frontiers in Applied Mathematics and Statistics

ORIGINAL RESEARCH
published: 16 May 2019
doi: 10.3389/fams.2019.00024

Least Square Approach to Out-of-Sample Extensions of Diffusion Maps

*Jianzhong Wang**

Department of Mathematics and Statistics, Sam Houston State University, Huntsville, TX, United States

Let $X = \mathbf{X} \cup \mathbf{Z}$ be a data set in $\mathbb{R}^D$, where $\mathbf{X}$ is the training set and $\mathbf{Z}$ the testing one. Assume that a kernel method produces a dimensionality reduction (DR) mapping $\mathfrak{F} : \mathbf{X} \to \mathbb{R}^d$ ($d \ll D$) that maps the high-dimensional data $\mathbf{X}$ to its row-dimensional representation $\mathbf{Y} = \mathfrak{F}(\mathbf{X})$. The out-of-sample extension of dimensionality reduction problem is to find the dimensionality reduction of X using the extension of $\mathfrak{F}$ instead of re-training the whole data set X. In this paper, utilizing the framework of reproducing kernel Hilbert space theory, we introduce a least-square approach to extensions of the popular DR mappings called Diffusion maps (Dmaps). We establish a theoretic analysis for the out-of-sample DR Dmaps. This analysis also provides a uniform treatment of many popular out-of-sample algorithms based on kernel methods. We illustrate the validity of the developed out-of-sample DR algorithms in several examples.

Keywords: out-of-sample extension, dimensionality reduction, reproducing kernel Hilbert space, least-square method, diffusion maps

AMS Subject Classification: 62-07, 42B35, 47A58, 30C40, 35P15

OPEN ACCESS

Edited by:
Ding-Xuan Zhou,
City University of Hong Kong,
Hong Kong

Reviewed by:
Bo Zhang,
Hong Kong Baptist University,
Hong Kong
Shao-Bo Lin,
Wenzhou University, China
Sui Tang,
Johns Hopkins University,
United States

***Correspondence:**
Jianzhong Wang
jzwang@shsu.edu

Specialty section:
This article was submitted to Mathematics of Computation and Data Science, a section of the journal Frontiers in Applied Mathematics and Statistics

Received: *02 March 2019*
Accepted: *25 April 2019*
Published: *16 May 2019*

Citation:
Wang J (2019) Least Square Approach to Out-of-Sample Extensions of Diffusion Maps. Front. Appl. Math. Stat. 5:24. doi: 10.3389/fams.2019.00024

1. INTRODUCTION

Recently, in many scientific and technological areas, we need to analyze and process high-dimensional data, such as speech signals, images and videos, text documents, stock trade records, and others. Due to the curse of dimensionality [1, 2], directly analyzing and processing high-dimensional data are often infeasible. Therefore, *dimensionality reduction* (DR) (see the books [3, 4]) becomes a critical step in high-dimensional data processing. DR maps high-dimensional data into a low-dimensional space so that the data process can be carried out on its low-dimensional representation. There exist many DR methods in literature. The famous linear method is *principle component analysis* (PCA) [5]. However, PCA cannot effectively reduce the dimension for the data set, which essentially resides on a nonlinear manifold. Therefore, to reduce the dimensions of such data sets, people employ non-linear DR methods [6–12], among which, the method of Diffusion Maps (Dmaps) introduced by Coifman and his research group [13, 14] have been proved attractive and effective. Adopting the ideas of the spectral clustering [15, 16] and Laplacian eigenmaps [17], Dmaps integrates them into a more conceptual framework—the geometric harmonics.

As a spectral method, Dmaps employs the diffusion kernel to define the similarity on a given data set $\mathbf{X} \subset \mathbb{R}^D$. The principal d-dimensional eigenspace ($d \ll D$) of the kernel provides the feature space of $\mathbf{X}$, so that a diffusing mapping $\mathfrak{F}$ maps $\mathbf{X}$ to the set $\mathbf{Y} = \mathfrak{F}(\mathbf{X})$, which is called a DR of $\mathbf{X}$.

Note that the mapping $\mathfrak{F}$ is constructed by the spectral decomposition of the kernel, which is data-dependent. If the set $\mathbf{X}$ is enlarged to $X = \mathbf{X} \cup \mathbf{Z}$ and we want to make DR of X by Dmaps, we have to retrain the set X in order to construct a new diffusing mapping. The retraining approach is often unpractical if the cardinality of X becomes very large, or the new data set $\mathbf{Z}$ comes as a time-stream.

Out-of-example DR extension method finds the DR of X by extending the diffusing mapping $\mathfrak{F}$ onto X. In most cases, we can assume that the new data set $\mathbf{Z}$ has the similar features as $\mathbf{X}$. Therefore, instead of retraining the whole set X, we realize the DR of X by extending the mapping $\mathfrak{F}$ from $\mathbf{X}$ to X only.

Lots of papers have introduced various out-of-example extension algorithms (see [14, 18, 19] and their references). However, the mathematical analysis on out-of-example extension is not studied sufficiently.

The main purpose of this paper is to give a mathematical analysis on the out-of-sample DR extension of Dmaps. In Wang [20], we preliminarily studied out-of-sample DR extensions for kernel PCA. Since the structure of kernels for Dmaps are different from kernel PCA, it needs a special analysis. In this paper we deal with the DR extensions of Dmaps in the framework of reproducing kernel Hilbert space (RKHS), in which Dmaps extension can be classified as the least square one.

The paper is organized as follows: In section 2, we introduce the general out-of-sample extensions in the RKHS framework. In section 3, we establish the least square out-of-sample DR extensions of Dmaps. In section 4, we give the mathematical analysis and algorithms for the Dmaps DR extension. In the last section, we give several examples for the extension.

2. PRELIMINARY

We first introduce some notions and notations. Let μ be a finite (positive) measure on a data set $X \subset \mathbb{R}^D$. We denoted by $L^2(X, \mu)$ the (real) Hilbert space on X, equipped with the inner product

$$\langle f, g \rangle_{L^2(X,\mu)} = \int_X f(x)g(x)d\mu(x), \quad f, g \in L^2(X, \mu).$$

Then, $\|f\|_{L^2(X,\mu)} = \sqrt{\langle f, f \rangle_{L^2(X,\mu)}}$. Later, we will abbreviate $L^2(X, \mu)$ to $L^2(X)$ (or L^2) if the measure μ (and the set X) is (are) not stressed.

Definition 1 *A function $k : X^2 \to \mathbb{R}$ is called a Mercer's kernel if it satisfies the following conditions:*

1. *k is symmetric: $k(x, y) = k(y, x)$;*
2. *k is positive semi-definite;*
3. *k is bounded on X^2, that is, there is an $M > 0$ such that $|k(x, y)| \le M, (x, y) \in X^2$.*

In this paper, we only consider Mercer's kernels. Hence, the term *kernel* will stand for Mercer's one. The kernel distance (associated with k) between two points $x, y \in X$ is defined by

$$d_k(x, y) = \sqrt{k(x, x) + k(y, y) - 2k(x, y)}.$$

A kernel k defines an RKHS H_k, in which the inner product satisfies [21]

$$\langle f(\cdot), k(x, \cdot) \rangle_{H_k} = f(x), \quad f \in H_k, x \in X. \tag{1}$$

Later, we will use H instead of H_k if the kernel k is not stressed. Recall that k has a dual identity. It derives the identity operator on H, as shown in 1, and also derives the following compact operator K on $L^2(X)$:

$$(Kf)(x) = \langle f(\cdot), k(x, \cdot) \rangle_{L^2} = \int_X k(x, y)f(y)d\mu(y), \quad f \in L^2.$$

In Wang [20], we proved that if

$$k(x, y) = \sum_{j=1}^{m} \phi_j(x)\phi_j(y),$$

where the set $\{\phi_1, \cdots, \phi_m\}$ is linearly independent, then the set is an o.n. basis of H. Therefore, for $f, g \in H$ with $f = \sum_j c_j\phi_j$ and $g = \sum_j d_j\phi_j$, we have $\langle f, g \rangle_{H_k} = \sum_j c_j d_j$.

