Appl. Comput. Harmon. Anal. 31 (2011) 44–58

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Orientability and diffusion maps

Amit Singer[a,*], Hau-tieng Wu[b]

[a] Department of Mathematics and PACM, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA
[b] Department of Mathematics, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA

ARTICLE INFO

Article history:
Received 27 June 2010
Revised 1 October 2010
Accepted 5 October 2010
Available online 8 October 2010
Communicated by Charles K. Chui

Keywords:
Diffusion maps
Orientability
Dimensionality reduction

ABSTRACT

One of the main objectives in the analysis of a high dimensional large data set is to learn its geometric and topological structure. Even though the data itself is parameterized as a point cloud in a high dimensional ambient space $\mathbb{R}^p$, the correlation between parameters often suggests the "manifold assumption" that the data points are distributed on (or near) a low dimensional Riemannian manifold $\mathcal{M}^d$ embedded in $\mathbb{R}^p$, with $d \ll p$. We introduce an algorithm that determines the orientability of the intrinsic manifold given a sufficiently large number of sampled data points. If the manifold is orientable, then our algorithm also provides an alternative procedure for computing the eigenfunctions of the Laplacian that are important in the diffusion map framework for reducing the dimensionality of the data. If the manifold is non-orientable, then we provide a modified diffusion mapping of its orientable double covering.

1. Introduction

The analysis of massive data sets is an active area of research with applications in many diverse fields. Compared with traditional data analysis, we now analyze larger data sets with more parameters; in many cases, the data is also noisy. Dealing with high dimensional, large, and possibly noisy data sets is the main focus of this area. Among many approaches, dimensionality reduction plays a central role. Its related researches and applications can be seen in areas such as statistics, machine learning, sampling theory, image processing, computer vision and more. Based on the observation that many parameters are correlated to each other, from the right scale of observation, a popular assumption is that the collected data is sampled from a low dimensional manifold embedded in the high dimensional ambient Euclidean space. Under this assumption, many algorithms have been proposed to analyze the data. Some examples are local linear embedding [14], Laplacian eigenmaps [2], Hessian eigenmaps [8], ISOMAP [17], and diffusion maps [6]. These non-linear methods are often found to be superior to traditional linear dimensionality reduction methods, such as principal component analysis (PCA) and classical multidimensional scaling (MDS), since they are able to capture the global non-linear structure while preserving the local linear structures. Besides dimensionality reduction, data analysts are often faced with problems such as classification, clustering and statistical inference. Clearly, the characteristics of the manifold are useful to solve such problems. A popular approach to extract the spectral geometry of the manifold is by constructing the Laplace operator and examining its eigenfunctions and eigenvalues [6]. The usefulness of this approach is validated by a series of theoretical works [12,3,9,6,16].

In this paper we consider the following question: Given n data points $x_1, x_2, \dots, x_n$ sampled from a low dimensional manifold $\mathcal{M}^d$ but viewed as points in a high dimensional ambient space $\mathbb{R}^p$, is it possible to determine whether or not the manifold is orientable? For example, can we decide that the Möbius band is non-orientable from just observing a finite number of sampled points? We introduce an algorithm that successfully determines if the manifold is orientable given a

[*] Corresponding author.
E-mail addresses: amits@math.princeton.edu (A. Singer), hauwu@math.princeton.edu (H.-t. Wu).

1063-5203/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.acha.2010.10.001

large enough number n of sampling points. Our algorithm is shown to be robust to noise in the sense that the data points are allowed to be located slightly off the manifold. If the manifold is determined to be orientable, then a byproduct of the algorithm are eigenvectors that can be used for dimensionality reduction, in an analogous way to the diffusion map framework. The algorithm, referred to as Orientable Diffusion Map (ODM), is summarized in Algorithm 1. If the manifold is determined to be non-orientable, then we show how to obtain a modified diffusion map embedding of its orientable double covering using the odd eigenfunctions of the Laplacian over the double cover.

2. The ODM algorithm

In this section we give a full description of ODM. There are three main ingredients to our algorithm: local PCA, alignment and synchronization.

Algorithm 1 Orientable Diffusion Map (ODM)

```
Require: Data points $x_1, x_2, \dots, x_n \in \mathbb{R}^p$.
1: Apply local PCA for each $x_i$ and its $N \ll n$ nearest neighbors to find a $p \times d$ matrix $O_i$ whose columns form an orthonormal basis to a subspace that approximates the tangent space $T_{x_i}\mathcal{M}$.
2: For nearby points $x_i$ and $x_j$ define $O_{ij} = \operatorname{argmin}_{O \in O(d)} \|O - O_i^T O_j\|_F$ given by $O_{ij} = UV^T$ via the SVD $O_i^T O_j = U\Sigma V^T$.
3: Define an $n \times n$ matrix $Z$ whose entries are $z_{ij} = \det O_{ij}$ for nearby points, and $z_{ij} = 0$ otherwise.
4: Define $\mathcal{Z} = D^{-1}Z$, where $D$ is diagonal with $D_{ii} = \sum_{j=1}^{n} |z_{ij}|$.
5: Compute the top eigenvector $v_1^{\mathcal{Z}}$ of $\mathcal{Z}$.
6: Examine the histogram of $v_1^{\mathcal{Z}}$ to decide the orientability of the manifold.
7: If the manifold is orientable, estimate the reflections as $\hat{z}_i = \operatorname{sign} v_1^{\mathcal{Z}}(i)$.
8: Compute the other eigenvectors $v_2, v_3, \dots, v_k$ of $\mathcal{Z}$ corresponding to the largest eigenvalues and embed the data points in $\mathbb{R}^{k-1}$ using
       $x_i \mapsto \hat{z}_i\big(\lambda_2^t v_2^{\mathcal{Z}}(i), \lambda_3^t v_3^{\mathcal{Z}}(i), \dots, \lambda_k^t v_k^{\mathcal{Z}}(i)\big), \quad \text{with } t > 0.$
```

2.1. Local PCA

The first ingredient of the ODM algorithm is local PCA. For every data point x_i we search for its N nearest neighbors $x_{i_1}, x_{i_2}, \dots, x_{i_N}$, where $N \ll n$. This set of neighbors is denoted $\mathcal{N}_{x_i}$. The nearest neighbors of x_i are located near the tangent space $T_{x_i}\mathcal{M}$ to the manifold at x_i, where deviations are possible due to curvature and noise. The result of PCA is an orthonormal basis to a d-dimensional subspace of $\mathbb{R}^p$ which approximates the basis of $T_{x_i}\mathcal{M}$, provided that $N \geqslant d + 1$. The basis found by PCA is represented as a $p \times d$ matrix O_i whose columns are orthonormal, i.e., $O_i^T O_i = I_{d \times d}$.

One problem that we may face in practice is lack of knowledge of the intrinsic dimension d. Estimating the dimensionality of the manifold from sampled points is an active area of research nowadays, and a multiscale version of the local PCA algorithm is presented and analyzed in [11]. We remark that such methods for dimensionality estimation can be easily incorporated in the first step of our ODM algorithm. Thus, we may allow N to vary from one data point to another. In this paper, however, we use the classical approach to local PCA, since it simplifies the presentation. We remark that there is no novelty in our usage of local PCA, and our detailed exposition of this step is merely for the sake of making this paper as self contained as possible. Readers who are familiar with PCA may quickly move on to the next subsection.

In our approach, we try to estimate the intrinsic dimension from the singular values of the mean-shifted local data matrix

$$X_i = [\, x_{i_1} - \mu \quad x_{i_2} - \mu \quad \dots \quad x_{i_N} - \mu \,]$$

of size $p \times N$, where $\mu = \frac{1}{N}\sum_{j=1}^{N} x_{i_j}$. We denote the singular values of X_i by $\sigma_{i,1} \geqslant \sigma_{i,2} \geqslant \dots \geqslant \sigma_{i,N}$. If the neighboring points in $\mathcal{N}_{x_i}$ are located exactly on $T_{x_i}\mathcal{M}$, then $\operatorname{rank} X_i = d$, and there are only d non-vanishing singular values (i.e., $\sigma_{i,d+1} = \dots = \sigma_{i,N} = 0$). In such a case, the dimension can be estimated as the number of non-zero singular values. In practice, however, due to the curvature effect, there may be more than d non-zero singular values. A common practice is to estimate the dimension as the number of singular values that account for high enough percentage of the variability of the data. That is, one sets a threshold γ between 0 and 1 (usually closer to 1 than to 0), and estimates the dimension as the smallest integer d_i for which

$$\frac{\sum_{j=1}^{d_i} \sigma_{i,j}^2}{\sum_{j=1}^{N} \sigma_{i,j}^2} > \gamma.$$

For example, setting $\gamma = 0.9$ means that d_i singular values account for at least 90% variability of the data, while $d_i - 1$ singular values account for less than 90%. We refer to the smallest integer d_i as the estimated local dimension of $\mathcal{M}$ at x_i. One possible way to estimate the dimension of the manifold would be to use the mean of the estimated local dimensions $d_1, \dots, d_n$, that is, $\hat{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$ (and then round it to the closest integer). The mean estimator minimizes the sum of squared errors $\sum_{i=1}^{n}(d_i-\hat{d})^2$. We estimate the intrinsic dimension of the manifold by the median value of all the d_i's, that is, we define the estimator $\hat{d}$ for the intrinsic dimension d as

$$\hat{d}=\operatorname{median}\{d_1,d_2,\ldots,d_n\}.$$

The median has the property that it minimizes the sum of absolute errors $\sum_{i=1}^{n}|d_i-\hat{d}|$ (originally due to Laplace). As a result, estimating the intrinsic dimension by the median is more robust to outliers compared to the mean estimator. In all proceeding steps of the algorithm we use the median estimator $\hat{d}$, but in order to facilitate the notation we write d instead of $\hat{d}$.

