Applied and Computational Harmonic Analysis 68 (2024) 101583

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

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

Spatiotemporal analysis using Riemannian composition of diffusion operators

Tal Shnitzer [a,*], Hau-Tieng Wu [b], Ronen Talmon [c]

[a] Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA, USA
[b] Department of Mathematics and Department of Statistical Science, Duke University, Durham, NC, USA
[c] Viterbi Faculty of Electrical and Computer Engineering, Technion - Israel Institute of Technology, Haifa, Israel

ARTICLE INFO

Communicated by Leslie F. Greengard

Keywords:
Manifold learning
Diffusion maps
Riemannian geometry
Symmetric positive-definite matrices

ABSTRACT

Multivariate time-series have become abundant in recent years, as many data-acquisition systems record information through multiple sensors simultaneously. In this paper, we assume the variables pertain to some geometry and present an operator-based approach for spatiotemporal analysis. Our approach combines three components that are often considered separately: (i) *manifold learning* for building operators representing the geometry of the variables, (ii) *Riemannian geometry of symmetric positive-definite matrices* for multiscale composition of operators corresponding to different time samples, and (iii) *spectral analysis* of the composite operators for extracting different dynamic modes. We propose a method that is analogous to the classical wavelet analysis, which we term Riemannian multi-resolution analysis (RMRA). We provide some theoretical results on the spectral analysis of the composite operators, and we demonstrate the proposed method on simulations and on real data.

1. Introduction

Multivariate time-series (temporal signals) have been studied in the statistics and signal processing societies for many years (e.g., see [37,81] for a non-exhaustive literature survey), and traditional analysis methods usually highly depend on some predefined models and do not consider the unknown nonlinear structure of the variables that often exists underlying the high-dimensional time samples. In order to accommodate contemporary data acquisitions and collections, large research activity has been devoted to developing spatiotemporal analysis that is specifically-designed to infer this underlying structure and/or take it explicitly into account. In the last decade, perhaps the most notable attempts to handle such signals with a geometry defined by graphs are graph signal processing [59,64,65,70], graph neural networks [44,66], and geometric deep learning [13]. Another prominent line of work is based on an operator-theoretic approach for dynamical systems analysis [8,14,45,67,82,87], where the time samples have a manifold structure. Still, despite these recent efforts, to the best of our knowledge, when there is a time-varying manifold structure underlying the time samples, only a few works are available, e.g. [2,32,33].

In this work, we propose a new multi-resolution spatiotemporal analysis of multivariate time-series. In contrast to standard multi-resolution analysis using wavelets defined on Euclidean space [22,53], we present an operator-based analysis approach combining manifold learning and Riemannian geometry, which we term *Riemannian multi-resolution analysis* (RMRA). Concretely, consider a multivariate time-series $\{\mathbf{x}_t\}$. Suppose the temporal propagation of the time-series at time step t can be modeled by two diffeomorphic

* Corresponding author.
E-mail address: talsd@mit.edu (T. Shnitzer).

https://doi.org/10.1016/j.acha.2023.101583
Received 3 February 2022; Received in revised form 21 July 2023; Accepted 8 August 2023
Available online 21 August 2023

manifolds $f_t : \mathcal{M}_t \to \mathcal{M}_{t+1}$, and suppose the corresponding pairs of time samples $(\mathbf{x}_t, \mathbf{x}_{t+1})$ are given by $\mathbf{x}_t[i] \in \mathcal{M}_t$ and $\mathbf{x}_{t+1}[i] = f_t(\mathbf{x}_t[i]) \in \mathcal{M}_{t+1}$, where $\mathbf{x}_t[i]$ is the ith entry of the sample $\mathbf{x}_t$ for $i = 1, \ldots, N$. Note that the entries of the samples $\mathbf{x}_t$ lie on a manifold, and therefore, each entry is typically high-dimensional. In other words, at each time t, we have N high-dimensional points that are distributed on the manifold $\mathcal{M}_t$. Our RMRA consists of the following steps. First, we construct a diffusion operator for each time sample $\mathbf{x}_t$, characterizing its underlying manifold $\mathcal{M}_t$. This step is performed using a manifold learning technique, diffusion maps [18], that facilitates a finite-dimensional matrix approximation of the Laplacian operator of the manifold based on the time sample. This approximation is informative because the Laplacian operator is known to bear the geometric information of the manifold [6,41]. Then, for each pair of temporally consecutive time frames $(\mathbf{x}_t, \mathbf{x}_{t+1})$, we present two composite operators based on "Riemannian combinations" of the two respective diffusion operators. Typically, diffusion operators are not symmetric, but they are similar to symmetric positive-definite (SPD) matrices. We could thus define diffusion operators as SPD matrices, whose space is endowed with a Riemannian structure. Therefore, taking into account this Riemannian manifold structure for the composition of the operators is natural. Indeed, we show, both theoretically and in practice, that one operator enhances common components that are expressed similarly in $\mathcal{M}_t$ and $\mathcal{M}_{t+1}$, while the other enhances common components that are expressed differently. These properties could be viewed as analogous to low-pass and high-pass filters in this setting, leading to a spatiotemporal decomposition of the multivariate time series into "low frequency" and "high frequency" components, by considering the common components expressed similarly (resp. differently) as the slowly (resp. rapidly) varying components.

To facilitate the multi-resolution analysis of the entire temporal sequence, the construction of the composite operators is recursively repeated at different time scales. Since the composite operators are viewed as low-pass and high-pass filters, the proposed framework can be viewed as analogous to the wavelet decomposition for time-varying manifolds in the following sense. At each iteration, the two consecutive time samples are "fused" using the composite operators, "decomposing" the multivariate time-series into two components: one that varies slowly and one that varies rapidly. The fast varying component is viewed as the "spectral feature" of the "first layer", and the slowly varying component is "downsampled", decomposed again in the next iteration using the composite operators into a slow component and a fast component. Again, the fast component leads to the "spectral feature" of the "second layer". By iterating this procedure, the multivariate time series is decomposed in multiple resolutions.

Broadly, the basic building block of our analysis, focusing on one time step, consists of two construction steps. First, given two consecutive time samples $(\mathbf{x}_t, \mathbf{x}_{t+1})$, we learn the underlying manifolds $\mathcal{M}_t$ and $\mathcal{M}_{t+1}$, and then, we study the (unknown) diffeomorphism f_t. We posit that this building block can serve as an independent analysis module by itself. Indeed, the setting of one time step we consider can be recast as a related multiview data analysis problem (see Section 2). Consider two diffeomorphic manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$ and the diffeomorphism $f : \mathcal{M}_1 \to \mathcal{M}_2$. Let $x \in \mathcal{M}_1$ and $y = f(x) \in \mathcal{M}_2$. The pair (x, y) could be considered as two views of some object of interest, providing distinct and complementary information. Applying the proposed two-step procedure to this case first learns the manifold of each view, and then, studies the diffeomorphism representing the relationship between the two views. In [68], the diffeomorphism was analyzed in terms of common and unique components, which were represented by the eigenvectors of the diffusion operators. Here, we further characterize these spectral common components. Roughly, the common components are classified into two kinds: components that are expressed similarly in the two manifolds in the sense that they have similar eigenvalues in both manifolds, and components that are expressed differently in the sense that they have different eigenvalues. Furthermore, we refine the analysis and in addition to considering *strictly common components*, i.e., the same eigenvectors in both manifolds, we also consider *weakly common components*, i.e., similar but not the same eigenvectors. In contrast to the local analysis presented in [68], we provide global and spectral analyses, showing that our method indeed extracts and identifies these different components.

We demonstrate the proposed RMRA on a dynamical system with a transitory double gyre configuration. We show that this framework is sensitive to the change rate of the dynamical system at different time-scales. Such a framework may be especially suitable for studying non-stationary multivariate time-series, particularly when there is a nontrivial geometric relationship among the multivariate coefficients. In addition, for the purpose of multimodal data analysis, we demonstrate that the proposed Riemannian composite operators enhance common structures in remote sensing data captured using hyperspectral and LiDAR sensors.

The remainder of this paper is organized as follows. In Section 2, we review related work. In Section 3, we present preliminaries. In Section 4, we present the proposed approach for multi-resolution spatiotemporal analysis using Riemannian composition of diffusion operators. Section 5 shows experimental results. In Section 6, we present spectral analysis of the proposed composite operators. Finally, in Section 7, we conclude the paper.

2. Related work

2.1. Manifold learning, diffusion maps, and diffusion wavelets

Manifold learning is a family of methods that consider data lying on some inaccessible manifold and provide a new low-dimensional representation of the data based on intrinsic patterns and similarities in the data [4,18,63,79]. From an algorithmic viewpoint, manifold learning techniques are broadly based on two stages. The first stage is the computation of a typically positive kernel that provides a notion of similarity between the data points. The second stage is the spectral analysis of the kernel, giving rise to an embedding of the data points into a low-dimensional space. Such a two-stage procedure results in aggregation of multiple pairwise similarities of data points, facilitating the extraction of the underlying manifold structure. This procedure was shown to be especially useful when there are limited high-dimensional data, plausibly circumventing the curse of dimensionality.

While the spectral analysis of the kernels has been the dominant approach and well investigated, recent work explores different directions as well. One prominent direction employs an operator-based analysis, which has led to the development of several key

methods. Arguably the first and most influential is diffusion maps [18],[1] where a transition matrix is constructed based on the kernel, forming a random walk on the dataset; such transition matrix is viewed as a diffusion operator on the data. There has been abundant theoretical support for diffusion maps. For example, it was shown in [4,40,71] that the operator associated with diffusion maps converges point-wisely to the Laplace-Beltrami operator of the underlying manifold, which embodies the geometric properties of the manifold, and its eigenfunctions form a natural basis for square integrable functions on the manifold. The spectral convergence of the eigenvalues and eigenvectors of the operator associated with diffusion maps to the eigenvalues and eigenfunctions of the Laplace-Beltrami operator was first explored in [5], and recently, the L^∞ spectral convergence with convergence rate was reported in [28]. See [28] and references therein for additional related work in this direction. The robustness of the diffusion maps operator was studied in [29,30], and recently, its behavior under different noise levels and kernel bandwidths was explored using random matrix theory [25].

The propagation rules associated with this diffusion operator are in turn used for defining a new distance, the so-called diffusion distance, which was shown to be useful and informative in many domains and applications [46,48,76,84]. This notion of diffusion promoted the development of well-designed and controlled anisotropic diffusions for various purposes, e.g., nonlinear independent component analysis [72], intrinsic representations [77], reduction of stochastic dynamical systems [27,73], and time-series forecasting [8,9,36,87] and filtering [69], to name but a few. In another line of work, the combination and composition of diffusion operators led to the development of new manifold learning techniques for learning multiple manifolds [47,51,54,68] as well as for time-series analysis [32].

A related line of work that considers multivariate time-series (high-dimensional temporal signals) introduces ways to define wavelets on graphs and manifolds, e.g., [19,38,62]. These techniques extend the classical wavelet analysis [53] from one or two dimensional Euclidean space to high-dimensional non-Euclidean spaces represented by graphs and manifolds. Specifically, diffusion wavelets [19] makes use of a hierarchy of diffusion operators with multiple well-designed diffusion scales organized in a dyadic tree. Importantly, none of these methods addresses an underlying manifold with a time-varying metric, but rather a fixed metric that exhibits different characteristics in different scales.

2.2. Manifold learning for sensor fusion

The basic building block of our RMRA is based on two diffeomorphic manifolds $\mathcal{M}_t$ and $\mathcal{M}_{t+1}$, which represent the temporal evolution at time t. A similar setting consisting of two diffeomorphic manifolds, say $\mathcal{M}_1$ and $\mathcal{M}_2$, has recently been investigate in the context of multimodal data analysis and sensor fusion.

The sensor fusion problem typically refers to the problem of harvesting useful information from data collected from multiple, often heterogeneous, sensors. Sensor fusion is a gigantic field. One line of work focuses on the extraction, analysis and comparison of the components expressed by the different sensors for the purpose of gaining understanding of the underlying scene [21,58,75]. However, due to the complex nature of such data, finding informative representations and metrics of these components by combining the information from the different sensors is challenging. Recently, several papers propose data fusion methods relying on manifold learning techniques and operator-based data analysis [23,31,43,47,51,68,78]. The basic idea is that data from different modalities or views are fused by constructing kernels that represent the data from each view and operators that combine those kernels. Different approaches are considered for the combination of kernels. Perhaps the most relevant to the present work is the alternating diffusion operator, which was introduced in [47,78] and shown to recover the common latent variables from multiple modalities. This operator is defined based on a product of two diffusion operators and then used for extracting a low dimensional representation of the common components shared by the different sensors. Other related approaches include different combinations of graph Laplacians [23,31], product of kernel density estimators and their transpose for nonparametric extension of canonical correlation analysis [55], and various other combinations of diffusion operators [43,51,68]. For a more comprehensive review of the different approaches, see [68] and references therein.

Largely, most existing sensor fusion algorithms, and particularly those based on kernel and manifold learning approaches, focus on the extraction and representation of the common components, in a broad sense. The sensor fusion framework proposed in [68] extends this scope and considers both the common components and the unique components of each sensor. Therefore, in the context of the present work, it could be used for the analysis of the basic building block consisting of two diffeomorphic manifolds. However, similarly to the other methods described above, the kernel combination in [68] is achieved through linear operations, thereby ignoring the prototypical geometry of the kernels. Conversely, in this work, by taking the Riemannian structure of SPD matrices into account, we propose a new geometry-driven combination of kernels and a systematic analysis. This aspect of our work could be viewed as an extension of [68] for the purpose of sensor fusion and multimodal data analysis, in addition to the new utility for multivariate time-series analysis.

3. Preliminaries

In this section we briefly present the required background for our method. For further details on the theory and derivations we refer the readers to [10] and [18].

[1] Laplacian eigenmaps could also be considered if the diffusion time is not taken into account [4].

3.1. Riemannian structure of SPD matrices

In many recent studies, representing the raw data using SPD matrices and taking into account their specific Riemannian geometry have shown promising results, e.g. in computer vision [7,83], for domain adaptation [85], on medical data [3] and in recognition tasks [39]. For example, Barachant et al. [3] proposed a support-vector-machine (SVM) classifier that takes into account the Riemannian geometry of the features, which are SPD covariance matrices, representing Electroencephalogram (EEG) recordings. They showed that their "geometry-aware" classifier obtains significantly better results compared with a classifier that simply vectorizes the covariance matrices.

Here, we consider the space of SPD matrices endowed with the so-called affine-invariant metric [60]. Using this particular Riemannian geometry results in closed-form expressions for useful properties and operations, such as the geodesic path connecting two points on the manifold [10] and the logarithmic map and the exponential map [60], which locally project SPD matrices onto the tangent space and back. While the focus is on the affine-invariant metric, which is arguably the most widely used, we remark that other geometries of SPD matrices exist, e.g., the log-Euclidean [1,61], the log-det [15,74], the log-Cholesky [49], and the Bures-Wasserstein [11,52], which could be considered as well. In the context of this work, since diffusion operators are strictly positive in principal but in practice often have negligible eigenvalues, one particular advantage of the affine-invariant geometry is its existing extensions to symmetric positive semi-definite (SPSD) matrices (see Section 3.2).