Let the spectral decomposition of k be the following:

$$k(x, y) = \sum_{j=1}^{m} \lambda_j v_j(x)v_j(y), \quad 0 \le m \le \infty, \tag{2}$$

where the eigenvalues are arranged decreasingly, $\lambda_1 \ge \cdots \ge \lambda_m > 0$, and the eigenfunctions $v_1, v_1, \cdots, v_m$, are normalized to satisfy

$$\langle v_i, v_j \rangle_{L^2(X)} = \delta_{i,j}.$$

Write $\gamma_i(x) = \sqrt{\lambda_i}v_i(x)$. Then, $\{\gamma_1, \cdots, \gamma_m\}$ is an o.n. basis of H, which is called the *canonic basis* of H. We also call $k(x, y) = \sum_{j=1}^{m} \gamma_j(x)\gamma_j(y)$ the *canonic decomposition* of k. By 2, we have

$$\gamma_j = \frac{1}{\lambda_j} \int_X k(x, y)\gamma_j(y)d\mu(y).$$

Thus, if $f \in H$ have the canonic representation $f = \sum_{j=1}^{m} c_j\gamma_j$, then, for any $g \in H$, the inner product $\langle f, g \rangle_H$ has the following integral form:

$$\langle f, g \rangle_H = \sum_{j=1}^{m} \frac{c_j}{\lambda_j} \int_X g(x)\gamma_j(x)d\mu(x).$$

To investigate the out-of-sample DR extension, we first recall some general results on function extensions. Let $X = \mathbf{X} \cup \mathbf{Z}$. To stress that a point $x \in X$ is also in $\mathbf{X}$, we use $\mathbf{x}$ instead of x. Similarly, we denote by $\mathbf{k}(\mathbf{x}, \mathbf{y})$ the restriction of $k(x, y)$ on $\mathbf{X}^2$. That is,

$$k(x, y) = \mathbf{k}(\mathbf{x}, \mathbf{y}), \quad (\mathbf{x}, \mathbf{y}) \in \mathbf{X}^2.$$

We also denote by $\mathbf{H}$ the RKHS associated with $\mathbf{k}$. Then a continuous map $\mathbf{E} : \mathbf{H} \to H$ is called an extension if

$$\mathbf{E}(\mathbf{f})(\mathbf{x}) = \mathbf{f}(\mathbf{x}), \quad \forall \mathbf{f} \in \mathbf{H}.$$

Correspondingly, we define the restriction $R : H \to \mathbf{H}$ by

$$R(f)(\mathbf{x}) = f(\mathbf{x}), \quad \forall f \in H, \mathbf{x} \in \mathbf{X}.$$

It is obvious that the extensions from $\mathbf{X}$ to X are not unique if $\mathbf{Z}$ is not empty. So, we define the set of all extensions of $\mathbf{f} \in \mathbf{H}$ by

$$A_{\mathbf{f}} = \{f \in H;\ R(f) = \mathbf{f}\},$$

and call $\hat{f} \in A_{\mathbf{f}}$ the *least-square extension* of $\mathbf{f}$ if

$$\|\hat{f}\|_H = \min_{f \in A_{\mathbf{f}}} \|f\|_H.$$

It is evident that the least-square extension of a function is unique. We denote by $\mathbf{T} : \mathbf{H} \to H$ the operator of the least-square extension.

In Wang [20], we already prove the following:

1. Let $\{\mathbf{v}_1, \cdots, \mathbf{v}_d\}$ be the canonic basis of $\mathbf{H}$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d > 0$ be the eigenvalues of the kernel $\mathbf{k}(\mathbf{x}, \mathbf{y})$. Then the least-square extension of $\mathbf{v}_j$ is

$$\hat{v}_j(x)(= \mathbf{T}(\mathbf{v}_j)(x)) = \frac{1}{\sigma_j} \int_{\mathbf{X}} k(x, \mathbf{y}) \mathbf{v}_j(\mathbf{y}) d\mu(\mathbf{y}),\ \ x \in X,\ \ 1 \leq j \leq d. \tag{3}$$

Therefore, for any $\mathbf{f} = \sum_{j=1}^{d} c_j \mathbf{v}_j \in \mathbf{H}$,

$$\hat{f}(x)(= \mathbf{T}(\mathbf{f})(x)) = \sum_{j=1}^{d} c_j \frac{1}{\sigma_j} \int_{\mathbf{X}} k(x, \mathbf{y}) \mathbf{v}_j(\mathbf{y}) d\mu(\mathbf{y}).$$

2. Let $\hat{H} = \mathbf{T}(\mathbf{H})$ and $\mathbf{T}^* : H \to \mathbf{H}$ be the joint operator of $\mathbf{T}$. Then $P = \mathbf{T}\mathbf{T}^*$ is an orthogonal projection from H to $\hat{H}$.

3. Let $\hat{k}(x, y)$ be the kernel of the RKHS $\hat{H}$. Then $k_0(x, y) = k(x, y) - \hat{k}(x, y)$ is a Mercer's kernel such that $k_0(x, y) = 0, (x, y) \in X^2 \setminus \mathbf{X}^2$. Denote by H_0 the RKHS associated with k_0. Then, $H = \hat{H} \bigoplus H_0$ and $\hat{H} \perp H_0$.

4. If $k(x, y)$ is a Gramian-type DR kernel [20], and $[\mathbf{v}_1(\mathbf{X}), \cdots, \mathbf{v}_d(\mathbf{X})]^T$ gives the DR of $\mathbf{X}$, then $[\hat{v}_1(X), \cdots, \hat{v}_d(X)]^T$ provides the least-square out-of-sample DR extension on X.

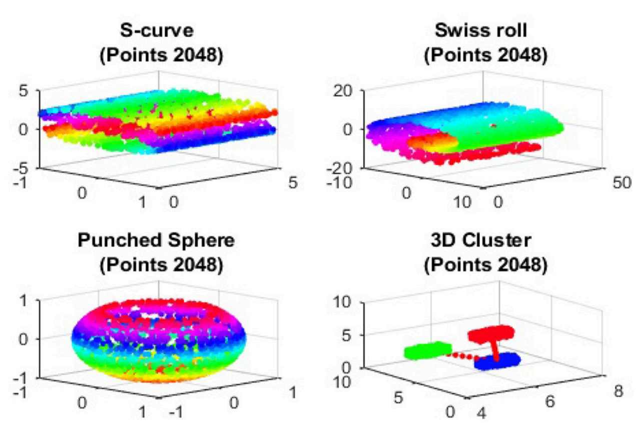

FIGURE 1 | S-curve, Swiss Roll, Punched Sphere, and 3D-Cluster.

3. LEAST-SQUARE OUT-OF-SAMPLE DR EXTENSIONS FOR DMAPS

The kernels of Dmaps are constructed based on the Gaussian kernel

$$w(x, y) = \exp\left(-\frac{\|x - y\|^2}{\epsilon}\right), \quad (x, y) \in X^2, \epsilon > 0.$$

The function

$$S(x) = \int_X w(x, y) d\mu(y)$$

defines a mass density on X, and $M = \int_X S(x) d\mu(x)$ is the total mass of X.

There are two important forms of the kernels of Dmaps: The *Graph-Laplacian* diffusion kernel and the *Laplace-Beltrami* one.

3.1. Dmaps With the Graph-Laplacian Kernel

We first discuss the least-square out-of-sample DR Extensions for the Dmaps with the Graph-Laplacian (GL) kernel. Normalizing the Gaussian kernel by $S(x)$, we obtain the following Graph-Laplacian diffusion kernel [4, 13]:

$$g(x, y) = \frac{w(x, y)}{\sqrt{S(x)S(y)}}.$$

This kernel relates to the data set X equipped with an undirected (weighted) graph. It is known that 1 is the greatest eigenvalue of $g(x, y)$ and its corresponding normalized eigenfunction is $\sqrt{\frac{S(x)}{M}}$.

Let H_g be the RKHS associated with the kernel g and $\{\phi_0, \cdots, \phi_m\}$ be its canonic basis, which suggest the following spectral decomposition of $g(x, y)$:

$$g(x, y) = \sum_{j=0}^{m} \lambda_j v_j(x) v_j(y),$$

where $1 = \lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_m > 0$ and $v_j(x) = \phi_j(x)/\sqrt{\lambda_j}$. Because $\phi_0 = \sqrt{\frac{S(x)}{M}}$ provides only the mass information of the data set, it should not reside on the feature space. Hence, we define the feature space as the RKHS associated with the kernel $\sum_{j=1}^{m} \phi_j(x)\phi_j(y)$, where ϕ_0 is removed.

Definition 2. *The mapping* $\Phi : X \to \mathbb{R}^m : \Phi(x) = [\phi_1(x), \cdots, \phi_m(x)]^T$ *is called the diffusion mapping and the data set* $\Phi(X) \subset \mathbb{R}^m$ *is called a DR of* X.

Remark. In Wang [20], we already pointed out that each orthogonal transformation of the set $\Phi(X)$ can also be considered as a DR of X. Hence, any non-canonical o.n. basis of the feature space also provides a DR mapping.

To study the out-of-sample extension, as what was done in the preceding section, we assume $X = \mathbf{X} \cup \mathbf{Z}$ and denote by $\mathbf{g}(\mathbf{x}, \mathbf{y})$ the Graph-Laplacian kernel on $\mathbf{X}$, that is,

$$\mathbf{g}(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{w}(\mathbf{x}, \mathbf{y})}{\sqrt{\mathbf{S}(\mathbf{x})\mathbf{S}(\mathbf{y})}},$$

where $\mathbf{S}(\mathbf{x})$ is the mass density on $\mathbf{X}$, and

$$\mathbf{w}(\mathbf{x},\mathbf{y}) = w(\mathbf{x},\mathbf{y}), \quad (\mathbf{x},\mathbf{y}) \in \mathbf{X}^2,$$

Assume that spectral decomposition of $\mathbf{g}$ is given by

$$\mathbf{g}(\mathbf{x},\mathbf{y}) = \sum_{j=0}^{d} \sigma_j \mathbf{v}_j(\mathbf{x})\mathbf{v}_j(\mathbf{y}). \tag{4}$$

Then the RKHS $H_{\mathbf{g}}$ associated with $\mathbf{g}$ has the canonic basis $\{\varphi_0, \varphi_1, \cdots, \varphi_d\}$:

$$\mathbf{g}(\mathbf{x},\mathbf{y}) = \sum_{j=0}^{d} \varphi_j(\mathbf{x})\varphi_j(\mathbf{y}),$$

where $\varphi_j = \sqrt{\sigma_j}\mathbf{v}_j$. Because $\mathbf{S}(\mathbf{x}) \neq S(\mathbf{x})$, in general,

$$\mathbf{g}(\mathbf{x},\mathbf{y}) \neq g(\mathbf{x},\mathbf{y}), \quad (\mathbf{x},\mathbf{y}) \in \mathbf{X}^2.$$

Hence, we cannot directly apply the extension technique in the preceding section to $\mathbf{g}$. Our main purpose in this subsection is to introduce the extension from $H_{\mathbf{g}}$ to H_g.