Suppose that the SVD of X_i is given by

$$X_i=U_i\Sigma_iV_i^T.$$

The columns of the $p\times N$ matrix U_i are orthonormal and are known as the left singular vectors

$$U_i=[\,u_{i_1}\quad u_{i_2}\quad\ldots\quad u_{i_N}\,].$$

We define the matrix O_i by the first d left singular vectors (corresponding to the largest singular values):

$$O_i=[\,u_{i_1}\quad u_{i_2}\quad\ldots\quad u_{i_d}\,].\tag{1}$$

Note that the columns of O_i form a numerical approximation to an orthonormal basis of the tangent plane $T_{x_i}\mathcal{M}$.

2.2. Alignment

The second ingredient of the ODM algorithm is alignment. Suppose x_i and x_j are two nearby points, satisfying[1] $x_j\in\mathcal{N}_{x_i}$ and $x_i\in\mathcal{N}_{x_j}$. Since $N\ll n$, the curvature effect is small, and the tangent spaces $T_{x_i}\mathcal{M}$ and $T_{x_j}\mathcal{M}$ are also close.[2] As a result, the matrix $O_i^TO_j$ is not necessarily orthogonal, and we define O_{ij} as its closest orthogonal matrix, i.e.,

$$O_{ij}=\underset{O\in O(d)}{\operatorname{argmin}}\big\|O-O_i^TO_j\big\|_F,\tag{2}$$

where $\|\cdot\|_F$ is the Hilbert–Schmidt norm (also known as the Frobenius norm). The minimization problem (2) has a simple solution via the singular value decomposition (SVD) of $O_i^TO_j$ [1]. Specifically, if

$$O_i^TO_j=U\Sigma V^T\tag{3}$$

is the SVD of $O_i^TO_j$, then

$$O_{ij}=UV^T\tag{4}$$

solves (2). The optimal orthogonal transformation O_{ij} can be either a rotation (i.e., an element of $SO(d)$), or a composition of a rotation and a reflection. Fig. 1 is an illustration of this procedure.

The determinant of O_{ij} classifies the two possible cases:

$$\det O_{ij}=\begin{cases}1 & \text{optimal transformation does not include a reflection,}\\ -1 & \text{optimal transformation includes a reflection.}\end{cases}\tag{5}$$

We refer to the process of finding the optimal orthogonal transformation between bases as alignment. Note that not all bases are aligned; only the bases of nearby points are aligned. We set $G=(V,E)$ to be the undirected graph with n vertices corresponding to the data points, where an edge between i and j exists iff their corresponding bases are aligned by the algorithm. We further encode the information about the reflections in a symmetric $n\times n$ matrix Z whose elements are given by

$$Z_{ij}=\begin{cases}\det O_{ij} & (i,j)\in E,\\ 0 & (i,j)\notin E.\end{cases}\tag{6}$$

That is, $Z_{ij}=1$ if no reflection is needed, $Z_{ij}=-1$ if a reflection is needed, and $Z_{ij}=0$ if the points are not nearby.

[1] Other notions of proximity are also possible. For example, x_i and x_j are nearby if they have at least c common points, that is, if $|\mathcal{N}_{x_i}\cap\mathcal{N}_{x_j}|\geqslant c$, where $c\geqslant 1$ is a parameter.

[2] This means that their Grassmannian distance defined via the operator norm $\|O_iO_i^T-O_jO_j^T\|$ is small.

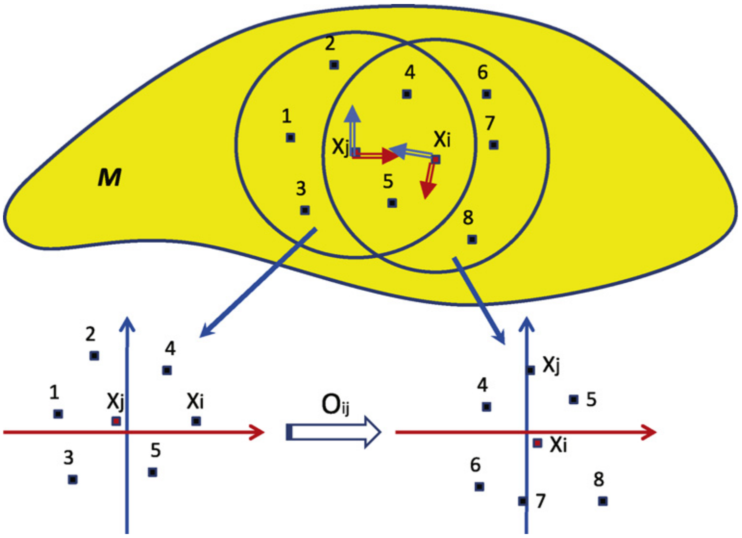

Fig. 1. The orthonormal bases of the tangent planes $T_{x_i}\mathcal{M}$ and $T_{x_i}\mathcal{M}$ determined by local PCA are marked on top of the manifold by double arrows. The lower part demonstrates the projection of the data points onto the subspaces spanned by the orthonormal bases (the tangent planes). The coloring of the axes (blue and red) correspond to the coloring of the basis vectors. In this case, O_{ij} is a composition of a rotation and a reflection. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.3. Synchronization

The third ingredient of the ODM algorithm is synchronization. If the manifold is orientable, then we can assign an orientation to each tangent space in a continuous way. For each x_i, the basis found by local PCA can either agree or disagree with that orientation. We may therefore assign a variable $z_i \in \{+1, -1\}$ that designates if the PCA basis at x_i agrees with the orientation ($z_i = 1$) or not ($z_i = -1$). Clearly, for nearby points x_i and x_j whose bases are aligned, we must have that

$$z_i z_j^{-1} = z_{ij}, \quad (i, j) \in E. \tag{7}$$

That is, if no reflection was needed in the alignment stage (i.e., $z_{ij} = 1$), then either both PCA bases at x_i and x_j agree with the orientation ($z_i = z_j = 1$) or both disagree with the orientation ($z_i = z_j = -1$). Similarly, if a reflection was needed ($z_{ij} = -1$), then it must be the case that $z_i = -z_j$. In practice, however, we are not given an orientation for the manifold, so the values of the z_i's are unknown. Instead, only the values of the z_{ij}'s are known to us after the alignment stage. Thus, we may try solving the system (7) to find the z_i's. A solution exists iff the manifold is orientable. The solution is unique up to a global sign change, i.e., if $(z_1, z_2, \ldots, z_n)$ is a solution, then $(-z_1, -z_2, \ldots, -z_n)$ is also a solution. If all equations in (7) are accurate and the graph G is connected, then finding a solution can be obtained in a straightforward manner by simply traversing a spanning tree of the graph from the root to the leafs.

In practice, however, due to curvature effects and noise, some of the equations in (7) may be incorrect and we would like to find a solution that satisfies as many equations as possible. This maximum satisfiability problem is also known as the synchronization problem over the group $\mathbb{Z}_2$ [15], and in [15,7] we proposed spectral algorithms to approximate its solution. In [15] we used the top eigenvector (corresponding to the largest eigenvalue) of Z to find the solution, while in [7] we normalized the matrix Z and used the top eigenvector of that normalized matrix. Here we use the algorithm in [7], since we find the normalization to give much improved results.

Specifically, we normalize the matrix Z by dividing the elements of each row by the number of non-zero elements in that given row, that is,

$$\mathcal{Z} = D^{-1} Z, \tag{8}$$

where D is a $n \times n$ diagonal matrix with $D_{ii} = \sum_{j=1}^{n} |z_{ij}|$. In other words, $D_{ii} = \deg(i)$, where $\deg(i)$ is the degree of vertex i in the graph G. We refer to $\mathcal{Z}$ as the reflection matrix. Note that although $\mathcal{Z}$ is not necessarily symmetric, it is similar to the symmetric matrix $D^{-1/2} Z D^{-1/2}$ through

$$\mathcal{Z} = D^{-1/2}\left(D^{-1/2} Z D^{-1/2}\right) D^{1/2}.$$

Therefore, the matrix $\mathcal{Z}$ has n real eigenvalues $\lambda_1^{\mathcal{Z}} > \lambda_2^{\mathcal{Z}} \geqslant \cdots \geqslant \lambda_n^{\mathcal{Z}}$ and n orthonormal eigenvectors $v_1^{\mathcal{Z}}, \ldots, v_n^{\mathcal{Z}}$, satisfying

$$\mathcal{Z} v_i^{\mathcal{Z}} = \lambda_i^{\mathcal{Z}} v_i^{\mathcal{Z}}.$$

We compute the top eigenvector $v_1^{\mathcal{Z}} \in \mathbb{R}^n$ of $\mathcal{Z}$, which satisfies

$$\mathcal{Z} v_1^{\mathcal{Z}} = \lambda_1^{\mathcal{Z}} v_1^{\mathcal{Z}}, \tag{9}$$

and use it to obtain estimators $\hat{z}_1, \ldots, \hat{z}_n$ for the unknown variables $z_1, \ldots, z_n$, in the following way:

$$\hat{z}_i = \operatorname{sign}\bigl(v_1^{\mathcal{Z}}(i)\bigr) = \frac{v_1^{\mathcal{Z}}(i)}{|v_1^{\mathcal{Z}}(i)|}, \quad i = 1, 2, \ldots, n. \tag{10}$$

The estimation by the top eigenvector is up to a global sign, since if $v_1^{\mathcal{Z}}$ is the top eigenvector of $\mathcal{Z}$ then so is $-v_1^{\mathcal{Z}}$.