Consider the set of symmetric matrices in $\mathbb{R}^{N\times N}$, denoted by $\mathcal{S}_N$. A symmetric matrix $\mathbf{W}\in\mathcal{S}_N$ is an SPD matrix if it has strictly positive eigenvalues. Let $\mathcal{P}_N$ denote the set of all $N\times N$ SPD matrices. The tangent space at any point in this set is the space of symmetric matrices $\mathcal{S}_N$. We denote the tangent space at $\mathbf{W}\in\mathcal{P}_N$ by $\mathcal{T}_{\mathbf{W}}\mathcal{P}_N$. In this work we consider the following affine-invariant metric in the tangent space at each matrix $\mathbf{W}\in\mathcal{P}_N$, which forms a differentiable Riemannian manifold [56]:

$$\langle \mathbf{D}_1,\mathbf{D}_2\rangle_{\mathbf{W}} = \left\langle \mathbf{W}^{-1/2}\mathbf{D}_1\mathbf{W}^{-1/2}, \mathbf{W}^{-1/2}\mathbf{D}_2\mathbf{W}^{-1/2}\right\rangle \tag{1}$$

where $\mathbf{D}_1,\mathbf{D}_2\in\mathcal{T}_{\mathbf{W}}\mathcal{P}_N$ denote matrices in the tangent space at $\mathbf{W}\in\mathcal{P}_N$ and $\langle\cdot,\cdot\rangle$ is given by the standard Frobenius inner product $\langle\mathbf{A},\mathbf{B}\rangle = \mathrm{Tr}\left(\mathbf{A}^T\mathbf{B}\right)$. Using this metric, there is a unique geodesic path connecting any two matrices $\mathbf{W}_1,\mathbf{W}_2\in\mathcal{P}_N$ [10], which is explicitly given by:

$$\gamma_{\mathbf{W}_1\to\mathbf{W}_2}(p) = \mathbf{W}_1^{1/2}\left(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}\right)^p\mathbf{W}_1^{1/2},\quad p\in[0,1]. \tag{2}$$

The arc-length of this geodesic path defines the Riemannian distance on the manifold,

$$d_R^2\left(\mathbf{W}_1,\mathbf{W}_2\right) = \left\|\log\left(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}\right)\right\|_F^2. \tag{3}$$

Using the Fréchet mean, we define the Riemannian mean of a set of matrices, $\mathbf{W}_1,\ldots,\mathbf{W}_n$, by $\arg\min_{\mathbf{W}\in\mathcal{P}_N}\sum_{i=1}^n d_R^2\left(\mathbf{W},\mathbf{W}_i\right)$. The Riemannian mean of two matrices is a special case, which coincides with the mid-point of the geodesic path connecting them and has the following closed form:

$$\gamma_{\mathbf{W}_1\to\mathbf{W}_2}(1/2) = \mathbf{W}_1^{1/2}\left(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}\right)^{1/2}\mathbf{W}_1^{1/2} \tag{4}$$

The mapping between the Riemannian manifold of SPD matrices and its tangent space is given by the exponential map and the logarithmic map. Each matrix $\mathbf{D}$ in the tangent space at $\mathbf{W}\in\mathcal{P}_N$ can be seen as the derivative of the geodesic connecting $\mathbf{W}$ and $\tilde{\mathbf{W}} = \mathrm{Exp}_{\mathbf{W}}(\mathbf{D})$, i.e., $\gamma_{\mathbf{W}\to\tilde{\mathbf{W}}}(p)$, at $p=0$. The exponential map in this setting has a known closed-form given by [56]:

$$\mathrm{Exp}_{\mathbf{W}}(\mathbf{D}) = \mathbf{W}^{1/2}\exp\left(\mathbf{W}^{-1/2}\mathbf{D}\mathbf{W}^{-1/2}\right)\mathbf{W}^{1/2}, \tag{5}$$

where $\mathrm{Exp}_{\mathbf{W}}(\mathbf{D})\in\mathcal{P}_N$ and $\exp(\cdot)$ is applied to the eigenvalues. The inverse of the exponential map is the logarithmic map, which is explicitly given by:

$$\mathrm{Log}_{\mathbf{W}}(\tilde{\mathbf{W}}) = \mathbf{W}^{1/2}\log\left(\mathbf{W}^{-1/2}\tilde{\mathbf{W}}\mathbf{W}^{-1/2}\right)\mathbf{W}^{1/2}, \tag{6}$$

where $\log(\cdot)$ is applied to the eigenvalues, $\tilde{\mathbf{W}}\in\mathcal{P}_N$, and $\mathrm{Log}_{\mathbf{W}}(\tilde{\mathbf{W}})\in\mathcal{T}_{\mathbf{W}}\mathcal{P}_N$.

3.2. An extension to SPSD matrices

In practice, the matrices of interest are often not strictly positive, but rather symmetric positive *semi*-definite (SPSD) matrices, that is symmetric matrices with non-negative eigenvalues. Below is a summary of a Riemannian geometry introduced in [12] that extends the affine-invariant metric. We remark that the existence of such an extension serves as an additional motivation to particularly consider the affine-invariant metric over the alternatives.

Let $\mathcal{S}^+(r,N)$ denote the set of $N\times N$ SPSD matrices of rank $r<N$, and let $\mathcal{V}_{N,r}$ denote the set of $N\times r$ matrices with orthonormal columns, i.e. $\mathbf{V}^T\mathbf{V} = \mathrm{I}_{r\times r},\ \forall\mathbf{V}\in\mathcal{V}_{N,r}$. Given an SPSD matrix $\mathbf{W}\in\mathcal{S}^+(r,N)$, we consider the following decomposition:

$$\mathbf{W} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T, \tag{7}$$

where $\mathbf{V}\in\mathcal{V}_{N,r}$ and $\mathbf{\Lambda}\in\mathcal{P}_r$. Consequently, the pair $(\mathbf{V},\mathbf{\Lambda})\in\mathcal{V}_{N,r}\times\mathcal{P}_r$ can be considered as a representation of the matrix $\mathbf{W}$. Based on this representation, the authors of [12] proposed to present vectors in the tangent space of $\mathcal{S}^+(r,N)$ at $\mathbf{W}$, denoted by $\mathcal{T}_{\mathbf{W}}\mathcal{S}^+(r,N)$,

by the infinitesimal variation $(\mathbf{\Delta}, \mathbf{D})$, where $\mathbf{\Delta} = \mathbf{V}_\perp \mathbf{B}$, $\mathbf{V}_\perp \in \mathcal{V}_{N,N-r}$, $\mathbf{V}_\perp \mathbf{V} = 0$, $\mathbf{B} \in \mathbb{R}^{(N-r)\times r}$ and $\mathbf{D} \in \mathcal{T}_{\mathbf{\Lambda}} \mathcal{P}_r = \mathcal{S}_r$. Using these tangent vectors, the metric on the tangent space of the SPSD manifold can be defined as a generalization of the metric on the manifold of SPD matrices:

$$\left\langle \left(\mathbf{\Delta}_1, \mathbf{D}_1\right), \left(\mathbf{\Delta}_2, \mathbf{D}_2\right) \right\rangle_{(\mathbf{V},\mathbf{\Lambda})} = \mathrm{Tr}\left(\mathbf{\Delta}_1^T \mathbf{\Delta}_2\right) + k \left\langle \mathbf{D}_1, \mathbf{D}_2 \right\rangle_{\mathbf{\Lambda}} \tag{8}$$

for some $k > 0$, where $\left(\mathbf{\Delta}_1, \mathbf{D}_1\right), \left(\mathbf{\Delta}_2, \mathbf{D}_2\right) \in \mathcal{T}_{\mathbf{W}} \mathcal{S}^+(r, N)$ and $\langle \cdot, \cdot \rangle_{\mathbf{\Lambda}}$ denotes the metric defined in (1) for SPD matrices.

Let $\mathbf{W}_1 = \mathbf{V}_1 \mathbf{\Lambda}_1 \mathbf{V}_1^T$ and $\mathbf{W}_2 = \mathbf{V}_2 \mathbf{\Lambda}_2 \mathbf{V}_2^T$ denote the decompositions of two SPSD matrices $\mathbf{W}_1, \mathbf{W}_2 \in \mathcal{S}^+(r, N)$, where $\mathbf{V}_1, \mathbf{V}_2 \in \mathcal{V}_{N,r}$ and $\mathbf{\Lambda}_1, \mathbf{\Lambda}_2 \in \mathcal{P}_r$. The closed-form expression of the geodesic path connecting any two such matrices in $\mathcal{S}^+(r, N)$ using the metric in (8) is unknown. However, the following approximation of it was proposed in [12]. Denote the singular value decomposition (SVD) of $\mathbf{V}_2^T \mathbf{V}_1$ by $\mathbf{O}_2 \mathbf{\Sigma} \mathbf{O}_1^T$, where $\mathbf{O}_1, \mathbf{O}_2 \in \mathbb{R}^{r\times r}$ and $\mathrm{diag}(\mathbf{\Sigma})$ are the cosines of the principal angles between $\mathrm{range}(\mathbf{W}_1)$ and $\mathrm{range}(\mathbf{W}_2)$, where $\mathrm{range}(\mathbf{W})$ denotes the column space of $\mathbf{W}$. Define $\mathbf{\Theta} = \arccos(\mathbf{\Sigma})$, which is a diagonal matrix of size $r \times r$ with the principal angles between the two subspaces on its diagonal. The approximation of the geodesic path connecting two points in $\mathcal{S}^+(r, N)$ is then given by:

$$\tilde{\gamma}_{\mathbf{W}_1 \to \mathbf{W}_2}(p) = \mathbf{U}_{\mathbf{W}_1 \to \mathbf{W}_2}(p) \mathbf{R}_{\mathbf{W}_1 \to \mathbf{W}_2}(p) \mathbf{U}^T_{\mathbf{W}_1 \to \mathbf{W}_2}(p), \quad p \in [0, 1] \tag{9}$$

where $\mathbf{R}_{\mathbf{W}_1 \to \mathbf{W}_2}(p)$ is the geodesic path connecting SPD matrices as defined in (2) calculated between the matrices $\mathbf{R}_1 = \mathbf{O}_1^T \mathbf{\Lambda}_1 \mathbf{O}_1$ and $\mathbf{R}_2 = \mathbf{O}_2^T \mathbf{\Lambda}_2 \mathbf{O}_2$, i.e. $\mathbf{R}_{\mathbf{W}_1 \to \mathbf{W}_2}(p) = \gamma_{\mathbf{R}_1 \to \mathbf{R}_2}(p)$, and $\mathbf{U}_{\mathbf{W}_1 \to \mathbf{W}_2}(p)$ is the geodesic connecting $\mathrm{range}(\mathbf{W}_1)$ and $\mathrm{range}(\mathbf{W}_2)$ on the Grassman manifold (the set of r dimensional subspaces of $\mathbb{R}^N$), defined by:

$$\mathbf{U}_{\mathbf{W}_1 \to \mathbf{W}_2}(p) = \mathbf{U}_1 \cos(\mathbf{\Theta} p) + \mathbf{X} \sin(\mathbf{\Theta} p) \tag{10}$$

where $\mathbf{U}_1 = \mathbf{V}_1 \mathbf{O}_1$, $\mathbf{U}_2 = \mathbf{V}_2 \mathbf{O}_2$ and $\mathbf{X} = \left(\mathrm{I} - \mathbf{U}_1 \mathbf{U}_1^T\right) \mathbf{U}_2 \left(\sin(\mathbf{\Theta})\right)^\dagger$, with $(\cdot)^\dagger$ denoting the Moore-Penrose pseudo-inverse.

3.3. Diffusion operator

As described in Section 2, most manifold learning methods, and particularly diffusion maps [18], are based on positive kernel matrices. Here, we briefly present the construction of such a kernel, which we term the diffusion operator, as proposed in [18]. In the sequel, we employ this diffusion operator in our framework to recover the geometry underlying each modality.

Given a set of N points, $\{\mathbf{x}[i]\}_{i=1}^N$, which are sampled from some hidden manifold $\mathcal{M}$ embedded in $\mathbb{R}^n$, consider the following affinity kernel matrix $\mathbf{K} \in \mathbb{R}^{N\times N}$, whose (i, j)th entry is given by:

$$\mathrm{K}[i, j] = \exp\left(-\frac{\|\mathbf{x}[i] - \mathbf{x}[j]\|_2^2}{\sigma^2}\right), \tag{11}$$

where $\|\cdot\|_2$ denotes the ℓ_2 norm and σ denotes the kernel scale, typically set to the median of the Euclidean distances between the sample points multiplied by some scalar. By Bochner's theorem, $\mathbf{K}$ is an SPD matrix. The kernel is normalized twice according to:

$$\begin{aligned} \widehat{\mathbf{W}} &= \widehat{\mathbf{D}}^{-1}\ \mathbf{K}\ \widehat{\mathbf{D}}^{-1} \\ \mathbf{W} &= \mathbf{D}^{-1/2}\ \widehat{\mathbf{W}}\ \mathbf{D}^{-1/2}, \end{aligned} \tag{12}$$

where $\widehat{\mathbf{D}}$ and $\mathbf{D}$ are diagonal matrices with $\widehat{\mathrm{D}}[i, i] = \sum_{j=1}^N \mathrm{K}[i, j]$ and $\mathrm{D}[i, i] = \sum_{j=1}^N \widehat{\mathrm{W}}[i, j]$, respectively.

The matrix $\mathbf{W}$ defined in (12) is similar to the diffusion operator considered in [18], which we call the diffusion maps operator for simplicity, with a normalization that removes the point density influence, given by $\mathbf{W}_{\mathrm{DM}} = \mathbf{D}^{-1} \widehat{\mathbf{W}}$. Due to this similarity, the matrix $\mathbf{W}$ and the diffusion maps operator $\mathbf{W}_{\mathrm{DM}}$ share the same eigenvalues and their eigenvectors are related by $\psi_{\mathrm{DM}} = \mathbf{D}^{-1/2} \psi$ and $\tilde{\psi}_{\mathrm{DM}} = \mathbf{D}^{1/2} \psi$, where ψ denotes an eigenvector of $\mathbf{W}$ and ψ_{DM} and $\tilde{\psi}_{\mathrm{DM}}$ denote the right and left eigenvectors of $\mathbf{W}_{\mathrm{DM}}$, respectively.

4. Riemannian multi-resolution analysis

We are ready to present our multi-resolution framework for multivariate time-series analysis from a manifold learning perspective. In this section, we focus on the algorithmic aspect, and in Section 6 we present the theoretical justification. We start by introducing the Riemannian composition of two operators that capture the relationship between two datasets sampled from two underlying diffeomorphic manifolds. Then, we generalize the setting to a sequence of datasets in time by presenting a wavelet-like analysis using the composite operators. Finally, we conclude this section with important implementation remarks.

4.1. Riemannian composition of two operators

Consider two datasets of N points denoted by $\{\mathbf{x}_1[i]\}_{i=1}^N, \{\mathbf{x}_2[i]\}_{i=1}^N$. Suppose there is some correspondence between the datasets and that they are ordered according to this correspondence, i.e., the two points $\mathbf{x}_1[i]$ and $\mathbf{x}_2[i]$ correspond. Such a correspondence could be the result of simultaneous recording from two, possibly different, sensors. We aim to recover the common structures in these datasets and characterize their expression in each dataset. Specifically, we consider two types of common components: common components that are expressed similarly in the two datasets and common components that are expressed differently.

