{D}_n^{-1}\boldsymbol{S}_n)^{2t}(i,j)$, $\|(\boldsymbol{D}_n^{-1}\boldsymbol{S}_n)^{2t}(i,j)\|_{HS}^2$, can be viewed as a measure of not only the number of paths of length $2t$ from i to j but also the amount of consistency of the vector status. Therefore, we define a new affinity between i and j by considering $\|(\boldsymbol{D}_n^{-1}\boldsymbol{S}_n)^{2t}(i,j)\|_{HS}^2$.

To understand this affinity, we consider the symmetric matrix $\widetilde{\boldsymbol{S}}_n = \boldsymbol{D}_n^{-1/2}\boldsymbol{S}_n\boldsymbol{D}_n^{-1/2}$ which is similar to $\boldsymbol{D}_n^{-1}\boldsymbol{S}_n$. Since $\widetilde{\boldsymbol{S}}_n$ is symmetric, it has a complete set of eigenvectors $v_1, v_2, \ldots, v_{nm'}$ and eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{nm'}$, where the eigenvalues are ordered by $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{nm'}$. A direct calculation of the HS norm of $\widetilde{\boldsymbol{S}}_n^{2t}(i,j)$ leads to:

$$\|\widetilde{\boldsymbol{S}}_n^{2t}(i,j)\|_{HS}^2 = \sum_{l,r=1}^{nm'} (\lambda_l\lambda_r)^{2t}\langle v_l(i), v_r(i)\rangle\langle v_l(j), v_r(j)\rangle. \tag{23}$$

The right hand side of $\|\widetilde{\boldsymbol{S}}_n^{2t}(i,j)\|_{HS}^2$ leads us to define the *vector diffusion mapping* (VDM) $V_t : \mathbb{G} \to \mathbb{R}^{(nm')^2}$ by

$$V_t : i \mapsto \left((\lambda_l\lambda_r)^t\langle v_l(i), v_r(i)\rangle\right)_{l,r=1}^{nm'}. \tag{24}$$

With this map, $\|\widetilde{\boldsymbol{S}}_n^{2t}(i,j)\|_{HS}^2$ becomes an inner product for the finite dimensional Hilbert space, that is,

$$\|\widetilde{\boldsymbol{S}}_n^{2t}(i,j)\|_{HS}^2 = \langle V_t(i), V_t(j)\rangle.$$

We thus define the new affinity between nodes i and j, referred to as *vector diffusion distance* (VDD), by

$$d_{\mathrm{VDD},t}(i,j) := \|V_t(i) - V_t(j)\|^2. \tag{25}$$

Furthermore, by a direct calculation we know that $|\lambda_l| \leq 1$ due to the following identity:

$$\boldsymbol{v}^T(\boldsymbol{I}_n \pm \widetilde{\boldsymbol{S}}_n)\boldsymbol{v} = \sum_{(i,j)\in E}\left\|\frac{\boldsymbol{v}[i]}{\sqrt{\sum_{l:(i,l)\in E} w_{il}}} \pm \frac{w_{ij}g_{ij}\boldsymbol{v}[j]}{\sqrt{\sum_{l:(j,l)\in E} w_{jl}}}\right\|^2 \geq 0,$$

for any $\boldsymbol{v} \in \mathbb{R}^{nm'}$. Note that we cannot guarantee that eigenvalues of $\widetilde{\boldsymbol{S}}_n$ are non-negative, and that is the main reason we define V_t through $\|\widetilde{\boldsymbol{S}}_n^{2t}(i,j)\|_{HS}^2$ rather than $\|\widetilde{\boldsymbol{S}}_n^{t}(i,j)\|_{HS}^2$.

We now come back to $\boldsymbol{D}_n^{-1}\boldsymbol{S}_n$. The eigenvector of $\boldsymbol{D}_n^{-1}\boldsymbol{S}_n$ associated with eigenvalue λ_l is $w_l = D^{-1/2}v_l$. So we define another VDM $V_t' : \mathbb{G} \to \mathbb{R}^{(nm')^2}$ by

$$V_t' : i \mapsto \left((\lambda_l\lambda_r)^t\langle w_l(i), w_r(i)\rangle\right)_{l,r=1}^{nm'},$$

so that $V_t'(i) = \frac{V_t(i)}{\sum_{l:(i,l)\in E} w_{il}}$. In other words, V_t' maps the data set in a Hilbert space upon proper normalization by the vertex degrees. The associated VDD is thus defined as $\|V_t'(i) - V_t'(j)\|^2$. For further discussion about the algorithm and the normalization, we refer the readers to [26].

4. Unify VDM, ODM, LE and DM from the principal bundle viewpoint

We state some known results relevant to VDM, ODM, LE and DM. Most of the results that have been obtained are of two types: either they provide the topological information about the data which is global in nature, or they concern the geometric information which aims to recover the local information of the data. Fix the undirected affinity graph $\mathbb{G} = (\mathbb{V}, \mathbb{E})$. When it is built from a point cloud randomly sampled from a Riemannian manifold $\iota : \mathrm{M} \hookrightarrow \mathbb{R}^p$ with the induced metric g, the main ingredient of LE and DM is the Laplace-Beltrami operator Δ_g of (M, g) [10]. It is well known that the Laplace-Beltrami operator Δ_g provides the topology and geometry information about M [13]. For example, the spectral embedding of M into the Hilbert space [5] preserves the geometric information of M and the dimension of the null space of Δ_g is the 0-Betti number of M. In the principal bundle setup, Δ_g is associated with the trivial line bundle. If we consider a different bundle, we obtain a different Laplacian operator, which provides different geometric/topological information [13]. For example, the core of VDM considered in [26] is the Connection Laplacian associated with the tangent bundle $T\mathrm{M}$, which provides not only the geodesic distance among nearby points (local information) but also the 1-Betti number mixed with the Ricci curvature of the manifold. In addition, the notion of synchronization of vector fields on $\mathbb{G}$ accompanied with translation group information can be analyzed by the graph Connection Laplacian [1].

4.1. **Principal Bundle Setup.** As the reader may have noticed, the appearance of VDM is almost the same as that of LE, DM and ODM discussed in [2, 3, 10, 14, 25]. This is not a coincidence if we take the notion of principal bundle and its connection into account, based on which we unify VDM, ODM, LE and DM in this section.

We make the following assumptions about the principal bundle setup.

Assumption 4.1. (A1) The base manifold M is d-dim, smooth and smoothly embedded in $\mathbb{R}^p$ via ι with induced metric g from the canonical metric of $\mathbb{R}^p$. If $\partial \mathrm{M} \neq \emptyset$, we assume that the boundary is smooth.
(A2) Fix a principal bundle $P(\mathrm{M},G)$ with a connection 1-form ω. Denote ρ to be the representation of G into $O(m')$ where $m'>0$ depending on the application.[2] Denote $\mathscr{E}:=\mathscr{E}(P(\mathrm{M},G),\rho,\mathbb{R}^{m'})$ to be the associated vector bundle with a fiber metric $g^{\mathscr{E}}$ and the metric connection $\nabla^{\mathscr{E}}$.[3]
(A3) The probability density function $p\in C^3(\mathrm{M})$ is uniformly bounded from below and above, that is, $0<p_m\leq p(x)\leq p_M<\infty$.

The following two special principal bundles and their associated vector bundles are directly related to ODM, LE and DM. The principal bundle for ODM is the non-trivial orientation bundle associated with the tangent bundle of a non-orientable manifold M, denoted as $P(\mathrm{M},\mathbb{Z}_2)$, where $\mathbb{Z}_2=\{-1,1\}$, and its associated vector bundle is $\mathscr{E}^{\mathrm{ODM}}=\mathscr{E}(P(\mathrm{M},\mathbb{Z}_2),\rho,\mathbb{R})$, where ρ is the representation of $\mathbb{Z}_2$ so that ρ satisfies $\rho(g)x=gx$ for all $g\in\mathbb{Z}_2$ and $x\in\mathbb{R}$. Note that $\mathbb{Z}_2\cong O(1)$. The principal bundle for LE and DM is $P(\mathrm{M},\{e\})$, where $\{e\}$ is the identify group, and its associated vector bundle is $\mathscr{E}^{\mathrm{DM}}=\mathscr{E}(P(\mathrm{M},\{e\}),\rho,\mathbb{R})$, where the representation ρ satisfies $\rho(e)x=x$ and $x\in\mathbb{R}$. In other words, it is the trivial bundle on M. Note that $\{e\}\cong SO(1)$.

Under the manifold setup assumption, the sampled data satisfy

Assumption 4.2. (B1) We sample a data cloud $\mathscr{X}=\{x_i\}_{i=1}^n$ from M according to the p.d.f. $p(x)$ under Assumption 4.1(A3).
(B2) For each $x_i\in\mathscr{X}$, sample $u_i\in P(\mathrm{M},G)$ according to a uniform distribution over G so that $\pi(g_i)=x_i$. Denote $\mathscr{G}=\{u_i:\mathbb{R}^{m'}\to E_{x_i}\}_{i=1}^n$.

The kernel and bandwidth used in the following sections satisfy:

Assumption 4.3. (K1) the kernel function $K\in C^2(\mathbb{R}^+)$ is a positive function so that $\inf_{x\in\mathbb{R}^+}K(x)=\delta>0$. Furthermore, $\mu_l^{(k)}:=\int_{\mathbb{R}^d}\|x\|^lK^{(k)}(\|x\|)\mathrm{d}x$, where $k=0,1,2$, $l\in\mathbb{N}\cup\{0\}$, and $K^{(k)}$ means the k-th order derivative of K. We assume $\mu_0^{(0)}=1$.
(K2) We assume that $0<\sqrt{h}<\min\{\tau,\mathrm{inj}(\mathrm{M})\}$, where $\mathrm{inj}(\mathrm{M})$ is the injectivity radius of M.

4.2. **Unify VDM, ODM, LE and DM.** Suppose Assumption 4.1 is satisfied and we are given $\mathscr{X}$ and $\mathscr{G}$ satisfying Assumption 4.2. The affinity graph $\mathbb{G}=(\mathbb{V},E)$ is constructed by the following way. Take $\mathbb{V}=\mathscr{X}$ and $E=\{(x_i,x_j)|x_i,x_j\in\mathscr{X},\|x_i-x_j\|_{\mathbb{R}^p}<\sqrt{h}\}$. Under this construction $\mathbb{G}$ is undirected by nature. The affinity between x_i and x_j is defined by $w_{ij}:=\widehat{K}_{h,\alpha,n}(x_i,x_j)$, where K is the kernel function satisfying Assumption 4.3. The translation group g_{ij} between x_i and x_j is constructed from $\mathscr{G}$ by

$$g_{ij}=u_i^{-1}/\!\!/_j^iu_j. \tag{26}$$

With the affinity graphs, the weight w_{ij} and the translation group g_{ij}, define the following $n\times n$ block matrix $\boldsymbol{P}_{h,\alpha,n}$ with $m'\times m'$ entries:

$$\boldsymbol{P}_{h,\alpha,n}(i,j)=\begin{cases} w_{ij}g_{ij} & (i,j)\in E,\\ 0_{m'\times m'} & (i,j)\notin E.\end{cases} \tag{27}$$

[2] We restrict ourselves to the orthogonal representation in order to obtain a symmetric matrix in the VDM algorithm. Indeed, if the translation of the vector status from x_i to x_j satisfies $u_j^{-1}u_i$, where $u_i,u_j\in P(\mathrm{M},G)$ and $\pi(u_i)=x_i$ and $\pi(u_j)=x_j$, the translation from x_j back to x_i should satisfy $u_i^{-1}u_j$, which is the inverse of $u_j^{-1}u_i$. To have a symmetric matrix in the end, we thus need $u_j^{-1}u_i=(u_i^{-1}u_j)^T$, which is satisfied only when G is represented into the orthogonal group. We refer the readers to Appendix A.1 for the reasoning based on the notion of connection.

[3] In general ρ can be the representation of G into $O(m')$ which acts on the tensor space $T_s^r(\mathbb{R}^{m'})$ of type (r,s) or others. But we consider $\mathbb{R}^{m'}=T_0^1(\mathbb{R}^{m'})$ to simplify the discussion.

Notice that the square matrix $\boldsymbol{P}_{h,\alpha,n}$ is symmetric since $w_{ij}=w_{ji}$ and $g_{ij}=g_{ji}^T$. Then define a $n\times n$ diagonal block matrix $\boldsymbol{D}_n$ with $m'\times m'$ entries, where the diagonal blocks are scalar multiples of the identity matrices given by

$$\boldsymbol{D}_{h,\alpha,n}(i,i)=\sum_{j:(i,j)\in E} w_{ij}\boldsymbol{I}_{m'}=\widehat{d}_{h,\alpha,n}(x)\boldsymbol{I}_{m'}. \tag{28}$$

Take $\boldsymbol{v}\in\mathbb{R}^{nm'}$. The matrix $\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}$ is thus an operator acting on $\boldsymbol{v}$ by

$$(\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}\boldsymbol{v})[i]=\frac{\sum_{j=1,j\neq i}^n \widehat{K}_{h,\alpha,n}(x_i,x_j)g_{ij}v_j}{\sum_{j=1,j\neq i}^n \widehat{K}_{h,\alpha,n}(x_i,x_j)}=\frac{1}{n-1}\sum_{j=1,j\neq i}^n \widehat{M}_{h,\alpha,n}(x_i,x_j)g_{ij}v_j. \tag{29}$$

With the eigenvalues and eigenvectors of $\boldsymbol{D}_{h,\alpha,n}^{-1/2}\boldsymbol{P}_{h,\alpha,n}\boldsymbol{D}_{h,\alpha,n}^{-1/2}$, the VDM and VDD are defined in the same way as that of (24) and (25). Recall that we interpret u_j as a linear mapping from of $\mathbb{R}^{m'}$ to E_{x_j} and view $\overline{X}_j:=u_j^{-1}X(x_j)\in\mathbb{R}^m$ as the coordinate of $X(x_j)\in E_{x_j}$ with related to the basis u_j; also recall the notation $\boldsymbol{X}:=B_{\mathscr{X}}\delta_{\mathscr{X}}X$ defined in (11). Then, consider the following quantity:

$$\begin{aligned}\Big(\frac{\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}-\boldsymbol{I}_n}{h}\boldsymbol{X}\Big)[i]=&\frac{1}{n-1}\sum_{j=1,j\neq i}^n \widehat{M}_{h,\alpha,n}(x_i,x_j)\frac{1}{h}(g_{ij}\boldsymbol{X}[j]-\boldsymbol{X}[i])\\ =&\frac{1}{n-1}\sum_{j=1,j\neq i}^n \widehat{M}_{h,\alpha,n}(x_i,x_j)\frac{1}{h}(u_i^{-1}/\!/_j^iX(x_j)-u_i^{-1}X(x_i)).\end{aligned} \tag{30}$$

Note that geometrically g_{ij} is closely related to the parallel transport from x_j to x_i. Indeed, by rewriting (9) as

$$\tau_0^{-1}\nabla_{\dot c(0)}X=\lim_{h\to 0}\frac{1}{h}\big\{\tau_h^{-1}X(c_h)-\tau_0^{-1}X(c_0)\big\}, \tag{31}$$

we see that the right hand side is exactly the term appearing in (30) by the definition of parallel transport. As will be shown explicitly in the next section, the VDM reveals the information about the manifold by accumulating the local information via taking the covariant derivative into account.

We unify ODM, LE and DM into the principal bundle framework by considering the special cases $\mathscr{E}^{\mathrm{ODM}}$ and $\mathscr{E}^{\mathrm{DM}}$. For ODM, we consider $\mathscr{E}^{\mathrm{ODM}}$. In this case, $\mathscr{G}$ is $\{u_i^{\mathrm{ODM}}\}_{i=1}^n$, the fiber E_x are isomorphic to $\mathbb{R}$ and $u_i^{\mathrm{ODM}}:\mathbb{R}\to E_i$. The translation group g_{ij}^{ODM} between x_i and x_j is constructed by

$$g_{ij}^{\mathrm{ODM}}=u_i^{\mathrm{ODM}\,-1}u_j^{\mathrm{ODM}}. \tag{32}$$

Note that in practice we sample u_j^{ODM} from sampling the frame bundle. Indeed, given x_i and $u_i\in O(\mathrm{M})$ so that $\pi(u_i)=x_i$, u_j^{ODM} is defined to be the orientation of u_i. Define a $n\times n$ matrix with scalar entries $\boldsymbol{P}_{h,\alpha,n}^{\mathrm{ODM}}$:

$$\boldsymbol{P}_{h,\alpha,n}^{\mathrm{ODM}}(i,j)=\begin{cases} w_{ij}g_{ij}^{\mathrm{ODM}} & (i,j)\in E,\\ 0 & (i,j)\notin E.\end{cases}$$

and a $n\times n$ diagonal matrix $\boldsymbol{D}_{h,\alpha,n}^{\mathrm{ODM}}$:

$$\boldsymbol{D}_{h,\alpha,n}^{\mathrm{ODM}}(i,i)=\sum_{j:(i,j)\in E} w_{ij}.$$

It has been shown in [25] that the orientability information of M can be obtained from analyzing $\boldsymbol{D}_{h,1,n}^{\mathrm{ODM}\,-1}\boldsymbol{P}_{h,1,n}^{\mathrm{ODM}}$. Furthermore, we get the orientable diffusion map (ODM) by taking the higher eigenvectors of $\boldsymbol{D}_{h,1,n}^{\mathrm{ODM}\,-1}\boldsymbol{P}_{h,1,n}^{\mathrm{ODM}}$ into account. We will prove in Appendix A.4 that any smooth, closed non-orientable manifold (M,g) has an orientable double covering embedded symmetrically inside $\mathbb{R}^p$ for some $p\in\mathbb{N}$. Combined this fact with the spectral convergence theorem that we will discuss later, we theoretically justify the claim in [25] that ODM can help reconstruct the orientable double covering.

For LE and DM, we consider $\mathscr{E}^{\mathrm{DM}}$. In this case, $m'=1$. Define a $n\times n$ matrix with scalar entries $\boldsymbol{P}_{h,\alpha,n}^{\mathrm{DM}}$:

$$\boldsymbol{P}_{h,\alpha,n}^{\mathrm{DM}}(i,j)=\begin{cases} w_{ij} & (i,j)\in E,\\ 0 & (i,j)\notin E.\end{cases}$$

and a $n\times n$ diagonal matrix $\boldsymbol{D}^{\text{DM}}_{h,\alpha,n}$:

$$\boldsymbol{D}^{\text{DM}}_{h,\alpha,n}(i,i)=\sum_{j:(i,j)\in E} w_{ij}.$$

Note that this is equivalent to ignoring the translation group in each vertex. Note that ${\boldsymbol{D}^{\text{DM}}_{h,0,n}}^{-1}\boldsymbol{P}^{\text{DM}}_{h,0,n}$ is the well-known graph Laplacian. With ${\boldsymbol{D}^{\text{DM}}_{h,\alpha,n}}^{-1}\boldsymbol{P}^{\text{DM}}_{h,\alpha,n}$ we can work with diffusion map which leads to dimension reduction. Recall that when we studied DM, we do not need the notion of translation group. This actually comes from the fact that functions defined on the manifold are actually sections of the trivial bundle of M – since the fiber $\mathbb{R}$ and M are "decoupled", we can directly take the algebraic relationship of $\mathbb{R}$ into consideration, so that it is not necessary to mention the bundle structure.

To sum up, we are able to unify the VDM, ODM, LE and DM by taking the principal bundle structure into account. In the next sections, we focus on the pointwise and spectral convergence of the corresponding operators, which justifies the definition of the VDD.

5. Pointwise and Spectral Convergence

With the above setup, we now do the asymptotic analysis under Assumption 4.1 and Assumption 4.2. In the whole proof, we systematically use a universal constant $R_M>0$ to bound $\|p^{(l)}\|_\infty$, $l=0,1,2,3$, the volume of M, the volume of ∂M, the curvature of M and ∂M and the second fundamental form introduced by the embedding ι, as well as their first few covariant derivatives. By assumption R_M is finite since $p\in C^3$, M, ∂M and ι are smooth and M is compact. From equation to equation, the exact constants might change, but we always use R_M to bound them. Also, we adapt the modified Landau notation to simplify the proof: the notation $X=O(C_1,C_2,\ldots,C_k)Y$ means that $|X|\leq\max\{C_1,C_2,\ldots,C_k\}Y$, where $C_i\geq 0$.

Definition 5.1. Define operators $T_{h,\alpha}:C(\mathscr{E})\to C(\mathscr{E})$ and $\widehat{T}_{h,\alpha,n}:C(\mathscr{E})\to C(\mathscr{E})$ as

$$T_{h,\alpha}X(y)=\int_{\text{M}} M_{h,\alpha}(y,x)\,/\!\!/_x^y X(x)p(x)\text{d}V(x)$$
$$\widehat{T}_{h,\alpha,n}X(y)=\frac{1}{n-1}\sum_{j=1,x_j\neq y}^{n}\widehat{M}_{h,\alpha,n}(y,x_j)\,/\!\!/_{x_j}^y X(x_j)$$

5.1. Pointwise Convergence.

We first show that asymptotically the matrix $\boldsymbol{D}^{-1}_{h,\alpha,n}\boldsymbol{P}_{h,\alpha,n}$ behaves like an integral operator.

Proposition 5.1. Suppose Assumption 4.1, Assumption 4.2 and Assumption 4.3 hold. For $X\in C(\mathscr{E})$ and for all $i=1,\ldots,n$, with high probability (w.h.p.) we have:

$$(B_{\mathscr{X}}\boldsymbol{D}^{-1}_{h,\alpha,n}\boldsymbol{P}_{h,\alpha,n}\boldsymbol{X})[i]=\widehat{T}_{h,\alpha,n}X(x_i)=T_{h,\alpha}X(x_i)+O(R_M,\|X\|_\infty)\frac{\sqrt{\log(n)}}{n^{1/2}h^{d/4-1/2}},$$

where $B_{\mathscr{X}}$ is defined in (10) and $\boldsymbol{X}$ is defined in (11).

Here the $\log(n)$ term shows up due to the normalization of the non-uniform sampling. Indeed, while normalizing the non-uniform sampling by (19), we introduce dependence, and handling this dependence is the resource of the $\log(n)$ term. Next, the integral operator $T_{h,\alpha}$ can be viewed as an approximation of identity in the following sense:

Proposition 5.2. Suppose Assumption 4.1 and Assumption 4.3 hold. For all $x_i\notin \text{M}_{\sqrt{h}}$ and $X\in C^4(\mathscr{E})$ we have

$$T_{h,\alpha}X(x)=X(x)+h\frac{\mu_2}{2d}\left(\nabla^2X(x)+\frac{2\nabla X(x)\cdot\nabla(p^{1-\alpha})(x)}{p^{1-\alpha}(x)}\right)+O(R_M,\|X^{(3)}\|_\infty)h^2,$$

where $\nabla X(x_i)\cdot\nabla(p^{1-\alpha})(x_i):=\sum_{l=1}^d\nabla_{\partial_l}X\nabla_{\partial_l}(p^{1-\alpha})$ and $\{\partial_l\}_{l=1}^d$ is an normal coordinate around x_i; when $x_i\in\text{M}_{\sqrt{h}}$,

$$T_{h,1}X(x)=X(x)+\frac{m_1^h}{m_0^h}P_{x,x_0}\nabla_{\partial_d}X(x_0)+O(R_M,\|X^{(2)}\|_\infty)h, \tag{33}$$

where $x_0 = \operatorname{argmin}_{y\in\partial \mathrm{M}} d(x_i,y)$, m_1^h and m_0^h are constants defined in (A.23) and (A.24), and ∂_d is the normal direction to the boundary at x_0.