To see why (10) is a good estimator, we first analyze it under the assumption that the matrix Z contains no errors on the relative reflections. Denoting by Υ the $n \times n$ diagonal matrix with ± 1 on its diagonal representing the correct reflections z_i, i.e. $\Upsilon_{ii} = z_i$, we can write the matrix Z as

$$Z = \Upsilon A \Upsilon^{-1}, \tag{11}$$

where A is the adjacency matrix of the graph G whose elements are given by

$$a_{ij} = \begin{cases} 1 & (i,j) \in E, \\ 0 & (i,j) \notin E, \end{cases} \tag{12}$$

because in the error-free case $z_{ij} = z_i z_j^{-1}$ for $(i,j) \in E$. The reflection matrix $\mathcal{Z}$ can now be written as

$$\mathcal{Z} = \Upsilon \bigl(D^{-1} A\bigr) \Upsilon^{-1}. \tag{13}$$

Hence, $\mathcal{Z}$ and $D^{-1}A$ share the same eigenvalues. Since the normalized discrete graph Laplacian $\mathcal{L}$ of the graph G is defined as

$$\mathcal{L} = I - D^{-1} A, \tag{14}$$

it follows that in the error-free case, the eigenvalues of $I - \mathcal{Z}$ are the same as the eigenvalues of $\mathcal{L}$. These eigenvalues are all non-negative, since $\mathcal{L}$ is similar to the positive semidefinite matrix $I - D^{-1/2} A D^{-1/2}$, whose non-negativity follows from the identity

$$v^T \bigl(I - D^{-1/2} A D^{-1/2}\bigr) v = \sum_{(i,j) \in E} \left(\frac{v(i)}{\sqrt{\deg(i)}} - \frac{v(j)}{\sqrt{\deg(j)}} \right)^2 \geqslant 0,$$

for any $v \in \mathbb{R}^n$. In other words,

$$1 - \lambda_i^{\mathcal{Z}} = \lambda_i^{\mathcal{L}} \geqslant 0, \quad i = 1, 2, \ldots, n, \tag{15}$$

where the eigenvalues of $\mathcal{L}$ are ordered in increasing order, i.e., $\lambda_1^{\mathcal{L}} \leqslant \lambda_2^{\mathcal{L}} \leqslant \cdots \leqslant \lambda_n^{\mathcal{L}}$, and the corresponding eigenvectors $v_i^{\mathcal{L}}$ satisfy $\mathcal{L} v_i^{\mathcal{L}} = \lambda_i^{\mathcal{L}} v_i^{\mathcal{L}}$. Furthermore, the sets of eigenvectors are related by

$$v_i^{\mathcal{Z}} = \Upsilon v_i^{\mathcal{L}}. \tag{16}$$

If the graph G is connected, then the eigenvalue $\lambda_1^{\mathcal{L}} = 0$ is simple and its corresponding eigenvector $v_1^{\mathcal{L}}$ is the all-ones vector $\mathbf{1} = (1, 1, \ldots, 1)^T$. Therefore,

$$v_1^{\mathcal{Z}} = \Upsilon \mathbf{1}, \tag{17}$$

and, in particular,

$$v_1^{\mathcal{Z}}(i) = z_i, \quad i = 1, 2, \ldots, n. \tag{18}$$

This implies that in the error-free case, (10) perfectly recovers the reflections. In other words, the eigenvector $v_1^{\mathcal{Z}}$ contains the orientation information of each tangent space. According to this orientation information, we can synchronize the orientations of all tangent spaces by flipping the orientation of the basis O_i whenever $v_1^{\mathcal{Z}}(i) < 0$. In conclusion, the above steps synchronize all the tangent spaces to agree on their orientations.

If the manifold is not orientable, then it is impossible to synchronize the orientation of the tangent spaces and the top eigenvector cannot be interpreted as before. We return to this interpretation later in the paper. For the moment, we just remark that the eigenvalues of A and Z (and similarly of $\mathcal{L}$ and $\mathcal{Z}$) contain information about the number of orientation preserving and orientation reversing closed loops. To that end, consider first $A^t(i,j)$, which is the number of paths of length t over the graph G that start at node i and end at node j. In particular, $A^t(i,i)$ is the number of closed loops that start and end at node i. Therefore,

$$\operatorname{Tr}\bigl(A^t\bigr) = \sum_{i=1}^{n} A^t(i,i) = \sum_{i=1}^{n} \bigl(\lambda_i^A\bigr)^t$$

is the number of closed loops of length t. Similarly,

$$\operatorname{Tr}\bigl(Z^t\bigr) = \sum_{i=1}^{n} Z^t(i,i) = \sum_{i=1}^{n} \bigl(\lambda_i^Z\bigr)^t$$

is the number of orientation preserving closed loops minus the number of orientation reversing closed loops of length t. If the manifold is orientable then all closed loops are orientation preserving and the two traces are the same, but if the manifold is non-orientable then there are orientation reversing closed loops and therefore the eigenvalues of A and Z must differ.

If the manifold is determined to be orientable, then we can use the computed eigenvectors of $\mathcal{Z}$ for dimensionality reduction in the following manner. Specifically, we use point-wise multiplication of the eigenvectors of $\mathcal{Z}$ by the estimated reflections to produce new vectors $\tilde{v}_1, \ldots, \tilde{v}_n$ that are given by

$$\tilde{v}_j(i) = \hat{z}_i v_j^{\mathcal{Z}}(i), \quad i, j = 1, \ldots, n. \tag{19}$$

The relation (16) between the eigenvectors of $\mathcal{Z}$ and $\mathcal{L}$, and the particular form of the top eigenvector $v_1^{\mathcal{Z}}$ given in (17) imply that the vectors $\tilde{v}_1, \ldots, \tilde{v}_n$ are eigenvectors of $\mathcal{L}$ with $\tilde{v}_j = v_j^{\mathcal{L}}$. The eigenvectors of $\mathcal{L}$ are used in the diffusion map framework to embed the data points onto a lower dimensional Euclidean space, where distances between points approximate the diffusion distance [6]. Similarly, we can embed the data points using the following rule:

$$x_i \mapsto \hat{z}_i\big(\lambda_2^t v_2^{\mathcal{Z}}(i), \lambda_3^t v_3^{\mathcal{Z}}(i), \ldots, \lambda_k^t v_k^{\mathcal{Z}}(i)\big) \in \mathbb{R}^{k-1}, \tag{20}$$

where $t \geqslant 0$ is a parameter.

If the data points are sampled from the uniform distribution over the manifold, then the eigenvectors of the discrete graph Laplacian $\mathcal{L}$ approximate the eigenfunctions of the Laplace–Beltrami operator [4], and it follows that the vectors $\tilde{v}_i$ that are used in our ODM embedding (20) also converge to the eigenfunctions of the Laplace–Beltrami operator. If the sampling process is from a different distribution, then the $\tilde{v}_i$'s approximate the eigenfunctions of the Fokker–Planck operator. By adapting the normalization rule suggested in [6] to our matrix Z, we can get an approximation of the eigenfunctions of the Laplace–Beltrami operator also for non-uniform distributions.

We remark that nothing prevents us from running the ODM algorithm in cases for which the sampling process from the manifold is noisy. We demonstrate the robustness of our algorithm through several numerical experiments in Section 4.

3. Orientable double covering

Every non-orientable manifold has an orientable double cover. Can we reconstruct the double covering from data points sampled only from the non-orientable manifold? In this section we give an affirmative answer to this question, by constructing a modified diffusion map embedding of the orientable double cover using only points that are sampled from the non-orientable manifold.

For simplicity of exposition, we start by making a simplifying assumption that we shall later relax. The assumption is that the orientable double cover has a symmetric isometric embedding in Euclidean space. That is, we assume that if $\mathcal{M}$ is the non-orientable manifold, and $\tilde{\mathcal{M}}$ is its orientable double cover, then $\tilde{\mathcal{M}}$ has a symmetric isometric embedding in $\mathbb{R}^q$, for some $q \geqslant 1$. Symmetric embedding means that for all $x \in \tilde{\mathcal{M}} \subset \mathbb{R}^q$, also $-x \in \tilde{\mathcal{M}}$. This assumption is motivated by examples: the orientable double cover of the Möbius band is the cylinder, the orientable double cover of $\mathbb{R}P^2$ is S^2 and the orientable double cover of the Klein bottle is T^2. The cylinder, S^2 and T^2 all have symmetric isometric embeddings in $\mathbb{R}^3$ ($q = 3$). We believe that this assumption holds true in general, but its proof is beyond the scope of this paper.[3] Our assumption means that we have a $\mathbb{Z}_2$ action on $\tilde{\mathcal{M}}$ given by $x \mapsto -x$. We note that this $\mathbb{Z}_2$ action is isometric and recall that if the group $\mathbb{Z}_2$ is properly discontinuously acting on $\tilde{\mathcal{M}}$, then $\tilde{\mathcal{M}}/\mathbb{Z}_2$ is non-orientable if and only if $\tilde{\mathcal{M}}$ has an orientation that is not preserved by the $\mathbb{Z}_2$ action.

By our assumption, the orientable double covering $\tilde{\mathcal{M}}$ can be isometrically embedded into $\mathbb{R}^q$ for some q so that $\mathbb{Z}_2 x = -x$ for all $x \in \tilde{\mathcal{M}}$ and $\tilde{\mathcal{M}}/\mathbb{Z}_2 = \mathcal{M}$. Denote the projection map as $\pi : \{x, -x\} \subset \tilde{\mathcal{M}} \mapsto [x] \in \mathcal{M}$. Under this assumption, each pair of antipodal points in $\tilde{\mathcal{M}}$ can be viewed as a single point in $\mathcal{M}$. In other words, for any point $[x] \in \mathcal{M}$, we can view it as two antipodal points $x, -x \in \tilde{\mathcal{M}}$ whose antipodal relationship is ignored.