Denote by H_w and $H_{\mathbf{w}}$ the RKHSs associated with the kernels w and $\mathbf{w}$, respectively. Because $w(\mathbf{x},\mathbf{y}) = \mathbf{w}(\mathbf{x},\mathbf{y})$ for $(\mathbf{x},\mathbf{y}) \in \mathbf{X}^2$, the extension technique in the preceding section can be applied.

Let $\mathbf{u}_j(\mathbf{x}) = \sqrt{\mathbf{S}(\mathbf{x})}\varphi_j(\mathbf{x})$ and $u_j(x) = \sqrt{S(x)}\phi_j(x)$. Then we have

$$\mathbf{w}(\mathbf{x},\mathbf{y}) = \sum_{j=0}^{d} \mathbf{u}_j(\mathbf{x})\mathbf{u}_j(\mathbf{y}), \quad w(x,y) = \sum_{j=0}^{m} u_j(x)u_j(y).$$

Lemma 3 *The least-square extension operator $\mathbf{T}: H_{\mathbf{w}} \to H_w$ has the following representation:*

$$\mathbf{T}(\mathbf{u}_j)(x) = \frac{1}{\sigma_j}\int_{\mathbf{X}} w(x,\mathbf{y})\frac{\mathbf{u}_j(\mathbf{y})}{\mathbf{S}(\mathbf{y})}d\mu(\mathbf{y}), \quad j = 0, 1, \cdots, d. \tag{5}$$

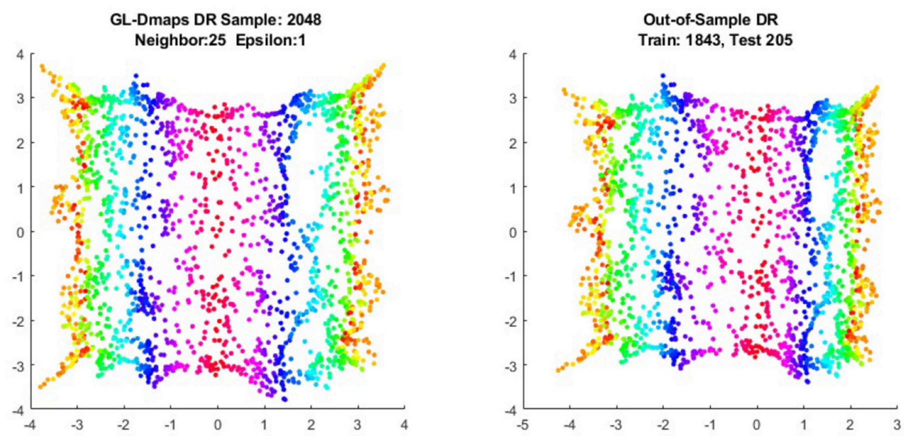

FIGURE 2 | GL out-of-sample extension for DR of S-Curve.

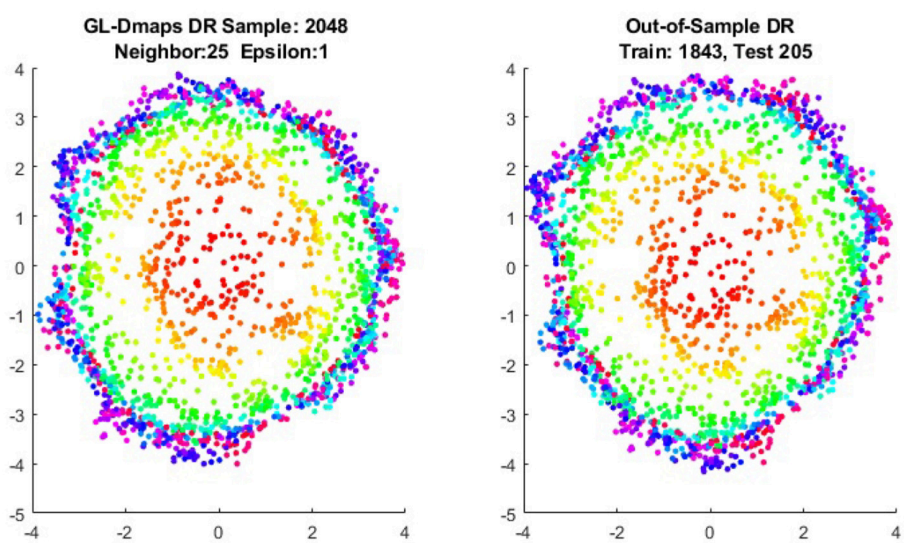

FIGURE 3 | GL out-of-sample extension for DR of Punched Sphere.

Proof. Because $\{\mathbf{u}_j\}_{j=0}^d$ is not a canonic o.n. basis of $H_\mathbf{w}$, we cannot directly apply the extension formula 3. Recall that the formula 3 can also be written as $\mathbf{T}(\mathbf{f})(x) = \langle \mathbf{f}, k(x,\cdot)\rangle_\mathbf{H}$. (In the considered case, the kernel w replaces k.) Note that

$$\langle \mathbf{u}_j, \mathbf{w}(\mathbf{x},\cdot)\rangle_{H_\mathbf{w}} = \mathbf{u}_j(\mathbf{x}) = \sqrt{\mathbf{S}(\mathbf{x})}\varphi_j(\mathbf{x}) = \sqrt{\mathbf{S}(\mathbf{x})}\frac{1}{\sigma_j}\int_\mathbf{X} \mathbf{g}(\mathbf{x},\mathbf{y})\varphi_j(\mathbf{y})\, d\mu(\mathbf{y}) = \frac{1}{\sigma_j}\int_\mathbf{X} \mathbf{w}(\mathbf{x},\mathbf{y})\frac{\mathbf{u}_j(\mathbf{y})}{\mathbf{S}(\mathbf{y})} d\mu(\mathbf{y}),$$

which implies that, for any $\mathbf{f} \in H_\mathbf{w}$, we have

$$\langle \mathbf{f}, \mathbf{u}_j\rangle_{H_\mathbf{w}} = \frac{1}{\sigma_j}\int_\mathbf{X} \mathbf{f}(\mathbf{y})\frac{\mathbf{u}_j(\mathbf{y})}{\mathbf{S}(\mathbf{y})} d\mu(\mathbf{y}).$$

Therefore, the formula $\mathbf{T}(\mathbf{u}_j)(x) = \langle w(x,\cdot), \mathbf{u}_j\rangle_{H_\mathbf{w}}$ yields 5. ■

We now write $\hat{u}_j = \mathbf{T}(\mathbf{u}_j)$ and define

$$\hat{w}(x,y) = \sum_{j=0}^{d} \hat{u}_j(x)\hat{u}_j(y).$$

Then the RKHS $H_{\hat{w}}$ associated with the kernel $\hat{w}$ is the extension of $H_\mathbf{w}$.

The function $S(x)$ induces the following multiplicator from H_g to H_w:

$$\mathfrak{S}_S(f)(x) = \sqrt{S(x)}f(x), \quad x \in X.$$

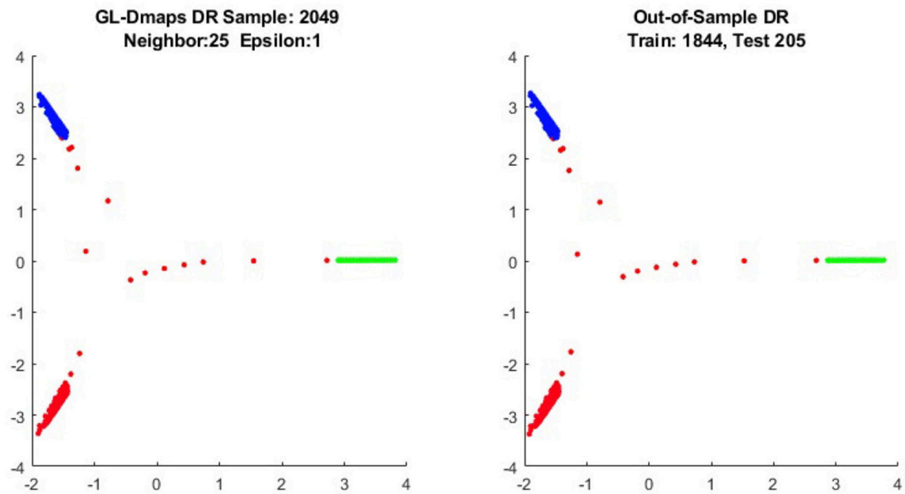

FIGURE 4 | GL out-of-sample extension for DR of 3D Cluster.