The proofs of Proposition 5.1 and Proposition 5.2 are postponed to the Appendix. These Propositions lead to the following pointwise convergence theorem.

Theorem 5.1. Suppose Assumption 4.1, Assumption 4.2 and Assumption 4.3 hold. With the same notations at those in Proposition 5.2, for all $x_i \notin \mathrm{M}_{\sqrt{h}}$ and $X \in C^4(\mathscr{E})$, w.h.p.

$$h^{-1}(\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}\boldsymbol{X}-\boldsymbol{X})[i] = \frac{\mu_2}{2m}u_i^{-1}\left\{\nabla^2 X(x_i)+\frac{2\nabla X(x_i)\cdot\nabla(p^{1-\alpha})(x_i)}{p^{1-\alpha}(x_i)}\right\}+O(R_M,\|X^{(3)}\|_\infty)(h+n^{-\frac{1}{2}}h^{-\frac{d+2}{4}}).$$

In particular, we obtain the Connection Laplacian by

$$\lim_{h\to 0}\lim_{n\to\infty} h^{-1}u_i(\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}\boldsymbol{X}-\boldsymbol{X})[i] = \frac{\mu_2}{2m}\nabla^2 X(x_i).$$

For all $x_i \in \mathrm{M}_{\sqrt{h}}$, we have w.h.p.

$$(\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}\boldsymbol{X})[i] = \rho(g_i)^{-1}P_{x_i,x_0}\left(X(x_0)+\frac{m_1^h}{m_0^h}\nabla_{\partial_d}X(x_0)\right)+O(R_M,\|X^{(2)}\|_\infty)(h+n^{-\frac{1}{2}}h^{-\frac{d-2}{4}}). \tag{34}$$

Note that the dominant term in (34), m_1^h/m_0^h, is of order $\sqrt{h}$, which asymptotically dominates h. A consequence of Theorem 5.1 and the spectral convergence theorem in the next section is that the eigenvectors of $\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}-\boldsymbol{I}_n$ are discrete approximations of the eigen-vector-fields of the Connection Laplacian operator with homogeneous Neumann boundary condition that satisfy

$$\begin{cases} \nabla^2 X(x) = -\lambda X(x), & \text{for } x\in \mathrm{M},\\ \nabla_{\partial_d}X(x) = 0, & \text{for } x\in\partial\mathrm{M}. \end{cases} \tag{35}$$

5.2. **Spectral Convergence.** As informative as the convergence results in Theorem 5.1 are, they are pointwise in nature and are not strong enough to guarantee the spectral convergence of our numerical algorithm so as to guarantee the information provided by VDD. In this section, we explore this problem and provide the spectral convergence theorem. We denote the spectrum of ∇^2 by $\{-\lambda_l\}_{l=0}^\infty$, where $0=\lambda_0\leq\lambda_1\leq\ldots$, and the corresponding eigenspaces by $E_l := \{X\in L^2(\mathscr{E}):\ \nabla^2 X = -\lambda_l X\}$, $l=0,1,\ldots$. It is well known [13] that $\dim(E_l)<\infty$, the eigen-vector-fields are smooth and form a basis for $L^2(\mathscr{E})$, that is, $L^2(\mathscr{E}) = \overline{\oplus_{l\in\mathbb{N}\cup\{0\}}E_l}$, where the over line means completion according to the measure associated with g. To simplify the statement and proof, we assume that λ_l for each l are simple and X_l is the normalized basis of E_l.[4] Note that in general λ_0 may not exist, for example, S^2 with the standard metric. The first theorem stating the spectral convergence of $(\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n})^{t/h}$ to $e^{t\nabla^2}$. Note that in the statement of the theorem, we use $\widehat{T}_{h,1,n}$ instead of $\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}$. As will be seen in the proof, they are equivalent under proper transformation.

Theorem 5.2. Assume Assumption 4.1, Assumption 4.2 and Assumption 4.3 hold. Fix $t>0$. Denote $\mu_{t,i,h,n}$ to be the i-th eigenvalue of $\widehat{T}_{h,1,n}^{t/h}$ with the associated eigenvector $Y_{t,i,h,n}$. Also denote $\mu_{t,i}>0$ to be the i-th eigenvalue of the heat kernel of the Connection Laplacian $e^{t\nabla^2}$ with the associated eigen-vector field $Y_{t,i}$. We assume that $\mu_{t,i}$ are simple and both $\mu_{t,i,h,n}$ and $\mu_{t,i}$ decrease as i increases, respecting the multiplicity. Fix $i\in\mathbb{N}$. Then there exists a sequence $h_n\to 0$ such that $\lim_{n\to\infty}\mu_{t,i,h_n,n}=\mu_{t,i}$ and $\lim_{n\to\infty}\|Y_{t,i,h_n,n}-Y_{t,i}\|_{L^2(\mathscr{E})}=0$ in probability.

Note that when $n<\infty$, $\mu_{t,i,h,n}$ may be negative but $\mu_{t,i}$ is always non-negative. The second theorem stating the spectral convergence of $h^{-1}(\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}-\boldsymbol{I}_{m'n})$ to ∇^2.

Theorem 5.3. Assume Assumption 4.1, Assumption 4.2 and Assumption 4.3 hold. Denote $\lambda_{i,h,n}$ to be the i-th eigenvalue of $h^{-1}(\widehat{T}_{h,1,n}-1)$ with the associated eigenvector $X_{i,h,n}$. Also denote $-\lambda_i$, where $\lambda_i>0$, to be the i-th eigenvalue of the Connection Laplacian ∇^2 with the associated eigen-vector field X_i. We assume that λ_i are simple and both $\lambda_{n,i}^h$

[4] When any of the eigenvalues is not simple, the statement and proof are complicated by introducing the notion of spectral projection [8].

and λ_i increase as i increases, respecting the multiplicity. Fix $i \in \mathbb{N}$. Then there exists a sequence $h_n \to 0$ such that $\lim_{n\to\infty} \lambda_{i,h_n,n} = \lambda_i$ and $\lim_{n\to\infty} \|X_{i,h_n,n} - X_i\|_{L^2(\mathscr{E})} = 0$ in probability.

Note that the statement and proof hold for the special cases associated with LE, DM and ODM. To prove Theorem 5.2 and Theorem 5.3, we need the following Lemma. This lemma essentially takes care of the pointwise convergence of a series of vector fields in the uniform norm on M.

Lemma 5.2. Suppose Assumption 4.1, Assumption 4.2 and Assumption 4.3 are satisfied. Denote

$$\mathscr{P}_h := \{K_h(x,\cdot), x \in \mathrm{M}\}, \tag{36}$$

$$\mathscr{K}_{h,\alpha} := \{K_{h,\alpha}(x,\cdot), x \in \mathrm{M}\}, \tag{37}$$

Then the above classes are Glivenko-Cantelli classes. Take $X \in C(\mathscr{E})$, take a measurable section $q_0 : \mathrm{M} \to P(\mathrm{M}, G)$ and denote

$$X \cdot \mathscr{M}_{h,\alpha} := \Big\{ M_{h,\alpha}(x,\cdot) q_0(x)^{-1} /\!\!/_{\cdot}^{x} X(\cdot), x \in \mathrm{M} \Big\} \tag{38}$$

as a $\mathbb{R}^{m'}$-valued function class. Then the above classes satisfies

$$\sup_{W \in X\cdot\mathscr{M}_{h,\alpha}} \|\mathbb{P}_n W - \mathbb{P} W\|_{\mathbb{R}^{m'}} \to 0 \tag{39}$$

a.s. when $n \to \infty$.

The above notations are chosen to be compatible with the matrix notation used in the VDM algorithm. Notice how (39) mimics the definition of the Glivenko-Cantelli classes.

Proof. We prove (39). The proof of (36) and (37) follows the same lines, or see [30, Proposition 11]. Fix $x \in \mathrm{M}$. It is well known that we can find a measurable function $q_x : \mathrm{M} \to G$ so that the section $q := q_0 q_x$, where $q(y) := q_x(y) \circ q_0(y)$ for all $y \in M$, is discontinuous on the cut locus of x, which is a set of measure zero. Define the $\mathbb{R}^{m'}$-valued function

$$W_x(y) := M_{h,\alpha}(x,y) q(x)^{-1} /\!\!/_{y}^{x} X(y)$$

is thus also discontinuous. Since $X \in C(\mathscr{E})$, M is compact, $\nabla^{\mathscr{E}}$ is metric and $q(x) : \mathbb{R}^{m'} \to E_x$ preserving the inner product, we know

$$\|W_x\|_{L^\infty} \le \delta^{-1} \|K\|_{L^\infty} \|q(x)^{-1} /\!\!/_{y}^{x} X(y)\|_{L^\infty} = \delta^{-1} \|K\|_{L^\infty} \|X\|_{L^\infty},$$

where the first inequality holds since $\sup_{x,y\in\mathrm{M}} |M_{h,\alpha}(x,y)| \le \delta^{-1}\|K\|_{L^\infty}$ by Assumption 4.3. Under Assumption 4.1, q_x is isometric pointwisely, so $X \cdot \mathscr{M}_{h,\alpha}$ is uniformly bounded.

Fix $x \in \mathrm{M}$. We now tackle the vector-valued function W_x component by component. Denote $W_x = (f_{x,1}, \dots, f_{x,m'})$ and fix $j = 1, \dots m'$. Consider the functional set

$$f_{x,j} \cdot \mathscr{M}_{h,\alpha} := \Big\{ M_{h,\alpha}(x,\cdot) f_{x,j}(\cdot), x \in \mathrm{M} \Big\}.$$

Fix $\varepsilon > 0$. Since $X \cdot \mathscr{M}_{h,\alpha}$ is uniformly bounded, we can choose finite ε-brackets $[l_{j,i}, u_{j,i}]$, where $i = 1, \dots, N(\varepsilon)$, $l_{j,i}$ and $u_{j,i}$ are continuous and independent of x, so that its union contains $f_{x,j} \cdot \mathscr{M}_{h,\alpha}$ and $\mathbb{P}|u_{j,i} - l_{j,i}| < \varepsilon$ for all $i = 1, \dots, N(j,\varepsilon)$. Then, for every $f \in f_{x,j} \cdot \mathscr{M}_{h,\alpha}$, there is a bracket $[l_{j,l}, u_{j,l}]$ such that

$$|\mathbb{P}_n f - \mathbb{P} f| \le |\mathbb{P}_n u_{j,l} - \mathbb{P} u_{j,l}| + \mathbb{P}|u_{j,l}(y) - f(y)| \le |\mathbb{P}_n u_{j,l} - \mathbb{P} u_{j,l}| + \varepsilon.$$

Hence we have

$$\sup_{f \in f_{x,j}\cdot\mathscr{M}_{\alpha}} |\mathbb{P}_n f - \mathbb{P} f| \le \max_{i=1,\dots,N(j,\varepsilon)} |\mathbb{P}_n u_{j,l} - \mathbb{P} u_{j,l}| + \varepsilon,$$

where the right hand side converges almost surely to ε by the strong law of large numbers and the fact that $N(j,\varepsilon)$ is finite. As a result, we have

$$|\mathbb{P}_n W_x - \mathbb{P} W_x| \le \sum_{j=1}^{m'} \sup_{f \in f_{x,j}\cdot\mathscr{M}_{h,\alpha}} |\mathbb{P}_n f - \mathbb{P} f| \le \sum_{l=1}^{m'} \max_{i=1,\dots,N(j,\varepsilon)} |\mathbb{P}_n u_{j,l} - \mathbb{P} u_{j,l}| + m'\varepsilon,$$

so that $\sup_{W_x\in X\cdot\mathscr{M}_{h,\alpha}}|\mathbb{P}_nW_x-\mathbb{P}W_x|$ is bounded by $m'\varepsilon$ almost surely as $n\to\infty$. Since m' is fixed and ε is arbitrary, we conclude the proof. □

With the Lemma, we can now finish the proof of Theorem 5.2 and Theorem 5.3.

Proof of Theorem 5.2 and Theorem 5.3. The proof consists of 3 steps. The boundary, the nonuniform sampling and the bundle structure are taken care simultaneously.

Step 1: Relationship between $\boldsymbol{D}^{-1}_{h,\alpha,n}\boldsymbol{P}_{h,\alpha,n}$ and $\widehat{T}_{h,\alpha,n}$.

We immediately have that

$$(B_{\mathscr{X}}\delta_{\mathscr{X}}\widehat{T}_{h,\alpha,n}X)[i]=\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\widehat{M}_{h,\alpha,n}(x_i,x_j)\,/\!/_{x_j}^{x_i}X(x_j)=(\boldsymbol{D}^{-1}_{h,\alpha,n}\boldsymbol{P}_{h,\alpha,n}\boldsymbol{X})[i].$$

Next we understand the relationship between the eigen-structure of $B_{\mathscr{X}}\delta_{\mathscr{X}}\widehat{T}_{h,\alpha,n}$ and $\boldsymbol{D}^{-1}_{h,\alpha,n}\boldsymbol{P}_{h,\alpha,n}$. Suppose X is an eigen-section of $\widehat{T}_{h,\alpha,n}$ with eigenvalue λ. By a direct calculation we know $\boldsymbol{X}=B_{\mathscr{X}}\delta_{\mathscr{X}}X$ is an eigenvector of $\boldsymbol{D}^{-1}_{h,\alpha,n}\boldsymbol{P}_{h,\alpha,n}$ with eigenvalue λ. Indeed, for all $i=1,\ldots,n$,

$$\begin{aligned}(\boldsymbol{D}^{-1}_{h,\alpha,n}\boldsymbol{P}_{h,\alpha,n}\boldsymbol{X})[i]=&\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\widehat{M}_{h,\alpha,n}(x_i,x_j)\,u_i^{-1}/\!/_{x_j}^{x_i}X(x_j)\\=u_i^{-1}\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}&\widehat{M}_{h,\alpha,n}(x_i,x_j)\,/\!/_{x_j}^{x_i}X(x_j)=u_i^{-1}\widehat{T}_{h,\alpha,n}X(x_i)=\lambda u_i^{-1}X(x_i)=\lambda\boldsymbol{X}[i].\end{aligned}$$

On the other hand, given an eigenvector $\boldsymbol{v}$ of $h^{-1}(\boldsymbol{D}^{-1}_{h,\alpha,n}\boldsymbol{P}_{h,\alpha,n}-\boldsymbol{I}_{nd})$ with eigenvalue λ, that is,

$$(\boldsymbol{D}^{-1}_{h,\alpha,n}\boldsymbol{P}_{h,\alpha,n}\boldsymbol{v})[i]=\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\widehat{M}_{h,\alpha,n}(x_i,x_j)\,u_i^{-1}/\!/_{ij}u_j\boldsymbol{v}_j=(1+h\lambda)\boldsymbol{v}[i].\tag{40}$$

When $h\lambda>-1$, we show that there is an eigen-vector field of $h^{-1}(\widehat{T}_{h,\alpha,n}-1)$ with eigenvalue λ. In order to show this fact, we note that if X is an eigen-vector field of $h^{-1}(\widehat{T}_{h,\alpha,n}-1)$ with eigenvalue λ so that $h\lambda>-1$, we have

$$X(x)=\frac{\widehat{T}_{h,\alpha,n}X(x)}{1+h\lambda}=\frac{\frac{1}{n-1}\sum_{j=1,x_j\neq x}^{n}\widehat{M}_{h,\alpha,n}(x,x_j)\,/\!/_{x_j}^{x}X(x_j)}{(1+h\lambda)},$$

which can be rewritten as

$$X(x_i)=\frac{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\widehat{M}_{h,\alpha,n}(x_i,x_j)\,/\!/_{x_j}^{x_i}u_ju_j^{-1}X(x_j)}{1+h\lambda}=\frac{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\widehat{M}_{h,\alpha,n}(x_i,x_j)\,/\!/_{x_j}^{x_i}u_j\overline{X}[j]}{1+h\lambda}.\tag{41}$$

The relationship in (41) leads us to consider the vector field

$$X_{\boldsymbol{v}}(x):=\frac{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\widehat{M}_{h,\alpha,n}(x,x_j)\,/\!/_{x_j}^{x}u_jv[j]}{1+h\lambda}$$

to be the related eigen-vector field associated with $\boldsymbol{v}$. To show that $X_{\boldsymbol{v}}$ is indeed an eigen-vector field of $h^{-1}(\widehat{T}_{h,\alpha,n}-1)$, we directly calculate:

$$\begin{aligned}\widehat{T}_{h,\alpha,n}X_{\boldsymbol{v}}(y)=&\frac{1}{n-1}\sum_{j=1,x_j\neq y}^{n}\widehat{M}_{h,\alpha,n}(y,x_j)\,/\!/_{x_j}^{y}X_{\boldsymbol{v}}(x_j)\\=&\frac{1}{n-1}\sum_{j=1,x_j\neq y}^{n}\widehat{M}_{h,\alpha,n}(y,x_j)\,/\!/_{x_j}^{y}\left(\frac{\frac{1}{n-1}\sum_{k=1}^{n}\widehat{M}_{h,\alpha,n}(x_j,x_k)\,/\!/_{x_k}^{x_j}u_k\boldsymbol{v}[k]}{1+h\lambda}\right)\\=&\frac{1}{1+h\lambda}\frac{1}{n-1}\sum_{j=1,x_j\neq y}^{n}\widehat{M}_{h,\alpha,n}(y,x_j)\,/\!/_{x_j}^{y}(1+h\lambda)u_j\boldsymbol{v}[j]=(1+h\lambda)X_{\boldsymbol{v}}(y),\end{aligned}$$

which leads to the fact that $X_{\boldsymbol{v}}$ is the eigen-vector field of $h^{-1}(\widehat{T}_{h,\alpha,n}-1)$ with eigenvalue λ since $h\lambda>-1$.

The above one to one relationship between eigenvalues and eigenfunctions of $\widehat{T}_{h,\alpha,n}$ and $\boldsymbol{D}^{-1}_{h,\alpha,n}\boldsymbol{P}_{h,\alpha,n}$ allows us to analyze the spectral convergence of $\boldsymbol{D}^{-1}_{h,\alpha,n}\boldsymbol{P}_{h,\alpha,n}$ to $T_{h,\alpha}$ by analyzing the spectral convergence of $\widehat{T}_{h,\alpha,n}$ to $T_{h,\alpha}$.

Step 2: compact convergence of $\widehat{T}_{h,\alpha,n}$ to $T_{h,\alpha}$.

We start from the definition of compact convergence of a series of operators [8, p122] adapted to our setup. We say $\widehat{T}_{h,\alpha,n}:C(\mathscr{E})\to C(\mathscr{E})$ compactly converges to $T_{h,\alpha}:C(\mathscr{E})\to C(\mathscr{E})$ if and only if

(C1) $\widehat{T}_{h,\alpha,n}$ converges to $T_{h,\alpha}$ pointwisely, that is, for all $X\in C(\mathscr{E})$, we have $\|\widehat{T}_{h,\alpha,n}X-T_{h,\alpha}X\|_{L^\infty(\mathscr{E})}\to 0$ as $n\to\infty$;
(C2) for any uniformly bounded sequence $\{X_l:\|X_l\|_{L^\infty}\leq 1\}_{l=1}^\infty\subset C(\mathscr{E})$, the sequence $\{(\widehat{T}_{h,\alpha,n}-T_{h,\alpha})X_l\}_{l=1}^\infty$ is relatively compact.

We prepare some bounds for the later use. Under the Assumption 4.3, for all $n\in\mathbb{N}$ and $x,y\in \mathrm{M}$ we have the following simple bounds:

$$\delta\leq|\widehat{p}_{h,n}(x)|\leq\|K\|_{L^\infty}, \tag{42}$$

$$\delta\|K\|^{-2\alpha}_{L^\infty}\leq|\widehat{K}_{h,\alpha,n}(x,y)|\leq\delta^{-2\alpha}\|K\|_{L^\infty}, \tag{43}$$

$$\delta\|K\|^{-2\alpha}_{L^\infty}\leq|\widehat{d}_{h,\alpha,n}(x)|\leq\delta^{-2\alpha}\|K\|_{L^\infty}, \tag{44}$$

$$\delta^{1+2\alpha}\|K\|^{-1-2\alpha}_{L^\infty}\leq|\widehat{M}_{h,\alpha,n}(x,y)|\leq\delta^{-1-2\alpha}\|K\|^{1+2\alpha}_{L^\infty}. \tag{45}$$

Now we show (C1). By a direct calculation we have

$$\|\widehat{T}_{h,\alpha,n}X-T_{h,\alpha}X\|_{L^\infty(\mathscr{E})}=\Big\|\mathbb{P}_n\widehat{M}_{h,\alpha,n}(y,\cdot)/\!/^y_\cdot X(\cdot)-\mathbb{P}M_{h,\alpha}(y,\cdot)/\!/^y_\cdot X(\cdot)\Big\|_{L^\infty(\mathscr{E})}$$

$$\leq\Big\|\mathbb{P}_n\widehat{M}_{h,\alpha,n}(y,\cdot)/\!/^y_\cdot X(\cdot)-\mathbb{P}_n\widehat{M}^{(d_{h,\alpha})}_{h,\alpha,n}(y,\cdot)/\!/^y_\cdot X(\cdot)\Big\|_{L^\infty(\mathscr{E})} \tag{46}$$

$$+\Big\|\mathbb{P}_n\widehat{M}^{(d_{h,\alpha})}_{h,\alpha,n}(y,\cdot)/\!/^y_\cdot X(\cdot)-\mathbb{P}_nM_{h,\alpha}(y,\cdot)/\!/^y_\cdot X(\cdot)\Big\|_{L^\infty(\mathscr{E})} \tag{47}$$

$$+\Big\|\mathbb{P}_nM_{h,\alpha}(y,\cdot)/\!/^y_\cdot X(\cdot)-\mathbb{P}M_{h,\alpha}(y,\cdot)/\!/^y_\cdot X(\cdot)\Big\|_{L^\infty(\mathscr{E})} \tag{48}$$