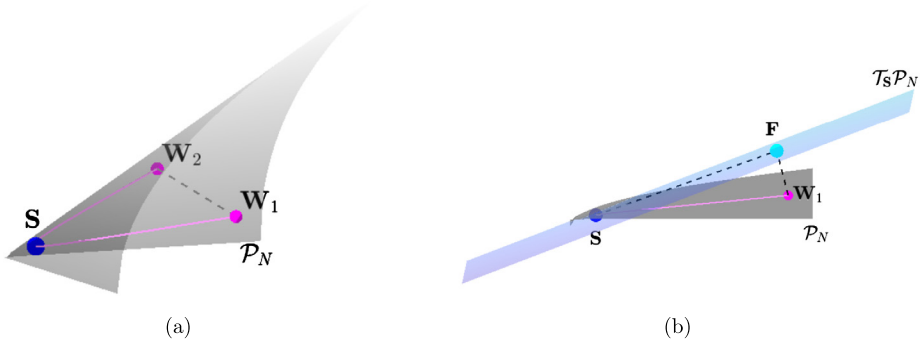


Fig. 1. Illustration of the definitions of the operators **S** and **F**. (a) Illustration of the operator **S** (blue point) on the geodesic path (solid line in magenta) connecting $\mathbf{W}_1$ and $\mathbf{W}_2$ (magenta points) on the manifold of SPD matrices (gray surface represents one level-set of the Riemannian manifold of SPD matrices). The dashed gray line denotes the shortest Euclidean path connecting the two matrices. (b) Illustration of the operator **F** (cyan point) on the tangent space at **S** (colored plane). Both plots present the same region in different orientations. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

To this end, we propose a two-step method. First, we assume that each dataset lies on some manifold and we characterize its underlying geometry using a diffusion operator constructed according to (12). This results in two SPD matrices denoted by $\mathbf{W}_1$ and $\mathbf{W}_2$ (see Section 3.1). Then, we propose to "fuse" the two datasets by considering compositions of $\mathbf{W}_1$ and $\mathbf{W}_2$ based on Riemannian geometry. In contrast to most previous studies that consider linear combinations involving addition, subtraction, and multiplication, e.g., [47,51,68], which often results in non-symmetric or non-positive matrices violating the fundamental geometric structure of diffusion operators, our Riemannian compositions yield symmetric and SPD matrices. Specifically, we define two operators by:

$$\mathbf{S}_p = \mathbf{W}_1 \#_p \mathbf{W}_2 = \mathbf{W}_1^{1/2} \left(\mathbf{W}_1^{-1/2} \mathbf{W}_2 \mathbf{W}_1^{-1/2} \right)^p \mathbf{W}_1^{1/2}, \tag{13}$$

$$\mathbf{F}_p = \mathrm{Log}_{\mathbf{S}_p} \left(\mathbf{W}_1 \right) = \mathbf{S}_p^{1/2} \log \left(\mathbf{S}_p^{-1/2} \mathbf{W}_1 \mathbf{S}_p^{-1/2} \right) \mathbf{S}_p^{1/2}, \tag{14}$$

where $0 \le p \le 1$ denotes the position along the geodesic path connecting $\mathbf{W}_1$ and $\mathbf{W}_2$ on the SPD manifold, $\mathbf{W}_1 \#_p \mathbf{W}_2$ with $p = 1/2$ denotes the midpoint on this geodesic, and $\mathrm{Log}_{\mathbf{S}_p}(\mathbf{W}_1)$ denotes the logarithmic map, projecting the matrix $\mathbf{W}_1$ onto the tangent space of the SPD manifold at point $\mathbf{S}_p = \mathbf{W}_1 \#_p \mathbf{W}_2$. The parameter p can be used to incorporate prior knowledge on the expected importance of each dataset, e.g., if we expect that the information in the first dataset is more meaningful, less noisy for example, we can set $0 < p < 1/2$. See [42] for a more extensive analysis of the effect of p. Note that a similar definition of the operator $\mathbf{S}_p$ was previously used with diffusion operators in [42] and analyzed in the context of multimodal data processing. In this work, we additionally define and analyze a new operator, $\mathbf{F}_p$, and then use both operators to construct the Riemannian multi-resolution analysis framework. Fig. 1 presents an illustration of the definitions of the operators on the Riemannian manifold of SPD matrices and its tangent space.

Intuitively, $\mathbf{S}_p$ describes the mean of the two matrices, so it enhances the components that are expressed similarly; that is, common eigenvectors with similar eigenvalues. Conversely, $\mathbf{F}_p$ can be seen as the difference of $\mathbf{S}_p$ and $\mathbf{W}_1$ along the geodesic connecting them, and therefore, it is related to the components expressed differently; that is, the common eigenvectors with different eigenvalues. In Section 6 we provide a theoretical justification for the above statements. Note that a recent work [20] defined an operator analogous to $\mathbf{S}_p$, as the exponent of a linear combination of multiple normalized Laplacians of different graphs with shared vertices. This also results in an SPD operator that can be interpreted as representing the average diffusion on the common set of vertices. However, in the context of our work, which aims to define a multi-resolution wavelet-like framework for time-varying manifolds, this setting lacks a natural formulation of differences between graphs or datasets.

We remark that $\mathbf{F}_p$ is symmetric but not positive-definite, since it is defined as a projection of an SPD matrix onto the tangent space, and that using $\mathbf{W}_2$ instead of $\mathbf{W}_1$ in the projection leads only to a change of sign. In addition, given $\mathbf{S}_p$ and $\mathbf{F}_p$, both SPD matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ can be reconstructed using the exponential map $\mathrm{Exp}_{\mathbf{S}_p}(\pm\mathbf{F}_p)$ (as defined in (5)). For simplicity of notations, we focus in the following on $\mathbf{F}_p$ and $\mathbf{S}_p$ with $p = 0.5$ and omit the notation of p. The extension to other values of p is straightforward.

In the second step, we propose new embeddings of the data points, representing the common components between the two datasets based on the operators **S** and **F**. Since both operators are symmetric, their eigenvalues and eigenvectors are real, and the eigenvectors are orthogonal. Denote the eigenvalues and eigenvectors of the operator **S** by $\lambda_n^{(\mathbf{S})}$ and $\psi_n^{(\mathbf{S})}$, respectively, and the eigenvalues and eigenvectors of the operator **F** by $\lambda_n^{(\mathbf{F})}$ and $\psi_n^{(\mathbf{F})}$, respectively, where $n = 1, \ldots, N$. The new embeddings are constructed based on the eigenvectors of **S** and **F** by taking the $M \le N$ leading eigenvectors, i.e. eigenvectors that correspond to the M largest eigenvalues (in absolute value for **F**), which are organized in decreasing order $\lambda_1^{(\mathbf{S})} \ge \lambda_2^{(\mathbf{S})} \ge \cdots \ge \lambda_M^{(\mathbf{S})}$ and $\lambda_1^{(\mathbf{F})} \ge \lambda_2^{(\mathbf{F})} \ge \cdots \ge \lambda_M^{(\mathbf{F})}$. The new embeddings are defined by:

$$(\mathbf{x}_1[i], \mathbf{x}_2[i]) \mapsto \Psi^{(\mathbf{S})}[i] = \left\{ \psi_1^{(\mathbf{S})}[i], \ldots, \psi_M^{(\mathbf{S})}[i] \right\} \tag{15}$$

$$(\mathbf{x}_1[i], \mathbf{x}_2[i]) \mapsto \Psi^{(\mathbf{F})}[i] = \left\{ \psi_1^{(\mathbf{F})}[i], \ldots, \psi_M^{(\mathbf{F})}[i] \right\}. \tag{16}$$

Algorithm 1 summarizes the above two-step operator and embedding construction. In Section 5.1, we demonstrate the properties of the operators **S** and **F** and the proposed embeddings on an illustrative toy example. In Section 6, we present some analysis.

As a final remark, we note that other SPD kernels and matrices may be considered instead of the proposed diffusion operators, e.g. covariance or correlation matrices, which can simply substitute $\mathbf{W}_1$ and $\mathbf{W}_2$ in the above definitions.

Algorithm 1 Operator composition and spectral embedding based on Riemannian geometry.

Input: Two datasets $\{\mathbf{x}_1[i]\}_{i=1}^N, \{\mathbf{x}_2[i]\}_{i=1}^N$; the embedding dimension M
Output: Operators $\mathbf{S}$ and $\mathbf{F}$ and new embeddings $\boldsymbol{\Psi}^{(\mathbf{S})}$ and $\boldsymbol{\Psi}^{(\mathbf{F})}$

Construct a diffusion operator for each dataset, $\mathbf{W}_\ell \in \mathbb{R}^{N\times N}$, $\ell = 1, 2$, according to (11) and (12)

Build operators $\mathbf{S}$ and $\mathbf{F}$:

$$\mathbf{S} = \mathbf{W}_1^{1/2}\left(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}\right)^{1/2}\mathbf{W}_1^{1/2}$$

$$\mathbf{F} = \mathbf{S}^{1/2}\log\left(\mathbf{S}^{-1/2}\mathbf{W}_1\mathbf{S}^{-1/2}\right)\mathbf{S}^{1/2}$$

Compute the eigenvalue decomposition of the operators $\mathbf{S}$ and $\mathbf{F}$

Take the M largest eigenvalues (in absolute value) and order them such that $\lambda_1^{(\mathbf{S})} \geq \lambda_2^{(\mathbf{S})} \geq \cdots \geq \lambda_M^{(\mathbf{S})}$ and $\lambda_1^{(\mathbf{F})} \geq \lambda_2^{(\mathbf{F})} \geq \cdots \geq \lambda_M^{(\mathbf{F})}$

Take the corresponding M eigenvectors of $\mathbf{S}$ and define:

$\boldsymbol{\Psi}^{(\mathbf{S})} = \left\{\psi_1^{(\mathbf{S})}, \ldots, \psi_M^{(\mathbf{S})}\right\} \in \mathbb{R}^{N\times M}$ ▷ Capture similarly expressed common components

Take the corresponding M eigenvectors of $\mathbf{F}$ and define:

$\boldsymbol{\Psi}^{(\mathbf{F})} = \left\{\psi_1^{(\mathbf{F})}, \ldots, \psi_M^{(\mathbf{F})}\right\} \in \mathbb{R}^{N\times M}$ ▷ Capture differently expressed common components

4.2. Operator-based analysis of a sequence of datasets

Let $\{\mathrm{x}_t[i]\}_{i=1}^N$ denote a temporal sequence of datasets, where $t = 1, \ldots, T = 2^m$, $m \in \mathbb{N}$, denotes time, and $\mathrm{x}_t[i] \in \mathbb{R}^d$ is the i-th point sampled at time t. Considering first just two consecutive datasets $\{\mathrm{x}_t[i]\}_{i=1}^N$ and $\{\mathrm{x}_{t+1}[i]\}_{i=1}^N$ is analogous to the setting presented in Section 4.1. Applying the same analysis gives rise to the operators $\mathbf{S}$ and $\mathbf{F}$ corresponding to $\{\mathrm{x}_t[i]\}_{i=1}^N$ and $\{\mathrm{x}_{t+1}[i]\}_{i=1}^N$, which facilitate the extraction of the two types of underlying common components. Unlike the general setting in Section 4.1, the temporal order of the two datasets considered here allows us to view the common components that are expressed similarly and extracted by $\mathbf{S}$ as the slowly changing components. Similarly, the common components that are expressed differently and extracted by $\mathbf{F}$ are considered as rapidly changing components.

The above description constitutes the basic building block of our analysis. With that in mind, we proceed to the construction of the proposed multi-resolution analysis of the entire sequence. At the first step, we build a diffusion operator according to (12) for the dataset $\{\mathrm{x}_t[i]\}_{i=1}^N$ at each time point t, resulting in T kernels $\mathbf{W}_t \in \mathbb{R}^{N\times N}$, $t = 1, \ldots, T$. Then, for every pair of consecutive time-points, $(2t-1, 2t)$, $t = 1, \ldots, T/2$, we construct the two operators[2] $\mathbf{S}_t^{(1)}$ and $\mathbf{F}_t^{(1)}$ according to (13) and (14) with $p = 0.5$. These $2\times T/2$ operators represent the fine level, denoted $\ell = 1$, of the multi-resolution framework and recover components which are common to consecutive time-frames. At coarser (higher) levels, i.e. for $\ell > 1$, the operators are constructed according to (13) and (14) (with $p = 0.5$) using the operators from the previous level as input. Specifically, at level $\ell > 1$, the operators are given by

$$\mathbf{S}_t^{(\ell)} = \mathbf{S}_{2t-1}^{(\ell-1)} \# \mathbf{S}_{2t}^{(\ell-1)} \tag{17}$$

$$\mathbf{F}_t^{(\ell)} = \mathrm{Log}_{\mathbf{S}_t^{(\ell)}}\left(\mathbf{S}_{2t-1}^{(\ell-1)}\right), \tag{18}$$

where $t = 1, \ldots, T/2^\ell$ and $\ell = 2, \ldots, \log_2 T$. At each level, only the operator $\mathbf{S}$ is used to construct the operators of the next level. The reason for this construction choice is that $\mathbf{S}$ enhances similarly expressed common components. In the present setting, the similarly expressed common components of consecutive time frames are in fact components that change slowly in time. Therefore, using the operators $\mathbf{S}_t^{(\ell-1)}$, $t = 1, \ldots, T/2^{\ell-1}$, to construct the operators of the coarser level, ℓ, has a smoothing effect. This is analogous to the construction of the ordinary wavelet decomposition, where the outputs of the low-pass filters at each level are used as inputs to the coarser level. In this analogy, the operator $\mathbf{F}_t^{(\ell)}$ can be viewed as a high-pass filter, since it enhances components that are expressed significantly different in consecutive time frames, i.e. rapidly changing components.

Similarly to the embedding defined based on $\mathbf{S}$ and $\mathbf{F}$ in Subsection 4.1, we define an embedding based on the eigenvectors of the operators $\mathbf{S}_t^{(\ell)}$ and $\mathbf{F}_t^{(\ell)}$ at different levels and time-frames. Denote the eigenvalues and eigenvectors of the operator $\mathbf{S}_t^{(\ell)}$ by $\lambda_n^{(\mathbf{S}_t^{(\ell)})}$ and $\psi_n^{(\mathbf{S}_t^{(\ell)})}$, respectively, and the eigenvalues and eigenvectors of the operator $\mathbf{F}_t^{(\ell)}$ by $\lambda_n^{(\mathbf{F}_t^{(\ell)})}$ and $\psi_n^{(\mathbf{F}_t^{(\ell)})}$, respectively, where $n = 1, \ldots, N$ and the eigenvalues are ordered in decreasing magnitude. The embedding at each level and each time frame is then defined by taking the $M \leq N$ leading eigenvectors of the operators as follows:

$$\mathbf{x}_t[i] \to \Psi^{(\mathbf{S}_t^{(\ell)})} = \left\{\psi_1^{(\mathbf{S}_t^{(\ell)})}[i], \ldots, \psi_M^{(\mathbf{S}_t^{(\ell)})}[i]\right\} \tag{19}$$

[2] Note that the operator underscore notation now denotes the time index rather than the geodesic curve parameter p as in Section 4.1.

$$\mathbf{x}_t[i] \to \Psi^{(\mathbf{F}_t^{(\ell)})} = \left\{ \psi_1^{(\mathbf{F}_t^{(\ell)})}[i], \ldots, \psi_M^{(\mathbf{F}_t^{(\ell)})}[i] \right\}, \tag{20}$$

where $t = 1, \ldots, T/2^\ell$ and $\ell = 2, \ldots, \log_2 T$. These embedding coordinates capture the slowly varying components ($\Psi^{(\mathbf{S}_t^{(\ell)})}$) and fast varying components ($\Psi^{(\mathbf{F}_t^{(\ell)})}$) at each time frame t and each level ℓ.

The proposed algorithm is summarized in Algorithm 2.

Algorithm 2 Riemannian multi-resolution analysis algorithm.

Input: A time-varying dataset $\{\mathrm{x}_t[i]\}_{i=1}^N$, $\mathrm{x}_t[i] \in \mathbb{R}^d$, $t = 1, \ldots, T$.

Output: Operators $\{\mathbf{S}_t^{(\ell)}, \mathbf{F}_t^{(\ell)}\}_{t=1}^{T/2^\ell}$ and new representations for each level $\{\Psi^{(\mathbf{S}_t^{(\ell)})}, \Psi^{(\mathbf{F}_t^{(\ell)})}\}_{t=1}^{T/2^\ell}$, where $\ell = 1, \ldots, \log_2 T$

Construct an SPD kernel representing each time point t, denoted by $\{\mathbf{W}_t\}_{t=1}^T$, according to (12).

for $t = 1 : T/2$ **do** ▷ Construct the operators for level 1
  $\mathbf{S}_t^{(1)} = \mathbf{W}_{2t-1} \#_{0.5} \mathbf{W}_{2t}$
  $\mathbf{F}_t^{(1)} = \mathrm{Log}_{\mathbf{S}_t^{(1)}} \left(\mathbf{W}_{2t-1} \right)$
end for
for $\ell = 2 : \log_2 T$ **do** ▷ Construct the operators for level ℓ
  for $t = 1 : T/2^\ell$ **do**
    $\mathbf{S}_t^{(\ell)} = \mathbf{S}_{2t-1}^{(\ell-1)} \#_{0.5} \mathbf{S}_{2t}^{(\ell-1)}$
    $\mathbf{F}_t^{(\ell)} = \mathrm{Log}_{\mathbf{S}_t^{(\ell)}} \left(\mathbf{S}_{2t-1}^{(\ell-1)} \right)$
  end for
end for

for $\ell = 1 : \log_2 T$ **do** ▷ Construct the new representations
  for $t = 1 : T/2^\ell$ **do**
    $\Psi^{(\mathbf{S}_t^{(\ell)})} = \left[\psi_1^{(\mathbf{S}_t^{(\ell)})}, \ldots, \psi_M^{(\mathbf{S}_t^{(\ell)})} \right]$
    $\Psi^{(\mathbf{F}_t^{(\ell)})} = \left[\psi_1^{(\mathbf{F}_t^{(\ell)})}, \ldots, \psi_M^{(\mathbf{F}_t^{(\ell)})} \right]$
    where $\psi^{(\mathbf{S}_t^{(\ell)})}$ and $\psi^{(\mathbf{F}_t^{(\ell)})}$ are the eigenvectors of $\mathbf{S}_t^{(\ell)}$ and $\mathbf{F}_t^{(\ell)}$.
  end for
end for

4.3. Implementation remarks

The numerical implementation of the proposed algorithm, particularly the Riemannian composition of operators, needs some elaboration. While the diffusion operators we consider in (12) are SPD matrices by definition, in practice, some of their eigenvalues could be close to zero numerically, forming in effect SPSD matrices instead. In order to address this issue, we propose an equivalent definition of the operators $\mathbf{S}$ and $\mathbf{F}$ for SPSD matrices of fixed rank, based on the Riemannian metric and mean that were introduced in [12].