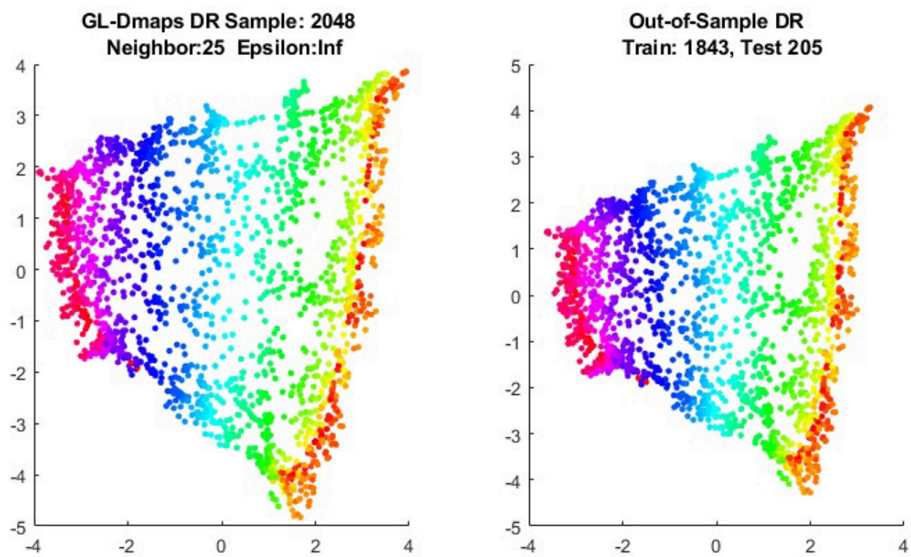

FIGURE 5 | GL out-of-sample extension for DR of Swiss Roll.

Similarly, the function $\mathbf{S}(\mathbf{x})$ induces the following multiplicator from $H_{\mathbf{g}}$ to $H_{\mathbf{w}}$:

$$\mathfrak{S}_{\mathbf{S}}(\mathbf{f})(\mathbf{x}) = \sqrt{\mathbf{S}(\mathbf{x})}\mathbf{f}(\mathbf{x}), \quad \mathbf{x} \in \mathbf{X}.$$

It is clear that the operator $\mathfrak{S}_S$ ($\mathfrak{S}_{\mathbf{S}}$) is an isometric mapping. With the aid of $\mathfrak{S}_S$ and $\mathfrak{S}_{\mathbf{S}}$, we define the *least-square extension* $\mathcal{T}$ from $H_{\mathbf{g}}$ to H_g by

$$\mathcal{T} = (\mathfrak{S}_S)^{-1} \circ \mathbf{T} \circ \mathfrak{S}_{\mathbf{S}}. \tag{6}$$

The following diagram shows the strategy of the out-of-sample extension using Graph-Laplacian diffusion mapping.

$$\begin{tikzcd}
H_{\mathbf{g}} \arrow[r, "\mathcal{T} = (\mathfrak{S}_S)^{-1} \circ \mathbf{T} \circ \mathfrak{S}_{\mathbf{S}}"] \arrow[d, "\mathfrak{S}_{\mathbf{S}}"'] & H_g \arrow[d, "\mathfrak{S}_S"] \\
H_{\mathbf{w}} \arrow[r, "\mathbf{T}"'] & H_w
\end{tikzcd}$$

We now derive the integral representation of the operator $\mathcal{T}$.

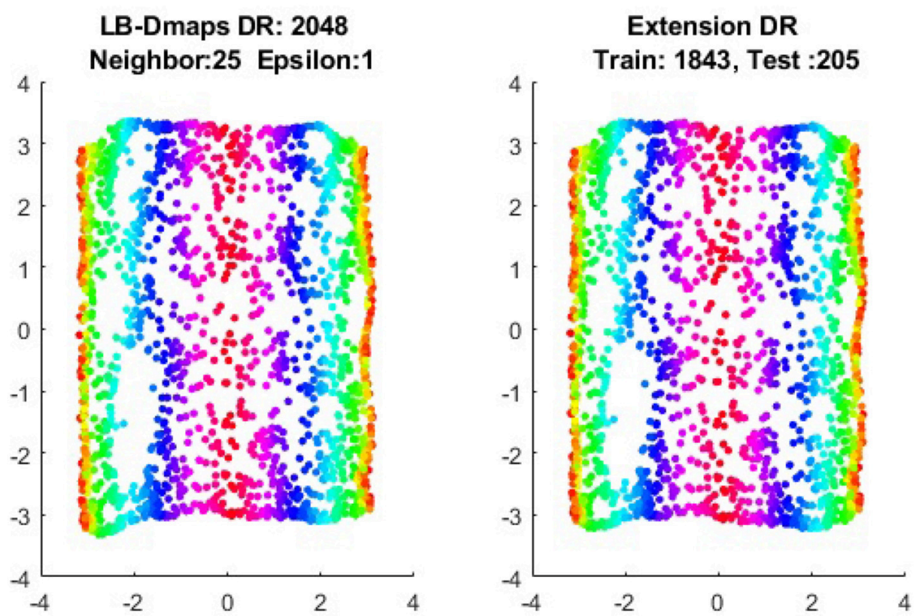

FIGURE 6 | LB out-of-sample extension for DR of S-Curve.

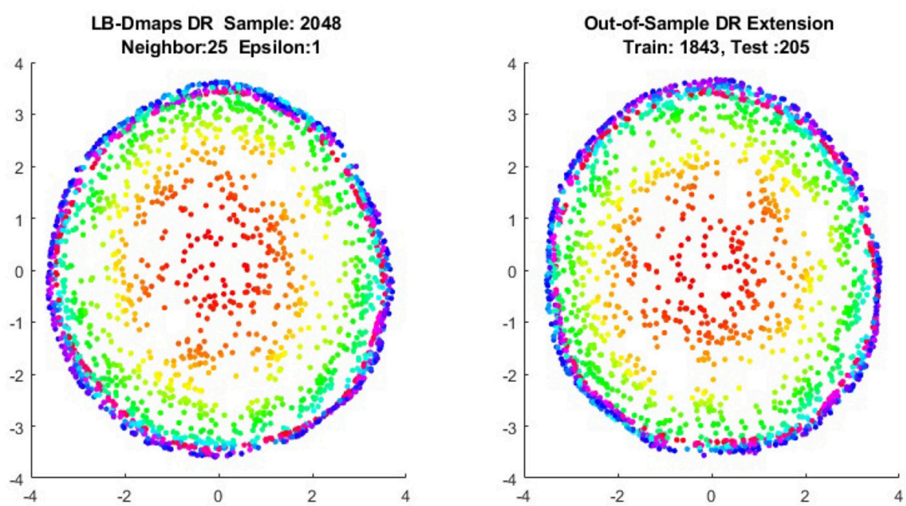

FIGURE 7 | LB out-of-sample extension for DR of Punched Sphere.

Lemma 4 *Let the canonic decomposition of* $\mathbf{g}$ *be given by 4 and* $\mathbf{f} = \sum_{j=0}^{d} c_j \varphi_j \in H_{\mathbf{g}}$. *Then*

$$\mathcal{T}(\mathbf{f})(x) = \sum_{j=0}^{d} c_j \hat{\phi}_j(x) = \sum_{j=0}^{d} \frac{c_j}{\sigma_j} \int_{\mathbf{X}} \frac{w(x, \mathbf{y})}{\sqrt{S(x)\mathbf{S}(\mathbf{y})}} \varphi_j(\mathbf{y}) d\mu(\mathbf{y}). \quad (7)$$

Its adjoint operator $\mathcal{T}^* : H_g \to H_{\mathbf{g}}$ *is given by*

$$\mathcal{T}^*(h)(\mathbf{x}) = h(\mathbf{x}) \sqrt{\frac{S(\mathbf{x})}{\mathbf{S}(\mathbf{x})}}, \quad h \in H_g. \quad (8)$$

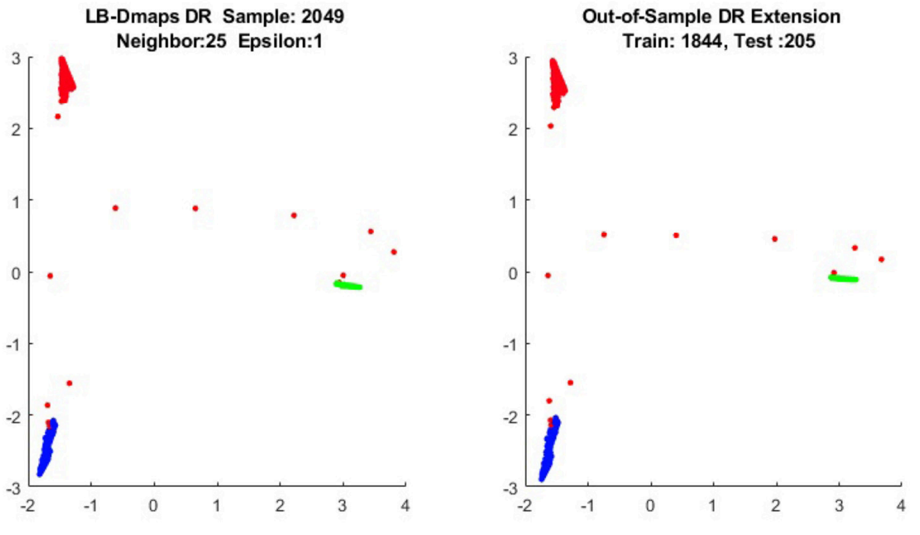

FIGURE 8 | LB out-of-sample extension for DR of 3D Cluster.