Rewrite (48) as $\sup_{W\in X\circ\mathscr{M}_{h,\alpha}}\|\mathbb{P}_nW-\mathbb{P}W\|_{\mathbb{R}^m}$. By Lemma 5.2, (48) converges to 0 a.s.. The convergence of (46) can be seen directly by

$$\begin{aligned}&\Big\|\mathbb{P}_n\widehat{M}_{h,\alpha,n}(y,\cdot)/\!/^y_\cdot X(\cdot)-\mathbb{P}_n\widehat{M}^{(d_{h,\alpha})}_{h,\alpha,n}(y,\cdot)/\!/^y_\cdot X(\cdot)\Big\|_{L^\infty}\leq\|X\|_{L^\infty}\Big\|\frac{\widehat{K}_{h,\alpha,n}(x,y)-K_{h,\alpha}(x,y)}{\widehat{d}_{h,\alpha,n}(x)}\Big\|_{L^\infty}\\
\leq&\|X\|_{L^\infty}\|K\|^{2\alpha-1}_{L^\infty}\|\widehat{K}_{h,\alpha,n}(x,y)-K_{h,\alpha}(x,y)\|_{L^\infty}\leq\|X\|_{L^\infty}\|K\|^{2\alpha}_{L^\infty}\Big\|\frac{1}{\widehat{p}^\alpha_{h,n}(x)\widehat{p}^\alpha_{h,n}(y)}-\frac{1}{p^\alpha_h(x)p^\alpha_h(y)}\Big\|_{L^\infty}\\
\leq&\frac{2\alpha\|X\|_{L^\infty}\|K\|^{2\alpha}_{L^\infty}}{\delta^{2\alpha+1}}\|\widehat{p}_{h,n}(y)-p_h(y)\|_{L^\infty}\leq\frac{2\alpha\|X\|_{L^\infty}\|K\|^{2\alpha}_{L^\infty}}{\delta^{2\alpha+1}}\sup_{f\in\mathscr{P}_h}\|\mathbb{P}_nf-\mathbb{P}f\|,\end{aligned}$$

where the second last inequality comes from the Taylor's expansion and the bounds of $p_h(x)$ and $\widehat{p}_{h,n}$, that is, $|\widehat{p}_{h,n}(y)^{-\alpha}-p_h(y)^{-\alpha}|\leq\alpha(\widehat{p}_{h,n}(x_i)-p_h(y))\delta^{-\alpha-1}$. By Lemma 5.2, (46) converges to 0 a.s. To get the convergence of (47), we do the following calculation.

$$\begin{aligned}&\Big\|\mathbb{P}_n\widehat{M}^{(d_{h,\alpha})}_{h,\alpha,n}(y,\cdot)/\!/^y_\cdot X(\cdot)-\mathbb{P}_nM_{h,\alpha}(y,\cdot)/\!/^y_\cdot X(\cdot)\Big\|_{L^\infty}\leq\|X\|_{L^\infty}\|K\|_{L^\infty}\Big\|\frac{1}{\widehat{d}_{h,\alpha,n}(x)}-\frac{1}{d_{h,\alpha}(x)}\Big\|_{L^\infty}\\
\leq&\|X\|_{L^\infty}\|K\|^{3-2\alpha}_{L^\infty}\|\widehat{d}_{h,\alpha,n}(x)-d_{h,\alpha}(x)\|_{L^\infty}\leq\|X\|_{L^\infty}\|K\|^{3-2\alpha}_{L^\infty}\|\widehat{d}_{h,\alpha,n}(x)-d_{h,\alpha}(x)\|_{L^\infty}.\end{aligned}$$

We bound the last term by

$$\|\widehat{d}_{h,\alpha,n}(x)-d_{h,\alpha}(x)\|_{L^\infty} \leq \|\widehat{d}_{h,\alpha,n}(x)-\widehat{d}^{(p_h)}_{h,\alpha,n}(x)\|_{L^\infty}+\|\widehat{d}^{(p_h)}_{h,\alpha,n}(x)-d_{h,\alpha}(x)\|_{L^\infty}$$
$$\leq \|\widehat{K}_{h,\alpha,n}(x,y)-K_{h,\alpha}(x,y)\|_{L^\infty}+\sup_{k\in\mathscr{K}_{h,\alpha}}\|\mathbb{P}_n k-\mathbb{P}k\| \leq \frac{2\alpha}{\delta^{3\alpha+1}}\sup_{f\in\mathscr{P}_h}\|\mathbb{P}_n f-\mathbb{P}f\|+\sup_{k\in\mathscr{K}_{h,\alpha}}\|\mathbb{P}_n k-\mathbb{P}k\|$$

which converges to 0 a.s. as $n\to\infty$ by Lemma 5.2.

Next we check the condition (C2). Since $T_{h,\alpha}$ is compact, the problem is reduced to show that $\widehat{T}_{h,\alpha,n}X_n$ is pre-compact for any given sequence of vector fields $\{X_1,X_2,\ldots\}\subset C(\mathscr{E})$ so that $\|X_n\|_{L^\infty}\leq 1$ for all $n\in\mathbb{N}$. We count on the Arzela-Ascoli theorem [11, IV.6.7] to finish the proof. A direct calculation leads to

$$\sup_{n\geq 1}\|\widehat{T}_{h,\alpha,n}X_n\|_{L^\infty(\mathscr{E})}=\sup_{n\geq 1}\left\|\frac{1}{n-1}\sum_{j=1,x_j\neq y}^{n}\widehat{M}_{h,\alpha,n}(y,x_i)/\!/_{x_i}^{y}X_n(x_i)\right\|_{L^\infty(\mathscr{E})}\leq\frac{\|K\|^{1+2\alpha}_{L^\infty}}{\delta^{1+2\alpha}},$$

where the second inequality comes from (45), which guarantees the uniform boundedness. Next we show the equi-continuity of $\widehat{T}_{h,\alpha,n}X_n$. For a given pair of close points $x\in\mathrm{M}$ and $y\in\mathrm{M}$, a direct calculation leads to

$$\|\widehat{T}_{h,\alpha,n}X_n(y)-/\!/_x^y\widehat{T}_{h,\alpha,n}X_n(x)\|=\left\|\frac{1}{n-1}\sum_{j=1,x_j\neq y}^{n}\widehat{M}_{h,\alpha,n}(y,x_i)/\!/_{x_i}^{y}X_n(x_i)-/\!/_x^y\frac{1}{n-1}\sum_{j=1,x_j\neq x}^{n}\widehat{M}_{h,\alpha,n}(x,x_i)/\!/_{x_i}^{x}X_n(x_i)\right\|$$
$$\leq\|X_n\|_{L^\infty(\mathscr{E})}\left\|\frac{1}{n-1}\sum_{j=1,x_j\neq y}^{n}\widehat{M}_{h,\alpha,n}(y,x_i)-\frac{1}{n-1}\sum_{j=1,x_j\neq x}^{n}\widehat{M}_{h,\alpha,n}(x,x_i)\right\|\leq\left\|\widehat{M}_{h,\alpha,n}(y,\cdot)-\widehat{M}_{h,\alpha,n}(x,\cdot)\right\|_{L^\infty(\mathrm{M})},$$

where the last inequality comes from the Holder's inequality. Note that by Assumption 4.3 and the compactness of M, $\widehat{p}_{h,n}$ is uniformly continuous on M and is uniformly bounded from below by (42). Therefore, $\widehat{K}_{h,\alpha,n}$ is uniformly continuous on $\mathrm{M}\times\mathrm{M}$, and hence $\widehat{d}_{h,\alpha,n}$ is uniformly continuous on M and is uniformly bounded from below by (44). As a result, $\widehat{\mathrm{M}}_{h,\alpha,n}$ is uniformly continuous on $\mathrm{M}\times\mathrm{M}$, which implies the equi-continuous of $\{\widehat{T}_{h,\alpha,n}X_n\}_{n\geq N}$.

Since the compact convergence implies the spectral convergence (see [8] or [30, Proposition 6]), we get the spectral convergence of $\widehat{T}_{h,\alpha,n}$ to $T_{h,\alpha}$.

Step 3(A): Spectral convergence of $T_{h,1}$ when $\partial\mathbf{M}=\emptyset$.

We assume $\frac{\mu_2}{d}=1$ to simplify the notation. Fix $l\geq 0$. For all $x\in\mathrm{M}$, by Proposition 5.2 we have uniformly

$$\frac{T_{\varepsilon,1}X_l(x)-X_l(x)}{h}=\nabla^2X_l(x)+O(R_M,\|X_l^{(3)}\|_{L^\infty(\mathscr{E})})h.$$

By the Sobolev embedding theory [22, Theorem 9.2], we know

$$\|X_l^{(3)}\|_{L^\infty(\mathscr{E})}\leq\|X_l\|_{H^{d/2+4}(\mathscr{E})}\leq\|(\nabla^2)^{d/4+2}X_l\|_{L^2(\mathscr{E})}=\lambda_l^{d/4+2},$$

where we choose $d/2+4$ for convenience. Thus, in the L^2 sense,

$$\left\|\frac{T_{h,1}X_l-X_l}{h}-\nabla^2X_l\right\|_{L^2(\mathscr{E})}=O(R_M)h\lambda_l^{d/4+2},$$

so that on $\overline{\oplus_{k\leq l}E_k}$ the following holds

$$T_{h,1}=I+h\nabla^2+O(R_M)h^2\lambda_l^{d/4+2}. \tag{49}$$

Next we show $T_{h,1}^{t/h}$ converges to $e^{t\nabla^2}$ for $t>0$ as $h\to 0$. When $h<\frac{1}{2\lambda_l}$, $I+h\nabla^2$ is invertible on $\overline{\oplus_{k\leq l}E_k}$ with norm $\frac{1}{2}\leq\|I+h\nabla^2\|<1$ and we have

$$(I+h\nabla^2)^{t/h}X_l=(1-t\lambda_l+t^2\lambda_l^2/2-th\lambda_l^2/2+\ldots)X_l$$

by the binomial expansion. On the other hand, we have

$$e^{t\nabla^2}X_l = (1 - t\lambda_l + t^2\lambda_l^2/2 + \ldots)X_l.$$

Therefore, when $h < \dfrac{1}{2\lambda_l}$ we have on $\overline{\oplus_{k\leq l}E_k}$

$$e^{t\nabla^2} = (I + h\nabla^2)^{t/h} + O(1)th\lambda_l^2. \tag{50}$$

Furthermore, for any $s > 0$, if B is a given bounded operator with norm $\|B\| < 1$, then we have

$$\|(I+B)^s - I\| = s(1+\|B\|)^{s-1}\|B\| \tag{51}$$

Now we put the above together. Fix $l \geq 0$. Take $h < \frac{1}{2\lambda_l}$ small enough so that

$$\|T_{h,1} - I - h\nabla^2\| \leq 1/2. \tag{52}$$

Then we have over $\overline{\oplus_{k\leq l}E_k}$:

$$\begin{aligned}
\|T_{h,1}^{t/h} - e^{t\nabla^2}\| &= \left\|(I + h\nabla^2 + O(R_M)h^2\lambda_l^{d/4+2})^{t/h} - \left(I + h\nabla^2\right)^{t/h} - O(1)th\lambda_l^2\right\| \\
&\leq \left\|\left(I + h\nabla^2\right)^{t/h}\right\| \left\|\left[I + \left(I + h\nabla^2\right)^{-1} O(R_M)h^2\lambda_l^{d/4+2}\right]^{t/h} - I - O(1)th\lambda_l^2\right\| \\
&\leq \left\|\left[I + \left(I + h\nabla^2\right)^{-1} O(R_M)h^2\lambda_l^{d/4+2}\right]^{t/h} - I - O(1)th\lambda_l^2\right\| \\
&\leq \frac{t}{h}2O(R_M)h^2\lambda_l^{d/4+2}(1 + 2O(R_M)h^2\lambda_l^{d/4+2})^{t/h-1} + O(1)th\lambda_l^2 \\
&\leq th\lambda_l^2\left(O(R_M)(1 + 2O(R_M)h^2\lambda_l^{d/4+2})^{t/h}\lambda_l^{d/4} + O(1)\right),
\end{aligned}$$

where the first equality comes from (49) and (50), the second inequality comes from the fact that $h < \dfrac{1}{2\lambda_l}$, and the third inequality comes from (51) and (52). Thus, when l is large enough, we have $\|T_{h,1}^{t/h} - e^{t\nabla^2}\| \leq O(R_M)(1 + 2O(R_M)h^2\lambda_l^{d/4+2})^{t/h}th\lambda_l^{d/4+2}$. We can thus choose h much smaller so that $h\lambda_l^{d/4+2} < h^{1/2}$, which is equivalent to $\lambda < h^{-2/(d+8)}$, and hence reach the fact that

$$\|T_{h,1}^{t/h} - e^{t\nabla^2}\| \leq O(R_M)(1 + 2O(R_M)h^2\lambda_l^{d/4+2})^{t/h}th^{1/2}$$

on $\mathscr{H}_h := \overline{\oplus_{l:\lambda_l<h^{-2/(d+8)}}E_l}$. Since $h^2\lambda_l^{d/4+2} < h^{3/2}$ under the assumption, as $h \to 0$, we get $(1 + 2O(R_M)h^2\lambda_l^{d/4+2})^{t/h}$ is bounded and hence we conclude the spectral convergence of $T_{h,1}^{t/h}$ while the boundary is empty.

Furthermore, for $X \in \mathscr{H}_h$, $\|X\|_{L^2(\mathscr{E})} = 1$ and $t = h$, we have

$$\begin{aligned}
\left\|h^{-1}(T_{h,1} - 1)X - \nabla^2X\right\|_{L^2(\mathscr{E})} &\leq \left\|h^{-1}(T_{h,1} - e^{h\nabla^2})X\right\|_{L^2(\mathscr{E})} + \left\|h^{-1}(e^{h\nabla^2} - 1)X - \nabla^2X\right\|_{L^2(\mathscr{E})} \\
&\leq O(R_M)(1 + 2O(R_M)h^2\lambda_l^{d/4+2})h^{1/2} + |h^{-1}(e^{-h\lambda_l} - 1) + \lambda_l|.
\end{aligned} \tag{53}$$

Thus, as $h \to 0$, $h^{-1}(T_{h,1} - 1)X$ converges to ∇^2X in $L^2(\mathscr{E})$.

Step 3(B): Spectral convergence of $T_{h,1}$ when $\partial\mathrm{M} \neq \emptyset$. We follow the same notation as that in Step 3-1. Fix $l \geq 0$. By Proposition 5.1 and the Sobolev embedding theory, for $x \in \mathrm{M}\backslash\mathrm{M}_{\sqrt{h}}$, we have

$$\frac{T_{\varepsilon,1}X_l(x) - X_l(x)}{h} = \nabla^2X_l(x) + O(R_M)h\lambda_l^{d/4+2};$$

for $x \in \mathrm{M}_{\sqrt{h}}$, by the Neuman's condition, we have

$$T_{h,1}X_l(x) = P_{x,x_0}X_l(x_0) + O(C, \|\nabla^2X_l\|_\infty)h = P_{x,x_0}X(x_0) + O(R_M)h\lambda_l^{d/4+3/2}.$$

Thus, in the L^2 sense,

$$\left\|\frac{T_{h,1}X_l - X_l}{h} - \nabla^2 X_l\right\|_{L^2(\mathscr{E})} = O(R_M)h^{1/4}\lambda_l^{d/4+3/2},$$

which implies that on $\overline{\oplus_{k\leq l}E_k}$, we have

$$T_{h,1} = I + h\nabla^2 + O(R_M)h^{5/4}\lambda_l^{d/4+3/2} \tag{54}$$

To show $T_{h,1}^{t/h}$ converges to $e^{t\nabla^2}$ for $t>0$ as $h\to 0$, note that (50) and (51) still hold. Take $h<\frac{1}{2\lambda_l}$ small enough so that

$$\|T_{h,1} - I - h\nabla^2\| \leq 1/2. \tag{55}$$

Then we have over $e\overline{\oplus_{k\leq l}E_k}$:

$$\begin{aligned}
\|T_{h,1}^{t/h} - e^{t\nabla^2}\| &= \left\|(I+h\nabla^2+O(R_M)h^{5/4}\lambda_l^{d/4+3/2})^{t/h} - \left(I+h\nabla^2\right)^{t/h} - O(1)th\lambda_l^2\right\| \\
&\leq \left\|\left[I+\left(I+h\nabla^2\right)^{-1}O(R_M)h^{5/4}\lambda_l^{d/4+3/2}\right]^{t/h} - I - O(1)th\lambda_l^2\right\| \\
&\leq th^{1/4}\lambda_l^{3/2}\left(O(R_M)(1+2O(R_M)h^{5/4}\lambda_l^{d/4+3/2})^{t/h}\lambda_l^{d/4} + O(1)\lambda_l^{1/2}\right),
\end{aligned}$$

where the first equality comes from (54) and (50), the second inequality comes from the fact that $h<\frac{1}{2\lambda_l}$, and the third inequality comes from (51) and (55). Thus, when l is large enough, we have $\|T_{h,1}^{t/h} - e^{t\nabla^2}\| \leq O(R_M)(1+2O(R_M)h^{5/4}\lambda_l^{d/4+3/2})^{t/h}th^{1/4}\lambda_l^{d/4+3/2}$. We can thus choose h much smaller so that $h^{1/4}\lambda_l^{d/4+3/2} < h^{1/8}$, which is equivalent to $\lambda_l < h^{-1/(2d+12)}$, and hence reach the fact that

$$\|T_{h,1}^{t/h} - e^{t\nabla^2}\| \leq O(R_M)(1+2O(R_M)h^{5/4}\lambda_l^{d/4+3/2})^{t/h}th^{1/8}$$

on $\mathscr{H}_h := \overline{\oplus_{l:\lambda_l<h^{-1/(2d+12)}}E_l}$. Since $h^{5/4}\lambda_l^{d/4+3/2} < h^{9/8}$ under the assumption, as $h\to 0$, we get $(1+2O(R_M)h^{5/4}\lambda_l^{d/4+3/2})^{t/h}$ is bounded and hence we conclude the spectral convergence of $T_{h,1}^{t/h}$ when the boundary is not empty.

Furthermore, for $X\in\mathscr{H}_h$, $\|X\|_{L^2(\mathscr{E})}=1$ and $t=h$, we have

$$\begin{aligned}
\left\|h^{-1}(T_{h,1}-1)X - \nabla^2 X\right\|_{L^2(\mathscr{E})} &\leq \left\|h^{-1}(T_{h,1}-e^{h\nabla^2})X\right\|_{L^2(\mathscr{E})} + \left\|h^{-1}(e^{h\nabla^2}-1)X - \nabla^2 X\right\|_{L^2(\mathscr{E})} \\
&\leq O(R_M)(1+2O(R_M)h^{5/4}\lambda_l^{d/4+3/2})h^{1/8} + |h^{-1}(e^{-h\lambda_l}-1)+\lambda_l|.
\end{aligned}$$

Thus, as $h\to 0$, $h^{-1}(T_{h,1}-1)X$ converges to $\nabla^2 X$ in $L^2(\mathscr{E})$.

Final step: Put everything together.

We prove Theorem 5.2 here. Denote $\mu_{t,i,h}$ to be the i-th eigenvalue of $T_{h,1}^{t/h}$ with the associated eigenvector $Y_{t,i,h}$. Fix $i\in\mathbb{N}$ and take $h_0>0$ so that $\mu_{t,i} < h_0^{-1}$. By Step 1, we know that the i-th eigenvalue of $\widehat{T}_{h,1,n}$ and $\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}$ are the same and their eigenvectors are related for all $h<h_0$. By Step 2, for all $h<h_0$, for any $\delta>0$, we can find $1/\delta>0$ so that for all $n>1/\delta$

$$P\{\|Y_{t,i,h,n} - Y_{t,i,h}\|_{L^2(\mathscr{E})} \geq \delta/2\} < \delta.$$

By step 3(A) (or Step 3(B) if $\partial \mathrm{M}\neq\emptyset$), for the given $\varepsilon>0$, there exists $h(\varepsilon)>0$ so that

$$\|Y_{t,i} - Y_{t,i,h(\varepsilon)}\|_{L^2(\mathscr{E})} < \varepsilon/2.$$

By putting the above together, for any $\varepsilon>0$ so that $h(\varepsilon)<h_0$, when $\delta<\varepsilon$ and $n>1/\delta$, we have

$$P\{\|Y_{t,i,h(\varepsilon),n} - Y_{t,i}\|_{L^2(\mathscr{E})} \geq \varepsilon\} \leq P\{\|Y_{t,i,h(\varepsilon),n} - Y_{t,i,h(\varepsilon)}\|_{L^2(\mathscr{E})} \geq \varepsilon/2\} \leq P\{\|Y_{t,i,h(\varepsilon),n} - Y_{t,i,h(\varepsilon)}\|_{L^2(\mathscr{E})} \geq \delta/2\} < \delta.$$

Let $\delta\to 0$, we conclude the convergence in probability by choosing h_n decreasing to 0. Since the proof for Theorem 5.3 is the same, we skip it. $\square$

5.3. **Conclusion.** In conclusion, the matrix $\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}$ (resp. $h^{-1}(\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}-\boldsymbol{I}_{m'n})$) considered in the VDM algorithm converges in the spectral sense to $e^{t\nabla^2}$ (resp. ∇^2). The above presentation demonstrates the essential part of the VDM algorithm – the principal bundle structure. However, in practice we often have only the samples of the manifold M. In the next Section, we show that even if we have only the samples of the manifold, we are able to get the Connection Laplacian of the tangent bundle by reconstructing the frame bundle $O(\mathrm{M})$.

6. EXTRACT MORE TOPOLOGICAL/GEOMETRIC INFORMATION FROM A POINT CLOUD

In Section 3, we study VDM under the assumption that we have access to the principal bundle structure of the manifold. However, in practice the knowledge of the principal bundle structure is not always available and we may only have the point cloud sampled from the manifold. Is it possible to reconstruct the principal bundle under this situation? The answer is yes if we restrict ourselves to a special principal bundle, the frame bundle, and the tangent bundle.

We summarize the proposed reconstruction algorithm in [26] below. Take a point cloud $\mathscr{X}=\{x_i\}_{i=1}^n$ sampled from M under Assumption 4.1 (A1), Assumption 4.1 (A3) and 4.2 (B1). The algorithm consists of the following three steps:

(a) reconstruct the embedded tangent bundle and the frame bundle from $\mathscr{X}$. It is possible since locally a manifold can be well approximated by an affine space up to the second order [26, 16, 17]. Thus, the embedded tangent bundle can be estimated by the local PCA algorithm with the kernel bandwidth $h_{\text{pca}}>0$ [26]. Denote O_y to be a $p\times d$ matrix which columns represent the basis of the estimated embedded tangent plane on $T_y\mathrm{M}$, where y may or may not in $\mathscr{X}$. Here $O_y:\mathbb{R}^d\to\iota_*T_y\mathrm{M}$ is the estimation of the frame bundle. When $y=x_i\in\mathscr{X}$, we use O_i to denote O_{x_i};

(b) estimate the parallel transport between tangent planes by aligning O_x and O_y by defining the alignment between x and y as

$$O_{x,y}=\operatorname*{argmin}_{O\in O(d)}\|O-O_x^TO_y\|_{HS}\in O(d),$$

where $\|\cdot\|_{HS}$ is the Hilbert-Schmidt norm. It is proved that $O_{x,y}$ is an approximation of the parallel transport from y to x [26, (B.6)] in the following sense:

$$O_xO_{x,y}\overline{X}_y\approx\iota_*/\!\!/_y^xX(y),$$

where $X\in C(T\mathrm{M})$ and $\overline{X}_y=O_y^T\iota_*X(y)\in\mathbb{R}^d$ is the coordinate of $X(y)$ with related to the estimated basis. Note that x and y may or may not be in $\mathscr{X}$. When both $x=x_i\in\mathscr{X}$ and $y=x_j\in\mathscr{X}$, we use O_{ij} to denote O_{x_i,x_j} and $\overline{X}_j$ to denote $\overline{X}_{x_j}$;

(c) run VDM mentioned in Section 3 based on the weighted graph built up from $\mathscr{X}$ and $\{O_{ij}\}$. We build up a block matrix $\boldsymbol{P}_{h,\alpha,n}^{\mathrm{O}}$ with $d\times d$ entries:

$$\boldsymbol{P}_{h,\alpha,n}^{\mathrm{O}}(i,j)=\begin{cases}\widehat{K}_{h,\alpha,n}(x_i,x_j)O_{ij} & (i,j)\in E,\\ 0_{d\times d} & (i,j)\notin E,\end{cases}$$

where $h>h_{\text{pca}}$ and the kernel K satisfies Assumption 4.3. Then define a $n\times n$ diagonal block matrix $\boldsymbol{D}_{h,\alpha,n}$ with $d\times d$ entries defined in (28). It is shown in [26, Theorem 5.3] that for $0\leq\alpha\leq1$ and $X\in C^3(T\mathrm{M})$,

$$\lim_{h\to0}\lim_{n\to\infty}\frac{1}{h}(\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}^{\mathrm{O}}\boldsymbol{X}-\boldsymbol{X})[i]=\frac{\mu_{2,0}}{2d}O_i^T\iota_*\left\{\nabla^2X(x_i)+\frac{2\nabla X(x_i)\cdot\nabla(p^{1-\alpha})(x_i)}{p^{1-\alpha}(x_i)}\right\}.$$

Note that the errors introduced in (a) and (b) accumulate and may influence the VDM algorithm in (c). However, as is shown in [26, Theorem 5.3], this influence disappears asymptotically. In this section we further study the spectral convergence of $\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}^{\mathrm{O}}$ which answers our question in the beginning – we are able to extract further geometric/topological information from the point cloud.