We can now offer an interpretation of the matrix Z whose entries are given in (6). Recall that we have n data points in $\mathbb{R}^p$ that are sampled from $\mathcal{M}$. We denote these points by $[x_1], \ldots, [x_n]$. Associate to each $[x_i] \in \mathcal{M}$ a point $x_i \in \tilde{\mathcal{M}} \subset \mathbb{R}^q$ such that $\pi(x_i) = [x_i]$ and such that the entries of Z satisfy the following conditions: $z_{ij} = 1$ if x_j is in the ϵ-neighborhood of x_i; $z_{ij} = -1$ if x_j is the ϵ-neighborhood of $\mathbb{Z}_2 x_i = -x_i$; and $z_{ij} = 0$ if the distance between x_j and x_i and the distance between x_j and $-x_i$ are both larger than ϵ. That is,

$$z_{ij} = \begin{cases} 1 & \text{if } \|x_i - x_j\|_{\mathbb{R}^q} < \epsilon, \\ -1 & \text{if } \|x_i + x_j\|_{\mathbb{R}^q} < \epsilon, \\ 0 & \text{otherwise.} \end{cases}$$

[3] We elude to a possible proof by noting that Nash embedding theorem, that guarantees that every Riemannian manifold can be isometrically embedded in Euclidean space, is based on modified Newton iterations. Thus, a possible roadmap for the proof would be to show that Whitney embedding theorem gives a symmetric embedding as an initialization, and then to show that the iterations in Nash theorem preserve symmetry.

Next, we build a $2n \times 2n$ matrix $\tilde{Z}$ as follows:

$$\tilde{Z} := \begin{bmatrix} Z & -Z \\ -Z & Z \end{bmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes Z, \tag{21}$$

where $\otimes$ is the Kronecker product. The matrix $\tilde{Z}$ corresponds to $2n$ points over $\tilde{\mathcal{M}}$, while originally we had only n such points. We identify the additional points, denoted $x_{n+1}, \ldots, x_{2n}$ as $x_{n+i} = -x_i$ for $i = 1, \ldots, n$. With this identification, it is easily verified that the entries of $\tilde{Z} = (\tilde{z}_{ij})_{i,j=1}^{2n}$ are also given by

$$\tilde{z}_{ij} = \begin{cases} 1 & \text{if } \|x_i - x_j\|_{\mathbb{R}^q} < \epsilon, \\ -1 & \text{if } \|x_i + x_j\|_{\mathbb{R}^q} < \epsilon, \\ 0 & \text{otherwise.} \end{cases}$$

In a sense, the $-Z$ in the $(1,2)$ and $(2,1)$ entries of $\tilde{Z}$ in (21) provides the lost information between $\pi^{-1}([x_i])$ and $\pi^{-1}([x_j])$. We also define the matrices $\tilde{\mathcal{Z}}$ and $\tilde{\mathcal{L}}$ as

$$\tilde{\mathcal{Z}} = \begin{bmatrix} \mathcal{Z} & -\mathcal{Z} \\ -\mathcal{Z} & \mathcal{Z} \end{bmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \mathcal{Z} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes D^{-1}Z, \tag{22}$$

and

$$\tilde{\mathcal{L}} = \begin{bmatrix} \mathcal{L} & -\mathcal{L} \\ -\mathcal{L} & \mathcal{L} \end{bmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \mathcal{L} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \left(I - D^{-1}Z\right). \tag{23}$$

It is clear that the matrix $\tilde{\mathcal{Z}}$ is a discretization[4] of the integral operator $\tilde{\mathcal{K}}_\epsilon : L^2(\tilde{\mathcal{M}}) \to L^2(\tilde{\mathcal{M}})$, given by

$$(\tilde{\mathcal{K}}_\epsilon f)(x) = \int\limits_{\tilde{\mathcal{M}}} \tilde{K}_\epsilon(x,y) f(y)\, dy, \tag{24}$$

where the kernel function $\tilde{K}_\epsilon$ is given by

$$\tilde{K}_\epsilon(x,y) = \begin{cases} \frac{1}{\operatorname{vol}(B_\epsilon(x) \cap \tilde{\mathcal{M}})} & \text{if } \|x - y\|_{\mathbb{R}^q} < \epsilon, \\ -\frac{1}{\operatorname{vol}(B_\epsilon(x) \cap \tilde{\mathcal{M}})} & \text{if } \|x + y\|_{\mathbb{R}^q} < \epsilon, \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

$B_\epsilon(x)$ is a ball of radius ϵ in $\mathbb{R}^q$ centered at x, and the volume is measured using the metric of $\tilde{\mathcal{M}}$ induced from $\mathbb{R}^q$. Clearly, the integral operator $\tilde{\mathcal{K}}_\epsilon$ commutes with the $\mathbb{Z}_2$ action:

$$\tilde{\mathcal{K}}_\epsilon \mathbb{Z}_2 f = \mathbb{Z}_2 \tilde{\mathcal{K}}_\epsilon f, \tag{26}$$

where $(\mathbb{Z}_2 f)(x) = f(-x)$. Moreover, the kernel of $\tilde{\mathcal{K}}_\epsilon$ includes all "even" functions. That is, if f satisfies $f(x) = f(-x)$ for all $x \in \tilde{\mathcal{M}}$, then $\tilde{\mathcal{K}}_\epsilon f = 0$. It follows that the eigenfunctions of $\tilde{\mathcal{K}}_\epsilon$ with non-vanishing eigenvalues must be "odd" functions that satisfy $f(x) = -f(-x)$.

Recall that the graph Laplacian $\mathcal{L}$ converges to the Laplace–Beltrami operator in the limit $n \to \infty$ and $\epsilon \to 0$ if the data points are sampled from the uniform distribution over the manifold. That is,

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{\epsilon} \mathcal{L} f(x) = m_2 \Delta_{\mathcal{M}} f(x), \tag{27}$$

where m_2 is a constant related to the second moment of the kernel function. Adapting this result to our case implies that we have the following convergence result for the matrix $\tilde{\mathcal{L}}$:

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{\epsilon} (\tilde{\mathcal{L}} f)(x) = \tilde{m}_2 \left[(\Delta_{\tilde{\mathcal{M}}} f)(x) - (\Delta_{\tilde{\mathcal{M}}} f)(-x) \right], \tag{28}$$

where $\tilde{m}_2$ is a constant. In other words, in the limit we recover the following operator:

$$\Delta_{\tilde{\mathcal{M}}} - \mathbb{Z}_2 \Delta_{\tilde{\mathcal{M}}}. \tag{29}$$

The Laplacian $\Delta_{\tilde{\mathcal{M}}}$ also commutes with the $\mathbb{Z}_2$ action, so by similar considerations, the kernel of $\Delta_{\tilde{\mathcal{M}}} - \mathbb{Z}_2 \Delta_{\tilde{\mathcal{M}}}$ contains all even functions, and the eigenfunctions with non-zero eigenvalues are odd functions.

[4] That is, $\tilde{\mathcal{Z}}$ converges to $\tilde{\mathcal{K}}_\epsilon$ in the limit $n \to \infty$.

Suppose $\phi_1, \phi_2, \ldots, \phi_{2n}$ and $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{2n} \geqslant 0$ are the eigenvectors and eigenvalues of $\tilde{\mathcal{L}}$. From the above discussion we have that $\lambda_{n+1} = \cdots = \lambda_{2n} = 0$ and $\phi_1, \ldots, \phi_n$ are odd vectors satisfying $\phi_l(i) = -\phi_l(n+i)$ $(1 \leqslant l \leqslant n)$. We define the map $\tilde{\Phi}_t$ for $t > 0$ as

$$\tilde{\Phi}_t : x_i \mapsto \left(\lambda_1^t \phi_1(i), \lambda_2^t \phi_2(i), \ldots, \lambda_n^t \phi_n(i)\right), \quad \text{for } i = 1, 2, \ldots, 2n. \tag{30}$$

We refer to $\tilde{\Phi}_t$ as a modified diffusion map. Note that $\tilde{\Phi}_t$ uses only the odd eigenfunctions of the Laplacian over $\tilde{\mathcal{M}}$, whereas the classical diffusion map uses all eigenfunctions, both odd and even.

Our construction shows that from the information contained in the matrix Z, which is built solely from the non-orientable manifold $\mathcal{M}$, we obtain the odd eigenfunctions of the Laplacian over the double cover $\tilde{\mathcal{M}}$. The odd eigenfunctions are extremely useful for embedding the double cover. For example, the double cover of $\mathbb{R}P^2$ is S^2 whose odd eigenfunctions are the odd degree spherical harmonics. In particular, the degree-1 spherical harmonics are the coordinate functions x, y and z (in $\mathbb{R}^3$), so truncating the modified diffusion map $\tilde{\Phi}_t$ to $\mathbb{R}^3$ results in an embedding of S^2 in $\mathbb{R}^3$. Similarly, the first 3 eigenvectors of $\tilde{\mathcal{L}}$ (or equivalently, $\tilde{\mathcal{Z}}$) built up from the Klein bottle and the Möbius band allow us to get a torus and a cylinder, respectively. These reconstruction results are shown in Section 4.