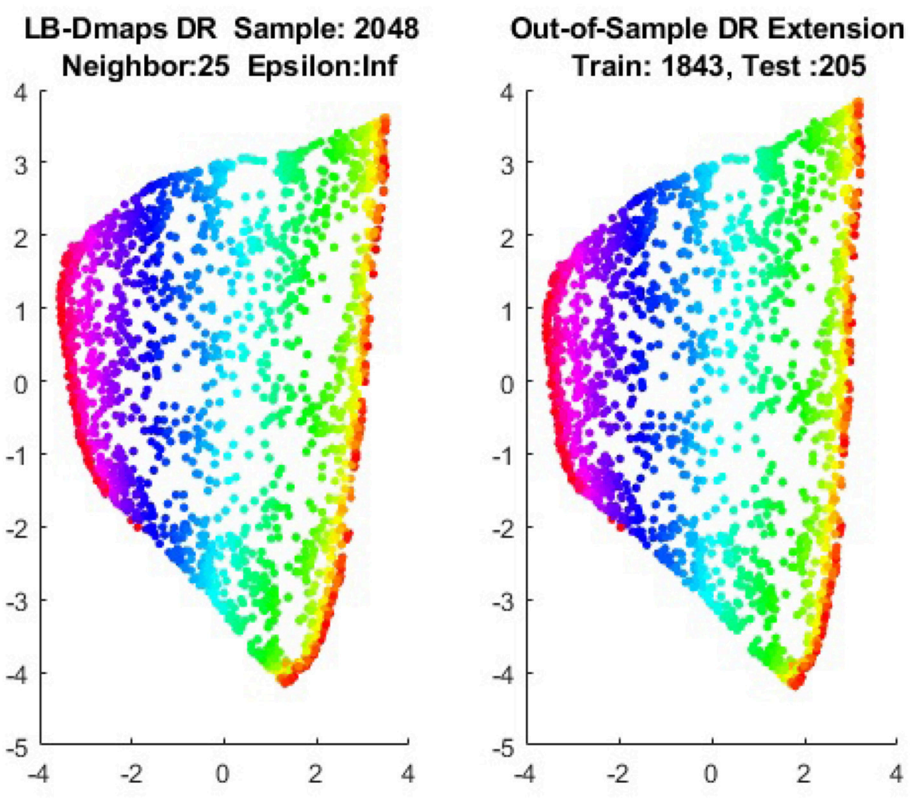

FIGURE 9 | LB out-of-sample extension for DR of Swiss Roll.

Proof. Write $\hat{\phi}_j = \mathcal{T}(\varphi_j)$. By 6, we have $\hat{\phi}_j(x) = \frac{\mathbf{T}(\mathbf{u}_j)(x)}{\sqrt{S(x)}}$. By Lemma 3, we obtain

$$\hat{\phi}_j(x) = \frac{1}{\sigma_j}\int_{\mathbf{X}} \frac{w(x,\mathbf{y})}{\sqrt{S(x)\mathbf{S}(\mathbf{y})}}\varphi_j(\mathbf{y})d\mu(\mathbf{y}), \tag{9}$$

which yields 7. Recall that $\frac{w(x,\mathbf{y})}{\sqrt{S(x)\mathbf{S}(\mathbf{y})}} = g(x,\mathbf{y})\sqrt{\frac{S(\mathbf{y})}{\mathbf{S}(\mathbf{y})}}$. For any $h \in H_g$, by $\langle h, g(\cdot,\mathbf{y})\rangle_{H_g} = h(\mathbf{y})$, we have

$$\langle h, \mathcal{T}(\mathbf{f})\rangle_{H_g} = \sum_{j=0}^{d}\frac{c_j}{\sigma_j}\int_{\mathbf{X}} h(\mathbf{y})\sqrt{\frac{S(\mathbf{y})}{\mathbf{S}(\mathbf{y})}}\varphi_j(\mathbf{y})d\mu(\mathbf{y}) = \left\langle \sqrt{\frac{S(\cdot)}{\mathbf{S}(\cdot)}}h(\cdot), \mathbf{f}\right\rangle_{H_{\mathbf{g}}},$$

which yields 8. ■

We now give the main theorem in this subsection.

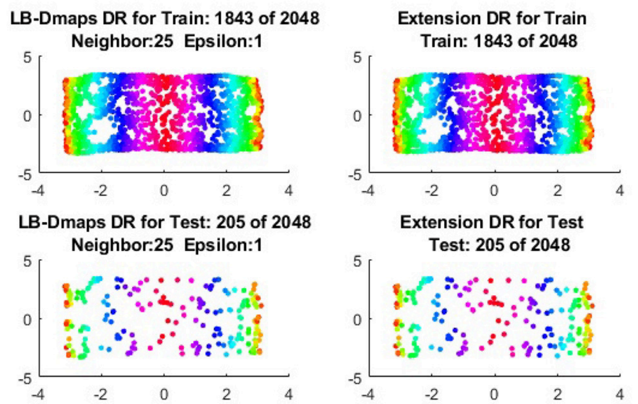

FIGURE 10 | Comparisons of DRs of training data and the testing data, respectively, for S-curve.

Theorem 5 *Let $\mathcal{T}$ be the operator defined in 6. Define $\hat{g}(x,y) = \sum_{j=0}^{d}\hat{\phi}_j(x)\hat{\phi}_j(y)$, where $\hat{\phi}_j = \mathcal{T}(\varphi_j)$, and let $H_{\hat{g}}$ be the RKHS associated with $\hat{g}$. Then,*

(i) $\mathcal{T}^*\mathcal{T} = I$ *on* $H_{\mathbf{g}}$.

(ii) $\{\hat{\phi}_0, \cdots, \hat{\phi}_d\}$ *is an orthonormal system in H_g, so that $H_{\hat{g}}$ is a subspace of H_g and $\mathcal{P} = \mathcal{T}\mathcal{T}^*$ is an orthogonal projection from H_g to $H_{\hat{g}}$. Therefore, we have $\mathcal{P}(\hat{\phi}_j) = \hat{\phi}_j$ and $\mathcal{T}^*(\hat{\phi}_j) = \varphi_j$.*

(iii) *The function $g_0(x,y) = g(x,y) - \hat{g}(x,y)$ is a Mercer's kernel. The RKHS H_{g_0} associated with g_0 is $(m-d)$ dimensional. Besides, $H_g = H_{\hat{g}} \bigoplus H_{g_0}$ and $H_{\hat{g}} \perp H_{g_0}$.*

(iv) *For any function $f \in H_{g_0}, f(\mathbf{x}) = 0, \mathbf{x} \in \mathbf{X}$.*

Proof. Recall that $\{\varphi_0, \varphi_1, \cdots, \varphi_d\}$ is an on. basis of $H_{\mathbf{g}}$. By 8 and 9, we have

$$\mathcal{T}^*\mathcal{T}(\varphi_j)(\mathbf{x}) = \hat{\phi}_j(\mathbf{x})\sqrt{\frac{S(\mathbf{x})}{\mathbf{S}(\mathbf{x})}} = \varphi_j(\mathbf{x}), \quad j = 0, 1, \cdots, d,$$

which yields $\mathcal{T}^*\mathcal{T}(\varphi_j) = \varphi_j$. Hence, $\mathcal{T}^*\mathcal{T} = I$ on $H_{\mathbf{g}}$. The proof of (i) is completed.

Note that

$$\langle \hat{\phi}_i, \hat{\phi}_j\rangle_{H_g} = \langle \varphi_i, \mathcal{T}^*\mathcal{T}(\varphi_j)\rangle_{H_{\mathbf{g}}} = \langle \varphi_i, \varphi_j\rangle_{H_{\mathbf{g}}} = \delta_{i,j},$$

which indicates that $\{\hat{\phi}_0(x), \cdots, \hat{\phi}_d(x)\}$ is an orthonormal system in H_g and $H_{\hat{g}}$ is a subspace of H_g. Because $\mathcal{P}^2 = \mathcal{P}$ and $\mathcal{P}(\hat{\phi}_j) = \hat{\phi}_j, j = 0, 1, \cdots, d$, $\mathcal{P}$ is an orthogonal projection from $H_{\tilde{g}}$ to $H_{\hat{g}}$, which proves (ii).

It is clear that (iii) is a direct consequence of (ii). Finally, we have

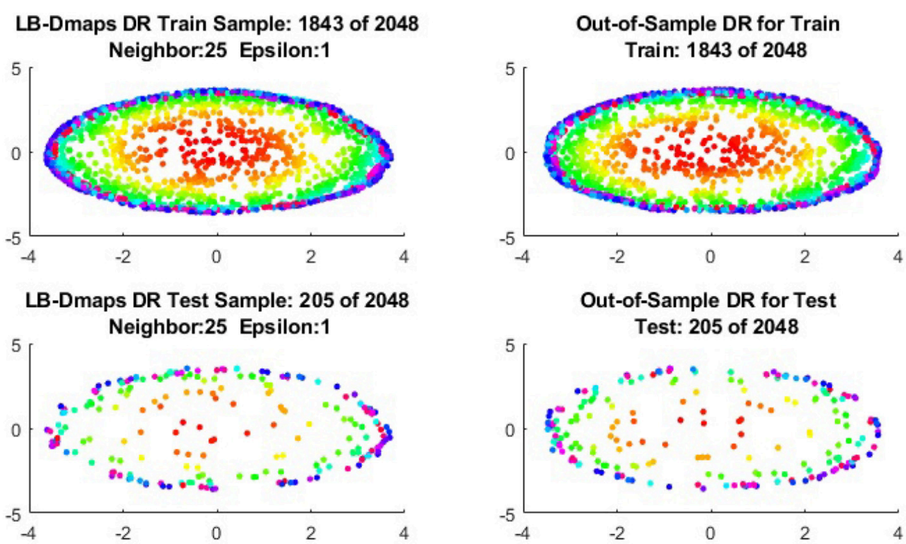

FIGURE 11 | Comparisons of DRs of training data and the testing data, respectively, for Punched Sphere.