Definition 6.1. Define operators $\widehat{T}_{h,\alpha,n}^{\mathrm{O}}:C(T\mathrm{M})\to C(T\mathrm{M})$ as

$$\widehat{T}_{h,\alpha,n}^{\mathrm{O}}X(y)=\iota_*^TO_y\frac{1}{n-1}\sum_{j=1,x_j\neq y}^{n}\widehat{M}_{h,\alpha,n}(y,x_j)\,O_{y,x_j}O_j^T\iota_*X(x_j)$$

The main result of this section is the following spectral convergence theorems stating the spectral convergence of $(\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}^{\mathrm{O}})^{t/h}$ to $e^{t\nabla^2}$ and $h^{-1}(\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}^{\mathrm{O}}-\boldsymbol{I}_{dn})$ to ∇^2.

Theorem 6.1. Assume Assumption 4.1 (A1), Assumption 4.1 (A3), Assumption 4.2 (B1) and Assumption 4.3 hold. Fix $t>0$. Denote $\mu_{t,i,h,n}^{\mathrm{O}}$ to be the i-th eigenvalue of $(\widehat{T}_{h,1,n}^{\mathrm{O}})^{t/h}$ with the associated eigenvector $Y_{t,i,h,n}^{\mathrm{O}}$. Also denote $\mu_{t,i}>0$ to be the i-th eigenvalue of the heat kernel of the Connection Laplacian $e^{t\nabla^2}$ with the associated eigen-vector field $Y_{t,i}$. We assume that both $\mu_{t,i,h,n}^{\mathrm{O}}$ and $\mu_{t,i}$ decrease as i increases, respecting the multiplicity. Fix $i\in\mathbb{N}$. Then there exists a sequence $h_n\to 0$ such that $\lim_{n\to\infty}\mu_{t,i,h_n,n}^{\mathrm{O}}=\mu_{t,i}$ and $\lim_{n\to\infty}\|Y_{t,i,h_n,n}^{\mathrm{O}}-Y_{t,i}\|_{L^2(TM)}=0$ in probability.

Theorem 6.2. Assume Assumption 4.1 (A1), Assumption 4.1 (A3), Assumption 4.2 (B1) and Assumption 4.3 hold. Denote $\lambda_{i,h,n}^{\mathrm{O}}$ to be the i-th eigenvalue of $h^{-1}(\widehat{T}_{h,1,n}^{\mathrm{O}}-1)$ with the associated eigenvector $X_{i,h,n}^{\mathrm{O}}$. Also denote $-\lambda_i$, where $\lambda_i>0$, to be the i-th eigenvalue of the Connection Laplacian ∇^2 with the associated eigen-vector field X_i. We assume that both $\lambda_{n,i}^{h}$ and λ_i increase as i increases, respecting the multiplicity. Fix $i\in\mathbb{N}$. Then there exists a sequence $h_n\to 0$ such that $\lim_{n\to\infty}\lambda_{i,h_n,n}^{\mathrm{O}}=\lambda_i$ and $\lim_{n\to\infty}\|X_{i,h_n,n}^{\mathrm{O}}-X_i\|_{L^2(TM)}=0$ in probability.

Before giving the proof, we make clear the role of ι. Suppose we know the embedded tangent plane $\iota_*T_x\mathrm{M}$ for all $x\in\mathrm{M}$ but not the embedding ι. Hence, when we estimate the translation group from the embedded frame bundle, the influence of the unknown ι cannot be removed. Denote the basis of the embedded tangent plane to be a $p\times d$ matrix Q_i. By [26, (B.68)], we are allowed to approximate the parallel transport from x_i to x_j from Q_i and Q_j:

$$/\!\!/_{x_j}^{x_i}X(x_j)\approx\iota_*^TQ_iQ_{ij}Q_j^T\iota_*X(x_j),$$

where $Q_{ij}=\mathrm{argmin}_{O\in O(d)}\|O-Q_i^TQ_j\|_{HS}$. Notice that we need Q_{ij} since in general $Q_i^TQ_j$ is not orthogonal due to the curvature. Furthermore, if the embedded tangent bundle information is also missing, we have to estimate it from the point cloud. Denote the estimated embedded tangent plane at $x\in\mathrm{M}$ by a $p\times d$ matrix O_x. It was shown in [26] that

$$O_x^T\iota_*X(x)\approx Q_i^T\iota_*X(x).$$

Proof of Theorem 6.1. We point out that in [26] we focus on the pointwise convergence so the error terms were simplified by the big O notations. To finish the proof, we have to trace these error terms. We recall the following results in [26]:

$$O_x^T\iota_*X(x)=Q_i^T\iota_*X(x)+O(h^{3/2}),\tag{56}$$

where $x\in\mathrm{M}$ and the $O(h^{3/2})$ term contains both the bias error and variance originating from the finite sample, and the constant solely depends on the curvatures of the manifold and their covariant derivatives. Indeed, the covariance matrix Ξ_i, $i=1,\ldots,n$, built up in the local PCA step is

$$\Xi_i=\frac{1}{n-1}\sum_{j\neq i}^{n}F_{i,j},$$

where $F_{i,j}$ are i.i.d random matrix of size $p\times p$

$$F_{i,j}=K\left(\frac{\|\iota(x_i)-\iota(x_j)\|_{\mathbb{R}^p}}{\sqrt{h_{\text{pca}}}}\right)(\iota(x_j)-\iota(x_i))(\iota(x_j)-\iota(x_i))^T,$$

so that its (k,l)-th entry

$$F_{i,j}(k,l)=K\left(\frac{\|\iota(x_i)-\iota(x_j)\|_{\mathbb{R}^p}}{\sqrt{h_{\text{pca}}}}\right)\langle\iota(x_j)-\iota(x_i),v_k\rangle\langle\iota(x_j)-\iota(x_i),v_l\rangle,$$

where v_l is the unit column vector with the l-th entry 1. Since $F_{i,j}$ are i.i.d. in j, we use F_i to denote the random matrix whose expectation is

$$\mathbb{E}F_i(k,l)=\int_{B_{\sqrt{h_{\text{pca}}}}(x_i)}K_{h_{\text{pca}}}(x_i,y)\langle\iota(y)-\iota(x_i),v_k\rangle\langle\iota(y)-\iota(x_i),v_l\rangle p(y)\mathrm{d}V(y),$$

where $0 < h_{\text{pca}} < h$. Fix $\varepsilon > 0$ and denote

$$p_{k,l}(n,\varepsilon) := \Pr\{|\Xi_i(k,l) - \mathbb{E}F_i(k,l)| > \varepsilon\}.$$

It has been shown in [26] that

$$p_{k,l}(n,\varepsilon) \leq \exp\left\{-\frac{(n-1)\varepsilon^2}{O(h_{\text{pca}}^{d/2+2}) + O(h_{\text{pca}})\varepsilon}\right\}.$$

when $k,l = 1,\ldots,d$;

$$p_{k,l}(n,\varepsilon) \leq \exp\left\{-\frac{(n-1)\varepsilon^2}{O(h_{\text{pca}}^{d/2+4}) + O(h_{\text{pca}}^2)\varepsilon}\right\},$$

when $k,l = d+1,\ldots,p$;

$$p_{k,l}(n,\varepsilon) \leq \exp\left\{-\frac{(n-1)\varepsilon^2}{O(h_{\text{pca}}^{d/2+3}) + O(h_{\text{pca}}^{3/2})\varepsilon}\right\},$$

for the other cases. Here the big O terms are bounded by R_M. Then, denote $\Omega_{n,\varepsilon_1,\varepsilon_2,\varepsilon_3}$ to be the event space that for all $i = 1,\ldots,n$, $|\Xi_i(k,l) - \mathbb{E}F_i(k,l)| < \varepsilon_1$ for all $k,l = 1,\ldots,d$, $|\Xi_i(k,l) - \mathbb{E}F_i(k,l)| < \varepsilon_2$ for all $k,l = d+1,\ldots,p$, $|\Xi_i(k,l) - \mathbb{E}F_i(k,l)| < \varepsilon_3$ for all $k = 1,\ldots,d$, $l = d+1,\ldots,p$ and $l = 1,\ldots,d$, $k = d+1,\ldots,p$. By a direct calculation we know the probability of $\Omega_{n,\varepsilon_1,\varepsilon_2,\varepsilon_3}$ is lower bounded by

$$\begin{aligned}1 - n\Big(d^2\exp\left\{-\frac{(n-1)\varepsilon_1^2}{O(h_{\text{pca}}^{d/2+2}) + O(h_{\text{pca}})\varepsilon_1}\right\} + (p-d)^2\exp\left\{-\frac{(n-1)\varepsilon_2^2}{O(h_{\text{pca}}^{d/2+3}) + O(h_{\text{pca}}^2)\varepsilon_2}\right\}\\ + p(p-d)\exp\left\{-\frac{(n-1)\varepsilon_3^2}{O(h_{\text{pca}}^{d/2+3}) + O(h_{\text{pca}}^{3/2})\varepsilon_3}\right\}\Big),\end{aligned}$$

which tends to 0 as $n \to \infty$. Thus, if we choose $\varepsilon_1 = \frac{\log(n)h_{\text{pca}}^{d/4+1}}{n^{1/2}}$, $\varepsilon_2 = \frac{\log(n)h_{\text{pca}}^{d/4+2}}{n^{1/2}}$ and $\varepsilon_3 = \frac{\log(n)h_{\text{pca}}^{d/4+3/2}}{n^{1/2}}$, with high probability,

$$O_i^T \iota_* X(x) = Q_i^T \iota_* X(x_i) + h_{\text{pca}}^{3/2} b_1 \iota_* X(x_i), \tag{57}$$

where $b_1 : \mathbb{R}^p \to \mathbb{R}^d$ is a bounded operator with the norm bounded by R_M. In addition to the bias introduced by ι and the curvature of M, b_1 is influenced by the finite sampling effect. Thus, conditional on $\Omega_{n,\varepsilon_1,\varepsilon_2,\varepsilon_3}$, O_i is deviated from Q_i up to order of $h_{\text{pca}}^{3/2}$.

Based on (57), conditional on $\Omega_{n,\varepsilon_1,\varepsilon_2,\varepsilon_3}$, we have [26, (B.76)]:

$$O_i^T O_i = Q_i^T Q_i + h^{3/2} b_2. \tag{58}$$

Here $b_2 : \mathbb{R}^d \to \mathbb{R}^d$ is a bounded operator with norm bounded by R_M. Note that since $h_{\text{pca}} < h$, the error introduced by local PCA step is absorbed in $h^{3/2}$. As a result, conditional on $\Omega_{n,\varepsilon_1,\varepsilon_2,\varepsilon_3}$ we have

$$\iota_*^T O_i O_{ij} B_i^T X(x_j) = /\!/_{x_j}^{x_i} X(x_j) + h^{3/2} b_3 X(x_j), \tag{59}$$

where $b_3 : T_{x_j}\mathrm{M} \to T_{x_i}\mathrm{M}$ is a bounded operator with norm bounded by R_M.

It should be emphasized that both O_i and O_{ij} are random in nature, and they are dependent to some extent. When conditional on $\Omega_{n,\varepsilon_1,\varepsilon_2,\varepsilon_3}$, the randomness is bounded and we are able to proceed.

Define operators $O_{\mathscr{X}} : V_{\mathscr{X}} \to T\mathrm{M}_{\mathscr{X}}$, $O_{\mathscr{X}}^T : T\mathrm{M}_{\mathscr{X}} \to V_{\mathscr{X}}$, $Q_{\mathscr{X}} : V_{\mathscr{X}} \to T\mathrm{M}_{\mathscr{X}}$ and $Q_{\mathscr{X}}^T : T\mathrm{M}_{\mathscr{X}} \to V_{\mathscr{X}}$ by

$$\begin{aligned}O_{\mathscr{X}}\boldsymbol{v} &:= [\iota_*^T O_1 \boldsymbol{v}[1], \ldots \iota_*^T O_n \boldsymbol{v}[n]] \in T\mathrm{M}_{\mathscr{X}},\\ O_{\mathscr{X}}^T\boldsymbol{w} &:= [(O_1^T \iota_* \boldsymbol{w}[1])^T, \ldots, (O_n^T \iota_* \boldsymbol{w}[n])^T]^T \in V_{\mathscr{X}},\\ Q_{\mathscr{X}}\boldsymbol{v} &:= [\iota_*^T Q_1 \boldsymbol{v}[1], \ldots \iota_*^T Q_n \boldsymbol{v}[n]] \in T\mathrm{M}_{\mathscr{X}},\\ Q_{\mathscr{X}}^T\boldsymbol{w} &:= [(Q_1^T \iota_* \boldsymbol{w}[1])^T, \ldots, (Q_n^T \iota_* \boldsymbol{w}[n])^T]^T \in V_{\mathscr{X}},\end{aligned}$$

where $\boldsymbol{w} \in T\mathrm{M}_{\mathscr{X}}$ and $\boldsymbol{v} \in V_{\mathscr{X}}$.

Note that $Q_{\mathscr{X}}\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}\boldsymbol{X}$ has been studied in Theorem 5.2, where $\boldsymbol{P}_{h,\alpha,n}$ is defined in (27) when the frame bundle information can be fully accessed. If we can control the difference between $Q_{\mathscr{X}}\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}\boldsymbol{X}$ and $O_{\mathscr{X}}\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}^{\mathrm{O}}\boldsymbol{X}$, by some modification of the proof of Theorem 5.2, we can conclude the Theorem. While conditional on $\Omega_{n,\varepsilon_1,\varepsilon_2,\varepsilon_3}$, by (58), for all $i = 1,\ldots,n$,

$$\left|Q_{\mathscr{X}}\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}\boldsymbol{X}[i] - O_{\mathscr{X}}\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}^{\mathrm{O}}\boldsymbol{X}[i]\right| = \left|\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\widehat{M}_{h,\alpha,n}(x_i,x_j)\big(/\!/_{x_j}^{x_i} - \iota_*^T O_i O_{ij} B_j^T\big)X(x_j)\right|$$
$$= h^{3/2}\left|\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\widehat{M}_{h,\alpha,n}(x_i,x_j)b_3X(x_j)\right| \leq R_M\frac{\|K\|_\infty^{2\alpha}}{\delta^{2\alpha}}\|X\|_\infty h^{3/2},$$

where the last inequality comes from $|\widehat{M}_{h,\alpha,n}(x,x_j)| \leq \dfrac{\|K\|_\infty^{2\alpha}}{\delta^{2\alpha}}$ for all $x \in \mathrm{M}$ and j,n (45).

Note that Step 1, Step 2 and Step 3 in the proof of Theorem 5.2 hold for $Q_{\mathscr{X}}\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}$. Suppose the eigenvalues of $O_{\mathscr{X}}\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}^{\mathrm{O}}$ and $Q_{\mathscr{X}}\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}$ are ordered in the descending way. As a result, conditional on $\Omega_{n,\varepsilon_1,\varepsilon_2,\varepsilon_3}$, by the perturbation theory, the i-th eigenvector of $O_{\mathscr{X}}\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}^{\mathrm{O}}$ is deviated from the i-th eigenvector of $Q_{\mathscr{X}}\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}$ by an error of order $h^{3/2}$, say, $C_ih^{3/2}$ for some $C_i > 0$.

Denote $\mu_{t,i,h}$ to be the i-th eigenvalue of $T_{h,1}^{t/h}$ with the associated eigenvector $Y_{t,i,h}$ and denote $\mu_{t,i,h,n}$ to be the i-th eigenvalue of $\widehat{T}_{h,1,n}^{t/h}$ with the associated eigenvector $Y_{t,i,h,n}$. Fix $i \in \mathbb{N}$ and take $h_0 > 0$ so that $\mu_{t,i} < h_0^{-1}$. By Step 1, we know that the i-th eigenvalue of $\widehat{T}_{h,1,n}^{\mathrm{O}}$ and $O_{\mathscr{X}}\boldsymbol{D}_{h,1,n}^{-1}\boldsymbol{P}_{h,1,n}^{O}$ are the same and their eigenvectors are related for all $h < h_0$. By Step 2, for all $h < h_0$, for any $\delta > 0$, we can find $1/\delta > 0$ so that for all $n > 1/\delta$

$$P\{\|Y_{t,i,h,n} - Y_{t,i,h}\|_{L^2(T\mathrm{M})} \geq \delta/2\} < \delta.$$

By step 3(A) (or Step 3(B) if $\partial\mathrm{M} \neq \emptyset$), for the given $\varepsilon > 0$, there exists $h(\varepsilon) > 0$ so that

$$\|Y_{t,i} - Y_{t,i,h(\varepsilon)}\|_{L^2(T\mathrm{M})} < \varepsilon/2.$$

By putting the above together, for any $\varepsilon > 0$ so that $h(\varepsilon) < h_0$, when $\delta < \varepsilon$ and $n > 1/\delta$, we have

$$\begin{aligned}
&P\{\|Y_{t,i,h(\varepsilon),n}^{\mathrm{O}} - Y_{t,i}\|_{L^2(T\mathrm{M})} \geq \varepsilon + C_ih(\varepsilon)^{3/2}\}\\
\leq &P\{\|Y_{t,i,h(\varepsilon),n}^{\mathrm{O}} - Y_{t,i,h(\varepsilon),n}\|_{L^2(T\mathrm{M})} + \|Y_{t,i,h(\varepsilon),n} - Y_{t,i}\|_{L^2(T\mathrm{M})} \geq \varepsilon + C_ih(\varepsilon)^{3/2}\}\\
\leq &P\{\|Y_{t,i,h(\varepsilon),n}^{\mathrm{O}} - Y_{t,i,h(\varepsilon),n}\|_{L^2(T\mathrm{M})} \geq C_ih(\varepsilon)^{3/2}\} + P\{\|Y_{t,i,h(\varepsilon),n} - Y_{t,i}\|_{L^2(T\mathrm{M})} \geq \varepsilon\}\\
\leq &1 - |\Omega_{n,\varepsilon_1,\varepsilon_2,\varepsilon_3}| + P\{\|Y_{t,i,h(\varepsilon),n} - Y_{t,i,h(\varepsilon)}\|_{L^2(T\mathrm{M})} \geq \varepsilon/2\}\\
\leq &1 - |\Omega_{n,\varepsilon_1,\varepsilon_2,\varepsilon_3}| + P\{\|Y_{t,i,h(\varepsilon),n} - Y_{t,i,h(\varepsilon)}\|_{L^2(T\mathrm{M})} \geq \delta/2\} < 1 - |\Omega_{n,\varepsilon_1,\varepsilon_2,\varepsilon_3}| + \delta.
\end{aligned}$$

Let $\delta \to 0$, or equivalently, $n \to \infty$, we conclude the convergence in probability by choosing h_n decreasing to 0, as $|\Omega_{n,\varepsilon_1,\varepsilon_2,\varepsilon_3}| \to 0$ when $n \to \infty$.

□

7. Acknowledgements

A. Singer was partially supported by Award Number DMS-0914892 from the NSF, by Award Number FA9550-09-1-0551 from AFOSR, by Award Number R01GM090200 from the NIGMS and by the Alfred P. Sloan Foundation. H.-T. Wu acknowledges support by AFOSR grant FA9550-09-1-0551, NSF grant CCF-0939370 and FRG grant DSM-1160319. H.-T. Wu thanks Afonso Bandeira for reading the first version of this manuscript.

REFERENCES

[1] A. S. Bandeira, A. Singer, and D. A. Spielman. A Cheeger Inequality for the Graph Connection Laplacian. *submitted*, 2013. arXiv:1204.3873 [math.SP].
[2] M. Belkin and P. Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural. Comput.*, 15(6):1373–1396, June 2003.
[3] M. Belkin and P. Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. In *Proceedings of the 18th Conference on Learning Theory (COLT)*, pages 486–500, 2005.
[4] M. Belkin and P. Niyogi. Convergence of Laplacian eigenmaps. In *Adv. Neur. In.: Proceedings of the 2006 Conference*, volume 19, page 129. The MIT Press, 2007.
[5] P. Bérard, G. Besson, and S. Gallot. Embedding Riemannian manifolds by their heat kernel. *Geom. Funct. Anal.*, 4:373–398, 1994.
[6] N. Berline, E. Getzler, and M. Vergne. *Heat Kernels and Dirac Operators*. Springer, 2004.
[7] R. L. Bishop and R. J. Crittenden. *Geometry of Manifolds*. Amer Mathematical Society, 2001.
[8] F. Chatelin. *Spectral Approximation of Linear Operators*. SIAM, 2011.
[9] M.-Y. Cheng and H.-T. Wu. Local linear regression on manifolds and its geometric interpretation. *submitted*, 2012. arXiv:1201.0327v3 [math.ST].
[10] R. R. Coifman and S. Lafon. Diffusion maps. *Appl. Comput. Harmon. Anal.*, 21(1):5–30, 2006.
[11] N. Dunford and J. T. Schwartz. *Linear operators, volume 1*. Wiley-Interscience, 1958.
[12] J. Frank. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. Oxford University Press, New York, 2nd edition, 2006.
[13] P. Gilkey. *The Index Theorem and the Heat Equation*. Princeton, 1974.
[14] M. Hein, J. Audibert, and U. von Luxburg. From graphs to manifolds - weak and strong pointwise consistency of graph Laplacians. In *Proceedings of the 18th Conference on Learning Theory (COLT)*, pages 470–485, 2005.
[15] S. A. Huettel, A. W. Song, and G. McCarthy. *Functional Magnetic Resonance Imaging*. Sinauer Associates, 2 edition, 2008.
[16] D.N. Kaslovsky and F.G. Meyer. Optimal tangent plane recovery from noisy manifold samples. *arXiv:1111.4601v1*, 2011.
[17] A.V. Little, Y.-M. Jung, and M. Maggioni. Multiscale estimation of intrinsic dimensionality of data sets. *Proc. AAAI*, 2009.
[18] J. Nash. Cˆ1 Isometric Imbeddings. *Annals of Mathematics*, 60(3):383–396, 1954.
[19] J. Nash. The Imbedding Problem for Riemannian Manifolds. *Annals of Mathematics*, 63(1):20–63, 1956.
[20] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. In *Twentieth Anniversary Volume:*, pages 1–23. Springer New York, 2009.
[21] M. Ovsjanikov, M. Ben-Chen, J. Solomon, A. Butscher, and L. Guibas. Functional Maps: A Flexible Representation of Maps Between Shapes. *ACM Transactions on Graphics*, 4(31), 2012.
[22] R. S. Palais. *Foundations of Global Non-linear Analysis*. W.A. Benjamin, Inc, 1968.
[23] H. Ronny and A. Singer. Representation Theoretic Patterns in Three-Dimensional Cryo-Electron Microscopy II The Class Averaging Problem. *Foundations of Computational Mathematics*, 11(5):589–616, 2011.
[24] A. Singer. From graph to manifold Laplacian: The convergence rate. *Appl. Comput. Harmon. Anal.*, 21(1):128–134, 2006.
[25] A. Singer and H.-T. Wu. Orientability and diffusion map. *Applied and Computational Harmonic Analysis*, 31(1):44–58, 2011.
[26] A. Singer and H.-T. Wu. Vector diffusion maps and the connection Laplacian. *Comm. Pure Appl. Math.*, 65(8):1067–1144, 2012.
[27] A. Singer, Z. Zhao, Y. Shkolnisky, and R. Hadani. Viewing angle classification of cryo-electron microscopy images using eigenvectors. *SIAM J. Imaging Sci.*, 4(2):723–759, 2011.
[28] S. Sternberg. *Lectures on Differential Geometry*. American Mathematical Society, 1999.
[29] A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes*. Springer-Verlag, 1996.
[30] U. von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, 36(2):555–586, April 2008.