Based on the approximated geodesic path in (9), we define the operators $\mathbf{S}$ and $\mathbf{F}$ for SPSD matrices as follows. First, define

$$\mathbf{S} = \tilde{\gamma}_{\mathbf{W}_1 \to \mathbf{W}_2}(0.5) = \mathbf{U}_{\mathbf{W}_1 \to \mathbf{W}_2}(0.5) \left(\mathbf{R}_1 \#_{0.5} \mathbf{R}_2 \right) \mathbf{U}_{\mathbf{W}_1 \to \mathbf{W}_2}^T(0.5). \tag{21}$$

Next, evaluate $\mathbf{V}_1^T \mathbf{V}_\mathbf{S} = \mathbf{O}_\mathbf{S} \tilde{\mathbf{\Sigma}} \tilde{\mathbf{O}}_1^T$ by SVD, where $\mathbf{\Lambda}_\mathbf{S}$ and $\mathbf{V}_\mathbf{S}$ denote the eigenvalues and eigenvectors of $\mathbf{S}$, respectively. Also, define $\mathbf{R}_\mathbf{S} := \mathbf{O}_\mathbf{S}^T \mathbf{\Lambda}_\mathbf{S} \mathbf{O}_\mathbf{S}$. Then, define

$$\mathbf{F} = \mathbf{U}_{\mathbf{S} \to \mathbf{W}_1}(1) \mathrm{Log}_{\mathbf{R}_\mathbf{S}} \left(\tilde{\mathbf{O}}_1^T \mathbf{\Lambda}_1 \tilde{\mathbf{O}}_1 \right) \mathbf{U}_{\mathbf{S} \to \mathbf{W}_1}^T(1) \tag{22}$$

where $\mathbf{U}_{\mathbf{S} \to \mathbf{W}_1}(p)$ is defined for matrices $\mathbf{S}$ and $\mathbf{W}_1$ correspondingly to (10) and the derivations leading to it. Intuitively, this can be viewed as applying the operators $\mathbf{S}$ and $\mathbf{F}$ to matrices expressed in the bases associated with the non-trivial SPD matrices of rank r, and then projecting the resulting operators back to the original space of SPSD matrices by applying $\mathbf{U}_{\mathbf{W}_1 \to \mathbf{W}_2}(p)$.

A summary of this derivation is presented in Algorithm 3. A demonstration of the properties of these new operators for SPSD matrices using a simple simulation of 4×4 matrices, similar to the one presented in Subsection 5.1 but with matrices of rank 3, could be found in Section 5.2.

5. Experimental results

In this section, we demonstrate the operators $\mathbf{S}$ and $\mathbf{F}$ on 3 examples. In Subsection 5.1 and Subsection 5.2, we present a toy example demonstrating the spectral properties of $\mathbf{S}$ and $\mathbf{F}$. These examples simulate Theorem 1 and Theorem 2. In Subsection 5.3, we apply our multi-resolution framework defined in Subsection 4.2, which is based on $\mathbf{S}$ and $\mathbf{F}$, to a 2D dynamical system. In Subsection 5.4, the operators $\mathbf{S}$ and $\mathbf{F}$ are applied to hyperspectral and LiDAR images to demonstrate their multimodal capabilities. Additional

Algorithm 3 Implementation of the operators for SPSD matrices.

Input: Two datasets with point correspondence $\{\mathbf{x}_1[i], \mathbf{x}_2[i]\}_{i=1}^N$
Output: Operators $\mathbf{S}$ and $\mathbf{F}$ and their eigenvectors: $\boldsymbol{\Psi}^{(\mathbf{S})}, \boldsymbol{\Psi}^{(\mathbf{F})}$

function SPSD-GEODESICS($\mathbf{G}_1, \mathbf{G}_2, p$) ▷ As defined in [12]
  Set $r = \min\{\text{rank}(\mathbf{G}_1), \text{rank}(\mathbf{G}_2)\}$.
  for $i \in [1,2]$ **do**
    Set $\mathbf{G}_i = \mathbf{V}_i \boldsymbol{\Lambda}_i \mathbf{V}_i^T$ ▷ Eigenvalue decomposition
    Set $\tilde{\mathbf{V}}_i = \mathbf{V}_i(:, 1:r)$
    Set $\tilde{\boldsymbol{\Lambda}}_i = \boldsymbol{\Lambda}_i(1:r, 1:r)$
  end for
  Set $[\mathbf{O}_1, \boldsymbol{\Sigma}, \mathbf{O}_2] = \text{SVD}(\tilde{\mathbf{V}}_2^T \tilde{\mathbf{V}}_1)$
  Set $\boldsymbol{\Theta} = \arccos(\boldsymbol{\Sigma})$
  for $i \in [1,2]$ **do**
    Set $\mathbf{U}_i = \tilde{\mathbf{V}}_i \mathbf{O}_i$
    Set $\mathbf{R}_i = \mathbf{O}_i^T \tilde{\boldsymbol{\Lambda}}_i \mathbf{O}_i$
  end for
  Compute $\mathbf{U}_{\mathbf{G}_1 \to \mathbf{G}_2}(p)$ ▷ According to (10)
  Compute $\mathbf{R}_{\mathbf{G}_1 \to \mathbf{G}_2}(p) = \mathbf{R}_1^{1/2} \left(\mathbf{R}_1^{-1/2} \mathbf{R}_2 \mathbf{R}_1^{-1/2}\right)^p \mathbf{R}_1^{1/2}$
  return $\mathbf{U}_{\mathbf{G}_1 \to \mathbf{G}_2}(p)$, $\mathbf{R}_{\mathbf{G}_1 \to \mathbf{G}_2}(p)$, $\mathbf{R}_{\mathbf{G}_1}$, $\mathbf{R}_{\mathbf{G}_2}$
end function

function MAIN
  Construct SPSD matrices for the two datasets $\mathbf{W}_1$ and $\mathbf{W}_2$ ▷ According to (12)
  SPSD-GEODESICS($\mathbf{W}_1, \mathbf{W}_2, 0.5$)
  $\mathbf{S} = \mathbf{U}_{\mathbf{W}_1 \to \mathbf{W}_2}(0.5) \mathbf{R}_{\mathbf{W}_1 \to \mathbf{W}_2}(0.5) \mathbf{U}_{\mathbf{W}_1 \to \mathbf{W}_2}^T(0.5)$
  SPSD-GEODESICS($\mathbf{S}, \mathbf{W}_1, 1$)
  $\mathbf{F} = \mathbf{U}_{\mathbf{S} \to \mathbf{W}_1}(1) \text{Log}_{\mathbf{R}_\mathbf{S}}(\mathbf{R}_{\mathbf{W}_1}) \mathbf{U}_{\mathbf{S} \to \mathbf{W}_1}^T(1)$ ▷ $\text{Log}_{\cdot}(\cdot)$ is defined as in (6)
end function

results demonstrating the properties of the operators $\mathbf{S}$ and $\mathbf{F}$ appear in Appendix B. In all experiments in the paper, we set $p = 1/2$ in order to maintain an equal influence of the different datasets and modalities. The code for the experiments is available on GitHub.

5.1. Illustrative toy example: SPD case

We demonstrate the properties of the composite operators $\mathbf{S}$ and $\mathbf{F}$, constructed in Algorithm 1, using a simple simulation of 4×4 matrices. Define two matrices, $\mathbf{M}_1 = \boldsymbol{\Psi} \boldsymbol{\Lambda}^{(1)} \boldsymbol{\Psi}^T$ and $\mathbf{M}_2 = \boldsymbol{\Psi} \boldsymbol{\Lambda}^{(2)} \boldsymbol{\Psi}^T$, with the following common eigenvectors:

$$\boldsymbol{\Psi} = \left[\psi_1, \psi_2, \psi_3, \psi_4\right] = \frac{1}{2} \begin{bmatrix} 1, & 1, & 1, & 1 \\ 1, & 1, -1, -1 \\ 1, -1, -1, & 1 \\ 1, -1, & 1, -1 \end{bmatrix} \tag{23}$$

and the following eigenvalues:

$$\boldsymbol{\Lambda}^{(1)} = \text{diag}\left(\left[\lambda_1^{(1)}, \lambda_2^{(1)}, \lambda_3^{(1)}, \lambda_4^{(1)}\right]\right) = \text{diag}([0.5, \ \ 1, \ 0.01, \ 0.2]) \tag{24}$$

$$\boldsymbol{\Lambda}^{(2)} = \text{diag}\left(\left[\lambda_1^{(2)}, \lambda_2^{(2)}, \lambda_3^{(2)}, \lambda_4^{(2)}\right]\right) = \text{diag}([0.01, \ 1, \ 0.5, \ \ 0.2]) \tag{25}$$

In this example, ψ_1 is a common eigenvector that is dominant in $\mathbf{M}_1$ and weak in $\mathbf{M}_2$, ψ_3 is a common eigenvector that is dominant in $\mathbf{M}_2$ and weak in $\mathbf{M}_1$ and ψ_2 and ψ_4 are common eigenvectors that are similarly expressed in both $\mathbf{M}_1$ and $\mathbf{M}_2$.

We construct the operators $\mathbf{S} = \mathbf{M}_1 \# \mathbf{M}_2$ and $\mathbf{F} = \text{Log}_\mathbf{S}(\mathbf{M}_1)$ and compute their eigenvalues and eigenvectors. Fig. 2 presents the 4 eigenvalues of $\mathbf{M}_1$, $\mathbf{M}_2$, $\mathbf{S}$ and $\mathbf{F}$, denoted by $\{\lambda_n^{(1)}\}_{n=1}^4$, $\{\lambda_n^{(2)}\}_{n=1}^4$, $\{\lambda_n^{(\mathbf{S})}\}_{n=1}^4$ and $\{\lambda_n^{(\mathbf{F})}\}_{n=1}^4$, respectively, in the left plots, and the corresponding eigenvectors in the right plots. This figure depicts that the two matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ share the same 4 eigenvectors (as defined) and that the resulting eigenvectors of $\mathbf{S}$ and $\mathbf{F}$ are similar to these 4 eigenvectors. Note that eigenvectors 2 and 4 of operator $\mathbf{F}$ are not identical to the eigenvectors of $\mathbf{M}_1$ and $\mathbf{M}_2$ due to numerical issues, which arise since these eigenvectors in $\mathbf{F}$ correspond to negligible eigenvalues. The left plots show that the eigenvalues of $\mathbf{S}$ and $\mathbf{F}$ capture the similarities and differences in the expression of the spectral components of $\mathbf{M}_1$ and $\mathbf{M}_2$. Specifically, since $\lambda_2^{(1)} = \lambda_2^{(2)}$ and $\lambda_4^{(1)} = \lambda_4^{(2)}$, the corresponding eigenvalues of $\mathbf{S}$ assume the same magnitude. In contrast, due to this equality, these eigenvalues correspond to negligible eigenvalues of $\mathbf{F}$. The two other eigenvectors, ψ_1 and ψ_3, correspond to eigenvalues that differ by an order of magnitude in the two matrices and are therefore the most dominant components in $\mathbf{F}$. In addition, note the opposite sign of eigenvalues $\lambda_1^{(\mathbf{F})}$ and $\lambda_3^{(\mathbf{F})}$, which indicates the source of the more dominant component, i.e. whether $\lambda_n^{(1)} > \lambda_n^{(2)}$ or $\lambda_n^{(1)} < \lambda_n^{(2)}$. These properties are proved and explained in more detail in Section 6.

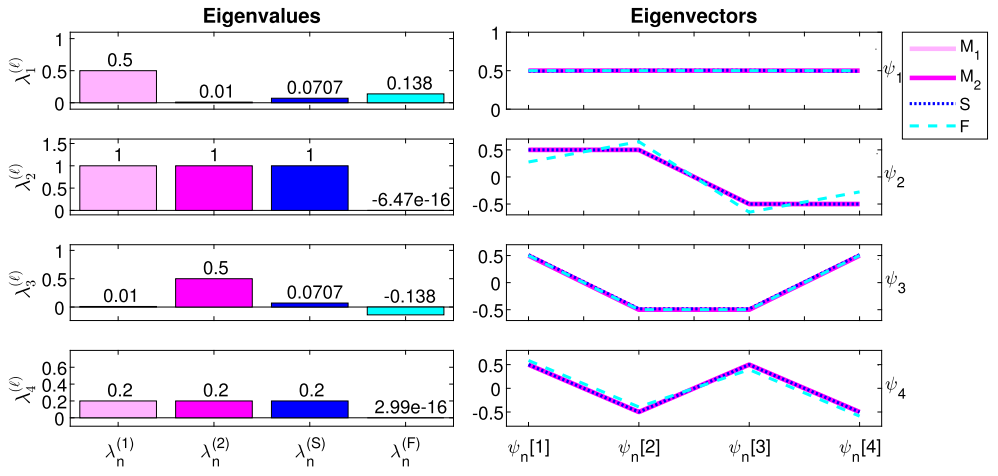


Fig. 2. Application of **S** and **F** to two 4×4 matrices with identical eigenvectors.

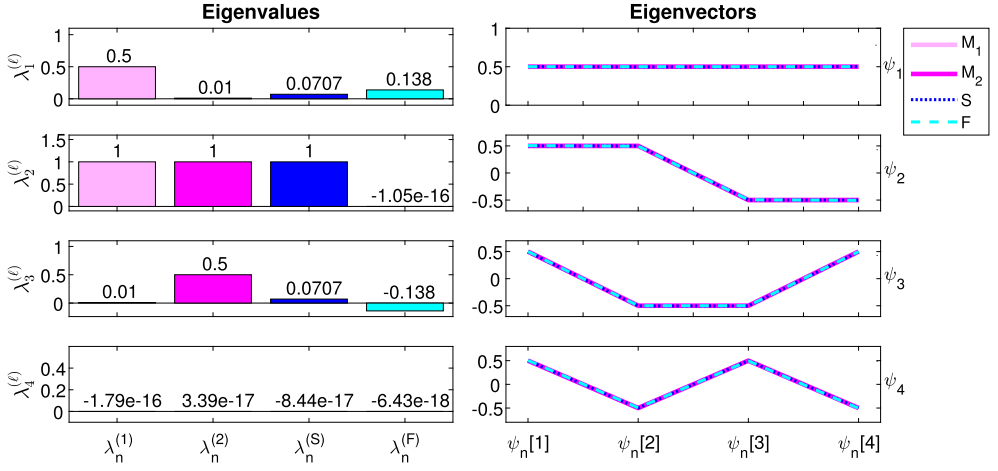


Fig. 3. Application of the operators **S** and **F** for SPSD matrices to two 4×4 matrices of rank 3 with identical eigenvectors.