We now return to the underlying assumption that the manifold $\tilde{\mathcal{M}}$ has a symmetric isometric embedding in $\mathbb{R}^q$ for some $q \geqslant 1$. While the assumption was useful to visualize the construction, we note that it is actually not necessary for the above derivation and interpretation. Indeed, the non-orientable manifold $\mathcal{M}$ has an abstract orientable double cover $\tilde{\mathcal{M}}$ with $\mathbb{Z}_2$ acting on it such that $\tilde{\mathcal{M}}/\mathbb{Z}_2 = \mathcal{M}$. While the assumption allowed us to make a concrete specification of the $\mathbb{Z}_2$ action (i.e., $\mathbb{Z}_2 x = -x$), this specification is not necessary in the abstract setup. The careful reader may check each and every step of the derivation and verify that they follow through also in the abstract sense. In particular, the matrix $\tilde{\mathcal{Z}}$ can be interpreted just as before, with $x_{n+i} = \mathbb{Z}_2 x_i \in \tilde{\mathcal{M}}$ for $\mathbb{Z}_2$ being the abstract action. Moreover, the Laplacian over $\tilde{\mathcal{M}}$ still commutes with $\mathbb{Z}_2$ and the matrix $\tilde{\mathcal{L}}$ converges to the operator given in (29). Thus, the modified diffusion mapping $\tilde{\Phi}_t$ given in (30) is an embedding of the abstract double cover. In the continuous setup, the eigenfunctions $\{\phi_n\}_{n=1}^{\infty}$ of the operator $\Delta_{\tilde{\mathcal{M}}} - \mathbb{Z}_2 \Delta_{\tilde{\mathcal{M}}}$ embed the abstract double cover $\tilde{\mathcal{M}}$ in the Hilbert space ℓ_2 through the mapping

$$x \mapsto \left(\lambda_1^t \phi_1(x), \lambda_2^t \phi_2(x), \ldots, \lambda_n^t \phi_n(x), \ldots\right), \quad \text{for } x \in \tilde{\mathcal{M}}, \tag{31}$$

where

$$(\Delta_{\tilde{\mathcal{M}}} - \mathbb{Z}_2 \Delta_{\tilde{\mathcal{M}}})\phi_n = \lambda_n \phi_n, \quad n = 1, 2, \ldots.$$

Finally, we show a useful “trick” to generate a point cloud over a non-orientable manifold without knowing any of the parametrizations that embed it in Euclidean space. For example, suppose we want to generate a point cloud over the Klein bottle, but we forgot its parametrization in $\mathbb{R}^4$ (the parametrization is given later in (32)). There is still a rather simple way to proceed that only requires the knowledge of the orientable double cover. That is, given a non-orientable manifold $\mathcal{M}$ from which we want to sample points in a Euclidean space, all we need to know is its orientable double cover $\tilde{\mathcal{M}}$ and its symmetric embedding in Euclidean space.[5] So, although we forgot the parametrization of the Klein bottle in $\mathbb{R}^4$, we assume that we still remember that its double cover is T^2 and the parametrization of T^2 in $\mathbb{R}^3$. From the symmetric embedding of T^2 in $\mathbb{R}^3$, we now describe how to get a diffusion map embedding of the Klein bottle in Euclidean space. In general, from the symmetric embedding of $\tilde{\mathcal{M}}$ we describe how to get the diffusion map embedding of $\mathcal{M}$.

We start by sampling n points uniformly from $\tilde{\mathcal{M}} \in \mathbb{R}^q$. Next, we double the number of points by setting $x_{n+i} = -x_i$ for $i = 1, \ldots, n$. We then build a $2n \times 2n$ matrix $\tilde{A} = (\tilde{a}_{ij})_{i,j=1}^{2n}$ given by

$$\tilde{a}_{ij} = \begin{cases} 1 & \text{if } \|x_i - x_j\|_{\mathbb{R}^q} < \epsilon, \\ 1 & \text{if } \|x_i + x_j\|_{\mathbb{R}^q} < \epsilon, \\ 0 & \text{otherwise}. \end{cases}$$

We see that the matrix $\tilde{A}$ does not distinguish between x and $\mathbb{Z}_2 x$, which leads to the previous identification $\{x, \mathbb{Z}_2 x\} = [x] \in \mathcal{M}$. After normalizing the matrix as before, it can be easily verified that it converges to an integral operator that has all the odd functions in its kernel and its non-zero eigenvalues correspond to even eigenfunctions. Moreover, in the limit of $n \to \infty$ and $\epsilon \to 0$ we get convergence to the following operator:

$$\Delta_{\tilde{\mathcal{M}}} + \mathbb{Z}_2 \Delta_{\tilde{\mathcal{M}}}.$$

This means that the computed eigenvectors approximate the even eigenfunctions of the Laplacian over $\tilde{\mathcal{M}}$. That is, we compute eigenfunctions that satisfy $\phi_l(x) = \phi_l(\mathbb{Z}_2 x)$ (for $l = 1, \ldots, n$). As a result, these eigenfunctions are only functions of $[x]$ and are the eigenfunctions of $\Delta_{\mathcal{M}}$. Thus, the mapping

$$[x_i] \mapsto \left(\lambda_1^t \phi_1(x_i), \ldots, \lambda_n^t \phi_n(x_i)\right), \quad \text{for } i = 1, \ldots, n,$$

is the diffusion mapping of $\mathcal{M}$.

[5] Here we actually need the assumption that the orientable double cover has a symmetric isometric embedding in Euclidean space.

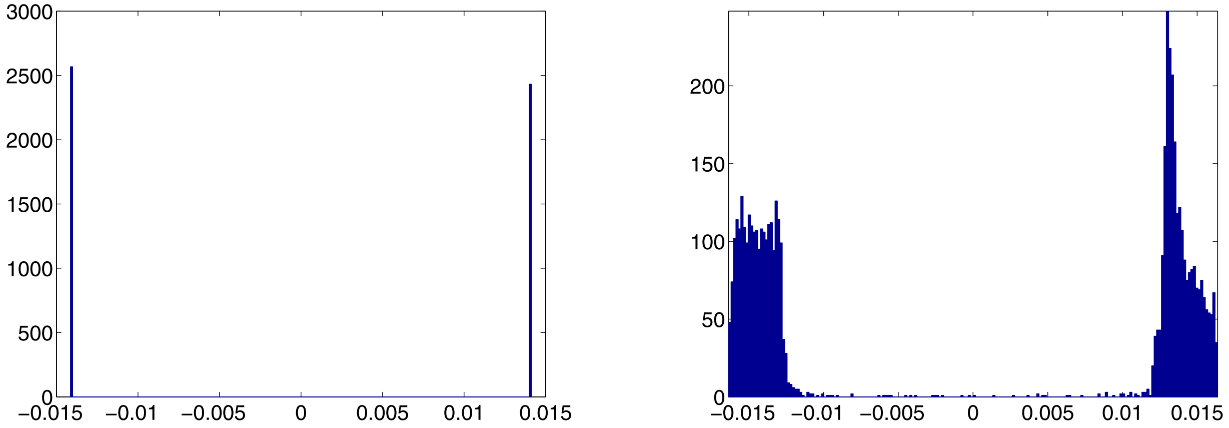

Fig. 2. Histogram of the values of the top eigenvector $v_1^{\mathcal{Z}}$. Left: clean S^2; Right: S^2 contaminated by 5 dB noise. Note that the eigenvector is normalized so its values concentrate around $\pm\frac{1}{\sqrt{n}}$ instead of ±1.

We remark that this construction can help us to build up more examples of non-orientable manifolds. For example, we can start from $S^{2d} \subset \mathbb{R}^{2d+1}$ and build up the diffusion map embedding of the non-orientable real projective space $\mathbb{R}P(2d)$ without knowing the parametrization of $\mathbb{R}P(2d)$.

4. Numerical examples

In this section we provide several numerical experiments demonstrating the application of ODM on simulative data. In our experiments, we measure the level of noise by the signal to noise ratio (SNR), which is defined in the following way. Suppose $s_1, s_2, \ldots, s_n \in \mathbb{R}^p$ is a set of "clean" data points that are located exactly on the manifold. This set of points is considered as the signal, and we define its sample variance as

$$\mathrm{Var}(\mathrm{Signal}) = \frac{1}{np} \sum_{i=1}^{n} \|s_i - \mu\|_{\mathbb{R}^p}^2,$$

where $\mu = \frac{1}{n} \sum_{i=1}^{n} s_i$ is the sample mean. The set of noisy data points $x_1, x_2, \ldots, x_n$ are located off the manifold and are generated using the following process:

$$x_i = s_i + \sigma \xi_i, \quad \text{for } i = 1, \ldots, n,$$

where $\sigma > 0$ is a fixed parameter, and $\xi \sim \mathcal{N}(0, I_{p\times p})$ are independently identically distributed (i.i.d.) standard multivariate normally distributed isotropic random variables. The SNR is measured in dB and is defined as

$$\mathrm{SNR}\,[\mathrm{dB}] = 10 \log_{10} \frac{\mathrm{Var}(\mathrm{Signal})}{\sigma^2}.$$

Our first example is the unit sphere S^2. This example demonstrates how to synchronize the orientation when the manifold is orientable. We sample $n = 5000$ points from S^2 uniformly without noise and run ODM with $N = 100$ nearest neighbors, and estimate the intrinsic dimension in the local PCA step using $\gamma = 0.8$. Then we sample another 5000 points from S^2 with $\mathrm{SNR} = 5$ dB and run ODM with the same parameters. In both cases, the intrinsic dimension is estimated correctly as $\hat{d} = 2$. The histograms of the top eigenvector $v_1^{\mathcal{Z}}$ of both cases are shown in Fig. 2. The results show the robustness of ODM in orientation synchronization.

Next we show how the ODM algorithm helps us to determine if a given manifold is orientable or not. We sample 5000 points from the Klein bottle embedded in $\mathbb{R}^4$ and $\mathbb{R}P(2)$ embedded in $\mathbb{R}^4$ without noise, respectively. Then run the ODM algorithm for these two data sets with $N = 200$ and $\gamma = 0.8$. The histograms of the top eigenvectors are shown in Fig. 3. It shows that when the underlying manifold is non-orientable, the values of the first eigenvector of the ODM do not cluster around two distinguished positive and negative values.