A.1. AN INTRODUCTION TO PRINCIPAL BUNDLE

In this appendix, we collect a few relevant and self-contained facts about the mathematical framework *principal bundle* which are used in the main text. We refer the readers to, for example [7, 6], for more general definitions which are not used in this paper.

We start from discussing the notion of *group action*, *orbit* and *orbit space*. Consider a set Y and a group G with the identity element e. The left group action of G on Y is a map from $G \times Y$ onto Y

$$G \times Y \to Y, \quad (g,x) \mapsto g \circ x \tag{A.1}$$

so that $(gh) \circ x = g \circ (h \circ x)$ is satisfied for all $g,h \in G$ and $x \in Y$ and $e \circ x = x$ for all x. The right group action can be defined in the same way. Note that we can construct a right action by composing with the inverse group operation, so in some scenarios it is sufficient to discuss only left actions. There are several types of group action. We call an action *transitive* if for any $x,y \in Y$, there exists a $g \in G$ so that $g \circ x = y$. In other words, under the group action we can jump between any pair of two points on Y, or $Y = G \circ x$ for any $x \in Y$. We call an action *effective* is for any $g,h \in G$, there exists x so that $g \circ x \neq h \circ x$. In other words, different group elements induce different permutations of Y. We call an action *free* if $g \circ x = x$ implies $g = e$ for all g. In other words, there is no fixed points under the G action, and hence the name free. If Y is a topological space, we call an action *totally discontinuous* if for every $x \in Y$, there is an open neighborhood U such that $(g \circ U) \cap U = \emptyset$ for all $g \in G$, $g \neq e$.

The *orbit* of a point $x \in Y$ is the set

$$Gx := \{g \circ x; g \in G\}.$$

The group action induces an equivalence relation. We say $x \sim y$ if and only if there exists $g \in G$ so that $g \circ x = y$ for all pairs of $x, y \in Y$. Clearly the set of orbits form a partition of Y, and we denote the set of all orbits as $Y/\sim$ or Y/G. We can thus define a projection map π by

$$Y \to Y/G, \quad x \mapsto Gx.$$

We call Y the *total space* or the *left G-space*, G the *structure group*, Y/G the *quotient space*, the *base space* or the *orbit space* of Y under the action of G and π the *canonical projection*.

We define a principal bundle as a special G-space which satisfies more structure. Note that the definitions given here are not the most general ones but are enough for our purpose.

Definition A.1.1 (Fibre bundle)**.** Let $\mathscr{F}$ and M be two smooth manifolds and π a smooth map from $\mathscr{F}$ to M. We say that $\mathscr{F}$ is a *fibre bundle* with fibre F over M if there is an open covering of M, denoted as $\{U_i\}$, and diffeomorphisms $\{\psi_i : \pi^{-1}(U_i) \to U_i \times F\}$ so that $\pi : \pi^{-1}(U_i) \to U_i$ is the composition of ψ_i with projection onto U_i.

By definition, $\pi^{-1}(x)$ is diffeomorphic to F for all $x \in$ M. We call $\mathscr{F}$ the total space of the fibre bundle, M is the base space, π the canonical projection, and F the fibre of $\mathscr{F}$. With the above algebraic setup, in a nutshell, the principal bundle is a special fibre bundle accompanied by a group action.

Definition A.1.2 (Principal bundle)**.** Let M be a smooth manifold and G a Lie group. A *principal bundle over M with structure group G* is a fibre bundle $P(\mathrm{M}, G)$ with fiber diffeomorphic to G, a smooth right action of G, denoted as $\circ$, on the fibres and a canonical projection $\pi : P \to \mathrm{M}$ so that

(1) π is smooth and $\pi(g \circ p) = \pi(p)$ for all $p \in P$ and $g \in G$;
(2) G acts freely and transitively;
(3) the diffeomorphism $\psi_i : \pi^{-1}(U_i) \to U_i \times G$ satisfies $\psi_i(p) = (\pi(p), \phi_i(p)) \in U_i \times G$ such that $\phi_i : \pi^{-1}(U_i) \to G$ satisfying $\phi_i(pg) = \phi_i(p)g$ for all $p \in \pi^{-1}(U_i)$ and $g \in G$.

Note that $\mathrm{M} = P(\mathrm{M}, G)/G$, where the equivalence relation is induced by G. From the view point of orbit space, $P(\mathrm{M}, G)$ is the total space, G is the structure group, and M is the orbit space of $P(\mathrm{M}, G)$ under the action of G. Intuitively, $P(\mathrm{M}, G)$ is composed of a bunch of sets diffeomorphic to G, all of which are pulled together under some rules.[5] We give some examples here:

Example. Consider $P(\mathrm{M}, G) = \mathrm{M} \times G$ so that G acts by $g \circ (x, h) = (x, hg)$ for all $(x, h) \in \mathrm{M} \times G$ and $g \in G$. We call such principal bundle *trivial*.

Example. A particular important example of the principal bundle is the *frame bundle*, denoted as $GL(\mathrm{M})$, which is the principal $GL(d, \mathbb{R})$-bundle with the base manifold a d-dim smooth manifold M. We construct $GL(\mathrm{M})$ for the purpose of completeness. Denote B_x to be the set of bases of the tangent space $T_x(\mathrm{M})$, that is, $B_x \cong GL(d, \mathbb{R})$ and $u_x \in B_x$ is a basis of $T_x\mathrm{M}$. Notice that we can view a basis of $T_x\mathrm{M}$ as an invertible linear map from $\mathbb{R}^d$ to $T_x\mathrm{M}$. Let $GL(\mathrm{M})$ be a set consisting of all bases at all points of M, that is, $GL(\mathrm{M}) := \{B_x; x \in \mathrm{M}\}$. Let $\pi : GL(\mathrm{M}) \to \mathrm{M}$ by $u_x \mapsto x$ for all $u_x \in B_x$ and $x \in$ M. Define the $GL(d, \mathbb{R})$ action of $L(\mathrm{M})$ by $g \circ u_x = v_x$, where $g = [g_{ij}]_{i,j=1}^d \in GL(d, \mathbb{R})$, $u_x = (X_1, \ldots, X_d) \in B_x$ and $v_x = (Y_1, \ldots, Y_d) \in B_x$ with $Y_i = \sum_{j=1}^d g_{ij} X_j$. By a direct calculation, $GL(d, \mathbb{R})$ acts on $GL(\mathrm{M})$ freely and transitively, and $\pi(g \circ u_x) = \pi(u_x)$. In a coordinate neighborhood U, $\pi^{-1}(U)$ is 1-1 corresponding with $U \times GL(d, \mathbb{R})$, which induces a differentiable structure on $GL(\mathrm{M})$. Thus $GL(\mathrm{M})$ is a principal $GL(d, \mathbb{R})$-bundle.

[5] These rules are referred to as *transition functions*.

If we are given a left G-space F, we can form a fibre bundle from $P(\mathrm{M},G)$ so that its fibre is diffeomorphic to F and its base manifold is M in the following way. By denoting the left G action on F by $\cdot$, we have

$$\mathscr{E}(P(\mathrm{M},G),\cdot,F) := P(\mathrm{M},G)\times_G F := P(\mathrm{M},G)\times F/G,$$

where the equivalence relation is defined as

$$(g\circ p, g^{-1}\cdot f)\sim(p,f)$$

for all $p\in P(\mathrm{M},G)$, $g\in G$ and $f\in F$. The canonical projection from $\mathscr{E}(P(\mathrm{M},G),\cdot,F))$ to M is denoted as $\pi_{\mathscr{E}}$:

$$\pi_{\mathscr{E}} : (p,f)\mapsto \pi(p),$$

for all $p\in P(\mathrm{M},G)$ and $f\in F$. We call $\mathscr{E}(P(\mathrm{M},G),\cdot,F)$ the *fibre bundle associated with $P(M,G)$ with standard fibre F* or the *associated fibre bundle* whose differentiable structure is induced from M. Given $p\in P(\mathrm{M},G)$, denote pf to be the image of $(p,f)\in P(\mathrm{M},G)\times F$ onto $\mathscr{E}(P(\mathrm{M},G),\cdot,F)$. By definition p is a diffeomorphism from F to $\pi_{\mathscr{E}}^{-1}(\pi(p))$ and

$$(g\circ p)f = p(g\cdot f).$$

Note that the associated fibre bundle $\mathscr{E}(P(\mathrm{M},G),\cdot,F)$ is a special fibre bundle and its fibre is diffeomorphic to F. When there is no danger of confusion, we denote $\mathscr{E} := \mathscr{E}(P(\mathrm{M},G),\cdot,F)$ to simply the notation.

Example. When $F=V$ is a vector space and the left G action on F is a linear representation, the associated fibre bundle is called the *vector bundle associated with the principal bundle $P(M,G)$ with fiber V*, or simply called the *vector bundle* if there is no danger of confusion. For example, take $F=\mathbb{R}^{m'}$, denote ρ to be a representation of G into $GL(m',\mathbb{R})$ and assume G acts on $\mathbb{R}^{m'}$ via the representation ρ. A particular example of interest is the tangent bundle $T\mathrm{M} := \mathscr{E}(P(\mathrm{M},GL(d,\mathbb{R})),\rho,\mathbb{R}^d)$ when M is a d-dim smooth manifold and the representation ρ is identity. The practical meaning of the frame bundle and its associated tangent bundle is change of coordinate. That is, if we view a point $u_x\in GL(\mathrm{M})$ as the basis of the fiber $T_x\mathrm{M}$, where $x=\pi(u_x)$, then the coordinate of a point on the tangent plane $T_x\mathrm{M}$ changes, that is, $v_x\to g\cdot v_x$ where $g\in GL(d,\mathbb{R})$ and $v_x\in\mathbb{R}^d$, according to the changes of the basis, that is, $g\to g\circ u_x$.

A (global) *section* of a fibre bundle $\mathscr{E}$ with fibre F over M is a smooth map

$$s:\mathrm{M}\to\mathscr{E}$$

so that $\pi(s(x))=x$ for all $x\in\mathrm{M}$. We denote $\Gamma(\mathscr{E})$ to be the set of sections; $C^l(\mathscr{E})$ to be the space of all sections with the l-th regularity, where $l\geq 0$. An important property of the principal bundle is that a principal bundle is trivial if and only if $C^0(P(\mathrm{M},G))\neq\emptyset$. In other words, all sections on a non-trivial principal bundle are discontinuous. On the other hand, there always exists a continuous section on the associated vector bundle $\mathscr{E}$.

Let V be a vector space. Denote $GL(V)$ to be the group of all invertible linear maps on V. If V comes with an inner product, then define $O(V)$ to be the group of all orthogonal maps on V with related to the inner product. From now on we focus on the vector bundle with fiber being a vector space V and the action $\cdot$ being a representation $\rho: G\to GL(V)$, that is, $\mathscr{E}(P(\mathrm{M},G),\rho,V)$.

To introduce the notion of *covariant derivative* on the vector bundle $\mathscr{E}$, we have to introduce the notion of *connection*. Denote $T\mathscr{E}$ to be the tangent bundle of the fibre bundle $\mathscr{E}$ and $T^*\mathscr{E}$ to be the cotangent bundle of $\mathscr{E}$. We call tangent vector X on $\mathscr{E}$ *vertical* if it is tangential to the fibers, that is, $X(\pi_{\mathscr{E}}^* f)=0$ for all $f\in C^\infty(\mathrm{M})$. Denote the bundle of vertical vectors as $V\mathscr{E}$, which is referred to as *the vertical bundle*, and is a subbundle of $T\mathscr{E}$. We call a vector field vertical if it is a section of the vertical bundle. Clearly the quotient of $T\mathscr{E}$ by its subbundle $V\mathscr{E}$ is isomorphic to $\pi^*T\mathrm{M}$, and hence we have a short exact sequence of vector bundles:

$$0\to V\mathscr{E}\to T\mathscr{E}\to \pi^*T\mathrm{M}\to 0. \tag{A.2}$$

However, there is no canonical splitting of this short exact sequence. A chosen splitting is called a *connection*.

Definition A.1.3 (Connection 1-form)**.** Let $P(\mathrm{M},G)$ be a principal bundle. A connection 1-form on $P(\mathrm{M},G)$ is an one form $\omega\in\Gamma(T^*P(\mathrm{M},G)\otimes VP(\mathrm{M},G))$ so that $\omega(X)=X$ for any $X\in\Gamma(VP(\mathrm{M},G))$ and is invariant under the action of G. The kernel of ω is called the *horizontal bundle* and is denoted as $HP(\mathrm{M},G)$

Note that $HP(\mathrm{M},G)$ is isomorphic to $\pi^*T\mathrm{M}$. In other words, a connection 1-form determines a splitting of (A.2), or the connection on $P(\mathrm{M},G)$. We call a section X_P of $HP(\mathrm{M},G)$ a *horizontal vector field*. Given $X\in\Gamma(T\mathrm{M})$, we say that X_P is the *horizontal lift* with respect to the connection on $P(\mathrm{M},G)$ of X if $X=\pi_*X_P$. Given a smooth curve $\tau:=c_t$, $t\in[0,1]$ on M and a point $u_0\in P(\mathrm{M},G)$, we call a curve $\tau^*=u_t$ on $P(\mathrm{M},G)$ the *(horizontal) lift of* c_t if the vector tangent to u_t is horizontal and $\pi(u_t)=c_t$ for $t\in[0,1]$. The existence of τ^* is an important property of the connection theory. We call u_t the *parallel displacement* of u_0 along the curve τ on M.

With the connection on $P(\mathrm{M},G)$, the connection on an associated vector bundle $\mathscr{E}$ with fiber V is determined. As a matter of fact, we define the connection, or $H\mathscr{E}$, to be the image of $HP(\mathrm{M},G)$ under the natural projection $P(\mathrm{M},G)\times V\to\mathscr{E}(P(\mathrm{M},G),\rho,V)$. Similarly, we call a section $X_{\mathscr{E}}$ of $H\mathscr{E}$ a *horizontal vector field*. Given $X\in\Gamma(T\mathrm{M})$, we say that $X_{\mathscr{E}}$ is the *horizontal lift* with respect to the connection on $\mathscr{E}$ of X if $X=\pi_{\mathscr{E}*}X_{\mathscr{E}}$. Given a smooth curve c_t, $t\in[0,1]$ on M and a point $v_0\in\mathscr{E}$, we call a curve v_t on $\mathscr{E}$ the *(horizontal) lift of* c_t if the vector tangent to v_t is horizontal and $\pi_{\mathscr{E}}(v_t)=c_t$ for $t\in[0,1]$. The existence of such horizontal life holds in the same way as that of the principal bundle. We call v_t the *parallel displacement* of v_0 along the curve τ on M. Note that we have interest in this connection on the vector bundle since it leads to the *covariant derivative* we have interest.

Definition A.1.4 (Covariant Derivative)**.** Take a vector bundle $\mathscr{E}$ associated with the principal bundle $P(\mathrm{M},G)$ with fiber V. The covariant derivative $\nabla^{\mathscr{E}}$ of a smooth section $X\in C^1(\mathscr{E})$ at $x\in\mathrm{M}$ in the direction $\dot{c}_0$ is defined as

$$\nabla^{\mathscr{E}}_{\dot{c}_0}X=\lim_{h\to 0}\frac{1}{h}[/\!/_{c_h}^{c_0}X(c_h)-X(x)], \tag{A.3}$$

where $c:[0,1]\to\mathrm{M}$ is a curve on M so that $c_0=x$ and $/\!/_{c_h}^{c_0}$ denotes the parallel displacement of X from c_h to c_0

Note that in general although all fibers of $\mathscr{E}$ are isomorphic to V, the notion of comparison among them is not provided. An explicit example demonstrating the derived problem is given in the appendix of [26]. However, with the parallel displacement based on the notion of connection, we are able to compare among fibers, and hence define the derivative. With the fact that

$$/\!/_{c_h}^{c_0}X(c_h)=u_0u_h^{-1}X(c_h), \tag{A.4}$$

where u_h is the horizontal lift of c_h to $P(\mathrm{M},G)$ so that $\pi(u_0)=x$, the covariant derivative (A.3) can be represented in the following format:

$$\nabla^{\mathscr{E}}_{\dot{c}_0}X=\lim_{h\to 0}\frac{1}{h}[u_0u_h^{-1}(X(c_h))-X(c_0)], \tag{A.5}$$

which is independent of the choice of u_0. To show (A.4), set $v:=u_h^{-1}(X(c_h))\in V$. Clearly $u_t(v)$, $t\in[0,h]$, is a horizontal curve in $\mathscr{E}$ by definition. It implies that $u_0v=u_0u_h^{-1}(X(c_h))$ is the parallel displacement of $X(c_h)$ along c_t from c_h to c_0. Thus, although the covariant derivatives defined in (9) and (A.5) are different in their appearances, they are actually equivalent. We can understand this definition in the frame bundle $GL(\mathrm{M}^d)$ and its associated tangent bundle. First, we find the coordinate of a point on the fiber $X(c_h)$, which is denoted as $u_h^{-1}(X(c_h))$, and then we put this coordinate $u_h^{-1}(X(c_h))$ to $x=c_0$ and map it back to the fiber $T_x\mathrm{M}$ by the basis u_0. In this way we can compare two different "abstract fibers" by comparing their coordinates. A more abstract definition of the covariant derivative, yet equivalent to the aboves, is the following. A covariant derivative of $\mathscr{E}$ is a differential operator

$$\nabla^{\mathscr{E}}:C^\infty(\mathscr{E})\to C^\infty(T^*\mathrm{M}\otimes\mathscr{E}) \tag{A.6}$$

so that the Leibniz's rule is satisfied, that is, for $X\in C^\infty(\mathscr{E})$ and $f\in C^\infty(\mathrm{M})$, we have

$$\nabla^{\mathscr{E}}(fX)=df\otimes X+f\nabla^{\mathscr{E}}X,$$

where d is the exterior derivative on M. Denote Λ^kT^*M (resp ΛT^*M) to be the bundle of k-th exterior differentials (resp. the bundle of exterior differentials), where $k\geq 1$. Given two vector bundles $\mathscr{E}_1$ and $\mathscr{E}_2$ on M with the covariant derivatives $\nabla^{\mathscr{E}_1}$ and $\nabla^{\mathscr{E}_2}$, we construct a covariant derivative on $\mathscr{E}_1\otimes\mathscr{E}_2$ by

$$\nabla^{\mathscr{E}_1\otimes\mathscr{E}_2}:=\nabla^{\mathscr{E}_1}\otimes 1+1\otimes\nabla^{\mathscr{E}_2}. \tag{A.7}$$

A fiber metric $g^{\mathscr{E}}$ in a vector bundle $\mathscr{E}$ is a positive-definite inner-product in each fiber V that varies smoothly on M. For any $\mathscr{E}$, if M is paracompact, $g^{\mathscr{E}}$ always exists. A connection in $P(\mathrm{M},G)$, and also its associated vector bundle $\mathscr{E}$, is called metric if

$$dg^{\mathscr{E}}(X_1,X_2) = g^{\mathscr{E}}(\nabla^{\mathscr{E}}X_1,X_2)+g^{\mathscr{E}}(X_1,\nabla^{\mathscr{E}}X_2),$$

for all $X_1,X_2\in C^\infty(\mathscr{E})$. We mainly focus on metric connection in this work. It is equivalent to say that the parallel displacement of $\mathscr{E}$ preserves the fiber metric. An important fact about the metric connection is that if a connection on $P(\mathrm{M},G)$ is metric given a fiber metric $g^{\mathscr{E}}$, than the covariant derivative on the associated vector bundle $\mathscr{E}$ can be equally defined from a sub-bundle $Q(\mathrm{M},H)$ of $P(\mathrm{M},G)$, which is defined as

$$Q(\mathrm{M},H) := \{p\in P(\mathrm{M},G) : g^{\mathscr{E}}(p(u),p(v)) = (u,v)\}, \tag{A.8}$$

where $(\cdot,\cdot)$ is an inner product on V and the structure group H is a closed subgroup of G. In other words, $p\in Q(\mathrm{M},H)$ is a linear map from V to $\pi_{\mathscr{E}}^{-1}(\pi(p))$ which preserves the inner product. A direct verification shows that the structure group of $Q(\mathrm{M},H)$ is

$$H := \{g\in G : \rho(g)\in O(V)\}\subset G. \tag{A.9}$$

Since orthogonal property is needed in our analysis, when we work with a metric connection on a principal bundle $P(\mathrm{M},G)$ given a fiber metric $g^{\mathscr{E}}$ on $\mathscr{E}(P(\mathrm{M},G),\rho,V)$, we implicitly assume we work with its sub bundle $Q(\mathrm{M},H)$. With the covariant derivative, we now define the Connection Laplacian. Assume M is a d-dim smooth Riemmanian manifold with the metric g. With the metric g we have an induced measure on M, denoted as $\mathrm{d}V$.[6] Denote $L^p(\mathscr{E})$, $1\le p<\infty$ to be the set of L^p integrable sections, that is, $X\in L^p(\mathscr{E})$ iff

$$\int |g_x^{\mathscr{E}}(X(x),X(x))|^{p/2}\mathrm{d}V(x)<\infty.$$

Denote $\mathscr{E}^*$ to be the dual bundle of $\mathscr{E}$, which is paired with $\mathscr{E}$ by $g^{\mathscr{E}}$, that is, the pairing between $\mathscr{E}$ and $\mathscr{E}^*$ is $\langle X,Y\rangle := g^{\mathscr{E}}(X,Y)$, where $X\in C^\infty(\mathscr{E})$ and $Y\in C^\infty(\mathscr{E}^*)$. The connection on the dual bundle $\mathscr{E}^*$ is thus defined by

$$d\langle X,Y\rangle = g^{\mathscr{E}}(\nabla^{\mathscr{E}}X,Y)+g^{\mathscr{E}}(X,\nabla^{\mathscr{E}^*}Y).$$

Recall that the Riemannian manifold (M,g) possesses a canonical connection referred to as the Levi-Civita connection ∇. Based on ∇ we define the connection $\nabla^{T^*M\otimes\mathscr{E}}$ on the tensor product bundle $T^*M\otimes\mathscr{E}$.