5.2. *Illustrative toy example: SPSD case*

Consider two matrices, $\mathbf{M}_1 = \mathbf{\Psi}\mathbf{\Lambda}^{(1)}\mathbf{\Psi}^T$ and $\mathbf{M}_2 = \mathbf{\Psi}\mathbf{\Lambda}^{(2)}\mathbf{\Psi}^T$, with $\mathbf{\Psi}$ defined in (23) and the following eigenvalues:

$$\mathbf{\Lambda}^{(1)} = \operatorname{diag}\left(\left[\lambda_1^{(1)}, \lambda_2^{(1)}, \lambda_3^{(1)}, \lambda_4^{(1)}\right]\right) = \operatorname{diag}([0.5,\ \ 1,\ 0.01,\ 0]) \tag{26}$$

$$\mathbf{\Lambda}^{(2)} = \operatorname{diag}\left(\left[\lambda_1^{(2)}, \lambda_2^{(2)}, \lambda_3^{(2)}, \lambda_4^{(2)}\right]\right) = \operatorname{diag}([0.01,\ 1,\ 0.5,\ \ 0]) \tag{27}$$

Note that the 4th eigenvalue is zero in both matrices resulting in SPSD matrices of rank 3.

We construct the operators **S** and **F** according to Algorithm 3 and compute their eigenvalues and eigenvectors. The results are presented in Fig. 3. Same as Fig. 2, Fig. 3 presents in the left plots the 4 eigenvalues of $\mathbf{M}_1$, $\mathbf{M}_2$, **S** and **F**, denoted by $\{\lambda_n^{(1)}\}_{n=1}^4$, $\{\lambda_n^{(2)}\}_{n=1}^4$, $\{\lambda_n^{(\mathbf{S})}\}_{n=1}^4$ and $\{\lambda_n^{(\mathbf{F})}\}_{n=1}^4$, respectively, and the corresponding eigenvectors in the right plots. Both matrices, $\mathbf{M}_1$ and $\mathbf{M}_2$, share the same 4 eigenvectors as depicted in the right plots, and the resulting eigenvectors of **S** and **F** are similar to these 4 eigenvectors. In this example, ψ_2 is a dominant component in both $\mathbf{M}_1$ and $\mathbf{M}_2$ with the same large eigenvalue. Therefore, similarly to the SPD case, the eigenvalue of **S** associated with this eigenvector remains large, whereas the eigenvalue of **F** associated with this eigenvector is negligible, as expected. In contrast, ψ_1 and ψ_3 are eigenvectors that are differently expressed in the two matrices (corresponding to eigenvalues 0.5 and 0.01), and therefore, they correspond to dominant eigenvalues in **F**. The left plot demonstrates that this behavior is indeed captured by the operators **S** and **F** for SPSD matrices. Moreover, the eigenvalues of the operators **S** and **F** for SPSD matrices are equal to the eigenvalues that were obtained by the operators for SPD matrices in a corresponding toy example, presented in Fig. 2. Note that all eigenvalues that correspond to ψ_4 are very close to zero, as expected due to the definition of the matrices $\mathbf{M}_1$ and $\mathbf{M}_2$.

5.3. Transitory double gyre flow

To demonstrate the proposed Riemannian multi-resolution analysis described in Section 4, we consider a variation of the transitory double gyre flow presented in [34,57].

We simulate a 2D dynamical system with coordinates (x_t, y_t) using the following equations:

$$\dot{x}_t = -\frac{\partial}{\partial y_t} H\left(x_t, y_t, t\right) \tag{28}$$

$$\dot{y}_t = \frac{\partial}{\partial x_t} H\left(x_t, y_t, t\right) \tag{29}$$

with the function:

$$H\left(x_t, y_t, t\right) = (1 - g(t)) H_1\left(x_t, y_t\right) + g(t) H_2\left(x_t, y_t\right) \tag{30}$$

$$H_1\left(x_t, y_t\right) = c_1 \sin(2\pi x_t) \sin(\pi y_t) \tag{31}$$

$$H_2\left(x_t, y_t\right) = c_2 \sin(\pi x_t) \sin(2\pi y_t) \tag{32}$$

$$g(t) = t^2(3 - 2t), \tag{33}$$

where $c_1 = 2$, $c_2 = 10$, $i = 1, ..., N$ and $t \in [0, 1]$.

These equations describe a double gyre pattern, which is horizontal at $t = 0$ and transitions into a vertical double gyre pattern at $t = 1$. Note that in our simulations we add the parameters c_1 and c_2 to the dynamics, which lead to a change in rate between the (slower) horizontal and (faster) vertical double gyre patterns. These parameters are added in order to demonstrate the time-varying multi-resolution properties of our analysis.

We generate $N = 2500$ trajectories with initial values uniformly distributed in $\left(x_0, y_0\right) \in [0, 1] \times [0, 1]$, where each trajectory has $T = 256$ time points on a discrete uniform time-grid with a step size of $\Delta t = 1/256$. We denote each of these trajectories with an index $i = 1, \ldots, N$ by a matrix $\mathbf{x}[i] \in \mathbb{R}^{2\times T}$, whose columns are the pair of time samples $(x_t[i], y_t[i])^T$. A short GIF file demonstrating the resulting trajectories is available on GitHub, where each point is colored according to its initial location along the x-axis to illustrate the point movement in time. The point movement demonstrated in this GIF exhibits two main structures: (i) points that rotate in two circular structures (transitioning from a horizontal setting into a vertical setting), which can be described as almost-invariant (coherent) sets as defined and captured by [34], and (ii) points that are located on the boundary of these almost-invariant sets and their movement changes significantly over time. Our goal in this example is to analyze these two movement types and to recover their different trajectories over time.

For this purpose, we construct an SPD kernel for each time frame t, denoted by $\mathbf{W}_t \in \mathbb{R}^{N\times N}$, according to (12) based on the distances between the points in that time frame, i.e., $\left\|(x_t[i] - x_t[j], y_t[i] - y_t[j])\right\|_2^2$, $i, j = 1, \ldots, N$, with σ set to 0.5 times the median of these distances. We then apply Algorithm 2 and obtain the multi-resolution representation and $\ell = \log_2(T) = 8$ levels of operators. We denote the operators of different levels and different time frames by $\mathbf{S}_r^{(\ell)} \in \mathbb{R}^{N\times N}$ and $\mathbf{F}_r^{(\ell)} \in \mathbb{R}^{N\times N}$, where $\ell = 1, \ldots, 8$ and $r = 1, \ldots, T/2^\ell$. Note that r is associated with the time-frame indices, e.g., at level $\ell = 6$, $r = \lceil t/2^\ell \rceil = 3$ corresponds to time points $t = 129, \ldots, 192$.

In the following, we focus on the second eigenvector of $\mathbf{S}_r^{(\ell)}$ and show that it indeed captures the common components and common trajectory behavior at the different operator levels. Fig. 4 presents the data-points colored according to the second eigenvector of $\mathbf{S}_r^{(\ell)}$, denoted by $\psi_2^{(\mathbf{S}_r^{(\ell)})}$ at levels $\ell = 8$, $\ell = 4$ and $\ell = 3$ and with r values corresponding to different time frames. Each time frame contains 2^ℓ time points. For visualization of the time frames, in the plots in Fig. 4 we present 8 evenly spaced time points from the entire time frame. The locations of the $N = 2500$ points are presented at each of the 8 time points and colored by the eigenvector describing the entire time frame. For example, Fig. 4 (a) presents the time frame of level $\ell = 8$, which covers the entire trajectory, containing $2^8 = 256$ time points. In this plot, we present the locations of the $N = 2500$ point at time points $t = 0, 32, 64, 96, 128, 160, 192, 224$, and the points are colored according to $\psi_2^{(\mathbf{S}_1^{(8)})}$. Note that $\psi_2^{(\mathbf{S}_1^{(8)})}$ describes the entire time frame $[0, 255]$, and that the changes along time in Fig. 4 (a), for example, are only due to the displacement of the points. In the plots in Fig. 4, each data point is colored according to its value in: (a) $\psi_2^{(\mathbf{S}_1^{(8)})}$, (b) $\psi_2^{(\mathbf{S}_4^{(4)})}$, (c) $\psi_2^{(\mathbf{S}_{10}^{(4)})}$, (d) $\psi_2^{(\mathbf{S}_7^{(3)})}$ and (e) $\psi_2^{(\mathbf{S}_8^{(3)})}$. Note that Fig. 4 (d) and Fig. 4 (e) present the preceding time frames of Fig. 4 (b). We maintained a consistent color coding in all time-instances, i.e., each point has the same color throughout its trajectory in time, and the most significant values (largest in absolute value) are colored in either yellow or blue.

In the figure we see the multi-resolution properties of the proposed framework. At the highest level, $\ell = 8$ in plot (a), the eigenvector of $\mathbf{S}_1^{(8)}$ captures the coherent circular structures, i.e. the almost-invariant sets, which change from a horizontal orientation at the beginning of the trajectory to a vertical orientation at the end. These structures are consistent with the ones described by [34]. In contrast, in plots (b)-(c) (level $\ell = 4$), the effect of the velocity change over time is apparent, demonstrating that our framework is capable of detecting such properties. These plots present two equal-length sub-segments of the trajectory: $t \in \{49, \ldots, 64\}$ in plot (b) and $t \in \{145, \ldots, 160\}$ in plot (c). Plot (c), which corresponds to the faster regime closer to the end of the trajectory, depicts that the circular structures are captured by the eigenvector, whereas in plot (b), which corresponds to the slower regime, these structures are not visible. Due to the increase in point movement velocity over time, the components that are similarly expressed over time in the sub-segment that is closer to the end of the trajectory (plot (c)) are mainly the almost-invariant sets, as captured by the operator from the highest level in plot (a). Conversely, in the slower regime, there are other components that are similarly expressed over short

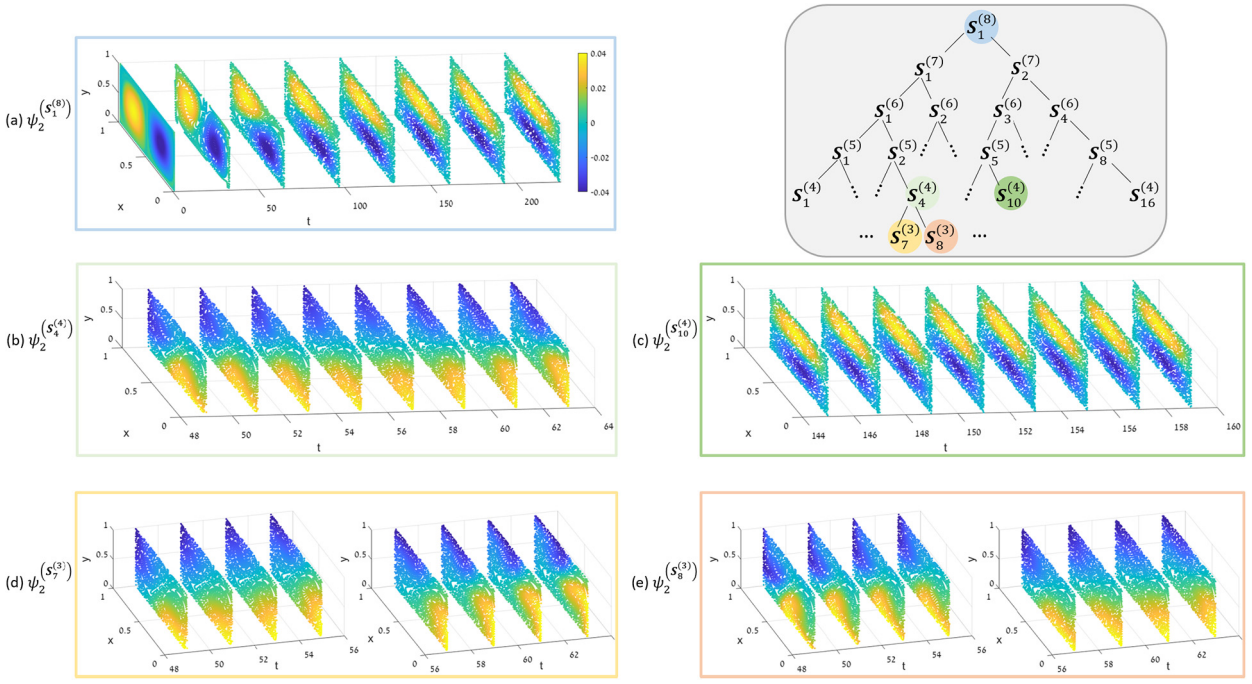


Fig. 4. Data points colored according to the second eigenvector of the operator $\mathbf{S}_r^{(\ell)}$ at different levels and time frames: (a) $\mathbf{S}_1^{(8)}$, (b) $\mathbf{S}_4^{(4)}$, (c) $\mathbf{S}_{10}^{(4)}$, (d) $\mathbf{S}_7^{(3)}$ and (e) $\mathbf{S}_8^{(3)}$. Plots (d) and (e) present the same 8 time points as in plot (b), where the points are colored according to a different eigenvector in each plot.

sub-segments in time, as captured by the eigenvector presented in plot (b). Plots (d) and (e) correspond to the two sub-segments $t \in \{49, \dots, 56\}$ and $t \in \{57, \dots, 64\}$, respectively, whose union is the sub-segment presented in plot (b). Note the similarity between the captured point dynamics in the two sub-segments of plots (d) and (e). This similarity explains the component emphasized by the eigenvector in plot (b), which is constructed based on these two sub-segments.

Note that the leading eigenvector of the operator $\mathbf{S}_r^{(\ell)}$ was omitted throughout this example since it mostly captures the common point distribution at the different time-frames. The point distribution is of less interest in this example since it provides a general geometric description of the problem setting rather than the common trajectory properties.

In the following we present the eigenvectors of the operators $\mathbf{F}$ and show that they indeed capture the time-varying trajectory behavior in consecutive time-frames.

Fig. 5 presents the data-points colored according to the leading eigenvectors of the respective operators $\mathbf{F}$ corresponding to the largest positive and negative (in absolute value) eigenvalues. In this setting, the eigenvectors corresponding to negative eigenvalues describe the components that are significantly more dominant in the first half of the time segment and the eigenvectors corresponding to positive eigenvalues describe components that are significantly more dominant in the second half of the segment, as we will show in Section 6 (Theorem 2). Similarly to Fig. 4, each plot in Fig. 5 describes one time frame containing 2^ℓ time points. For visualization, we present the $N = 2500$ data points locations at 8 evenly spaced time points along the time frame, and color them according to the eigenvector describing the entire time frame.

Fig. 5 (a) and Fig. 5 (b) present the eigenvectors of the operator $\mathbf{F}_1^{(8)}$ corresponding to the smallest negative eigenvalue in plot (a) and to the largest positive eigenvalue in plot (b). These plots depict that $\mathbf{F}_1^{(8)}$ captures the differences between the slower point movement in $t \in \{1, \dots, 128\}$ and the faster point movement in $t \in \{129, \dots, 256\}$. Due to the change in point movement velocity over time, the component describing the circular structures (the almost-invariant sets) is significantly more dominant in the sub-segment from the faster regime ($t \in \{129, \dots, 256\}$) than the slower regime ($t \in \{1, \dots, 128\}$), as captured by the eigenvector in plot (b). In contrast, in the slower regime other components are dominant (as demonstrated also by Fig. 4 (b)), leading to different structures being emphasized in Fig. 5 (a), which mostly captures the boundary points. Fig. 5 (c) and Fig. 5 (d) correspond to the time segment $t \in \{129, \dots, 256\}$. Plot (c) presents the leading eigenvector of the operator $\mathbf{F}_2^{(7)}$ with a negative eigenvalue, describing the components that are more dominant in the slower regime ($t \in \{129, \dots, 192\}$) and plot (d) presents the leading eigenvector with a positive eigenvalue, describing the components that are more dominant in the faster regime ($t \in \{193, \dots, 256\}$). Note that both plots (c) and (d) emphasize circular structures, however, the structures in plot (d) are smaller than the ones in plot (b) and are approximately complemented by the structures in plot (c). This behavior implies that our framework decomposes the almost-invariant sets into smaller components in short sub-segments (at lower operator-tree levels), and therefore, indicates that the proposed method indeed captures meaningful dynamical information in different time-scales. Fig. 5 (e) presents the eigenvector of $\mathbf{F}_1^{(6)}$ (describing $t \in [\{1, \dots, 64\}$) with the largest negative eigenvalue. This plot depicts that in the slower regime (at the beginning of the trajectory) the operator $\mathbf{F}$ highlights high-resolution fine components of the point movement dynamics.