We proceed to considering the application of ODM for dimensionality reduction, looking at data points sampled from the bell-shaped, closed, non-intersected curve, which is parameterized by $C(\theta) = (x(\theta), y(\theta))$, $\theta \in [0, 2\pi]$, where

$$x(\theta) = \cos(\theta),$$
$$y(\theta) = \begin{cases} \sin(\theta) & \text{for } \theta \in [0, \pi/4] \cup [3\pi/4, 5\pi/4] \cup [7\pi/4, 2\pi], \\ \sin(\theta)\left(1 - e^{-(2\theta-\pi)^2} + e^{-(\pi/2)^2}\right) & \text{otherwise.} \end{cases}$$

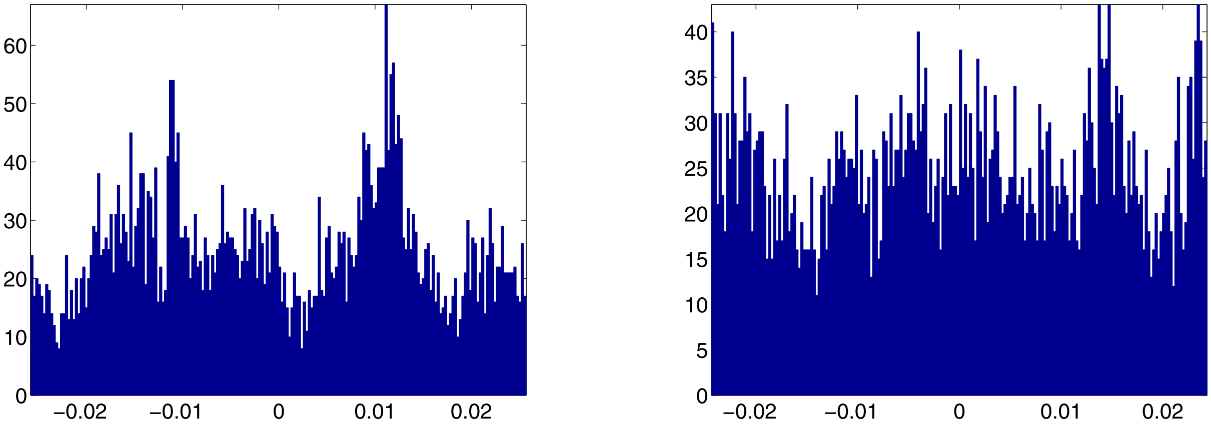

Fig. 3. Histogram of the values of the top eigenvector $v_1^{\mathbb{Z}}$. Left: Klein bottle; Right: $\mathbb{R}P^2$.

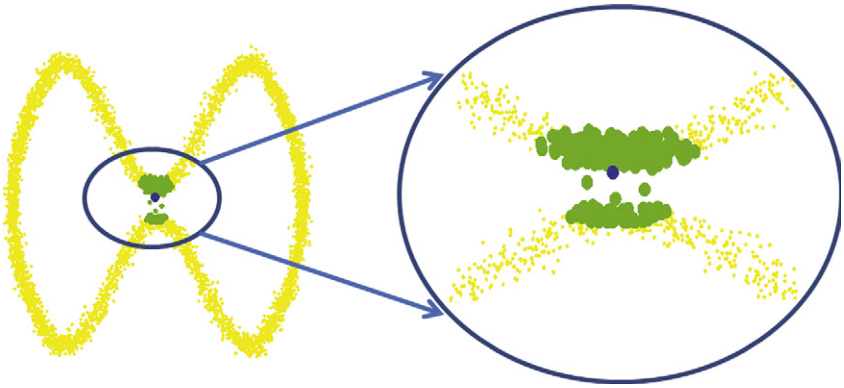

Fig. 4. The bell shaped curve: the green points are the nearest neighbors of the blue point. The SNR is 30 dB. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The curve is depicted in Figs. 4 and 5. Note that the isthmus of $C(\theta)$ (near $\theta = \frac{\pi}{2}$ and $\theta = \frac{3\pi}{2}$) imposes a difficultly whenever the noise is too large or the sampling rate is too low. This difficulty is due to the identity of the nearest neighbors, as illustrated in Fig. 4. It reflects the fact that the nearest neighbors are neighbors in the extrinsic sense (when the distance is measured by the Euclidean distance), but not in the intrinsic sense (when the distance is measured by the geodesic distance). To cope with this difficulty, we first cluster the nearest neighbors in order to find the “true” neighbors in the geodesic distance sense, prior to applying PCA. For clustering the points in $\mathcal{N}_{x_i}$ we apply the spectral clustering algorithm [13], and for PCA we consider only the points that belong to the cluster of x_i.

We applied ODM to 5000 points from $C(\theta)$ without noise, with parameters $N = 200$ and $\gamma = 0.8$, and then embed the data points into $\mathbb{R}^2$ using the second and third eigenvectors and (20). We repeat the same steps for noisy data sets with SNR $= 30$ dB and SNR $= 24$ dB. Fig. 5 shows the original data set, the diffusion map (DM) embedding, and the ODM embedding. In the DM algorithm, we used the Gaussian kernel $\exp\{-\|x_i - x_j\|^2/\sigma\}$ with $\sigma = 0.5$. As expected, the embedding results of ODM and DM are comparable.

Our next example is the Swiss roll, a manifold often used by machine learners as a testbed for dimensionality reduction methods. The Swiss roll model is the following manifold S embedded in $\mathbb{R}^3$:

$$\begin{aligned} x(t,s) &= \frac{3\pi(1+2t)}{8}\cos\frac{3\pi(1+2t)}{2}, \\ y(t,s) &= s, \\ z(t,s) &= \frac{3\pi(1+2t)}{8}\sin\frac{3\pi(1+2t)}{2}, \end{aligned}$$

for $(t,s) \in [0,1] \times [0,21]$. The Swiss roll model exhibits the same difficulty we encountered before with the bell-shape curve, namely, that the geodesically distant points might be close in the Euclidean sense (see Fig. 6). Moreover, unlike the manifolds considered before, the Swiss roll has a boundary. To cope with these problems, we again apply the spectral clustering algorithm [13] prior to the local PCA step of the ODM algorithm.

We sampled $n = 7000$ points from the Swiss roll without noise, run ODM with parameters $N = 40$ and $\gamma = 0.8$, and then embedded the data points into $\mathbb{R}^2$ by the second and third eigenvectors using (20). We repeated the same steps for data

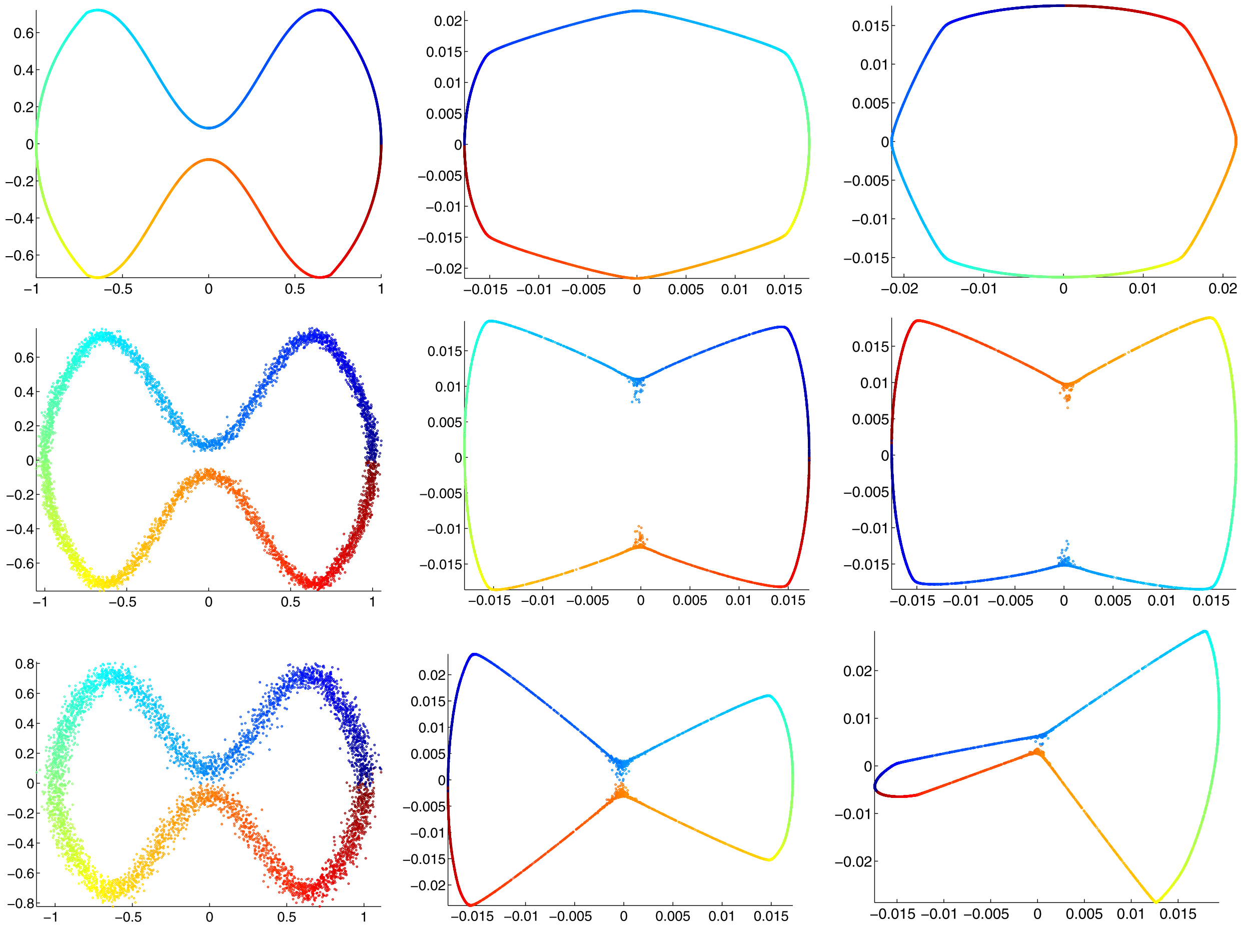

Fig. 5. From left to right: the input data set consisting of $n = 5000$ points, embedding by diffusion map, and embedding by ODM. Both the data set and the embeddings are colored by the parameter θ. The rows from top to bottom correspond to clean data, SNR $= 30$ dB and SNR $= 24$ dB.