Definition A.1.5. Take the Riemannian manifold (M,g), the vector bundle $\mathscr{E} := \mathscr{E}(P(\mathrm{M},G),\rho,V)$ and its connection $\nabla^{\mathscr{E}}$. The Connection Laplacian on $\mathscr{E}$ is defined as $\nabla^2 : C^\infty(\mathscr{E})\to C^\infty(\mathscr{E})$ by

$$\nabla^2 := -\mathrm{tr}(\nabla^{T^*M\otimes\mathscr{E}}\nabla^{\mathscr{E}}),$$

where $\mathrm{tr} : C^\infty(T^*M\otimes T^*M\otimes\mathscr{E})\to C^\infty(\mathscr{E})$ by contraction with the metric g.

If we take the normal coordinate $\{\partial_i\}_{i=1}^d$ around $x\in M$, for $X\in C^\infty(\mathscr{E})$, we have

$$\nabla^2X(x) = -\sum_{i=1}^d \nabla_{\partial_i}\nabla_{\partial_i}X(x).$$

Given compactly supported smooth sections $X,Y\in C^\infty(\mathscr{E})$, a direct calculation leads to

$$\mathrm{tr}\big[\nabla(g^{\mathscr{E}}(\nabla^{\mathscr{E}}X,Y))\big] = \mathrm{tr}\big[g^{\mathscr{E}}(\nabla^{T^*M\otimes\mathscr{E}}\nabla^{\mathscr{E}}X,Y)+g^{\mathscr{E}}(\nabla^{\mathscr{E}}X,\nabla^{\mathscr{E}}Y)\big] = g^{\mathscr{E}}(\nabla^2X,Y)+\mathrm{tr}g^{\mathscr{E}}(\nabla^{\mathscr{E}}X,\nabla^{\mathscr{E}}Y).$$

By the divergence theorem, the left hand side disappears after integration over M, and we obtain $\nabla^2 = -\nabla^{\mathscr{E}*}\nabla^{\mathscr{E}}$. Similarly we can show that ∇^2 is self-adjoint. We refer the readers to [13] for further properties of ∇^2, for example the ellipticity, its heat kernel, and its application to the index theorem.

[6]To obtain the most geometrically invariant formulations, we may consider the density bundles as is considered in [6, Chapter 2]. We choose not to do that in order to simplify the discussion.

A.2. [PROOF OF PROPOSITION 5.1]

The proof is a direct generalization of that of [26, Theorem B.3] to the principal bundle structure. Note that in [26, Theorem B.3] only the uniform sampling p.d.f. case was discussed. Here we offer the proof for the nonuniform p.d.f. in order to trace the error term in the proof of the spectral convergence theorem.

Proof of Proposition 5.1. Take $x_i \in \mathrm{M}$ and $u_i \in P(\mathrm{M},G)$ so that $\pi(u_i) = x_i$. By definition we have

$$(\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}\boldsymbol{X})[i] = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} \widehat{M}_{h,\alpha,n}(x_i,x_j)g_{ij}\boldsymbol{X}[j] = \frac{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\frac{K_h(x_i,x_j)}{\widehat{p}_{h,n}(x_j)^{\alpha}}g_{ij}\boldsymbol{X}[j]}{\frac{1}{n-1}\sum_{l=1,l\neq i}^{n}\frac{K_h(x_i,x_l)}{\widehat{p}_{h,n}(x_l)^{\alpha}}}.$$

To evaluate the right hand side, we expand it as

$$\text{(A.10)} \qquad \frac{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\frac{K_h(x_i,x_j)}{p_h(x_j)^{\alpha}}g_{ij}\boldsymbol{X}[j]}{\frac{1}{n-1}\sum_{l=1,l\neq i}^{n}\frac{K_h(x_i,x_l)}{p_h(x_l)^{\alpha}}}$$

$$\text{(A.11)} \qquad +\frac{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}(\widehat{p}_{h,n}(x_j)^{-\alpha}-p_h(x_j)^{-\alpha})K_h(x_i,x_j)g_{ij}\boldsymbol{X}[j]}{\frac{1}{n-1}\sum_{l=1,l\neq i}^{n}\frac{K_h(x_i,x_l)}{p_h(x_l)^{\alpha}}}$$

$$\text{(A.12)} \qquad +\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\frac{K_h(x_i,x_j)}{\widehat{p}_{h,n}(x_j)^{\alpha}}g_{ij}\boldsymbol{X}[j]\left(\frac{1}{\frac{1}{n-1}\sum_{l=1,l\neq i}^{n}\frac{K_h(x_i,x_l)}{\widehat{p}_{h,n}(x_l)^{\alpha}}}-\frac{1}{\frac{1}{n-1}\sum_{l=1,l\neq i}^{n}\frac{K_h(x_i,x_l)}{p_h(x_l)^{\alpha}}}\right).$$

The (A.10) term can be handled by the same manner as that in [26]. Indeed, we expect the following approximation:

$$\frac{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\frac{K_h(x_i,x_j)}{p_h(x_j)^{\alpha}}g_{ij}\boldsymbol{X}[j]}{\frac{1}{n-1}\sum_{l=1,l\neq i}^{n}\frac{K_h(x_i,x_l)}{p_h(x_l)^{\alpha}}} \approx u_i^{-1}T_{h,\alpha}X(x_i),$$

since $\frac{K_h(x_i,x_j)}{p_h(x_j)^{\alpha}}g_{ij}\boldsymbol{X}[j]$ are i.i.d and $\frac{K_h(x_i,x_l)}{p_h(x_l)^{\alpha}}$ are i.i.d.. Thus, the trick in [24] can be applied and we get

$$\Pr\left\{\left\|\frac{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\frac{K_h(x_i,x_j)}{p_h(x_j)^{\alpha}}g_{ij}\boldsymbol{X}[j]}{\frac{1}{n-1}\sum_{l=1,l\neq i}^{n}\frac{K_h(x_i,x_l)}{p_h(x_l)^{\alpha}}}-u_i^{-1}T_{h,\alpha}X(x_i)\right\|>\varepsilon\right\} \leq C_1\exp\left\{\frac{-C_2(n-1)\varepsilon^2h^{d/2}}{2h\left[\|\nabla|_{y=x_i}/\!\!/_y^{x_i}X(y)\|^2+O(h)\right]}\right\},$$

where $\varepsilon > 0$ and C_1 and C_2 are some constants (related to d and vol(M)). Denote $\Omega_{\varepsilon,1}$ to be the event space so that $\left\|\frac{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\frac{K_h(x_i,x_j)}{p_h(x_j)^{\alpha}}g_{ij}\boldsymbol{X}[j]}{\frac{1}{n-1}\sum_{l=1,l\neq i}^{n}\frac{K_h(x_i,x_l)}{p_h(x_l)^{\alpha}}}-u_i^{-1}T_{h,\alpha}X(x_i)\right\| \leq \varepsilon$ holds for all $i = 1,\ldots,n$. Note that since there are n points,

$$|\Omega_{\varepsilon,1}| > 1-C_1 n\exp\left\{\frac{-C_2(n-1)\varepsilon^2h^{d/2}}{2h\left[\|\nabla|_{y=x_i}/\!\!/_y^{x_i}X(y)\|^2+O(h)\right]}\right\}.$$

To bound (A.11), we have to prepare some bounds. Fix j. Since $\widehat{p}_{h,n}(x_j) := \frac{1}{n-1}\sum_{k=1,k\neq j}^{n}K_h(x_j,x_k)$, and the random variables $K_h(x_j,x_k)$ are i.i.d, we simply apply the large deviation theory and get

$$\Pr\left\{\left|\widehat{p}_{h,n}(x_j)-p_h(x_j)\right|>\varepsilon\right\} \leq C_3\exp\left\{\frac{-C_4(n-1)\varepsilon^2h^{d/2}}{2h\left[\|\nabla|_{y=x_i}p_h(y)\|^2+O(h)\right]}\right\},$$

where C_3 and C_4 are some constants. This leads to

$$\text{(A.13)} \qquad \Pr\left\{\left|\widehat{p}_{h,n}(x_j)-p_h(x_j)\right|>\varepsilon \text{ for all } j=1,\ldots,n\right\} \leq C_3 n\exp\left\{\frac{-C_4(n-1)\varepsilon^2h^{d/2}}{2h\left[\|\nabla|_{y=x_i}p_h(y)\|^2+O(h)\right]}\right\}.$$

Denote $\Omega_{\varepsilon,2}$ to be the event space so that $\left|\widehat{p}_{h,n}(x_j)-p_h(x_j)\right|\leq\varepsilon$ for all $j=1,\ldots,n$. Note that

$$|\Omega_{\varepsilon,2}|>1-C_3 n\exp\left\{\frac{-C_4(n-1)\varepsilon^2 h^{d/2}}{2h\,[\|\nabla|_{y=x_i}p_h(y)\|^2+O(h)]}\right\}.$$

Also, by Assumption 4.3 and (42), we have for all $x\in\mathrm{M}$

$$\begin{aligned}\delta&\leq|p_h(x)|\leq\|K\|_{L^\infty}\\ \delta&\leq|\widehat{p}_{h,n}(x)|\leq\|K\|_{L^\infty}.\end{aligned}\tag{A.14}$$

Furthermore, under $\Omega_{\alpha,2}$, by Taylor's expansion and (A.14) we have

$$|\widehat{p}_{h,n}(x_i)^{-\alpha}-p_h(x_i)^{-\alpha}|\leq\alpha|\widehat{p}_{h,n}(x_i)-p_h(x_i)|\delta^{-\alpha-1}\leq\alpha\delta^{-\alpha-1}\varepsilon.$$

With these bounds, under $\Omega_{\varepsilon,2}$ (A.11) is bounded by $\alpha\delta^{-\alpha-2}\|K\|_\infty^{1+\alpha}\|X\|_\infty\varepsilon$.

For (A.12), by the same argument, under $\Omega_{\varepsilon,2}$ we have the following bound:

$$\left|\frac{1}{\frac{1}{n-1}\sum_{l=1,l\neq i}^n\frac{K_h(x_i,x_l)}{\widehat{p}_{h,n}(x_l)^\alpha}}-\frac{1}{\frac{1}{n-1}\sum_{l=1,l\neq i}^n\frac{K_h(x_i,x_l)}{p_h(x_l)^\alpha}}\right|=\left|\frac{\frac{1}{n-1}\sum_{l=1,l\neq i}^n K_h(x_i,x_l)\left(\frac{1}{p_h(x_l)^\alpha}-\frac{1}{\widehat{p}_{h,n}(x_l)^\alpha}\right)}{\frac{1}{n-1}\sum_{l=1,l\neq i}^n\frac{K_h(x_i,x_l)}{\widehat{p}_{h,n}(x_l)^\alpha}\frac{1}{n-1}\sum_{l=1,l\neq i}^n\frac{K_h(x_i,x_l)}{p_h(x_l)^\alpha}}\right|\leq\alpha\|K\|_\infty^{1+2\alpha}\delta^{-\alpha-3}\varepsilon.$$

Hence, (A.12) is bounded by $\alpha\delta^{-2\alpha-3}\|K\|_\infty^{2+2\alpha}\|X\|_\infty\varepsilon$ under $\Omega_{\varepsilon,2}$.

Putting the above together, under $\Omega_{\varepsilon,1}\cap\Omega_{\varepsilon,2}$, $(\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}\boldsymbol{X})[i]$ deviates from $u_i^{-1}T_{h,\alpha}X(x_i)$ by an error bounded by $\varepsilon(1+\alpha\delta^{-\alpha-2}\|K\|_\infty^{1+\alpha}\|X\|_\infty+\alpha\delta^{-2\alpha-3}\|K\|_\infty^{2+2\alpha}\|X\|_\infty)$. Here we know

$$|\Omega_{\varepsilon,1}\cap\Omega_{\varepsilon,2}|>1-C_1 n\exp\left\{\frac{-C_2(n-1)\varepsilon^2h^{d/2}}{2h\left[\|\nabla|_{y=x_i}/\!\!/_y^{x_i}X(y)\|^2+O(h)\right]}\right\}-C_3 n\exp\left\{\frac{-C_4(n-1)\varepsilon^2h^{d/2}}{2h\,[\|\nabla|_{y=x_i}p_h(y)\|^2+O(h)]}\right\}.$$

As a result, this large deviation bound implies that w.h.p. the variance term is $O(R_M,\|X\|_\infty)\frac{\sqrt{\log(n)}}{n^{1/2}h^{d/4-1/2}}$, where the $\log(n)$ term comes from (A.13). In other words, w.h.p.

$$u_i(\boldsymbol{D}_{h,\alpha,n}^{-1}\boldsymbol{P}_{h,\alpha,n}\boldsymbol{X})[i]=T_{h,\alpha}X(x_i)+O(R_M,\|X\|_\infty)\frac{\sqrt{\log(n)}}{n^{1/2}h^{d/4-1/2}}$$

for all $i=1,\ldots,n$, and we finish the proof. □

A.3. [Proof of Theorem 5.2]

The proof is a direct generalization of that of [26, Theorem B.3] to the principal bundle structure. Note that in [26, Theorem B.3] the constants of error terms are not explicitly shown and we make them explicit here in order to evaluate the optimal spectral convergence rate. First we elaborate the error term regarding the non-uniform p.d.f.. Denote $\widetilde{B}_{\sqrt{h}}(x)=\iota^{-1}(B_{\sqrt{h}}^{\mathbb{R}^p}(x)\cap\iota(\mathrm{M}))$.

Lemma A.3.1 (Lemma 8 in [10])**.** Suppose $f\in C^3(\mathrm{M})$ and $x\notin\mathrm{M}_{\sqrt{h}}$, then

$$\int_{\widetilde{B}_{\sqrt{h}}(x)}h^{-d/2}K_h(x,y)f(y)\mathrm{d}V(y)=f(x)+h\frac{\mu_2}{d}\left(\frac{\Delta f(x)}{2}+w(x)f(x)\right)+O(R_M,\|f'''\|_\infty)h^2,$$

where $w(x)=s(x)+\frac{m_3'z(x)}{24|S^{d-1}|}$ and $z(x)=\int_{S^{d-1}}\|\Pi(\theta,\theta)\|\mathrm{d}\theta$ are both bounded by R_M.

Proof. The proof is exactly the same as that of [10, Lemma 8] except the explicit expansion of the error term. Since the main point is the uniform bound of the third derivative of the embedding function ι and f on M, we simply list the calculation steps:

$$
\begin{aligned}
&\int_{\widetilde{B}_{\sqrt{h}}(x)} K_h(x,y) f(y)\mathrm{d}V(y) = \int_{\widetilde{B}_{\sqrt{h}}(x)} K\Big(\frac{\|x-y\|_{\mathbb{R}^p}}{\sqrt{h}}\Big) f(y)\mathrm{d}V(y)\\
=&\int_{\widetilde{B}_{\sqrt{h}}(x)} \left[K\left(\frac{t}{\sqrt{h}}\right) + K'\left(\frac{t}{\sqrt{h}}\right)\frac{\|\Pi(\theta,\theta)\|t^3}{24\sqrt{h}} + O(R_M)\frac{t^6}{h}\right]\left[f(x)+\nabla_\theta f(x)t+\nabla^2_{\theta,\theta}f(x)\frac{t^2}{2}+O(\|f'''\|_\infty)t^3\right]\mathrm{d}V(y)\\
=&\int_{S^{d-1}}\int_0^{\sqrt{h}} \left[K\left(\frac{t}{\sqrt{h}}\right) + K'\left(\frac{t}{\sqrt{h}}\right)\frac{\|\Pi(\theta,\theta)\|t^3}{24\sqrt{h}} + O(R_M)\frac{t^6}{h}\right]\\
&\times\left[f(x)+\nabla_\theta f(x)t+\nabla^2_{\theta,\theta}f(x)\frac{t^2}{2}+O(\|f'''\|_\infty)t^3\right]\left[t^{d-1}+\mathrm{Ric}(\theta,\theta)t^{d+1}+O(R_M)t^{d+2}\right]\mathrm{d}t\mathrm{d}\theta.
\end{aligned}
$$

By the direct expansion, the regularity assumption and the compactness of M, we conclude the Lemma. □

Proof of Theorem 5.2. By Lemma A.3.1, we get

$$
p_h(y) = p(y) + h\frac{\mu_2}{d}\left(\frac{\Delta p(y)}{2} + w(y)p(y)\right) + O(R_M)h^2,
$$

which leads to

$$
\frac{p(y)}{p_h^\alpha(y)} = p^{1-\alpha}(y)\left[1-\alpha h\frac{\mu_2}{d}\left(w(y)+\frac{\Delta p(y)}{2p(y)}\right)\right] + O(R_M)h^2. \tag{A.15}
$$

Plug (A.15) into the numerator of $T_{h,\alpha}X(x)$:

$$
\begin{aligned}
&\int_{\widetilde{B}_{\sqrt{h}}(x)} K_{h,\alpha}(x,y) /\!/_y^x X(y)p(y)\mathrm{d}V(y) = p_h^{-\alpha}(x)\int_{\widetilde{B}_{\sqrt{h}}(x)} K_h(x,y) /\!/_y^x X(y)p_h^{-\alpha}(y)p(y)\mathrm{d}V(y)\\
=& p_h^{-\alpha}(x)\int_{\widetilde{B}_{\sqrt{h}}(x)} K_h(x,y) /\!/_y^x X(y)p^{1-\alpha}(y)\left[1-\alpha h\frac{\mu_2}{d}\left(w(y)+\frac{\Delta p(y)}{2p(y)}\right)\right]\mathrm{d}V(y) + O(R_M)h^{d/2+2}\\
:=& p_h^{-\alpha}(x)A - \frac{\mu_2}{d}\alpha p_h^{-\alpha}(x)hB + O(R_M)h^{d/2+2}.
\end{aligned}
$$

where

$$
\begin{cases}
A := \int_{\widetilde{B}_{\sqrt{h}}(x)} K_h(x,y) /\!/_y^x X(y)p^{1-\alpha}(y)\mathrm{d}V(y),\\
B := \int_{\widetilde{B}_{\sqrt{h}}(x)} K_h(x,y) /\!/_y^x X(y)p^{1-\alpha}(y)\left(w(y)+\frac{\Delta p(y)}{2p(y)}\right)\mathrm{d}V(y).
\end{cases}
$$

We evaluate A and B by changing the integration variables to the polar coordinates, and the odd monomials in the integral vanish because the kernel is symmetric. Thus, applying Taylor's expansion to A leads to:

$$
\begin{aligned}
A =& \int_{S^{d-1}}\int_0^{\sqrt{h}} \left[K\left(\frac{t}{\sqrt{h}}\right) + K'\left(\frac{t}{\sqrt{h}}\right)\frac{\|\Pi(\theta,\theta)\|t^3}{24\sqrt{h}} + O(R_M)\frac{t^6}{h}\right]\\
&\times\left[X(x)+\nabla_\theta X(x)t+\nabla^2_{\theta,\theta}X(x)\frac{t^2}{2}+O(\|X^{(3)}\|_\infty)t^3\right]\\
&\times\left[p^{1-\alpha}(x)+\nabla_\theta(p^{1-\alpha})(x)t+\nabla^2_{\theta,\theta}(p^{1-\alpha})(x)\frac{t^2}{2}+O(R_M)t^3\right]\left[t^{d-1}+\mathrm{Ric}(\theta,\theta)t^{d+1}+O(R_M)t^{d+2}\right]\mathrm{d}t\mathrm{d}\theta,
\end{aligned}
$$

which after rearrangement becomes

$$
\begin{aligned}
A =& p^{1-\alpha}(x)X(x)\int_{S^{d-1}}\int_0^{\sqrt{h}}\left\{K\left(\frac{t}{\sqrt{h}}\right)\left[1+\mathrm{Ric}(\theta,\theta)t^2\right]t^{d-1} + K'\left(\frac{t}{\sqrt{h}}\right)\frac{\|\Pi(\theta,\theta)\|t^{d+2}}{24\sqrt{h}}\right\}\mathrm{d}t\mathrm{d}\theta\\
&+ p^{1-\alpha}(x)\int_{S^{d-1}}\int_0^{\sqrt{h}} K\left(\frac{t}{\sqrt{h}}\right)\nabla^2_{\theta,\theta}X(x)\frac{t^{d+1}}{2}\mathrm{d}t\mathrm{d}\theta + X(x)\int_{S^{d-1}}\int_0^{\sqrt{h}} K\left(\frac{t}{\sqrt{h}}\right)\nabla^2_{\theta,\theta}(p^{1-\alpha})(x)\frac{t^{d+1}}{2}\mathrm{d}t\mathrm{d}\theta\\
&+ \int_{S^{d-1}}\int_0^{\sqrt{h}} K\left(\frac{t}{\sqrt{h}}\right)\nabla_\theta X(x)\nabla_\theta(p^{1-\alpha})(x)t^{d+1}\mathrm{d}t\mathrm{d}\theta + O(R_M,\|X^{(3)}\|_\infty)h^{d/2+2}.
\end{aligned}
$$

Following the same argument as that in [26], we have

$$\int_{B_{\sqrt{h}}(0)} \frac{1}{h^{d/2}} K'\left(\frac{t}{\sqrt{h}}\right) \frac{\|\Pi(\theta,\theta)\| t^{d+2}}{24\sqrt{h}} \mathrm{d}t\mathrm{d}\theta = \frac{h^{d/2+1} m_3' z(x)}{24|S^{d-1}|},$$

$$\int_{S^{d-1}} \nabla^2_{\theta,\theta} X(x) \mathrm{d}\theta = \frac{|S^{d-1}|}{d} \nabla^2 X(x) \quad \text{and} \quad \int_{S^{d-1}} \mathrm{Ric}(\theta,\theta) d\theta = \frac{|S^{d-1}|}{d} s(x).$$

Therefore, A become

$$\begin{aligned} &h^{d/2} p^{1-\alpha}(x) \left\{ \left(1 + \frac{h\mu_2}{d} \frac{\Delta(p^{1-\alpha})(x)}{2p^{1-\alpha}(x)} + \frac{h\mu_2}{d} w(x)\right) X(x) + \frac{h\mu_2}{2d} \nabla^2 X(x) \right\} \\ &\quad + h^{d/2+1} \frac{\mu_2}{d} \nabla X(x) \cdot \nabla (p^{1-\alpha})(x) + O(R_M, \|X^{(3)}\|_\infty) h^{d/2+2} \end{aligned}$$

Next, we expand B by the same argument as that in [26]. Denote $Q(y) = p^{1-\alpha}(y) \left(w(y) + \frac{\Delta p(y)}{2p(y)}\right) \in C^2(\mathrm{M})$ to simplify notation and we have

$$\begin{aligned} B =& \int_{B_{\sqrt{h}}(x)} K_h(x,y) /\!/_y^x X(y) Q(y) \mathrm{d}V(y) \\ =& \int_{S^{d-1}} \int_0^{\sqrt{h}} \left[K\left(\frac{t}{\sqrt{h}}\right) + O(R_M) \frac{t^3}{\sqrt{h}} \right] \left[X(x) + \nabla_\theta X(x) t + O(\|X^{(3)}\|_\infty) t^2 \right] \\ &\times \left[Q(x) + \nabla_\theta Q(x) t + O(t^2) \right] \left[t^{d-1} + \mathrm{Ric}(\theta,\theta) t^{d+1} + O(R_M) t^{d+2} \right] \mathrm{d}t\mathrm{d}\theta \\ =& h^{d/2} X(x) Q(x) + O(R_M, \|X^{(3)}\|_\infty) h^{d/2+1}. \end{aligned}$$

In conclusion, the numerator of $T_{h,\alpha} X(x)$ becomes

$$\begin{aligned} &h^{d/2} \frac{p^{1-\alpha}(x)}{p_h^\alpha(x)} \left\{ 1 + \frac{h\mu_2}{d} \left[\frac{\Delta(p^{1-\alpha})(x)}{2p^{1-\alpha}(x)} - \alpha \frac{\Delta p(x)}{2p(x)} \right] \right\} X(x) \\ &+ h^{d/2+1} \frac{\mu_2}{d p_h^\alpha(x)} \left\{ \frac{p^{1-\alpha}(x)}{2} \nabla^2 X(x) + \nabla X(x) \cdot \nabla (p^{1-\alpha})(x) \right\} + O(R_M, \|X^{(3)}\|_\infty) h^{d/2+2}. \end{aligned}$$

Similar calculation of the denominator of the $T_{h,\alpha} X(x)$ gives

$$h^{d/2} p_h^{-\alpha}(x) p^{1-\alpha}(x) \left\{ 1 + h \frac{\mu_2}{d} \left(\frac{\Delta(p^{1-\alpha})(x)}{2p^{1-\alpha}(x)} - \alpha \frac{\Delta p(x)}{2p(x)} \right) \right\} + O(R_M) h^{d/2+2}.$$

Putting all the above together, we have when $x \in \mathrm{M} \backslash \mathrm{M}_{\sqrt{h}}$,

$$T_{h,\alpha} X(x) = X(x) + h \frac{\mu_2}{2d} \left(\nabla^2 X(x) + \frac{2\nabla X(x) \cdot \nabla (p^{1-\alpha})(x)}{p^{1-\alpha}(x)} \right) + O(R_M, \|X^{(3)}\|_\infty) h^2.$$

Next we consider the case when $x \in \mathrm{M}_{\sqrt{h}}$. The proof is again almost the same as that in [26]. Suppose $\min_{y \in \partial \mathrm{M}} d(x,y) = \tilde{h}$. Choose a normal coordinate $\{\partial_1, \ldots, \partial_d\}$ on the geodesic ball $B_{h^{1/2}}(x)$ around x so that $x_0 = \exp_x(\tilde{h} \partial_d(x))$. Due to Gauss Lemma, we know $\mathrm{span}\{\partial_1(x_0), \ldots, \partial_{d-1}(x_0)\} = T_{x_0} \partial \mathrm{M}$ and $\partial_d(x_0)$ is outer normal at x_0.