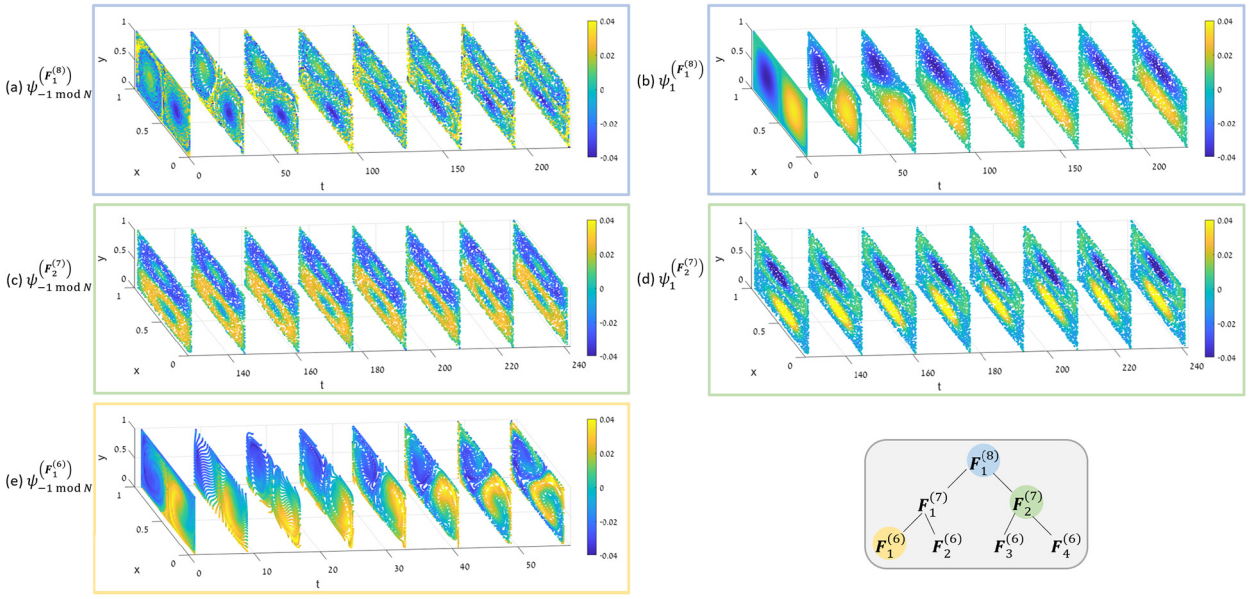


Fig. 5. Data points colored according to the eigenvector of (a-b) $\mathbf{F}_1^{(8)}$ with the smallest negative and largest positive eigenvalues, (c-d) $\mathbf{F}_2^{(7)}$ with the smallest negative and largest positive eigenvalues, and (e) $\mathbf{F}_1^{(6)}$ with the smallest negative eigenvalue.

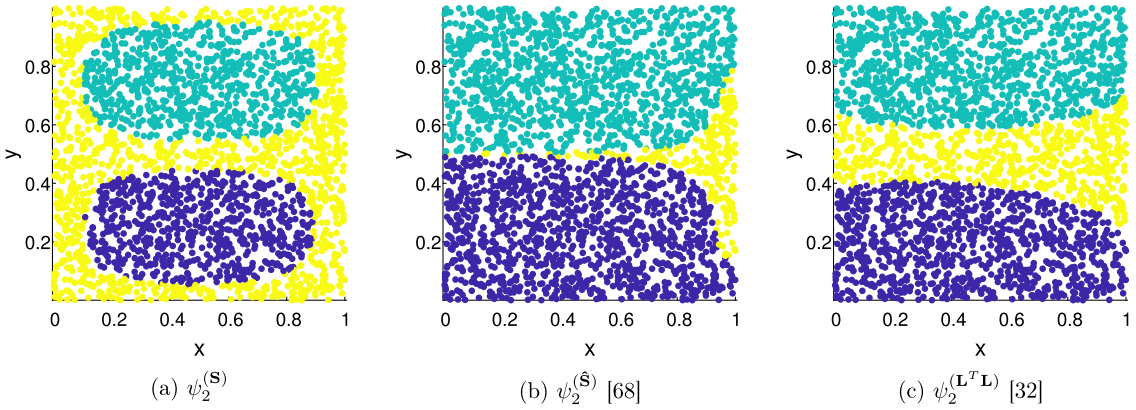


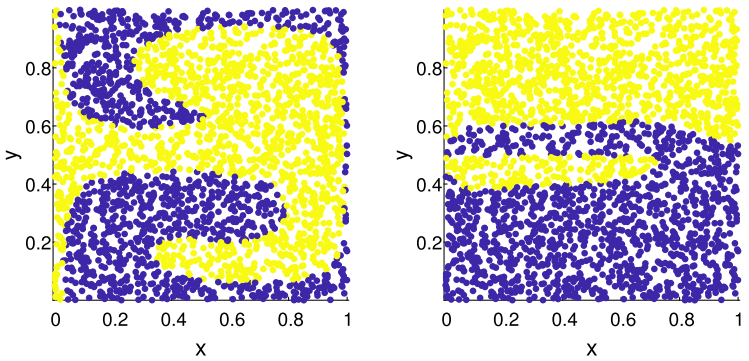
Fig. 6. Clustering of trajectory points based on eigenvectors of the proposed operator $\mathbf{S}$ in plot (a), and the operators from [68] in plot (b) and from [32] in plot (c). All operators are constructed by combining two time frames at $t = 250$ and $t = 256$.

We remark that different choices of the kernel scale in (12) lead to different resolutions. For example, taking a smaller kernel scale leads to a slower "convergence" to the almost-invariant sets of the representations obtained at the different levels of the operator $\mathbf{S}_r^{(\ell)}$, as well as an enhancement of finer structures captured by the operator $\mathbf{F}_r^{(\ell)}$.

In order to evaluate our framework with respect to previous work, we compare the operators $\mathbf{S}$ and $\mathbf{F}$, which serve as the building blocks of our algorithm, with related operators: (i) the dynamic Laplacian [32], which was shown to recover coherent sets from multiple time-frames of dynamical systems, and (ii) symmetric and anti-symmetric diffusion operators that were shown to recover similar and different components in multimodal data [68]. In [32], with a slight abuse of notation and in analogy to (13), the dynamic Laplacian is defined by $\mathbf{L}^\top\mathbf{L}$, where $\mathbf{L} = \mathbf{W}_1\mathbf{W}_2$. The common and difference operators in [68] are defined by $\hat{\mathbf{S}} = \mathbf{W}_1\mathbf{W}_2^\top + \mathbf{W}_2\mathbf{W}_1^\top$ and $\hat{\mathbf{A}} = \mathbf{W}_1\mathbf{W}_2^\top - \mathbf{W}_2\mathbf{W}_1^\top$, respectively, which are analogous to the operators $\mathbf{S}$ and $\mathbf{F}$ in (13) and (14).

Fig. 6 presents point clustering using k-means applied to the second eigenvector of the 3 operators used for recovering similarities: the proposed operator $\mathbf{S}$ in plot (a), the operator $\hat{\mathbf{S}}$ from [68] in plot (b) and the operator $\mathbf{L}^\top\mathbf{L}$ from [32] in plot (c). All 3 operators were constructed from time frames $t = 250$ and $t = 256$. We see in this figure that the proposed formulation of the operator is significantly better at capturing the almost-invariant sets (the circular structures). Note that the results presented here for $\mathbf{L}^\top\mathbf{L}$ are different than those in [32], since we take into account only two close time frames, whereas in [32] the operator is constructed using all the points along the trajectory.

In Fig. 7, we present the point clustering obtained by applying k-means to the second eigenvector of the 2 operators used for recovering differences: the proposed operator $\mathbf{F}$ in plot (a) and the operator $\hat{\mathbf{A}}$ from [68] in plot (b). In contrast to plot (b), in plot (a) we clearly see the swirl of the flow from and to the invariant sets (outward and inward).



(a) $\psi_2^{(\mathbf{F})}$ (b) $\psi_2^{(\hat{\mathbf{A}})}$ [68]

Fig. 7. Clustering of trajectory points based on eigenvectors of the proposed operator $\mathbf{F}$ in plot (a) and operator $\hat{\mathbf{A}}$ from [68] in plot (b).

In addition to the differences demonstrated in Fig. 6 and Fig. 7, another crucial advantage of our formulation relates to the construction of the multi-resolution framework and its theoretical justification presented in Section 6. The operators in [68] and [32] do not have any theoretical guarantees in such an operator-tree construction and may not be suitable to such a setting. Indeed, we report that a similar operator-tree constructed using the operators from [68] did not exhibit the expected behavior and no meaningful representations were obtained.

We conclude by noting that such a multi-resolution analysis of the dynamics may be especially useful in applications where the parameters of interest are inaccessible, e.g., for oceanic current analysis based on ocean drifters data [2,35], since the data is represented using non-linear kernels.

5.4. Hyperspectral and LiDAR imagery

In Section 4.1, we consider two datasets and present the two Riemannian composite operators $\mathbf{S}$ and $\mathbf{F}$. Later, in Section 4.2, these two datasets are considered as two consecutive sets in a temporal sequence of datasets, and the two Riemannian composite operators are used as a basic building block for our Riemannian multi-resolution analysis. Alternatively, similarly to the setting in [47,68], the two datasets could arise from simultaneous observations from two views or modalities. Here, we demonstrate the properties of the Riemannian composite operators $\mathbf{S}$ and $\mathbf{F}$ on real remote sensing data obtained by two different modalities.

We consider data from the 2013 IEEE GRSS data fusion contest,[3] which includes a hyperspectral image (HSI) with 144 spectral bands ($380-1050$ nm range) and a LiDAR Digital Surface Model of the University of Houston campus and its neighboring urban area in Houston, Texas. The data from both modalities have the same spatial resolution of 2.5 m. This data was previously considered in the context of manifold learning in [58].

We focus on two 60×90 image patches from the full dataset, in order to reduce computation time of the operators and their eigenvalue decomposition. We first preprocess the 60×90 LiDAR image and each image in the $60\times 90\times 144$ HSI data. The preprocessing stage includes dividing each 60×90 image by its standard deviation and removing outliers: for the LiDAR image, pixel values larger than the 99th percentile were removed, and for the HSI data, in each image, pixel values larger than the 95th percentile or smaller than the 5th percentile were removed. In both modalities the outliers were replaced by their nearest non-outlier values. Fig. 8 (a) and Fig. 9 (a) present the two LiDAR image patches after preprocessing, and Fig. 8 (b) and Fig. 9 (b) present the two average HSI image patches after preprocessing.

We apply the operators $\mathbf{S}$ and $\mathbf{F}$ to this data in order to analyze the scene properties captured by both the LiDAR and the HSI sensors and extract the similarities and differences between them. This can be viewed as a manifold-driven component analysis.

We construct the operators according to Algorithm 3, where the LiDAR image and the HSI images are defined as two datasets with point correspondence between them given by the pixel location. We reshape both datasets such that $\mathbf{x}_1 \in \mathbb{R}^{N\times 1}$ is the reshaped LiDAR image and $\mathbf{x}_2 \in \mathbb{R}^{N\times 144}$ is the reshaped HSI images, where $N = 5400$. The resulting kernels, $\mathbf{W}_1$ and $\mathbf{W}_2$, and operators, $\mathbf{S}$ and $\mathbf{F}$, are matrices of size $N\times N$.

We begin with an analysis of the first chosen image patch, shown in Fig. 8 (a) and (b). To depict the advantages of applying the proposed operators, we visually compare the eigenvectors of the kernels, $\mathbf{W}_1$ and $\mathbf{W}_2$, with the eigenvectors of the operators $\mathbf{S}$ and $\mathbf{F}$.

Fig. 8 (c-k) presents the absolute values of the leading eigenvectors of $\mathbf{S}$ in (c), $\mathbf{W}_1$ in (d-e), $\mathbf{W}_2$ in (h-i), and of $\mathbf{F}$ that correspond to the largest positive eigenvalues in (f-g) and largest negative (in absolute value) eigenvalues in (j-k). All eigenvectors are reshaped into images of size 60×90. The absolute value of the eigenvectors is presented in order to emphasize the dominant structures in the images and the differences between the leading eigenvectors of the two kernels and the leading eigenvectors of the operators $\mathbf{S}$ and $\mathbf{F}$.

[3] http://www.grss-ieee.org/community/technical-committees/data-fusion/2013-ieee-grss-data-fusion-contest/.

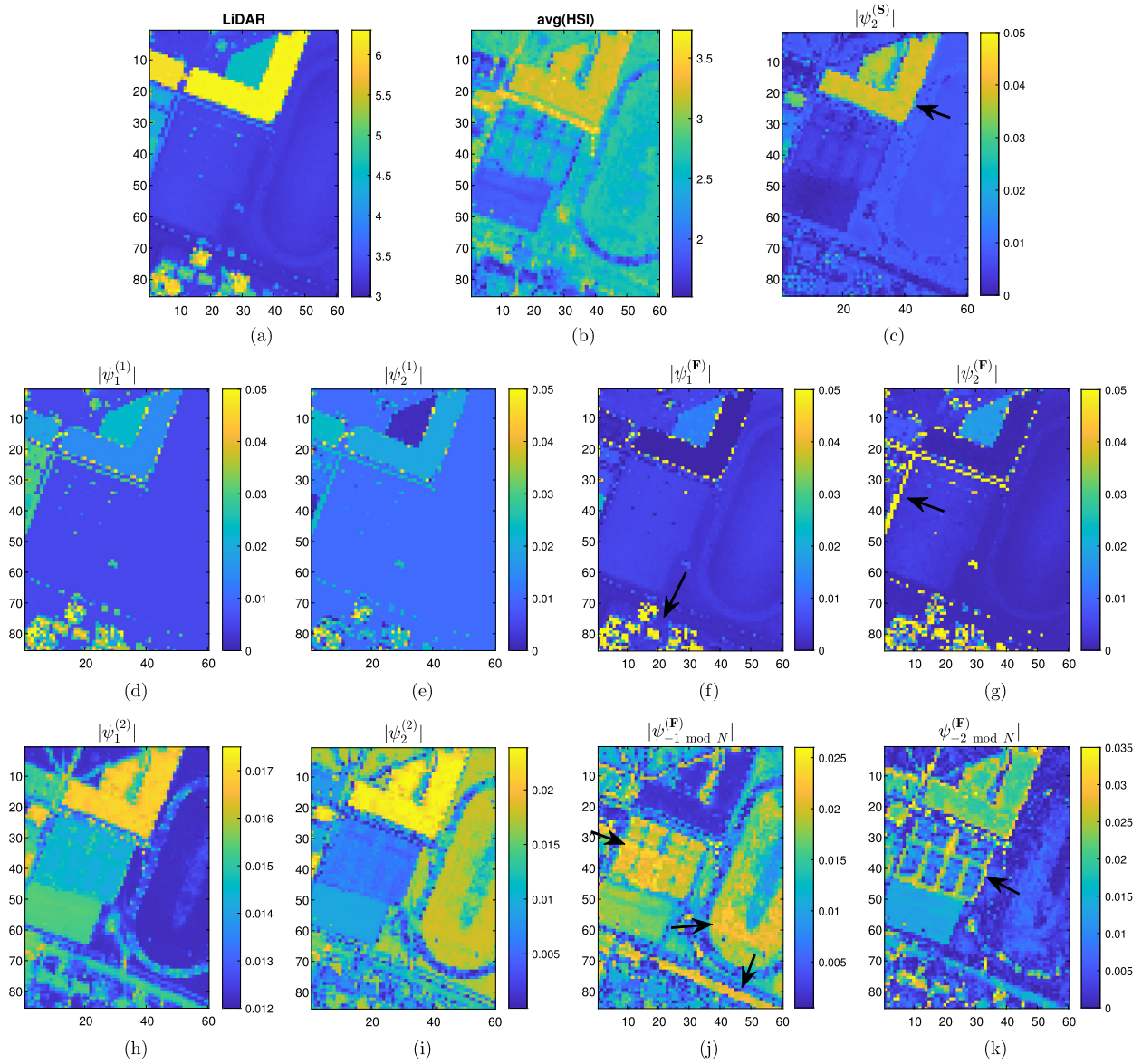


Fig. 8. The two chosen image patches of (a) the LiDAR image and (b) the HSI image after preprocessing, along with the leading eigenvectors of (c) $\mathbf{S}$, (d-e) $\mathbf{W}_1$ (LiDAR), (f-g) $\mathbf{F}$ corresponding to its 2 largest positive eigenvalues, (h-i) $\mathbf{W}_2$ (HSI), (j-k) $\mathbf{F}$ corresponding to its two smallest negative eigenvalues.