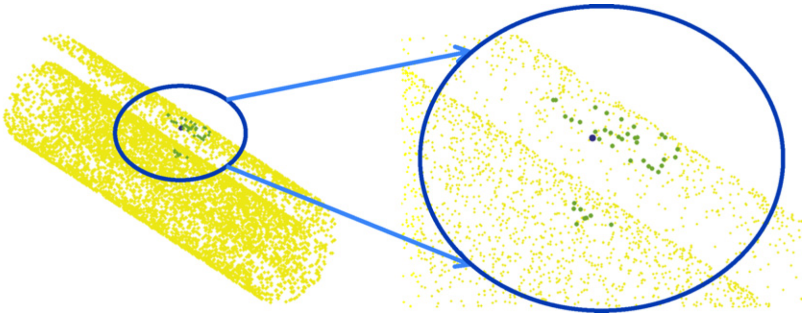

Fig. 6. The Swiss roll: the yellow points are the original data, while the green points are nearest neighbors of the blue point. The points are sampled from the Swiss roll without any noise. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sets with SNR values 50 dB, 30 dB and 25 dB. Fig. 7 shows the original data, the embedding by DM, and the embedding by ODM. In the DM algorithm, we used the Gaussian kernel with $\sigma = 5$.

In conclusion, we see that besides determining the orientability of the manifold, ODM can also be used as a dimensionality reduction method.

In the last part of this section we demonstrate results about recovering the orientable double covering of a non-orientable manifold as discussed in Section 3. First we demonstrate the orientable double covering of $\mathbb{R}P(2)$, Klein bottle and Möbius band in Fig. 8. For $\mathbb{R}P(2)$, we use the following embedding inside $\mathbb{R}^4$:

$$(x, y, z) \mapsto \big(xy, xz, y^2 - z^2, 2yz\big),$$

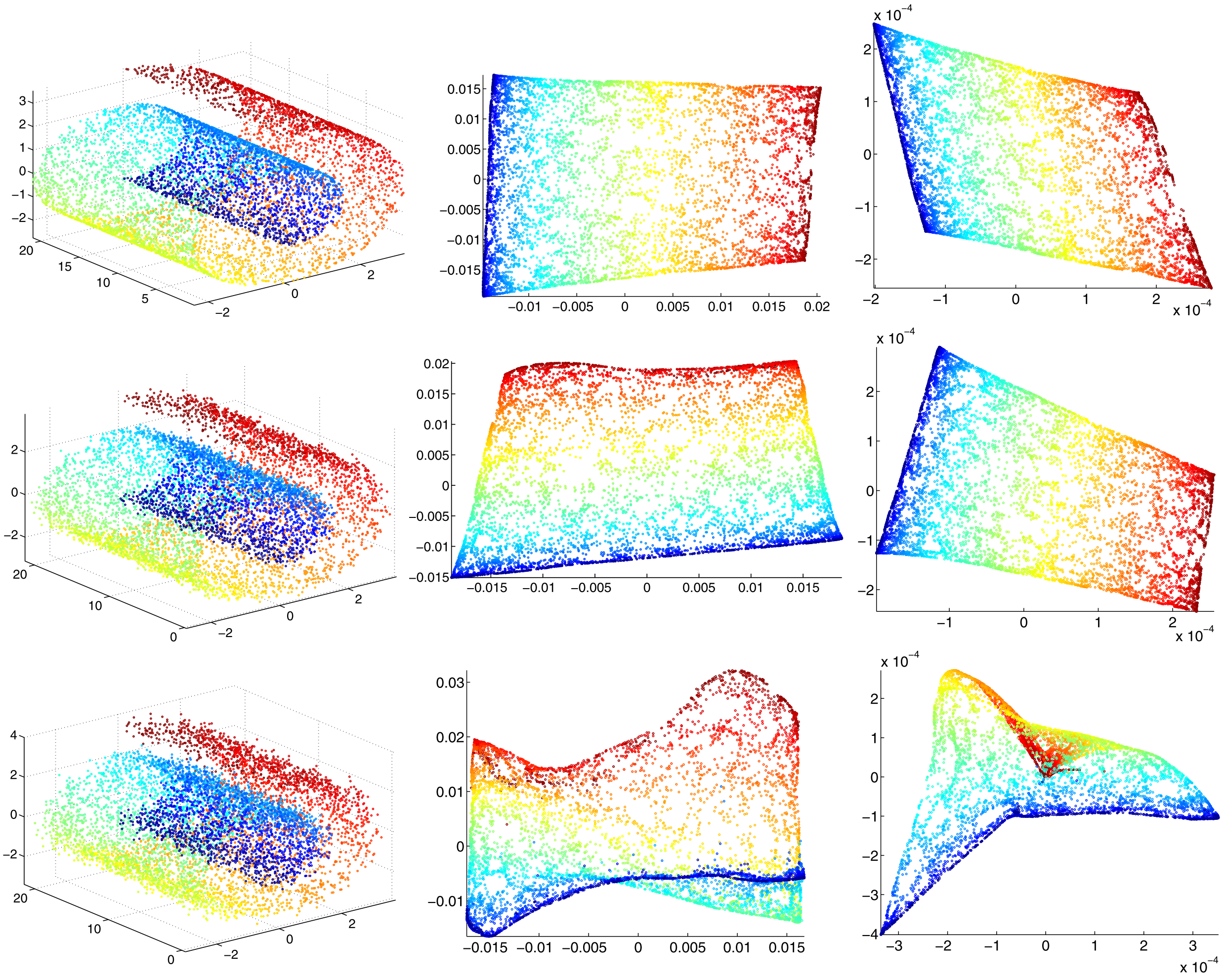

Fig. 7. From left to right: the input data, the embedding by diffusion map, and the embedding by ODM. The data sets and the embeddings are colored by the parameter t. From top to bottom: clean data set, SNR = 30 dB and SNR = 25 dB.

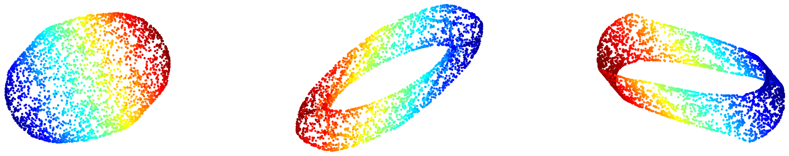

Fig. 8. Left: the orientable double covering of $\mathbb{R}P(2)$, which is S^2; Middle: the orientable double covering of the Klein bottle, which is T^2; Right: the orientable double covering of the Möbius strip, which is a cylinder.

where $(x, y, z) \in S^2 \subset \mathbb{R}^3$. We sample 8000 points uniformly from S^2 and map the sampling points to $\mathbb{R}P(2)$ by the embedding, which gives us a non-uniform sampling data set. Then we evaluate eigenvectors of the $16\,000 \times 16\,000$ sparse matrix $\tilde{\mathcal{Z}}$ constructed from the data set with $N = 40$ and get the approximation of the coordinates of the orientable double covering which allows us to get the orientable double covering (Fig. 8, left panel).

For the Klein bottle, we apply the following embedding inside $\mathbb{R}^4$:

$$\begin{cases} x = (2\cos v + 1)\cos u, \\ y = (2\cos v + 1)\sin u, \\ z = 2\sin v\cos(u/2), \\ w = 2\sin v\sin(u/2), \end{cases} \tag{32}$$

where $(u, v) \in [0, 2\pi] \times [0, 2\pi]$. We sample 8000 points uniformly from $[0, 2\pi] \times [0, 2\pi]$ and map the sampling points to the Klein bottle by the embedding, which gives us a non-uniform sampling data set over the Klein bottle. Then we run the same procedure as in $\mathbb{R}P(2)$ with $N = 40$ and get the result (Fig. 8, middle panel).

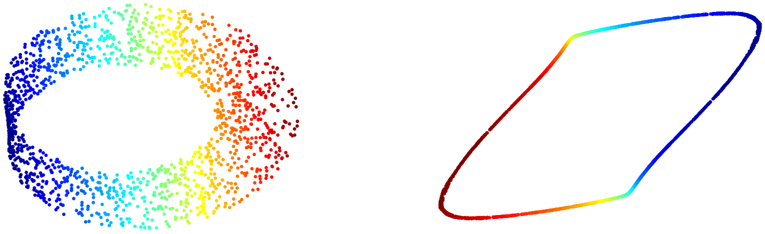

Fig. 9. Left: the Möbius band embedded in $\mathbb{R}^3$ with $b = 1$; Right: the algorithm failed to recover the cylinder.

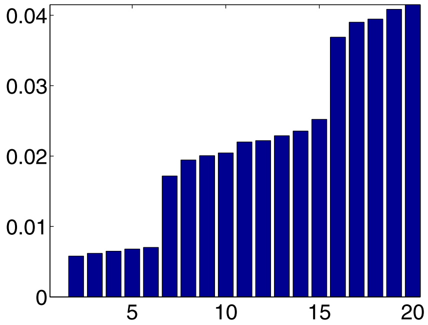

Fig. 10. Bar plot of the eigenvalues of the graph Laplacian of the abstract $\mathbb{R}P(2)$ constructed from gluing antipodal points of S^2. The multiplicities $1, 5, 9, \ldots$ correspond to the even degree spherical harmonics.

For the Möbius band, we use the following parametrization:

$$\begin{cases} x = (a + bv\cos(0.5u))\cos(u), \\ y = (a + bv\cos(0.5u))\sin(u), \\ z = bv\sin(0.5u), \end{cases}$$

where $a = 3$, $b = 5$ and $(u, v) \in [-1, 1] \times [0, 2\pi]$. As above, we sample 8000 points uniformly from $[-1, 1] \times [0, 2\pi]$ and map the sampling points to the Möbius band by the parametrization, which gives us a non-uniform sampling data set. Then we run the same procedure with $N = 20$ and get the result (Fig. 8, right panel). It is worth mentioning that when b is small, e.g., $b = 1$, we fail to recover the cylinder, as can be seen in Fig. 9. This difficulty comes from the lack of information in the width of the band when b is too small.