We focus first on the integral appearing in the numerator of $T_{h,1} X(x)$:

$$\int_{B_{\sqrt{h}}(x) \cap \mathrm{M}} \frac{1}{h^{d/2}} K_{h,1}(x,y) /\!/_y^x X(y) p(y) \mathrm{d}V(y).$$

We divide the integral domain $\exp_x^{-1}(B_{\sqrt{h}}(x) \cap \mathrm{M})$ into slices S_η defined by

$$S_\eta = \{(\boldsymbol{u}, \eta) \in \mathbb{R}^d : \|(u_1, \ldots, u_{d-1}, \eta)\| < \sqrt{h}\},$$

where $\eta \in [-h^{1/2}, h^{1/2}]$ and $\boldsymbol{u} = (u_1, \ldots, u_{d-1}) \in \mathbb{R}^{d-1}$. By Taylor's expansion and (A.15), the numerator of $T_{h,1}X$ becomes

$$\begin{aligned}
&\int_{B_{\sqrt{h}}(x)\cap \mathrm{M}} K_{h,1}(x,y) /\!\!/_y^x X(y) p(y) \mathrm{d}V(y) \\
=&p_h^{-1}(x) \int_{S_\eta} \int_{-\sqrt{h}}^{\sqrt{h}} K\left(\frac{\sqrt{\|\boldsymbol{u}\|^2+\eta^2}}{\sqrt{h}}\right) \\
&\times \left(X(x) + \sum_{i=1}^{d-1} u_i \nabla_{\partial_i} X(x) + \eta \nabla_{\partial_d} X(x) + O(\|X^{(2)}\|_\infty) h\right) \left[1 - h\frac{\mu_2}{d}\left(w(y) + \frac{\Delta p(y)}{2p(y)}\right) + O(R_M) h^2\right] \mathrm{d}\eta \mathrm{d}\boldsymbol{u} \\
=&p^{-1}(x) \int_{S_\eta} \int_{-\sqrt{h}}^{\sqrt{h}} K\left(\frac{\sqrt{\|\boldsymbol{u}\|^2+\eta^2}}{\sqrt{h}}\right) \\
&\times \left(X(x) + \sum_{i=1}^{d-1} u_i \nabla_{\partial_i} X(x) + \eta \nabla_{\partial_d} X(x) + O(\|X^{(2)}\|_\infty) h\right) \mathrm{d}\eta \mathrm{d}\boldsymbol{u} + O(R_M, \|X^{(2)}\|_\infty) h.
\end{aligned} \tag{A.16}$$

Note that in general the integral domain S_η is not symmetric with related to $(0, \ldots, 0, \eta)$, so we will try to symmetrize S_η by defining the symmetrized slices:

$$\tilde{S}_\eta = \cap_{i=1}^{d-1} (R_i S_\eta \cap S_\eta),$$

where $R_i(u_1, \ldots, u_i, \ldots, \eta) = (u_1, \ldots, -u_i, \ldots, \eta)$. Recall the following relationship [26, (B.23)]:

$$\partial_l(\exp_x(t\theta)) = /\!\!/_x^y \partial_l(x) + \frac{t^2}{6} /\!\!/_x^y (\mathscr{R}(\theta, \partial_l(x))\theta) + O(R_M) t^3,$$

where $\theta \in T_x \mathrm{M}$ and $t \ll 1$, which leads to

$$P_{x_0,x} \partial_l(x) = \partial_l(x_0) + O(R_M) h, \tag{A.17}$$

for all $l = 1, \ldots, d$. Also note that up to error $O(R_M) h^{3/2}$, we can express $\partial \mathrm{M} \cap B_{h^{1/2}}(x)$ by a homogeneous degree 2 polynomial with variables $\{P_{x_0,x}\partial_1(x), \ldots, P_{x_0,x}\partial_{d-1}(x)\}$. Thus the difference between $\tilde{S}_\eta$ and S_η is $O(R_M)h$, and hence (A.16) becomes:

$$p^{-1}(x) \int_{\tilde{S}_\eta} \int_{-\sqrt{h}}^{\sqrt{h}} K\left(\frac{\sqrt{\|\boldsymbol{u}\|^2+\eta^2}}{\sqrt{h}}\right) \left(X(x) + \sum_{i=1}^{d-1} u_i \nabla_{\partial_i} X(x) + \eta \nabla_{\partial_d} X(x) + O(\|X^{(2)}\|_\infty) h\right) \mathrm{d}\eta \mathrm{d}\boldsymbol{u} + O(R_M, \|X^{(2)}\|_\infty) h. \tag{A.18}$$

By Taylor's expansion we have for all $i = 1, \ldots, d$

$$\nabla_{\partial_i} X(x) = P_{x,x_0}(\nabla_{\partial_i} X(x_0)) + O(\|X^{(2)}\|_\infty) \tilde{h}^{1/2}. \tag{A.19}$$

By plugging (A.19) into (A.18), (A.16) is further reduced to:

$$\begin{aligned}
&p^{-1}(x) \int_{\tilde{S}_\eta} \int_{-\sqrt{h}}^{\sqrt{h}} K\left(\frac{\sqrt{\|\boldsymbol{u}\|^2+\eta^2}}{\sqrt{h}}\right) \\
&\quad \times \left(X(x) + P_{x,x_0}\Big(\sum_{i=1}^{d-1} u_i \nabla_{\partial_i} X(x_0) + (\eta - \tilde{h}) \nabla_{\partial_d} X(x_0)\Big) + O(\|X^{(2)}\|_\infty) h\right) \mathrm{d}\eta \mathrm{d}\boldsymbol{u} + O(R_M, \|X^{(2)}\|_\infty) h
\end{aligned} \tag{A.20}$$

The symmetry of the kernel implies that for $i = 1, \ldots, d-1$,

$$\int_{\tilde{S}_\eta} K\left(\frac{\sqrt{\|\boldsymbol{u}\|^2+\eta^2}}{\sqrt{h}}\right) u^i \mathrm{d}\boldsymbol{u} = 0, \tag{A.21}$$

and hence the numerator of $T_{1,h}X(x)$ becomes

$$p^{-1}(x) m_0^h \Big(X(x) + \frac{m_1^h}{m_0^h} P_{x,x_0} \nabla_{\partial_d} X(x_0)\Big) + O(R_M, \|X^{(2)}\|_\infty) h, \tag{A.22}$$

where

$$m_0^h := \int_{\tilde{S}_\eta} \int_{-\sqrt{h}}^{\sqrt{h}} K\left(\frac{\sqrt{\|u\|^2+\eta^2}}{\sqrt{h}}\right) \mathrm{d}\eta\, \mathrm{d}x = O(1)h^{d/2} \tag{A.23}$$

and

$$m_1^h := \int_{\tilde{S}_\eta} \int_{-\sqrt{h}}^{\sqrt{h}} K\left(\frac{\sqrt{\|u\|^2+\eta^2}}{\sqrt{h}}\right) (\eta-\tilde{h}) \mathrm{d}\eta\, \mathrm{d}x = O(1)h^{d/2+1/2}. \tag{A.24}$$

Similarly, the denominator of $T_{h,1}X$ can be expanded as:

$$\int_{B_{\sqrt{h}}(x)\cap \mathrm{M}} K_{h,1}(x,y)p(y)\mathrm{d}V(y) = p^{-1}(x)m_0^h + O(1)h^{d/2+1/2}, \tag{A.25}$$

which together with (A.22) gives us the following asymptotic expansion:

$$T_{h,1}X(x) = X(x) + \frac{m_1^h}{m_0^h} P_{x,x_0} \nabla_{\partial_d} X(x_0) + O(R_M, \|X^{(2)}\|_\infty)h, \tag{A.26}$$

which finish the proof. □

A.4. Symmetric Isometric Embedding

Suppose we have a closed, connected and smooth d-dim Riemannian manifold (M,g) with free isometric $\mathbb{Z}_2 := \{1,z\}$ action on it. Note that M can be viewed as a principal bundle $P(\mathrm{M}/\mathbb{Z}_2,\mathbb{Z}_2)$ with the group $\mathbb{Z}_2$ as the fiber. Without loss of generality, we assume the diameter of M is less than 1. The eigenfunctions $\{\phi_j\}_{j\geq 0}$ of the Laplace-Beltrami operator Δ_M are known to form an orthonormal basis of $L^2(\mathrm{M})$, where $\Delta_\mathrm{M}\phi_j = -\lambda_j\phi_j$ with $\lambda_j \geq 0$. Denote E_λ the eigenspace of Δ_M with eigenvalue λ. Since $\mathbb{Z}_2$ commutes with Δ_M, E_λ is a representation of $\mathbb{Z}_2$, where the action of z on ϕ_j is defined by $z\circ\phi_j(x) := \phi_j(z\circ x)$.

We claim that all the eigenfunctions of Δ_M are either even or odd. Indeed, since $\mathbb{Z}_2$ is an abelian group and all the irreducible representations of $\mathbb{Z}_2$ are real, we know $z\circ\phi_i = \pm\phi_i$ for all $i\geq 0$. We can thus distinguish two different types of eigenfunctions:

$$\phi_i^e(z\circ x) = \phi_i^e(x) \quad \text{and} \quad \phi_i^o(z\circ x) = -\phi_i^o(x),$$

where the superscript e (resp. o) means even (resp. odd) eigenfunctions.

It is well known that the heat kernel $k(x,y,t)$ of Δ_M is a smooth function over x and y and analytic over $t>0$, and can be written as

$$k(x,y,t) = \sum_i e^{-\lambda_i t}\phi_i(x)\phi_i(y),$$

we know for all $t>0$ and $x\in\mathrm{M}$, $\sum_j e^{-\lambda_j t}\phi_j(x)\phi_j(x) < \infty$. Thus we can define a family of maps by exceptionally taking odd eigenfunctions into consideration:

$$\begin{array}{rccll} \Psi_t^o: & \mathrm{M} & \to & \ell^2 & \quad\text{for } t>0, \\ & x & \mapsto & \{e^{-\lambda_j t/2}\phi_j^o(x)\}_{j\geq 1} & \end{array}$$

Lemma A.4.1. For $t>0$, the map Ψ_t^o is an embedding of M into ℓ^2.

Proof. If $x_n\to x$, we have by definition

$$\begin{aligned} \|\Psi_t^o(x_n)-\Psi_t^o(x)\|_{\ell^2}^2 &= \sum_j \left| e^{-\lambda_j t/2}\phi_j^o(x_n) - e^{-\lambda_j t/2}\phi_j^o(x)\right|^2 \\ &\leq \sum_j \left| e^{-\lambda_j t/2}\phi_j^o(x_n) - e^{-\lambda_j t/2}\phi_j^o(x)\right|^2 + \sum_j \left| e^{-\lambda_j t/2}\phi_j^e(x_n) - e^{-\lambda_j t/2}\phi_j^e(x)\right|^2 \\ &= k(x_n,x_n,t)+k(x,x,t)-2k(x_n,x,t), \end{aligned}$$

which goes to 0 as $n\to\infty$ due to the smoothness of the heat kernel. Thus Ψ_t^o is continuous.

Since the eigenfunctions $\{\phi_j\}_{j\geq 0}$ of the Laplace-Beltrami operator form an orthonormal basis of $L^2(\mathrm{M})$, it follows that they separate points. We now show that odd eigenfunctions are enough to separate points. Given $x \neq y$ two distinct points on M, we can find a small enough neighborhood N_x of x that separates it from y. Take a characteristic odd function f such that $f(x) = 1$ on N_x, $f(z\circ x) = -1$ on $z\circ N_x$ and 0 otherwise. Clearly we know $f(x) \neq f(y)$. Since f is odd, it can be expanded by the odd eigenfunctions:

$$f = \sum_j a_j \phi_j^o.$$

Hence $f(x) \neq f(y)$ implies that there exists α such that $\phi_\alpha^o(x) \neq \phi_\alpha^o(y)$.

Suppose we have $\Psi_t^o(x) = \Psi_t^o(y)$, then $\phi_i^o(x) = \phi_i^o(y)$ for all i. By the above argument we conclude that $x = y$, that is, Ψ_t^o is an 1-1 map. To show that Ψ_t^o is an immersion, consider a neighborhood N_x so that $N_x \cap z\circ N_x = \emptyset$. Suppose there exists $x \in \mathrm{M}$ so that $\mathrm{d}\Psi_t^o(X) = 0$ for $X \in T_x\mathrm{M}$, which implies $\mathrm{d}\phi_i^o(X) = 0$ for all i. Thus by the same argument as above we know $\mathrm{d}f(X) = 0$ for all $f \in C_c^\infty(N_x)$, which implies $X = 0$.

In conclusion, Ψ_t^o is continuous and 1-1 immersion from M, which is compact, onto $\Psi_t^o(\mathrm{M})$, so it is an embedding. $\square$

Note that $\Psi_t^o(\mathrm{M})$ is symmetric with respect to 0, that is, $\Psi_t^o(z\circ x) = -\Psi_t^o(x)$. However, it is not an isometric embedding and the embedded space is of infinite dimension. Now we construct an isometric symmetric embedding of M to a finite dimensional space by extending the Nash embedding theorem [18, 19]. We start from considering an open covering of M in the following way. Since Ψ_t^o, $t > 0$, is an embedding of M into ℓ^2, for each given $p \in \mathrm{M}$, there exists d odd eigenfunctions $\{\phi^o_{i^p_j}\}_{j=1}^d$ so that

$$\begin{aligned} &v_p : x \in \mathrm{M} \mapsto (\phi^o_{i^p_1}(x), \ldots, \phi^o_{i^p_d}(x)) \in \mathbb{R}^d \\ &v_{z\circ p} : z\circ x \in \mathrm{M} \mapsto -(\phi^o_{i^p_1}(x), \ldots, \phi^o_{i^p_d}(x)) \in \mathbb{R}^d \end{aligned} \tag{A.27}$$

are of full rank at p and $z\circ p$. We choose a small enough neighborhood N_p of p so that $N_p \cap z\circ N_p = \emptyset$ and v_p and $v_{z\circ p}$ are embedding of N_p and $z\circ N_p$. It is clear that $\{N_p,\, z\circ N_p\}_{p\in\mathrm{M}}$ is an open covering of M.

With the open covering $\{N_p,\, z\circ N_p\}_{p\in\mathrm{M}}$, it is a well known fact [28] that there exists an atlas of M

$$\mathscr{A} = \{(V_j, h_j),\ (z\circ V_j, h_j^z)\}_{j=1}^L \tag{A.28}$$

where $V_j \subset M$, $z\circ V_j \subset M$, $h_j : \mathrm{M} \to \mathbb{R}^d$, $h_j^z : \mathrm{M} \to \mathbb{R}^d$, so that the following holds and the symmetry is taken into account:

(a) $\mathscr{A}$ is a locally finite refinement of $\{N_p,\, z\circ N_p\}_{p\in\mathrm{M}}$, that is, for every V_i (resp. $z\circ V_i$), there exists a $p_i \in \mathrm{M}$ (resp. $z\circ p_i \in \mathrm{M}$) so that $V_i \subset N_{p_i}$ (resp. $z\circ V_i \subset z\circ N_{p_i}$),
(b) $h_j(V_j) = B_2$, $h_j^z(z\circ V_j) = B_2$, and $h_j(x) = h_j^z(z\circ x)$ for all $x \in V_j$,
(c) for the p_i chosen in (a), there exists $\phi^o_{i_{p_i}}$ so that $\phi^o_{i_{p_i}}(x) \neq \phi^o_{i_{p_i}}(z\circ x)$ for all $x \in V_i$.
(d) $\mathrm{M} = \cup_j \left(h_j^{-1}(B_1) \cup (h_j^z)^{-1}(B_1)\right)$. Denote $O_j = h_j^{-1}(B_1)$.

where $B_r = \{x \in \mathbb{R}^d : \|x\| < 1\}$. We fix the point $p_i \in \mathrm{M}$ when we determine $\mathscr{A}$, that is, if $V_i \in \mathscr{A}$, we have a unique $p_i \in \mathrm{M}$ so that $V_i \subset N_{p_i}$. Note that (c) holds since Ψ_t^o, $t > 0$, is an embedding of M into ℓ^2 and the eigenfunctions of Δ_M are smooth. We will fix a partition of unity $\{\eta_i \in C_c^\infty(V_i),\ \eta_i^z \in C_c^\infty(z\circ V_i)\}$ subordinate to $\{V_j,\, z\circ V_j\}_{j=1}^L$. Due to symmetry, we have $\eta_i(x) = \eta_i^z(z\circ x)$ for all $x \in V_i$. To ease notation, we define

$$\psi_i(x) = \begin{cases} \eta_i(x) & \text{when } x \in V_i \\ \eta_i^z(x) & \text{when } x \in z\circ V_i \end{cases} \tag{A.29}$$

so that $\{\psi_i\}_{i=1}^L$ is a partition of unit subordinate to $\{V_i \cup z\circ V_i\}_{i=1}^L$.

Lemma A.4.2. There exists a symmetric embedding $\tilde{u} : \mathrm{M}^d \hookrightarrow \mathbb{R}^N$ for some $N \in \mathbb{N}$.

Proof. Fix V_i and hence $p_i \in \mathrm{M}$. Define

$$u_i : x \in \mathrm{M} \mapsto (\phi^o_{i_{p_i}}(x), v_{p_i}(x)) \in \mathbb{R}^{d+1},$$

where v_{p_i} is defined in (A.27). Note that u_i is of full rank at p_i. Due to symmetry, the assumption (c) and the fact that $V_i \cap z \circ V_i = \emptyset$, we can find a rotation $R_i \in SO(d+1)$ and modify the definition of u_i:

$$u_i : x \mapsto R_i(\phi^o_{i_{p_i}}(x), v_{p_i}(x)),$$

which is an embedding of $V_i \cup z \circ V_i$ onto $\mathbb{R}^{d+1}$ so that $u_i(V_i \cup z \circ V_i)$ does not meet all the axes of $\mathbb{R}^{d+1}$. Note that since $v_{z\circ p}(z \circ x) = -v_p(x)$ and $\phi^o_{i_{p_i}}(z \circ x) = -\phi^o_{i_{p_i}}(x)$, we have $u_i(z \circ x) = -u_i(x)$. Define

$$\bar{u} : x \mapsto (u_1(x), \ldots, u_L(x)).$$

Since locally $d\bar{u}$ is of full rank and

$$\bar{u}(z \circ x) = (u_1(z \circ x), \ldots, u_L(z \circ x)) = -(u_1(x), \ldots, u_L(x)) = -\bar{u}(x),$$

$\bar{u}$ is clearly an symmetric immersion from M to $\mathbb{R}^{L(d+1)}$. Denote

$$\varepsilon = \min_{i=1,\ldots,L} \min_{x \in V_i \cup z\circ V_i} \min_{k=1,\ldots,d+1} \langle u_i(x), e_k \rangle,$$

where $\{e_k\}_{k=1,\ldots,d+1}$ is the canonical basis of $\mathbb{R}^{d+1}$. By the construction of u_i, $\varepsilon > 0$.

By the construction of the covering $\{O_i \cup g \circ O_i\}_{i=1}^L$, we know $L \geq 2$. We claim that by properly perturbing $\bar{u}$ we can generate a symmetric 1-1 immersion from M to $\mathbb{R}^{L(d+1)}$.

Suppose $\bar{u}$ is 1-1 in $W \subset \mathrm{M}$, which is invariant under $\mathbb{Z}_2$ action by the construction of $\bar{u}$. Consider a symmetric closed subset $K \subset W$. Let $O_i^1 = W \cap (O_i \cup g \circ O_i)$ and $O_i^2 = (\mathrm{M} \backslash K) \cap (O_i \cup g \circ O_i)$. Clearly $\{O_i^1, O_i^2\}_{i=1}^L$ is an covering of M. Consider a partition of unity $\mathscr{P} = \{\theta_\alpha\}$ subordinate to this covering so that $\theta_\alpha(z \circ x) = \theta_\alpha(x)$ for all α. Index $\mathscr{P}$ by integer numbers so that for all $i > 0$, we have $\mathrm{supp}\theta_i \subset O_i^2$.