Fig. 8 (c) presents the absolute values of the leading eigenvector of $\mathbf{S}$ and depicts that the operator $\mathbf{S}$ indeed recovers common structures strongly expressed in both images. Specifically, this figure mostly highlights an 'L'-shaped building at the top of the image, which is the most dominant structure (represented by the high pixel values) in both modalities. Fig. 8 (d-k) depicts that the eigenvectors of the operator $\mathbf{F}$ capture and enhance differently expressed common structures. Consider for example the most dominant structures (with highest absolute values) in the LiDAR image presented in Fig. 8 (a). These structures include the 'L'-shaped building at the top of the image and trees at the bottom. Both structures are represented by high values in the eigenvectors of $\mathbf{W}_1$ in Fig. 8 (d-e). However, in Fig. 8 (f-g), which presents the leading eigenvectors of $\mathbf{F}$ with positive eigenvalues, only the trees are significantly highlighted, whereas the 'L'-shaped building is significantly attenuated. This is due to the differences between the two modalities, since the HSI image highlights this 'L' shaped building but not the trees. Other structures exhibiting such properties are marked by black arrows in Fig. 8 (f-g). In addition, Fig. 8 (h-k) depicts that the structures which are dominant only in the HSI images are emphasized by the eigenvectors of $\mathbf{F}$ corresponding to negative eigenvalues, whereas structures that are dominant in both modalities are significantly attenuated. Examples for such structures are marked by black arrows in Fig. 8 (j-k).

We repeat the presentation for the second image patch, shown in Fig. 9 (a) and (b). Fig. 9 (c-g) presents the absolute value of the leading eigenvector of $\mathbf{W}_1$ in plot (c), of $\mathbf{W}_2$ in plot (d), of $\mathbf{F}$ with a positive eigenvalue in plot (e), of $\mathbf{F}$ with a negative eigenvalue in plot (f) and of $\mathbf{S}$ in plot (g).

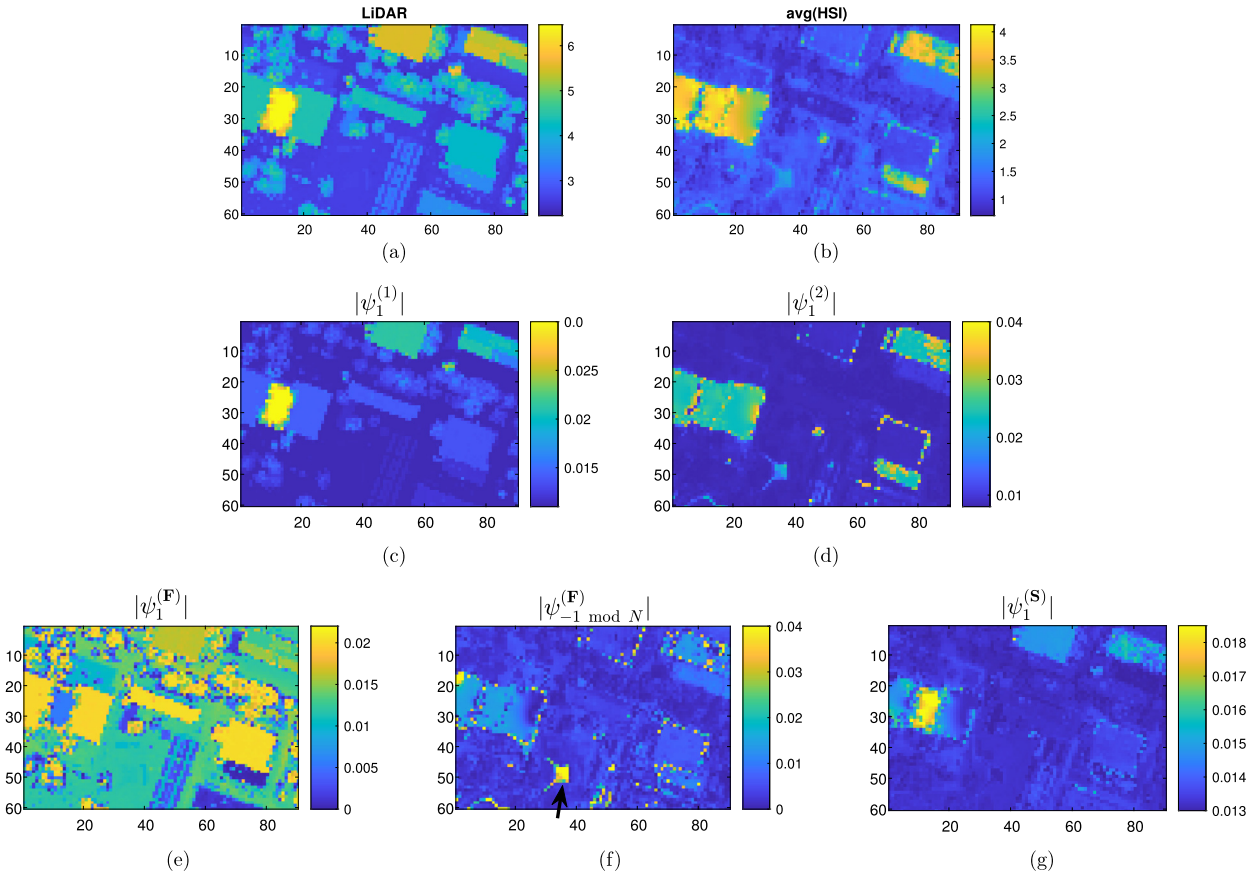


Fig. 9. The two chosen image patches of (a) the LiDAR image and (b) the HSI image after preprocessing along with the leading eigenvectors of (c) $\mathbf{W}_1$ (LiDAR data), (d) $\mathbf{W}_2$ (HSI data), (e) $\mathbf{F}$ corresponding to the largest positive eigenvalue, (f) $\mathbf{F}$ corresponding to the smallest negative eigenvalue and (g) $\mathbf{S}$.

Note the dominant structures (with high absolute values) in the leading eigenvectors of the two modalities in Fig. 9 (c) and (d). The dominant structures of the eigenvector representing the LiDAR image in plot (c) include buildings and trees, whereas in plot (d), which relates to the HSI image, only some of the building appear with high intensities and the trees are not clearly visible. This corresponds to the data presented in Fig. 9 (a) and (b). The leading eigenvector of $\mathbf{F}$ with a positive eigenvalue, presented in Fig. 9 (e), captures buildings and trees that are expressed more dominantly in the LiDAR image compared with the HSI image. In addition, the structures that are dominant in both modalities appear to be less dominant in this plot. For example, the "attenuated" structures in plot (e) include the small rectangular roof part around pixels $(x, y) = (10, 25)$ (where x denotes the horizontal axis and y denotes the vertical axis) and the building around pixels $(x, y) = (80, 10)$. Conversely, the leading eigenvector of $\mathbf{F}$ with a negative eigenvalue, presented in Fig. 9 (f), significantly enhances a specific location, marked by a black arrow in this plot, that is clearly visible in the HSI image presented in Fig. 9 (b) but barely visible in Fig. 9 plot (a). Note that this building is not represented by high pixel values in the raw HSI average image and therefore a simple subtraction between the two images will not lead to a similar emphasis of the building.

The leading eigenvector of $\mathbf{S}$, presented in Fig. 9 (g), captures some combination of the structures that are dominant in both modalities.

To summarize this example, we showed that the operator $\mathbf{F}$ captures common components that are expressed strongly only by one of the modalities and that the sign of the eigenvalues of $\mathbf{F}$ indicates in which modality the component is stronger. In addition, we showed that the operator $\mathbf{S}$ captures some combination of the dominant components in both modalities.

6. Spectral analysis

To provide theoretical justification to the proposed RMRA framework for spatiotemporal analysis presented in Section 4, we analyze the operators $\mathbf{S}$ and $\mathbf{F}$ defined in (13) and (14) and show that they admit the desired properties. Specifically, we show that the operator $\mathbf{S}$ enhances common eigenvectors that are expressed similarly in two consecutive time frames in the sense that they have similar eigenvalues. In addition, we show that the operator $\mathbf{F}$ enhances common eigenvectors that are expressed differently in two consecutive time frames in the sense that they have different eigenvalues.

In the following theoretical analysis we focus on two cases: (i) $\mathbf{W}_1$ and $\mathbf{W}_2$ have *strictly* common components, i.e., some of the eigenvectors of the two matrices are identical, and (ii) $\mathbf{W}_1$ and $\mathbf{W}_2$ have *weakly* common components, i.e. some of the eigenvectors of the two matrices differ by a small perturbation.

6.1. Strictly common components

Given that some of the eigenvectors of matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ are identical, we show in the following that for these identical eigenvectors, the operator $\mathbf{S}$ enhances the eigenvectors that have similar dominant eigenvalues and the operator $\mathbf{F}$ enhances the eigenvectors that have significantly different eigenvalues.

We begin by reiterating a theorem from [42] along with its proof, which shows that the eigenvectors that are similarly expressed in both matrices, $\mathbf{W}_1$ and $\mathbf{W}_2$, correspond to the largest eigenvalues of $\mathbf{S}$.

Theorem 1. *Consider a vector ψ, which is an eigenvector of both $\mathbf{W}_1$ and $\mathbf{W}_2$ with possibly different eigenvalues: $\mathbf{W}_1\psi = \lambda^{(1)}\psi$ and $\mathbf{W}_2\psi = \lambda^{(2)}\psi$. Then, ψ is also an eigenvector of $\mathbf{S} = \mathbf{W}_1^{1/2}\left(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}\right)^{1/2}\mathbf{W}_1^{1/2}$ with the corresponding eigenvalue:*

$$\lambda^{(\mathbf{S})} = \sqrt{\lambda^{(1)}\lambda^{(2)}} \tag{34}$$

Proof. From (13) we have:

$$\begin{aligned}
\mathbf{S}\psi &= \mathbf{W}_1^{1/2}\left(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}\right)^{1/2}\mathbf{W}_1^{1/2}\psi \\
&= \mathbf{W}_1^{1/2}\left(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}\right)^{1/2}\sqrt{\lambda^{(1)}}\psi \\
&= \mathbf{W}_1^{1/2}\sqrt{\lambda^{(2)}/\lambda^{(1)}}\sqrt{\lambda^{(1)}}\psi \\
&= \sqrt{\lambda^{(1)}\lambda^{(2)}}\psi
\end{aligned}$$

where the transition before last is due to $\mathbf{W}_1^{-1/2}\mathbf{W}^{(2)}\mathbf{W}_1^{-1/2}\psi = (\lambda^{(2)}/\lambda^{(1)})\psi$. □

This result implies that strictly common components that are dominant and similarly expressed in both datasets (with similar large eigenvalues) are dominant in $\mathbf{S}$ (have a large eigenvalue $\lambda^{(\mathbf{S})}$), i.e. if $\lambda^{(1)} \approx \lambda^{(2)}$ then $\lambda^{(\mathbf{S})} \approx \lambda^{(1)}, \lambda^{(2)}$.

We derive a similar theoretical analysis for the operator $\mathbf{F}$.

Theorem 2. *Consider a vector ψ, which is an eigenvector of both $\mathbf{W}_1$ and $\mathbf{W}_2$ with possibly different eigenvalues: $\mathbf{W}_1\psi = \lambda^{(1)}\psi$ and $\mathbf{W}_2\psi = \lambda^{(2)}\psi$. Then ψ is also an eigenvector of $\mathbf{F} = \mathbf{S}^{1/2}\log\left(\mathbf{S}^{-1/2}\mathbf{W}_1\mathbf{S}^{-1/2}\right)\mathbf{S}^{1/2}$ with the corresponding eigenvalue:*

$$\lambda^{(\mathbf{F})} = \frac{1}{2}\sqrt{\lambda^{(1)}\lambda^{(2)}}(\log(\lambda^{(1)}) - \log(\lambda^{(2)})) \tag{35}$$

Proof. From (14) we have:

$$\begin{aligned}
\mathbf{F}\psi &= \mathbf{S}^{1/2}\log\left(\mathbf{S}^{-1/2}\mathbf{W}_1\mathbf{S}^{-1/2}\right)\mathbf{S}^{1/2}\psi \\
&= \mathbf{S}^{1/2}\log\left(\mathbf{S}^{-1/2}\mathbf{W}_1\mathbf{S}^{-1/2}\right)(\lambda^{(1)}\lambda^{(2)})^{0.25}\psi \\
&= \mathbf{S}^{1/2}(0.5\log(\lambda^{(1)}) - 0.5\log(\lambda^{(2)}))(\lambda^{(1)}\lambda^{(2)})^{0.25}\psi \\
&= \frac{1}{2}\sqrt{\lambda^{(1)}\lambda^{(2)}}(\log(\lambda^{(1)}) - \log(\lambda^{(2)}))\psi
\end{aligned}$$

where the transition before last is due to $\mathbf{S}^{-1/2}\mathbf{W}_1\mathbf{S}^{-1/2}\psi = \sqrt{\lambda^{(1)}/\lambda^{(2)}}\psi$ and the application of log to the matrix multiplication, which is equivalent to applying log to its eigenvalues, leading to $\log(\sqrt{\lambda^{(1)}/\lambda^{(2)}}) = 0.5\log(\lambda^{(1)}) - 0.5\log(\lambda^{(2)})$. □

This result indicates that the strictly common components of the two datasets are also expressed by $\mathbf{F}$ and that their order is determined by their relative expression in each dataset. For example, if ψ is an eigenvector of both $\mathbf{W}_1$ and $\mathbf{W}_2$ and corresponds to equal eigenvalues, $\lambda^{(1)} = \lambda^{(2)}$, then this component is part of the null space of $\mathbf{F}$; if $\lambda^{(1)} \neq \lambda^{(2)}$, then ψ corresponds to a nonzero eigenvalue in $\mathbf{F}$. Note that $\lambda^{(\mathbf{F})}$ depends also on the multiplication by $\sqrt{\lambda^{(1)}\lambda^{(2)}}$.

Another notable result of Theorem 2 is that the sign of the eigenvalues of $\mathbf{F}$ indicates in which dataset their corresponding eigenvector is more dominant. For example, if ψ is an eigenvector of both $\mathbf{W}_1$ and $\mathbf{W}_2$ that has a large corresponding eigenvalue in $\mathbf{W}_1$ but a small eigenvalue in $\mathbf{W}_2$ ($\lambda^{(1)} \gg \lambda^{(2)}$), then the corresponding eigenvalue in $\mathbf{F}$ is large and positive. Conversely, if ψ is more dominant in $\mathbf{W}_2$, then its corresponding eigenvalue in $\mathbf{F}$ is large (in absolute value) and negative.

An example depicting these properties is presented in Subsection 5.1, which demonstrates that the eigenvalues of the operators $\mathbf{S}$ and $\mathbf{F}$ are indeed equal to the expected values based on Theorem 1 and Theorem 2.

It is important to note that when $\mathbf{W}_1$ and $\mathbf{W}_2$ are defined with a Gaussian kernel, as in (11), the kernel scale choice may affect the eigenvalue decay and thus affect the highlighted components by $\mathbf{S}$ and $\mathbf{F}$, as indicated by Theorem 1 and Theorem 2. In our experiments we have focused on a simple approach to reduce such tuning issues by normalizing distances by their median, leading to cancellation of global scaling differences. The effect of this approach is studied in [26] under a random matrix framework, which showed the importance of choosing a proper bandwidth. However, such a tuning approach does not directly solve the issue when two modalities have significantly different characteristics and outliers, as shown in [24] for kernel sensor fusion algorithms. Another approach is adaptively tuning the kernel scale locally. For example, in [50], a scale is chosen for each dimension by optimizing over the resulting intrinsic dimension of the data. Other methods locally scale the kernel bandwidth for each pair of points, e.g., [9,16,86]. Such methods allow for a better choice of the kernel scale and can minimize the variability of the bandwidths of $\mathbf{W}_\ell$ between different modalities, and therefore, reduce the effect of the bandwidth on $\mathbf{S}$ and $\mathbf{F}$.