Last, we demonstrate how to get a non-orientable manifold by constructing the $\tilde{\mathcal{A}}$ matrix as discussed in Section 3. We start from showing the spectrum of $\mathbb{R}P(2)$ constructed by gluing the antipodal points of S^2 together. We build up $\tilde{\mathcal{A}}$ by taking $N = 50$ from 16 000 data points sampled from S^2 embedded in $\mathbb{R}^3$:

$$\begin{cases} x = \sin(\theta)\cos(\phi), \\ y = \sin(\theta)\sin(\phi), \\ z = \cos(\theta), \end{cases}$$

which is the approximation of the Laplace–Beltrami operator over $S^2/\mathbb{Z}_2 \cong \mathbb{R}P(2)$. The spectrum of the operator is shown in Fig. 10.

Then we demonstrate the reconstruction of the orientable double covering of an abstract non-orientable manifold, by which we mean an abstract non-orientable manifold not yet embedded inside any Euclidean space. We start from sampling 16 000 points from a cylinder C embedded in $\mathbb{R}^3$, which is parameterized by:

$$\begin{cases} x = 2\cos\theta, \\ y = 2\sin\theta, \\ z = a, \end{cases}$$

where $(\theta, a) \in [0, 2\pi] \times [-2, 2]$, and run the following steps. First we fix $N = 50$, and build up $\tilde{\mathcal{A}}$, the approximation of the Laplace–Beltrami operator over $C/\mathbb{Z}_2$ by gluing the antipodal points together. Note that $C/\mathbb{Z}_2$ is an abstract Möbius band up to now. Second, we embed $C/\mathbb{Z}_2$ by applying diffusion map, where we take the second to fifth eigenvectors of $\tilde{\mathcal{A}}$, that is, we embed $C/\mathbb{Z}_2$ into $\mathbb{R}^4$. Finally we run ODM with $N = 50$ over the embedded $C/\mathbb{Z}_2$ to reconstruct the double covering. The result, shown in Fig. 11, shows that we recover back a cylinder. Similarly, we run the same process with the same parameters $N = 50$ and the second to fifth eigenvectors of $\tilde{\mathcal{A}}$ to embed $T^2/\mathbb{Z}_2$ and $S^2/\mathbb{Z}_2$ into $\mathbb{R}^4$, and then run ODM with $N = 50$, which results in the reconstruction of the orientable double covering of an abstract Klein bottle and $\mathbb{R}P(2)$. The results are shown in Figs. 12 and 13. Using different number of eigenvectors for embedding, for example, from the second to sixth or from the second to seventh eigenvectors give us similar results.

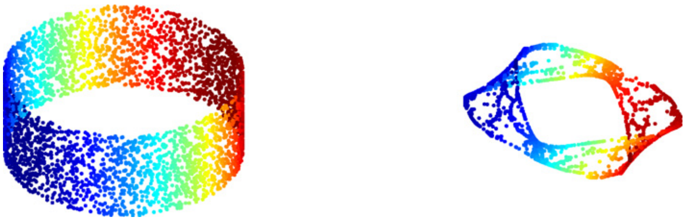

Fig. 11. Left: the cylinder embedded in $\mathbb{R}^3$; Right: the recovered cylinder.

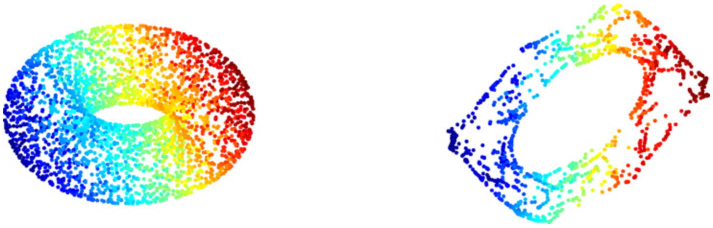

Fig. 12. Left: the torus embedded in $\mathbb{R}^3$; Right: the recovered torus.

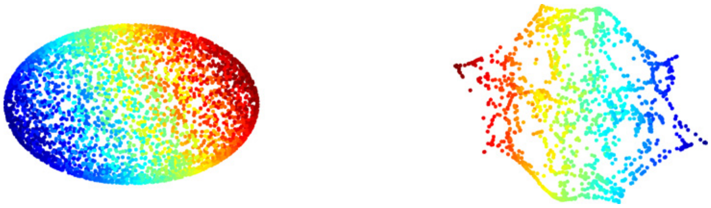

Fig. 13. Left: S^2 embedded in $\mathbb{R}^3$; Right: the recovered S^2.

Note that in the above three examples shown in Figs. 11, 12 and 13, the recovered objects look different from the original one. This distortion mainly comes from the following step: the embedding by diffusion map chosen in the demonstration is not isometric but just an approximation of an isometric embedding.

5. Summary and discussion

We introduced an algorithm for determining the orientability of a manifold from scattered data. If the manifold is orientable then the algorithm synchronizes the orientations and can further be used for dimensionality reduction in a similar fashion to the diffusion map framework.

The algorithm consists of three steps: local PCA, alignment and synchronization. In the synchronization step we construct a matrix $\mathcal{Z}$, which for an orientable manifold is shown to be similar to the normalized Laplacian $\mathcal{L}$. For non-orientable manifolds this similarity does not hold anymore and we show how to obtain an embedding of the orientable double covering using the eigenvectors of the Kronecker product $\tilde{\mathcal{Z}} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \mathcal{Z}$. In particular, we show that the eigenvectors of $\tilde{\mathcal{Z}}$ approximate the odd eigenfunctions of the Laplacian over the double covering. We also show how to use this observation in order to embed an abstract non-orientable manifold in Euclidean space given its orientable double covering.

We comment that the reflection matrix $\mathcal{Z}$ can also be built by taking a kernel function into consideration, like the weighted graph Laplacians discussed in [6]. For example, we may replace the definition (6) of Z by

$$Z_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{\sigma}} \det O_{ij} & (i, j) \in E, \\ 0 & (i, j) \notin E, \end{cases} \tag{33}$$

and

$$\mathcal{Z} = D^{-1} Z, \tag{34}$$

where D is a $n \times n$ diagonal matrix with $D_{ii} = \sum_{j=1}^{n} |z_{ij}|$. All the results in this paper also apply to this kernelized reflection matrix.

One possible application of our framework is the geometrical and topological study of high contrast local patches that are extracted from natural images [10]. Computational topology methods were used in [5] to conclude that this data set has the topology of the Klein bottle. This result may perhaps be confirmed using the methods presented here.

Finally, we remark that while the reflection matrix only uses the information in $\det O_{ij}$, it seems natural also to take into account all the information in O_{ij}. This modification will be the subject of a separate publication.

Acknowledgments

The problem of reconstructing the double covering of a non-orientable manifold was proposed to us by Peter Jones and Ronald Coifman during a discussion at Yale University. 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 FHWA grant DTFH61-08-C-00028.

References

[1] K.S. Arun, T.S. Huang, S.D. Blostein, Least-squares fitting of two 3-D point sets, IEEE Trans. Patt. Anal. Mach. Intell. 9 (5) (1987) 698–700.
[2] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.
[3] M. Belkin, P. Niyogi, Towards a theoretical foundation for Laplacian-based manifold methods, in: Proceedings of the 18th Conference on Learning Theory (COLT), 2005, pp. 486–500.
[4] M. Belkin, P. Niyogi, Convergence of Laplacian eigenmaps, in: Adv. Neural Inf. Process. Syst., vol. 19, MIT Press, 2007.
[5] G. Carlsson, T. Ishkhanov, V. de Silva, A. Zomorodian, On the local behavior of spaces of natural images, Int. J. Comput. Vision 76 (1) (2008) 1–12.
[6] R.R. Coifman, S. Lafon, Diffusion maps, Appl. Comput. Harmon. Anal. 21 (1) (2006) 5–30.
[7] M. Cucuringu, Y. Lipman, A. Singer, Sensor network localization by eigenvector synchronization over the Euclidean group, submitted for publication.
[8] D.L. Donoho, C. Grimes, Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data, Proc. Natl. Acad. Sci. USA 100 (10) (2003) 5591–5596.
[9] M. Hein, J. Audibert, 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), 2005, pp. 470–485.
[10] A.B. Lee, K.S. Pedersen, D. Mumford, The non-linear statistics of high-contrast patches in natural images, Int. J. Comput. Vision 54 (1–3) (2003) 83–103.
[11] A.V. Little, J. Lee, Y.M. Jung, M. Maggioni, Estimation of intrinsic dimensionality of samples from noisy low-dimensional manifolds in high dimensions with multiscale SVD, in: 2009 IEEE/SP 15th Workshop on Statistical Signal Processing, 2009, pp. 85–88.
[12] B. Nadler, S. Lafon, R.R. Coifman, I.G. Kevrekidis, Diffusion maps, spectral clustering and reaction coordinates of dynamical systems, Appl. Comput. Harmon. Anal. 21 (1) (2006) 113–127.
[13] A.Y. Ng, M.I. Jordan, Y. Weiss, On spectral clustering: Analysis and an algorithm, in: T.K. Leen, T.G. Dietterich, V. Tresp (Eds.), Adv. Neural Inf. Process. Syst., vol. 14, MIT Press, Cambridge, MA, 2002.
[14] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[15] A. Singer, Angular synchronization by eigenvectors and semidefinite programming, Appl. Comput. Harmon. Anal., doi:10.1016/j.acha.2010.02.001, in press.
[16] A. Singer, From graph to manifold Laplacian: The convergence rate, Appl. Comput. Harmon. Anal. 21 (1) (2006) 128–134.
[17] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.