We will inductively define a sequence $\tilde{u}_k$ of immersions by properly choosing constants $b_i \in \mathbb{R}^{L(d+1)}$:

$$\tilde{u}_k = \bar{u} + \sum_{i=1}^k b_i s_i \theta_i,$$

where $s_i \in C_c^\infty(\mathrm{M})$ so that $\mathrm{supp}(s_i) \subset N_i \cup z \circ N_i$ and

$$s_i(x) = \begin{cases} 1 & \text{when } x \in V_i \\ -1 & \text{when } x \in z \circ V_i \end{cases}.$$

Note that u_k by definition will be symmetric. Suppose u_k is properly defined to become an immersion and $\|\tilde{u}_j - \tilde{u}_{j-1}\|_{C^\infty} < 2^{-j-2}\varepsilon$ for all $j \leq k$.

Denote

$$D_{k+1} = \{(x,y) \in \mathrm{M} \times \mathrm{M} : s_{k+1}(x)\theta_{k+1}(x) \neq s_{k+1}(y)\theta_{k+1}(y)\},$$

which is of dimension $2d$. Define $G_{k+1} : D_{k+1} \to \mathbb{R}^{L(d+1)}$ as

$$G_{k+1}(x,y) = \frac{\tilde{u}_k(x) - \tilde{u}_k(y)}{s_{k+1}(x)\theta_{k+1}(x) - s_{k+1}(y)\theta_{k+1}(y)}.$$

Since G_{k+1} is differentiable and $L \geq 2$, by Sard's Theorem $G_{k+1}(D_{k+1})$ is of measure zero. By choosing $b_{k+1} \notin G_{k+1}(D_{k+1})$ small enough, $\tilde{u}_{k+1}$ can be made an immersion and $\|\tilde{u}_{k+1} - \tilde{u}_k\| < 2^{-k-3}\varepsilon$. In this case $\tilde{u}_{k+1}(y_1) = \tilde{u}_{k+1}(y_2)$ implies

$$b_{k+1}(s_{k+1}(x)\theta_{k+1}(x) - s_{k+1}(y)\theta_{k+1}(y)) = \tilde{u}_k(x) - \tilde{u}_k(y).$$

Since $b_{k+1} \notin G_{k+1}(D_{k+1})$, this can happen only if $s_{k+1}(x)\theta_{k+1}(x) = s_{k+1}\theta_{k+1}(y)$ and $\tilde{u}_k(x) = \tilde{u}_k(y)$.

Define

$$\tilde{u} = \tilde{u}_L.$$

By definition $\tilde{u}$ is a symmetric immersion and differs from $\bar{u}$ by $\varepsilon/2$ in C^∞.

Now we claim that $\tilde{u}$ is 1-1. Suppose $\tilde{u}(x)=\tilde{u}(y)$. Note that by the construction of b_j this implies $s_L(x)\theta_L(x)=s_L(y)\theta_L(y)$ and $u_{L-1}(x)=u_{L-1}(y)$. Inductively we have $\overline{u}(x)=\overline{u}(y)$ and $s_j(x)\theta_j(x)=s_j(y)\theta_j(y)$ for all $j>0$. Suppose $x\in W$ but $y\notin W$, then $s_j(y)\theta_j(y)=s_j(x)\theta_j(x)=0$ for all $j>0$, which is impossible. Suppose both x and y are outside W, then there are two cases to discuss. First, if x and y are both inside V_i for some i, then $s_j(x)\theta_j(x)=s_j(y)\theta_j(y)$ for all $j>0$ and $\overline{u}(x)=\overline{u}(y)$ imply $x=y$ since $\overline{u}$ embeds V_i. Second, if $x\in V_i\backslash V_j$ and $y\in V_j\backslash V_i$ where $i\neq j$, then $s_j(x)\theta_j(x)=s_j(y)\theta_j(y)$ for all $j>0$ is impossible. In conclusion, $\tilde{u}$ is 1-1.

Since M is compact and $\tilde{u}$ is continuous, we conclude that $\tilde{u}$ is a symmetric embedding of M into $\mathbb{R}^{L(d+1)}$.

□

The above Lemma shows that we can always find a symmetric embedding of M into $\mathbb{R}^{L(d+1)}$ for some $L>0$. The next Lemma helps us to show that we can further find a symmetric embedding of M into $\mathbb{R}^p$ for some $p>0$ which is isometric. We define $s_p:=\frac{p(p+1)}{2}$ in the following discussion.

Lemma A.4.3. There exists a symmetric smooth map Φ from $\mathbb{R}^p$ to $\mathbb{R}^{s_p+p}$ so that $\partial_i\Phi(x)$ and $\partial_{ij}\Phi(x)$, $i,j=1,\dots p$, are linearly independent as vectors in $\mathbb{R}^{s_p+p}$ for all $x\neq 0$.

Proof. Denote $x=(x_1,\dots x_p)\in\mathbb{R}^p$. We define the map Φ from $\mathbb{R}^p$ to $\mathbb{R}^{s_p+p}$ by

$$\Phi: x\mapsto\left(x_1\ ,\ \dots,x_p\ ,\ x_1\frac{e^{x_1}+e^{-x_1}}{2}\ ,\ x_1\frac{e^{x_2}+e^{-x_2}}{2}\ ,\dots,\ x_p\frac{e^{x_p}+e^{-x_p}}{2}\right).$$

where $i,j=1,\dots,p$ and $i\neq j$. It is clear that Φ is a symmetric smooth map, that is, $\Phi(-x)=-\Phi(x)$. Note that

$$\partial_{ij}\left(x_k\frac{e^{x_\ell}+e^{-x_\ell}}{2}\right)=\delta_{jk}\frac{e^{x_i}-e^{-x_i}}{2}+\delta_{ik}\frac{e^{x_j}-e^{-x_j}}{2}+x_k\delta_{j\ell}\frac{e^{x_i}+e^{-x_i}}{2}$$

Thus when $x\neq 0$, for all $i=1,\dots,p$, $\partial_i\Phi(x)$ and $\partial_{ij}\Phi(x)$, $i,j=1,\dots p$, are linearly independent as vectors in $\mathbb{R}^{s_p+p}$. □

Combining Lemma A.4.2 and A.4.3, we know there exists a symmetric embedding $u:\mathrm{M}^d\hookrightarrow\mathbb{R}^{s_{L(d+1)}+L(d+1)}$ so that $\partial_i u(x)$ and $\partial_{ij}u(x)$, $i,j=1,\dots,d$, are linearly independent as vectors in $\mathbb{R}^{s_{L(d+1)}+L(d+1)}$ for all $x\in\mathrm{M}$. Indeed, we define

$$u=\Phi\circ\tilde{u}.$$

Clearly u is a symmetric embedding of M into $\mathbb{R}^{s_{L(d+1)}+L(d+1)}$. Note that $\tilde{u}(x)\neq 0$ otherwise $\tilde{u}$ is not an embedding. Moreover, by the construction of $\tilde{u}$, we know $u_i(V_i\cup z\circ V_i)$ is away from the axes of $\mathbb{R}^{L(d+1)}$ by $\varepsilon/2$, so the result.

Next we control the metric on $u(\mathrm{M})$ induced by the embedding. By properly scaling u, we have $g-\mathrm{d}u^2>0$. We will assume properly scaled u in the following.

Lemma A.4.4. Given the atlas $\mathscr{A}$ defined in (A.28), there exists $\xi_i\in C^\infty(V_i,\,\mathbb{R}^{s_d+d})$ and $\xi_i^z\in C^\infty(z\circ V_i,\,\mathbb{R}^{s_d+d})$ so that $\xi_i^z-\xi_i>cI_{s_d+d}$ for some $c>0$ and

$$g-\mathrm{d}u^2=\sum_{j=1}^m\eta_j^2\mathrm{d}\xi_j^2+\sum_{j=1}^m(\eta_j^z)^2(\mathrm{d}\xi_j^z)^2.$$

Proof. Fix V_i. By applying the local isometric embedding theorem [28] we have smooth maps $x_i:h_i(V_i)\hookrightarrow\mathbb{R}^{s_d+d}$ and $x_i^z:h_i^z(z\circ V_i)\hookrightarrow\mathbb{R}^{s_d+d}$ so that

$$(h_i^{-1})^*g=\mathrm{d}x_i^2\quad\text{ and }\quad((h_i^z)^{-1})^*g=(\mathrm{d}x_i^z)^2,$$

where $\mathrm{d}x_i^2$ (resp. $(\mathrm{d}x_i^z)^2$) means the induced metric on $h_i(V_i)$ (resp. $h_i^z(z\circ V_i)$) from $\mathbb{R}^{s_d+d}$. Note that the above relationship is invariant under affine transformation of x_i and x_i^z. By assumption (b) of $\mathscr{A}$ we have $h_i(x)=h_i^z(z\circ x)$ for all $x\in V_i$, so we modify x_i and x_i^z so that that

$$x_i^z=x_i+c_iI_{s_d+d},$$

where $c_i>0$, $I_{s_d+d}=(1,\dots,1)^T\in\mathbb{R}^{s_d+d}$, and $x_i(B_1)\cap x_i^z(B_1)=\emptyset$. Denote $c=\max_{i=1}^L\{c_i\}$ and further set

$$x_i^z=x_i+cI_{s_d+d}$$

for all i. By choosing x_i and x_i^z in this way, we have embedded V_i and $z\circ V_i$ simultaneously into the same Euclidean space. Note that

$$g = h_i^*(h_i^{-1})^*g = \mathrm{d}(x_i\circ h_i)^2$$

on V_i and

$$g = (h_i^z)^*((h_i^z)^{-1})^*g = \mathrm{d}(x_i^z\circ h_i^z)^2$$

on $z\circ V_i$. Thus, by defining $\xi_i = x_i\circ h_i$ and $\xi_i^z = x_i^z\circ h_i^z$, and applying the partition of unity with (A.29), we have the results. □

Theorem A.4.5. Any smooth, closed manifold (M,g) with free isometric $\mathbb{Z}_2$ action admits a smooth symmetric, isometric embedding in $\mathbb{R}^p$ for some $p\in\mathbb{N}$.

Proof. By the remark following Lemma A.4.2 and A.4.3 we have a smooth embedding $u:\mathrm{M}\hookrightarrow\mathbb{R}^N$ so that $g-\mathrm{d}u^2>0$, where $N = s_{L(d+1)}+L(d+1)$. By Lemma A.4.4, with atlas $\mathscr{A}$ fixed we have

$$g-\mathrm{d}u^2 = \sum_j \eta_j^2\mathrm{d}\xi_j^2 + \sum_j(\eta_j^z)^2(\mathrm{d}\xi_j^z)^2.$$

where $\xi_i^z-\xi_i = cI_{s_d+d}$. Denote $c = \frac{(2\ell+1)\pi}{\lambda}$, where λ and ℓ will be determined later. To ease the notion, we define

$$\gamma_i(x) = \begin{cases} \xi_i(x) & \text{when } x\in N_i\\ \xi_i^z(x) & \text{when } x\in g\circ N_i\end{cases}.$$

Then by the definition (A.29) we have

$$g-\mathrm{d}u^2 = \sum_{j=1}^L \psi_j^2\mathrm{d}\gamma_j^2.$$

Given $\lambda>0$ we can define the following map $u_\lambda:\mathrm{M}\to\mathbb{R}^{2L}$:

$$u_\lambda = \left(\frac{1}{\lambda}\psi_i\cos(\lambda\gamma_i),\frac{1}{\lambda}\psi_i\sin(\lambda\gamma_i)\right)_{i=1}^L,$$

where $\cos(\lambda\gamma_i)$ means taking cosine on each entry of $\lambda\gamma_i$. Set ℓ so that $\frac{(2\ell+1)\pi}{\lambda}>1$ and we claim that u_λ is a symmetric map. Indeed,

$$\psi_i(z\circ x)\cos(\lambda\gamma_i(z\circ x)) = \psi_i(x)\cos\left(\lambda\left(\gamma_i(x)+\frac{(2\ell+1)\pi}{\lambda}\right)\right) = -\psi_i(x)\cos(\lambda\gamma_i(x)).$$

and

$$\psi_i(z\circ x)\sin(\lambda\gamma_i(z\circ x)) = \psi_i(x)\sin\left(\lambda\left(\gamma_i(x)+\frac{(2\ell+1)\pi}{\lambda}\right)\right) = -\psi_i(x)\sin(\lambda\gamma_i(x)).$$

Direct calculation gives us

$$g-\mathrm{d}u^2 = \mathrm{d}u_\lambda^2 - \frac{1}{\lambda^2}\sum_{j=1}^L\mathrm{d}\psi_j^2.$$

We show that when λ is big enough, there exists a smooth symmetric embedding w so that

$$\mathrm{d}w^2 = \mathrm{d}u^2 - \frac{1}{\lambda^2}\sum_i^L\mathrm{d}\psi_i^2. \tag{A.30}$$

Since for all $\lambda>0$ we can find a ℓ so that u_λ is a symmetric map without touching ψ_i, we can thus choosing λ as large as possible so that (A.30) is solvable. The solution w provides us with a symmetric isometric embedding $(w,u_\lambda):\mathrm{M}\hookrightarrow\mathbb{R}^{N+2L}$ so that we have

$$g = \mathrm{d}u_\lambda^2+\mathrm{d}w^2.$$

Now we solve (A.30). Fix V_i and its relative $p\in V_i$. Suppose $w = u+a^2v$ is the solution where $a\in C_c^\infty(V_i)$ with $a=1$ on suppη. We claim if $\varepsilon:=\lambda^{-1}$ is small enough we can find a smooth map $v:N_i\to\mathbb{R}^N$ so that Equation A.30 is solved on V_i.

Equation A.30 can be written as

$$\mathrm{d}(u+a^2v)^2=\mathrm{d}u^2-\frac{1}{\lambda^2}\sum_i^L \mathrm{d}\psi_i^2. \tag{A.31}$$

which after expansion is

$$\begin{aligned}\partial_j(a^2\partial_i u\cdot v)+\partial_i(a^2\partial_j u\cdot v)-2a^2\partial_{ij}u\cdot v+a^4\partial_i v\cdot\partial_j v+\partial_i(a^3\partial_j a|v|^2)+\partial_j(a^3\partial_i a|v|^2)&\\ =-\frac{1}{\lambda^2}\mathrm{d}\psi_i^2+2a^2(\partial_i a\partial_j a+a\partial_{ij}a)|v|^2&\end{aligned} \tag{A.32}$$

To simplify this equation we will solve the following Dirichlet problem:

$$\begin{cases}\Delta(a\partial_i v\cdot\partial_j v)=\partial_i(a\Delta v\cdot\partial_j v)+\partial_j(a\Delta v\cdot\partial_i v)+r_{ij}(v,a)\\ a\partial_i v\cdot\partial_j v|_{\partial V_i}=0\end{cases}$$

where

$$r_{ij}=\Delta a\partial_i v\cdot\partial_j v-\partial_j a\Delta v\cdot\partial_j v-\partial_j a\partial_i v\Delta v+2\partial_\ell a\partial_\ell(\partial_i v\cdot\partial_j v)+2a(\partial_{i\ell}v\cdot\partial_{j\ell}v-\Delta v\cdot\partial_{ij}v)$$

By solving this equation and multiplying it by a^3, we have

$$\begin{aligned}a^4\partial_i v\cdot\partial_j v=&\partial_i(a^3\Delta^{-1}(a\Delta v\cdot\partial_j v))+\partial_j(a^3\Delta^{-1}(a\Delta v\cdot\partial_i v))-3a^2\partial_i a\Delta^{-1}(a\Delta v\cdot\partial_j v)\\ &-3a^2\partial_j a\Delta^{-1}(a\Delta v\cdot\partial_i v)+a^3\Delta^{-1}r_{ij}(v,a)\end{aligned} \tag{A.33}$$

Plug Equation (A.33) into Equation (A.32) we have

$$\partial_j(a^2\partial_i u\cdot v-a^2N_i(v,a))+\partial_i(a^2\partial_j u\cdot v-a^2N_j(v,a))-2a^2\partial_{ij}u\cdot v=-\frac{1}{\lambda^2}\mathrm{d}\psi_i^2-2a^2M_{ij}(v,a), \tag{A.34}$$

where for $i,j=1,\ldots d$

$$\begin{cases}N_i(v,a)=-a\Delta^{-1}(a\Delta v\cdot\partial_i v)-a\partial_i a|v|^2\\ M_{ij}(v,a)=\frac{1}{2}a\Delta^{-1}r_{ij}(v,a)-(a\partial_{ij}a+\partial_i a\partial_j a)|v|^2-\frac{3}{2}(\partial_i a\Delta^{-1}(a\Delta v\cdot\partial_j v))+\partial_j a\Delta^{-1}(a\Delta v\cdot\partial_i v)\end{cases}$$

Note that by definition and the regularity theory of elliptic operator, we know both $N_i(\cdot,a)$ and $M_{ij}(\cdot,a)$ are maps in $C^\infty(V_i)$. We will solve Equation (A.34) through solving the following differential system:

$$\begin{cases}\partial_i u\cdot v &=& N_i(v,a)\\ \partial_{ij}u\cdot v &=& -\frac{1}{\lambda^2}\mathrm{d}\psi_i^2-M_{ij}(v,a).\end{cases} \tag{A.35}$$

Since by construction we know u has linearly independent $\partial_i u$ and $\partial_{ij}u$, $i,j=1,\ldots,d$, we can solve the under-determined linear system (A.35) by

$$v=E(u)F(v,h), \tag{A.36}$$

where

$$E(u)=\left[\begin{pmatrix}\partial_i u\\ \partial_{ij}u\end{pmatrix}^T\begin{pmatrix}\partial_i u\\ \partial_{ij}u\end{pmatrix}\right]^{-1}\begin{pmatrix}\partial_i u\\ \partial_{ij}u\end{pmatrix}^T$$

and

$$F(v,\varepsilon)=\left(N_i(v,a),-\frac{1}{\lambda^2}\mathrm{d}\psi_i^2-M_{ij}(v,a)\right)^T=\left(N_i(v,a),-\varepsilon^2\mathrm{d}\psi_i^2-M_{ij}(v,a)\right)^T.$$

Next we will apply contraction principle to show the existence of the solution v. Substitute $v=\mu v'$ for some $\mu\in\mathbb{R}$ to be determined later. By the fact that $N_i(0,a)=0$ and $M_{ij}(0,a)=0$ we can rewrite Equation (A.36) as

$$w=\mu E(u)F(v',0)+\frac{1}{\mu}E(u)F(0,\varepsilon).$$

Set

$$\Sigma=\left\{w\in C^{2,\alpha}(V_i,\mathbb{R}^N);\|w\|_{2,\alpha}\leq 1\right\}\quad\text{and}\quad Tw=\mu E(u)F(v',0)+\frac{1}{\mu}E(u)F(0,\varepsilon).$$

By taking

$$\mu=\left(\frac{\|E(u)F(0,\varepsilon)\|_{2,\alpha}}{\|E(u)\|_{2,\alpha}}\right)^{1/2},$$

we have

$$\|Tw\|_{2,\alpha} \leq \mu \|E(u)\|_{2,\alpha}\|F(v',0)\|_{2,\alpha} + \frac{1}{\mu}\|E(u)F(0,\varepsilon)\|_{2,\alpha} = C_1(\|E(u)\|_{2,\alpha}\|E(u)F(0,\varepsilon)\|_{2,\alpha})^{1/2},$$

where C_1 depends only on $\|a\|_{4,\alpha}$. Thus T maps Σ into Σ if $\|E(u)\|_{2,\alpha}\|E(u)F(0,\varepsilon)\|_{2,\alpha} \leq 1/C_1^2$. This can be achieved by taking ε small enough, that is, by taking λ big enough.

Similarly we have

$$\begin{aligned}\|Tw_1 - Tw_2\|_{2,\alpha} \leq & \mu \|E(u)\|_{2,\alpha}\|F(w_1,0) - F(w_2,0)\|_{2,\alpha} \\ \leq & C_2\|w_1 - w_2\|_{2,\alpha}(\|E(u)\|_{2,\alpha}\|E(u)F(0,\varepsilon)\|_{2,\alpha})^{1/2}.\end{aligned}$$

Then if $\|E(u)\|_{2,\alpha}\|E(u)F(0,\varepsilon)\|_{2,\alpha} \leq \frac{1}{C_1^2+C_2^2}$ we show that T is a contraction map. By the contraction mapping principle, we have a solution $v \in \Sigma$.

Further, since we have

$$v = \mu^2 E(u)F(w,0) + E(u)F(0,\varepsilon),$$

by definition of μ we have

$$\|v\|_{2,\alpha} \leq C\|E(u)F(0,\varepsilon)\|_{2,\alpha},$$

where C is independent of u and v. Thus by taking ε small enough, we can not only make $w = u + a^2 v$ satisfy Equation (A.30) but also make w an embedding. Thus we are done with the patch V_i.

Now we take care V_i's companion $z \circ V_i$. Fix charts around $x \in V_i$ and $z \circ x \in z \circ V_i$ so that $y \in V_i$ and $g \circ y \in z \circ V_i$ have the same coordinates for all $y \in V_i$. Working on these charts we have

$$\partial_j u = \partial_j(\Phi \circ \tilde{u}) = \partial_\ell \Phi \partial_j \tilde{u}^\ell$$

and

$$\partial_{ij} u = \partial_{ij}(\Phi \circ \tilde{u}) = \partial_{k\ell}\Phi \partial_i \tilde{u}^k \partial_j \tilde{u}^\ell + \partial_\ell \Phi \partial_{ij}\tilde{u}^\ell.$$

Note that since the first derivative of Φ is an even function while the second derivative of Φ is an odd function and $\tilde{u}(g \circ y) = -\tilde{u}(y)$ for all $y \in N_i$, we have

$$E(u)(z \circ x) = -E(u)(x).$$

Moreover, we have $N_i(v,a) = N_i(-v,a)$ and $M_{ij}(v,a) = M_{ij}(-v,a)$ for all $i,j = 1,\ldots,d$. Thus in $g \circ N_i$, we have $-v$ as the solution to Equation (A.35) and $w - a^2 v$ as the modified embedding. After finishing the perturbation of V_i and $z \circ V_i$, the modified embedding is again symmetric.

Inductively we can perturb the embedding of V_i for all $i = 1,\ldots,L$. Since there are only finite patches, by choosing ε small enough, we finish the proof. □

Note that we do not show the optimal dimension p of the embedded Euclidean space but simply show the existence of the symmetric isometric embedding. How to take the symmetry into account in the optimal isometric embedding will be reported in the future work.

Corollary A.4.1. Any smooth, closed non-orientable manifold (M,g) has an orientable double covering embedded symmetrically inside $\mathbb{R}^p$ for some $p \in \mathbb{N}$.

Proof. It is well known that the orientable double covering of M has isometric free $\mathbb{Z}_2$ action. By applying Theorem A.4.5 we get the result. □

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, FINE HALL, WASHINGTON ROAD, PRINCETON NJ, USA

E-mail address: amits@math.princeton.edu

DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY, STANFORD CA 94305 USA

E-mail address: hauwu@stanford.edu