$\mathcal{P}(f) = 0$ for $f \in H_{g_0}$, which yields $\mathcal{T}^*(f) = 0$. Therefore, $f(\mathbf{x}) = 0, \mathbf{x} \in \mathbf{X}$. The proof of (iv) is completed, ■

By Definition 2, the mapping $\Phi : \Phi(\mathbf{x}) = [\varphi_1(\mathbf{x}), \cdots, \varphi_d(\mathbf{x})]^T$ is a diffusion mapping from $\mathbf{X}$ to $\mathbb{R}^d$ and the set $\Phi(\mathbf{X})$ is a DR of $\mathbf{X}$. We now give the following definition.

Definition 6 *Let $\mathcal{T}$ be the operator defined in 6 and $\hat{\phi}_j = \mathcal{T}(\varphi_j)$. Then the set $\hat{\Phi}(X) = [\hat{\phi}_1(X), \cdots, \hat{\phi}_d(X)]^T \subset \mathbb{R}^d$ is called the* least-square out-of-sample DR extension *of the Dmaps with the Graph-Laplacian kernel.*

A DR extension on X is called *exact* if it is equal to a DR of X as defined in Definition 2 (see [20]). The following corollary is a direct consequence of Theorem 5.

Corollary 7 The least-square out-of-sample DR extension given by $\mathcal{T}$ from $H_{\mathbf{g}}$ to H_g is exact if and only if $\dim(H_g) = \dim(H_{\mathbf{g}})$, or equivalently, $H_{g_0} = \{0\}$.

3.2. Dmaps With the Laplace-Beltrami Kernel

The discussion on the out-of-sample DR extension of Dmaps with the Laplace-Beltrami (BL) kernel is similar to that in the previous subsection. Hence, in this subsection, we only outline the main results, skipping the details. We start the discussion from the asymmetrically normalized kernel

$$m(x, y) = \frac{1}{S(x)} w(x, y),$$

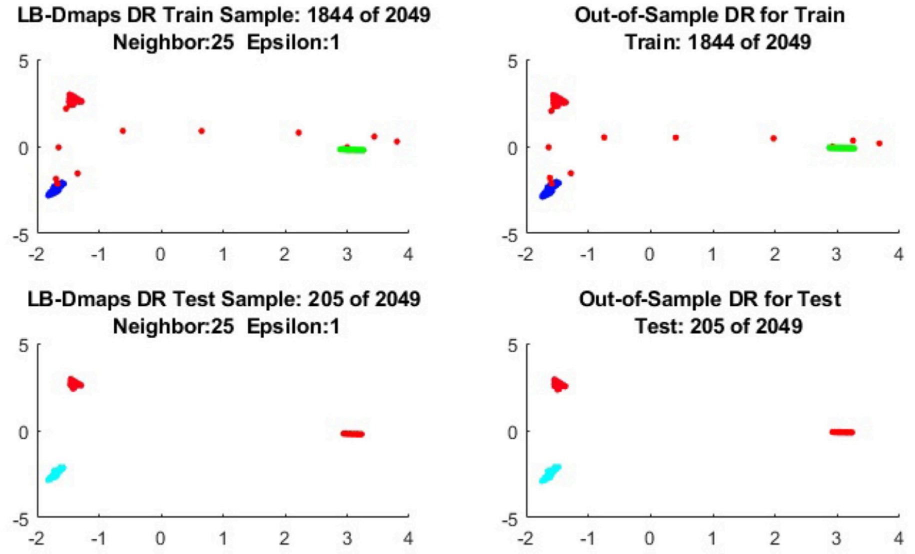

FIGURE 12 | Comparisons of DRs of training data and the testing data, respectively, for 3D Cluster.

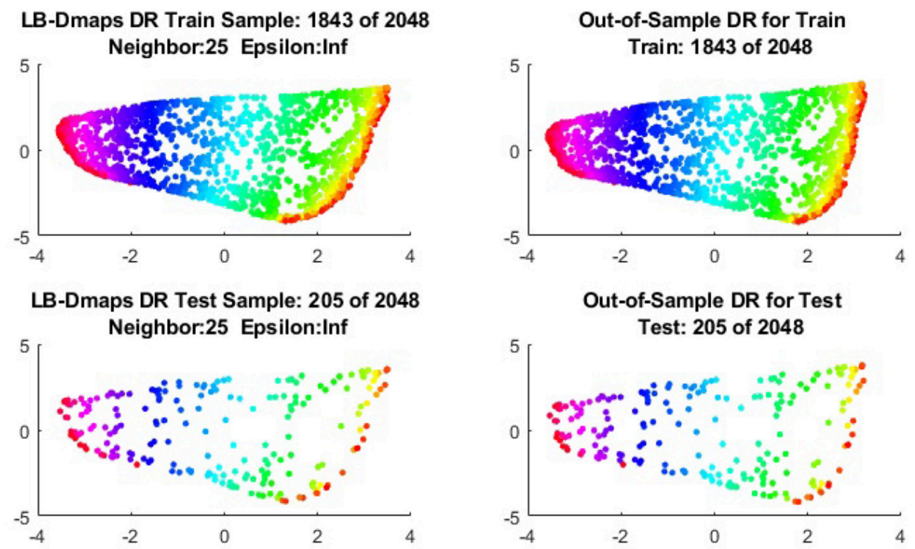

FIGURE 13 | Comparisons of DRs of training data and the testing data, respectively, for Swiss Roll.

which defines a random walk on the data set X such that $m(x,y)$ is the probability of the walk from the node x to the node y after a unit time. From the viewpoint of the random walk, we naturally modify the Gaussian kernel $w(x,y)$ to the following:

$$a(x,y) = \frac{w(x,y)}{S(x)S(y)}.$$

Then, we normalize it to

$$b(x,y) = \frac{a(x,y)}{\sqrt{P(x)P(y)}} = \frac{w(x,y)}{\sqrt{R(x)R(y)}},$$

where

$$P(x) = \int_X a(x,y)d\mu(y), \quad R(x) = S^2(x)P(x).$$

We call $b(x,y)$ the *Laplace-Beltrami* kernel of Dmaps, which relates to the data set X sampled from a manifold in $\mathbb{R}^D$. The greatest eigenvalue of $b(x,y)$ is also 1, which corresponds to the normalized eigenfunction $\sqrt{\frac{P(x)}{L}}$, where

$$L = \int_X P(x)d\mu(x).$$

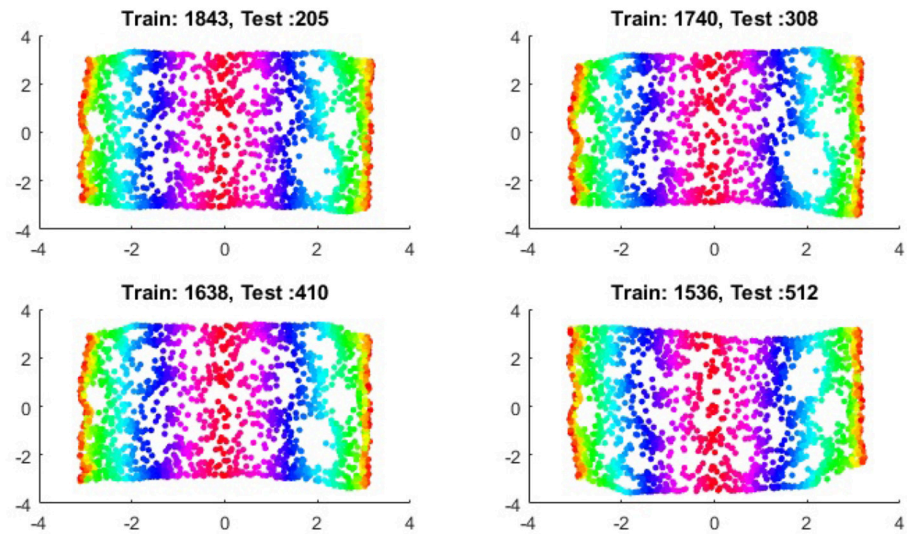

FIGURE 14 | LB out-of-sample extension for different sizes of the test sets of S-curve (I).

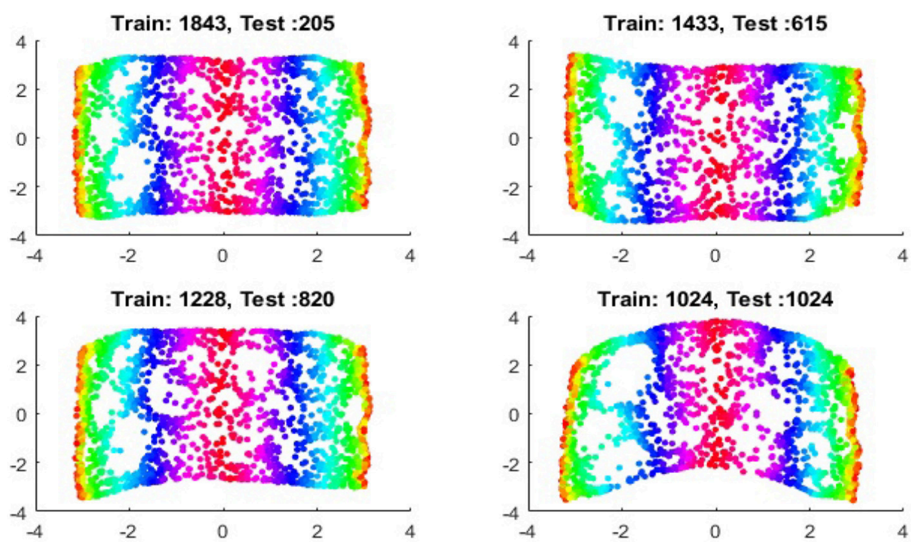

FIGURE 15 | LB out-of-sample extension for different sizes of the test sets of S-curve (II).

Let H_b be the RKHS associated with b and assume that the spectral decomposition of b is

$$b(x,y) = \sum_{j=0}^{m} \beta_j q_j(x) q_j(y) = \sum_{j=0}^{m} \psi_j(x)\psi_j(y),$$

where $1 = \beta_0 \geq \beta_1 \geq \cdots \geq \beta_m > 0$ and $\psi_j(x) = \sqrt{\beta_j} q_j(x)$. Similar to the discussion in the previous subsection, since $\psi_0 = \sqrt{\frac{P(x)}{L}}$ does not contains any feature of the data set, we exclude it from the feature space.

Definition 8 *The mapping* $\Psi : X \to \mathbb{R}^m : \Psi(x) = [\psi_1(x), \cdots, \psi_m(x)]^T$ *is called the Laplace-Beltrami diffusion mapping and the data set* $\Psi(X) \subset \mathbb{R}^m$ *is called a DR of* X *associated with Laplace-Beltrami Dmaps.*

We new assume again $X = \mathbf{X} \cup \mathbf{Z}$ and denote by $\mathbf{b}(\mathbf{x},\mathbf{y})$ the Laplace-Beltrami kernel on $\mathbf{X}$. Assume that spectral decomposition of $\mathbf{b}$ is

$$\mathbf{b}(\mathbf{x},\mathbf{y}) = \sum_{j=0}^{d} \gamma_j \mathbf{q}_j(\mathbf{x}) \mathbf{q}_j(\mathbf{y}). \tag{10}$$

Then the RKHS $H_{\mathbf{b}}$ associated with $\mathbf{b}$ has the canonic basis $\{\omega_0, \omega_1, \cdots, \omega_d\}$, where $\omega_j = \sqrt{\gamma_j}\mathbf{q}_j$.

Define the multiplicator from H_b to H_w by

$$\mathfrak{S}_R(f)(x) = \sqrt{R(x)} f(x), \quad x \in X,$$

and the multiplicator from $H_{\mathbf{b}}$ to $H_{\mathbf{w}}$ by

$$\mathfrak{S}_{\mathbf{R}}(\mathbf{f})(\mathbf{x}) = \sqrt{\mathbf{R}(\mathbf{x})}\mathbf{f}(\mathbf{x}), \quad \mathbf{x} \in \mathbf{X}.$$

The operator $\mathfrak{S}_R$ ($\mathfrak{S}_{\mathbf{R}}$) is an isometric mapping. We now define the *least-square extension* $\mathcal{M}$ from $H_{\mathbf{b}}$ to H_b by

$$\mathcal{M} = (\mathfrak{S}_R)^{-1} \circ \mathbf{T} \circ \mathfrak{S}_{\mathbf{R}}. \tag{11}$$

The integral representation of $\mathcal{M}$ is give by the following lemma:

Lemma 9 *Let* $\{\omega_0, \omega_1, \cdots, \omega_d\}$ *be the canonic basis of* $\mathbf{b}$. *Write* $\hat{\psi}_j = \mathcal{M}(\omega_j)$. *Then*

$$\hat{\psi}_j(x) = \frac{1}{\gamma_j} \int_{\mathbf{X}} \frac{w(x,\mathbf{y})}{\sqrt{R(x)\mathbf{R}(\mathbf{y})}} \omega_j(\mathbf{y}) d\mu(\mathbf{y}). \tag{12}$$

Particularly, for $\mathbf{f} = \sum_{j=0}^{d} c_j \omega_j \in H_{\mathbf{b}}$, *we have*

$$\mathcal{M}(\mathbf{f})(x) = \sum_{j=0}^{d} c_j \hat{\psi}_j(x) = \sum_{j=0}^{d} \frac{c_j}{\gamma_j} \int_{\mathbf{X}} \frac{w(x,\mathbf{y})}{\sqrt{R(x)\mathbf{R}(\mathbf{y})}} \omega_j(\mathbf{y}) d\mu(\mathbf{y}).$$

Its adjoint operator $\mathcal{M}^* : H_b \to H_{\mathbf{b}}$ *is given by*

$$\mathcal{M}^*(h)(\mathbf{x}) = h(\mathbf{x}) \sqrt{\frac{P(\mathbf{x})}{\mathbf{P}(\mathbf{x})}}, \quad h \in H_b.$$

Since the proof is similar to that for Lemma 4, we skip it here.

Theorem 10 *Let* $\mathcal{M}$ *be the operator defined in 11. Define* $\hat{b}(x,y) = \sum_{j=0}^{d} \hat{\psi}_j(x)\hat{\psi}_j(y)$, *where* $\hat{\psi}_j = \mathcal{M}(\omega_j)$, *and let* $H_{\hat{b}}$ *be the RKHS associated with* $\hat{b}$. *Then,*

1. $\mathcal{M}^*\mathcal{M} = I$ *on* $H_{\mathbf{b}}$.
2. $\{\hat{\psi}_0, \cdots, \hat{\psi}_d\}$ *is an orthonormal system in* H_b, *so that* $H_{\hat{b}}$ *is a subspace of* H_b *and* $\mathcal{Q} = \mathcal{M}\mathcal{M}^*$ *is an orthogonal projection from* H_b *to* $H_{\hat{b}}$. *Therefore, we have* $\mathcal{Q}(\hat{\psi}_j) = \hat{\psi}_j$ *and* $\mathcal{M}^*(\hat{\psi}_j) = \omega_j$.
3. *The function* $b_0(x,y) = b(x,y) - \hat{b}(x,y)$ *is a Mercer's kernel. The RKHS* H_{b_0} *associated with* b_0 *is* $(m-d)$ *dimensional. Besides,* $H_b = H_{\hat{b}} \bigoplus H_{b_0}$ *and* $H_{\hat{b}} \perp H_{b_0}$.
4. *For any function* $f \in H_{b_0}, f(\mathbf{x}) = 0, \mathbf{x} \in \mathbf{X}$.

We skip the proof of Theorem 10 because it is similar to that for Theorem 5. We now give the following definition:

Definition 11 *Let* $\mathcal{M}$ *be the operator defined in 11 and* $\hat{\psi}_j = \mathcal{M}(\omega_j)$. *Then the set* $\hat{\Psi}(X) = [\hat{\psi}_1(X), \cdots, \hat{\psi}_d(X)]^T \subset \mathbb{R}^d$ *is called the* least-square out-of-sample DR extension *of the Dmaps with the Laplace-Beltrami kernel.*

Corollary 12 The least-square out-of-sample DR extension given by $\mathcal{M}$ from $H_{\mathbf{b}}$ to H_b is exact if and only if $\dim(H_b) = \dim(H_{\mathbf{b}})$, or equivalently, $H_{b_0} = \{0\}$.

3.3. Algorithms for Out-of-Sample DR Extension of Dmaps

In this subsection, we present the algorithm for out-of-sample DR extension of Dmaps. The algorithm contains two parts. In the first part, we construct the DR for $\mathbf{X}$ by 4 and 10. In the second part, we extend the DR to the set X, by 9 and 12.

In the algorithm, we represent the data sets $\mathbf{X}$, $\mathbf{Z}$, and $X = \mathbf{X} \cup \mathbf{Z}$ as the $D \times N$, $D \times M$, and $D \times (N+M)$ matrices, respectively, so that $X = [\mathbf{X}, \mathbf{Z}]$. We assume the measure $d\mu(x) = dx$. Write $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N]$, $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_M]$, and $X = [x_1, \cdots, x_{(N+M)}]$, where $x_j = \mathbf{x}_j, 1 \leq j \leq N$ and $x_j = \mathbf{z}_{j-N}, N+1 \leq j \leq N+M$. Then we represent all kernels by matrices and all functions by vectors. For example, $\mathbf{w}$ is now represented by the $N \times N$ matrix with $\mathbf{w}_{i,j} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/\epsilon)$. To treat GL-map and LB-map in a uniform way, we write $\mathbf{S}_i = \sum_j \mathbf{w}_{i,j}$ and define

$$\mathbf{d}_i = \begin{cases} \sqrt{\mathbf{S}_i}, & \text{for GL-map} \\ \sqrt{\mathbf{S}_i \sum_j (\mathbf{w}_{i,j}/\mathbf{S}_j)}, & \text{for LB-map} \end{cases}$$

Then we set either kernel on $\mathbf{X}$ as the $N \times N$ matrix $\mathbf{k}$ with

$$\mathbf{k}_{i,j} = \frac{\mathbf{w}_{i,j}}{\mathbf{d}_i \mathbf{d}_j}.$$

The pseudo-code is given in Algorithm 1.

Algorithm 1: Out-of-Sample DR Extension for Dmaps

Require: Training data set $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N]$; testing data set $\mathbf{Z} = [\mathbf{z}_1, \mathbf{x}_2, \cdots, \mathbf{z}_M]$; kernel parameter ϵ for creating the diffusion kernel $W(\mathbf{x}, \mathbf{y}) = \exp(\|\mathbf{x} - \mathbf{y}\|^2/\epsilon)$; and optional threshold $\eta > 0$ for constructing sparse kernel forms of GL or LB.

Ensure: Out-of-sample DR extension Y for X using Dmaps.

Part I. Create DR of $\mathbf{X}$.

Construct the Gaussian kernel

$$\mathbf{w}_{i,j} = \begin{cases} W(\mathbf{x}_i, \mathbf{x}_j), & W(\mathbf{x}_i, \mathbf{x}_j) \geq \eta \\ 0, & W(\mathbf{x}_i, \mathbf{x}_j) < \eta \end{cases}$$

Compute the mass functions on $\mathbf{X}$: $\mathbf{S}_i = \sum_j \mathbf{w}_{i,j}$.

if kernel form = GL **then** $\mathbf{d}_i = \sqrt{\mathbf{S}_i}$;

else $\mathbf{d}_i = \mathbf{S}_i \sqrt{\sum_j (\mathbf{w}_{i,j}/\mathbf{S}_j)}$.

end if

Construct the kernel $\mathbf{k}$ for Dmaps on $\mathbf{X}$: $\mathbf{k}_{i,j} = \mathbf{w}_{i,j}/(\mathbf{d}_i \mathbf{d}_j)$.

Make the spectral decomposition of $\mathbf{k}$: $\mathbf{k} = U^T \Sigma U = \Phi^T \Phi$, where $\Phi = \sqrt{\Sigma} U \in \mathbb{R}^{(d+1)\times N}$.

Let $\mathbf{Y}$ be the matrix obtained by removing the first row of Φ. $\triangleright$ The column set of $\mathbf{Y}$ is the DR of $\mathbf{X}$.

Part II. Make out-of-sample DR extension on $X = \mathbf{X} \cup \mathbf{Z}$.

Compute the mass functions on X: $S_i = \sum_j w_{i,j}$.

if kernel form = GL **then** $d_i = \sqrt{S_i}$;

else $d_i = S_i \sqrt{\sum_j (w_{i,j}/S_j)}$.

end if

Extend $\mathbf{w}$ to w and $\mathbf{k}$ to k on X.

Set $\mathbf{D} = \text{diag}(\mathbf{d}_1, \cdots, \mathbf{d}_N)$ and $D_{\mathbf{X}} = \text{diag}(d_1, \cdots, d_N)$.

Compute $Y_{\mathbf{X}} = \mathbf{Y}\mathbf{D}^{-1} D_{\mathbf{X}}$.

Set $\Sigma_d = \text{diag}(\sigma_1, \cdots, \sigma_d)$, $D_{\mathbf{Z}} = \text{diag}(d_{N+1}, \cdots, d_{N+M})$, and $w_{\mathbf{Z}} = [w_{i,j}]_{i=1,j=N+1}^{N,N+M}$.

Compute $Y_{\mathbf{Z}} = \Sigma_d^{-1} \mathbf{Y}\mathbf{D}^{-1} w_{\mathbf{Z}} D_{\mathbf{Z}}^{-1}$.

Set $Y = [Y_{\mathbf{X}}, Y_{\mathbf{Z}}]$.

4. ILLUSTRATIVE EXAMPLES

In this section, we give several illustrative examples to show the validity of the Dmaps out-of-sample extensions. We employ four benchmark artificial data sets, S-curve, Swiss roll, punched sphere, and 3D cluster, in our samples. The graphs of these four data sets are give in **Figure 1**.

4.1. Out-of-Sample Extension by Graph-Laplacian Mapping

We first show the examples for the out-of-sample extensions provided by Graph-Laplacian mapping for the four benchmark figures. We set the size of each of these data sets by $|X| = 2,048$. When the out-of-example algorithm is applied, we choose the size of the training data set to be $|\mathbf{X}| = 1,843$, which is 90% of the all samples, and choose the size of the testing set $|\mathbf{Z}| = 205$, which is 10% of all samples. The parameters for the Graph-Laplacian kernel are set as follows: For obtaining the sparse kernel, we choose 25 nearest neighbors for every node, and assign the diffusion parameter $\epsilon = 1$ for S-curve, Punched Sphere, and 3D Cluster, while assign $\epsilon = \infty$ for Swiss Roll. We compare the DR result of the whole set X obtained by out-of-example extension with that obtained without out-of-example extension in the **Figures 2–5**. The figures show that the DRs obtained by out-of-sample extensions are satisfactory.

4.2. Out-of-Sample Extension by Laplace-Beltrami Mapping

We now show the examples for the out-of-sample extensions provided by Laplace-Beltrami mapping for the same four benchmark figures. We set the same sizes for $|X|$,$|\mathbf{X}|$, and $|\mathbf{Z}|$, respectively. The parameters for the Laplace-Beltrami kernel are also set the same as for Graph-Laplacian kernel. The results of the comparisons are give in **Figures 6–9**.

To give more detailed comparisons, in **Figures 10–13**, we show the DRs of the training data and the testing data obtained by out-of-extensions and without extensions, respectively, for LB mapping.

It is a common sense that if we reduce the size of the training set while increase the size for the testing set, the out-of-sample extension will introduce larger errors for DR. **Figures 14–15** show that, in a relative large scope, say, the size of the testing set is no greater than the size of the training set. the out-of-sample extension still produces the acceptable results.

DATA AVAILABILITY

No datasets were generated or analyzed for this study.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

1. Bellman R. *Adaptive Control Processes: A Guided Tour.* Princeton, NJ: Princeton University Press (1961).
2. Scott DW, Thompson JR. Probability density estimation in higher dimensions. In: Gentle JE, editor. *Computer Science and Statistics: Proceedings of the Fifteenth Symposium on the Interface.* Amsterdam; New York, Ny; Oxford: North Holland-Elsevier Science Publishers (1983). p. 173–9.
3. Lee JA, Verleysen M. *Nonlinear Dimensionality Reduction.* New York, NY: Springer (2007).
4. Wang JZ. *Geometric Structure of High-Dimensional Data and Dimensionality Reduction.* Beijing; Berlin; Heidelberg: Higher Educaiton Press; Springer (2012).
5. Jolliffe IT. *Principal Component Analysis.* Springer Series in Statistics. Berlin: Springer-Verlag (1986).
6. Zhang ZY, Zha HY. Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment. *SIAM J Sci Comput.* (2004) **26**:313–38.
7. Schölkopf B, Smola A, Müller K-R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* (1998) **10**:1299–319.

8. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science.* (2000) **290**:2323–6. doi: 10.1126/science.290.5500.2323
9. Donoho DL, Grimes C. Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data. *Proc Natl Acad Sci USA.* (2003) **100**:5591–6. doi: 10.1073/pnas.1031596100
10. Shmueli Y, Sipola T, Shabat G, Averbuch A. Using affinity perturbations to detect web traffic anomalies. In: *The 11th International Conference on Sampling Theory and Applications* (Bremen) (2013).
11. Shmueli Y, Wolf G, Averbuch A. Updating kernel methods in spectral decomposition by affinity perturbations. *Linear Algebra Appl.* (2012) **437**:1356–65. doi: 10.1016/j.laa.2012.04.035
12. Balasubramanian M, Schwaartz E, Tenenbaum J, de Silva V, Langford J. The isomap algorithm and topological staility. *Science* (2002) **295**:7. doi: 10.1126/science.295.5552.7a
13. Coifman RR, Lafon S. Diffusion maps. *Appl Comput Harmon Anal.* (2006) **21**:5–30. doi: 10.1016/j.acha.2006.04.006
14. Coifman RR, Lafon S. Geometric harmonics: a novel tool for multiscale out-of-sample extension of empirical functions. *Appl Comput Harmon Anal.* (2006) **2**:31–52. doi: 10.1016/j.acha.2005.07.005
15. Ng A, Jordan M, Weiss Y. On spectral clustering: analysis and an algorithm. *Adv Neural Inform Process Syst.* (2001) **14**:849–56.
16. Shi J, Malik J. Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Mach Intell.* (2000) **22**:888–905. doi: 10.1109/34.868688
17. Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* (2003) **15**:1373–96. doi: 10.1162/089976603321780317
18. Aizenbud Y, Bermanis A, Averbuch A. PCA-based out-of-sample extension for dimensionality reduction. *arXiv: 1511.00831* (2015).
19. Bengio Y, Paiement J, Vincent P, Delalleau O, Le Roux N, Ouimet M. Out-of-sample extensions for LLE, Isomap, MDS, eigenmaps, and spectral clustering. In: Thrun S, Saul L, Schölkopf B, editors. *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press (2004).
20. Wang JZ. Mathematical analysis on out-of-sample extensions. *Int J Wavelets Multiresol Inform Process.* (2018) **16**:1850042. doi: 10.1142/S021969131850042X
21. Aronszajn N. Theory of reproducing kernels. *Trans Amer Math Soc.* (1950) **68**:337–404.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.