6.2. Weakly common components

To further demonstrate the power of the operators $\mathbf{S}$ and $\mathbf{F}$, we provide stability analysis by investigating how a small variation of the common eigenvector affects the results. For this purpose, we make use of the concept of a pseudo-spectrum [80]. While pseudo-spectra is typically used to provide an analytic framework for investigating non-normal matrices and operators, here we apply it to symmetric matrices for the purpose of analysis of nearly but non-common eigenvectors. We begin by recalling three equivalent definitions of the ϵ-pseudo-spectrum as presented in [80].

Definition 1 *(Pseudo-spectrum).* Given a matrix $\mathbf{M} \in \mathbb{R}^{N\times N}$, the following definitions of the ϵ-pseudo-spectrum are equivalent for a small $\epsilon > 0$:

1. $\sigma_\epsilon(\mathbf{M}) = \left\{ \lambda \in \mathbb{R} : \left\| (\lambda \mathrm{I} - \mathbf{M})^{-1} \right\| \geq \epsilon^{-1} \right\}$
2. $\sigma_\epsilon(\mathbf{M}) = \left\{ \lambda \in \mathbb{R} : \lambda \in \sigma(\mathbf{M} + \mathbf{E}) \text{ for a } \mathbf{E} \text{ with } \|\mathbf{E}\| \leq \epsilon \right\}$
3. $\sigma_\epsilon(\mathbf{M}) = \left\{ \lambda \in \mathbb{R} : \exists v \in \mathbb{R}^N \text{ with } \|v\|_2 = 1 \text{ s.t. } \|(\mathbf{M} - \lambda \mathrm{I})v\|_2 \leq \epsilon \right\}$

where $\sigma(\mathbf{M})$ denotes the set of eigenvalues of $\mathbf{M}$, I denotes the identity matrix, $\|\cdot\|_2$ denotes the ℓ_2 norm, and $\|\cdot\|$ denotes the induced operator norm. Moreover, we term a vector v that adheres to definition 3 an ϵ-pseudo-eigenvector.

The following theorem is the counterpart of Theorem 1 for the case where the eigenvector is slightly perturbed.

Theorem 3. *Suppose there exists an eigenpair $\lambda^{(k)}$ and $\psi^{(k)}$ of $\mathbf{W}_k$ for $k = 1, 2$ so that $\psi^{(1)} = \psi^{(2)} + \psi_{\epsilon_\mathbf{S}}$, where $\left\|\psi_{\epsilon_\mathbf{S}}\right\|_2 \leq \frac{\sqrt{\lambda^{(2)}}}{\tilde{\lambda}^{(2)}_{max}\sqrt{\lambda^{(1)}}}\epsilon_\mathbf{S}$ for a small $\epsilon_\mathbf{S} > 0$, where $\tilde{\lambda}^{(2)}_{max} = \|\mathbf{W}_2 - \lambda^{(2)}\mathrm{I}\|$. Then, we have*

$$\sqrt{\lambda^{(1)}\lambda^{(2)}} \in \sigma_{\epsilon_\mathbf{S}}(\mathbf{S}). \tag{36}$$

Specifically, we have

$$\left\| (\mathbf{S} - \sqrt{\lambda^{(1)}\lambda^{(2)}}\mathrm{I})\psi^{(1)} \right\|_2 \leq \epsilon_\mathbf{S},$$

and $\psi^{(1)}$ is a corresponding $\epsilon_\mathbf{S}$-pseudo-eigenvector of $\mathbf{S}$.

Informally, this theorem implies that when $\psi^{(1)}$ is a slight perturbation of $\psi^{(2)}$, then $\psi^{(1)}$ is "almost" an eigenvector of $\mathbf{S}$ with a corresponding eigenvalue $\sqrt{\lambda^{(1)}\lambda^{(2)}}$. Equivalently, $\psi^{(2)}$ can also be shown to be an $\epsilon_\mathbf{S}$-pseudo-eigenvector of $\mathbf{S}$ with the same corresponding eigenvalue, which suggests that the operator $\mathbf{S}$ is "stable" to finite perturbations. We remark that since $\|\mathbf{W}_2\| = 1$, we have that $\tilde{\lambda}^{(2)}_{max} = \max(1 - \lambda^{(2)}, \lambda^{(2)}) \in [0.5, 1)$, guaranteeing that the perturbation of the eigenvector is small. We also remark that our numerical study shows that the bound for the ϵ-pseudo-eigenvalues and ϵ-pseudo-eigenvectors of $\mathbf{S}$ is tight.

Proof. By Proposition 1 (see Appendix A), we have

$$\mathbf{S} = \mathbf{W}_1^{1/2}\left(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}\right)^{1/2}\mathbf{W}_1^{1/2} = (\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2}\mathbf{W}_1. \tag{37}$$

Since $\psi^{(1)}$ is an eigenvector of $\mathbf{W}_1$ with an eigenvalue $\lambda^{(1)}$, we have

$$\mathbf{S}\psi^{(1)} = \left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)^{1/2}\mathbf{W}_1\psi^{(1)} = \lambda^{(1)}\left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)^{1/2}\psi^{(1)}. \tag{38}$$

Therefore, it is sufficient to show that $\psi^{(1)}$ is an ϵ-pseudo-eigenvector of $\left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)^{1/2}$. By a direct expansion, we have

$$\begin{aligned}\mathbf{W}_2\mathbf{W}_1^{-1}\psi^{(1)} &= \frac{1}{\lambda^{(1)}}\mathbf{W}_2\psi^{(1)} = \frac{1}{\lambda^{(1)}}\mathbf{W}_2(\psi^{(2)} + \psi_{\epsilon_\mathbf{S}}) \\ &= \frac{\lambda^{(2)}}{\lambda^{(1)}}\psi^{(2)} + \frac{1}{\lambda^{(1)}}\mathbf{W}_2\psi_{\epsilon_\mathbf{S}} = \frac{\lambda^{(2)}}{\lambda^{(1)}}\psi^{(1)} + \frac{1}{\lambda^{(1)}}(\mathbf{W}_2 - \lambda^{(2)}\mathrm{I})\psi_{\epsilon_\mathbf{S}},\end{aligned} \tag{39}$$

where the last transition is obtained by replacing $\psi^{(2)}$ with $\psi^{(1)} - \psi_{\epsilon_{\mathbf{S}}}$. By reorganizing the elements in (39), we have

$$\left(\mathbf{W}_2\mathbf{W}_1^{-1} - \frac{\lambda^{(2)}}{\lambda^{(1)}}\mathrm{I}\right)\psi^{(1)} = \frac{1}{\lambda^{(1)}}(\mathbf{W}_2 - \lambda^{(2)}\mathrm{I})\psi_{\epsilon_{\mathbf{S}}}\,. \tag{40}$$

The left hand side can be decomposed into:

$$\left(\mathbf{W}_2\mathbf{W}_1^{-1} - \frac{\lambda^{(2)}}{\lambda^{(1)}}\mathrm{I}\right)\psi^{(1)} = \left((\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2} + \sqrt{\frac{\lambda^{(2)}}{\lambda^{(1)}}}\mathrm{I}\right)\left((\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2} - \sqrt{\frac{\lambda^{(2)}}{\lambda^{(1)}}}\mathrm{I}\right)\psi^{(1)}, \tag{41}$$

and since $(\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2} + \sqrt{\lambda^{(2)}/\lambda^{(1)}}\mathrm{I}$ is positive definite, (40) and (41) lead to

$$\left((\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2} - \sqrt{\frac{\lambda^{(2)}}{\lambda^{(1)}}}\mathrm{I}\right)\psi^{(1)} = \frac{1}{\lambda^{(1)}}\left((\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2} + \sqrt{\frac{\lambda^{(2)}}{\lambda^{(1)}}}\mathrm{I}\right)^{-1}(\mathbf{W}_2 - \lambda^{(2)}\mathrm{I})\psi_{\epsilon_{\mathbf{S}}}. \tag{42}$$

Multiplying both sides by $\lambda^{(1)}$, we have

$$\begin{aligned}\left(\mathbf{S} - \sqrt{\lambda^{(1)}\lambda^{(2)}}\mathrm{I}\right)\psi^{(1)} &\underset{(38)}{=} \lambda^{(1)}(\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2}\psi^{(1)} - \sqrt{\lambda^{(1)}\lambda^{(2)}}\psi^{(1)} \\ &= \lambda^{(1)}\left((\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2} - \sqrt{\frac{\lambda^{(2)}}{\lambda^{(1)}}}\mathrm{I}\right)\psi^{(1)} \\ &\underset{(42)}{=} \left((\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2} + \sqrt{\frac{\lambda^{(2)}}{\lambda^{(1)}}}\mathrm{I}\right)^{-1}(\mathbf{W}_2 - \lambda^{(2)}\mathrm{I})\psi_{\epsilon_{\mathbf{S}}}.\end{aligned} \tag{43}$$

Applying the ℓ_2 norm leads to:

$$\begin{aligned}\left\|\left(\mathbf{S} - \sqrt{\lambda^{(1)}\lambda^{(2)}}\mathrm{I}\right)\psi^{(1)}\right\|_2 &\le \left\|\left((\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2} + \sqrt{\frac{\lambda^{(2)}}{\lambda^{(1)}}}\mathrm{I}\right)^{-1}\right\|\left\|\mathbf{W}_2 - \lambda^{(2)}\mathrm{I}\right\|\left\|\psi_{\epsilon_{\mathbf{S}}}\right\|_2 \\ &\le \frac{1}{\sigma_{\min}\left((\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2}\right) + \sqrt{\lambda^{(2)}/\lambda^{(1)}}}\tilde{\lambda}^{(2)}_{max}\frac{\sqrt{\lambda^{(2)}}}{\tilde{\lambda}^{(2)}_{max}\sqrt{\lambda^{(1)}}}\epsilon_{\mathbf{S}} \\ &= \frac{\sqrt{\lambda^{(2)}/\lambda^{(1)}}}{\sigma_{\min}\left((\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2}\right) + \sqrt{\lambda^{(2)}/\lambda^{(1)}}}\epsilon_{\mathbf{S}} \\ &\le \frac{\sqrt{\lambda^{(2)}/\lambda^{(1)}}}{\sqrt{\lambda^{(2)}/\lambda^{(1)}}}\epsilon_{\mathbf{S}} = \epsilon_{\mathbf{S}},\end{aligned} \tag{44}$$

where $\sigma_{\min}$ denotes the minimum eigenvalue. Thus, $\sqrt{\lambda^{(1)}\lambda^{(2)}}$ is an $\epsilon_{\mathbf{S}}$-pseudo-eigenvalue of $\mathbf{S}$, where $\psi^{(1)}$ is a corresponding $\epsilon_{\mathbf{S}}$-pseudo-eigenvector. □

Note that the above proof shows that $\psi^{(1)}$ is a pseudo-eigenvector of $\mathbf{S}$, when $\mathbf{S}$ is defined as the midpoint of the geodesic curve connecting $\mathbf{W}_1$ and $\mathbf{W}_2$ (by setting $p = 0.5$). However, due to the decomposition in (41), the proof is not compatible with definitions of $\mathbf{S}$ at other points $p \in (0,1)$ along the geodesic path. For such cases, a different proof is required, specifically, without using the algebraic relationship in (41) that leads to (42). In the following statement, which is the counterpart of Theorem 2 for the case where the eigenvector is not strictly common, we control the "pseudo" part by a straightforward perturbation argument.

Theorem 4. *For $k = 1, 2$, consider the eigendecomposition $\mathbf{W}_k = \mathbf{U}_k\mathbf{L}_k\mathbf{U}_k^\top \in \mathbb{R}^{N\times N}$, where $\mathbf{L}_k := diag(\lambda_1^{(k)}, \ldots, \lambda_N^{(k)})$ so that $\lambda_1^{(1)} \ge \ldots \ge \lambda_N^{(1)}$ and $\mathbf{U}_k := \begin{bmatrix}\psi_1^{(k)} & \ldots & \psi_N^{(k)}\end{bmatrix} \in \mathcal{V}_{N,N}$. Assume the above eigendecomposition satisfies $\mathbf{U}_2 = \mathbf{U}_1 + \epsilon\mathbf{A}$, where $\|\mathbf{A}\| = 1$, $\epsilon > 0$ is a small constant, and $c^{-1} \le \ell_i := \lambda_i^{(2)}/\lambda_i^{(1)} \le c$ for some constant $c \ge 1$ for all $i = 1, \ldots, N$. For any $i = 1, \ldots, N$, denote the spectral gap $\gamma_i := \min_{k, \ell_k \ne \ell_i}|\ell_i - \ell_k|$.*

Fix j. Then, for the j-th eigenpair of $\mathbf{W}_1$, when ϵ is sufficiently small, we have

$$\left\|\left(\mathbf{F} - 0.5\sqrt{\lambda_j^{(1)}\lambda_j^{(2)}}\log\left(\frac{\lambda_j^{(1)}}{\lambda_j^{(2)}}\right)\mathrm{I}\right)\psi_j^{(1)}\right\|_2 = O(\epsilon)\,,$$

where the implied constant depends on $\frac{\sqrt{c}\ln c}{\min_i\left(\gamma_i\sqrt{\lambda_i^{(1)}}\right)}$.

Proof. By Proposition 1 (see Appendix A), we rewrite the operator $\mathbf{F} = \mathbf{S}^{1/2} \log\left(\mathbf{S}^{-1/2}\mathbf{W}_1\mathbf{S}^{-1/2}\right)\mathbf{S}^{1/2}$ as $\mathbf{F} = \log\left(\mathbf{W}_1\mathbf{S}^{-1}\right)\mathbf{S}$. Denote

$$\hat{\lambda}_{\mathbf{F}} = 0.5\sqrt{\lambda_j^{(1)}\lambda_j^{(2)}}\log\left(\frac{\lambda_j^{(1)}}{\lambda_j^{(2)}}\right).$$

By Theorem 3 and the definition of an $\epsilon_{\mathbf{S}}$-pseudo-eigenvector (see Proposition 2 in Appendix A), there exists $\mathbf{E}_{\mathbf{S}}$ with a sufficiently small norm, such that

$$\mathbf{S}\psi_j^{(1)} = \left(\sqrt{\lambda_j^{(1)}\lambda_j^{(2)}}\mathbf{I} - \mathbf{E}_{\mathbf{S}}\right)\psi_j^{(1)},$$

and therefore, we have

$$\begin{aligned}\left(\mathbf{F} - \hat{\lambda}_{\mathbf{F}}\mathbf{I}\right)\psi_j^{(1)} &= \left(\log\left(\mathbf{W}_1\mathbf{S}^{-1}\right)\mathbf{S} - \hat{\lambda}_{\mathbf{F}}\mathbf{I}\right)\psi_j^{(1)} \\ &= \sqrt{\lambda_j^{(1)}\lambda_j^{(2)}}\left(\log\left(\mathbf{W}_1\mathbf{S}^{-1}\right) - \frac{\hat{\lambda}_{\mathbf{F}}}{\sqrt{\lambda_j^{(1)}\lambda_j^{(2)}}}\mathbf{I}\right)\psi_j^{(1)} - \log\left(\mathbf{W}_1\mathbf{S}^{-1}\right)\mathbf{E}_{\mathbf{S}}\psi_j^{(1)}. \end{aligned} \tag{45}$$

Since both $\mathbf{W}_1$ and $\mathbf{W}_2$ are positive definite, they are invertible and also $\left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)^{1/2}$ is invertible. Therefore, by (37), $\mathbf{W}_1\mathbf{S}^{-1} = \mathbf{W}_1\mathbf{W}_1^{-1}\left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)^{-1/2} = \left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)^{-1/2}$. Substituting this relationship into (45) yields

$$\left(\mathbf{F} - \hat{\lambda}_{\mathbf{F}}\mathbf{I}\right)\psi_j^{(1)} = -\frac{\sqrt{\lambda_j^{(1)}\lambda_j^{(2)}}}{2}\left(\log\left(\mathbf{W}_2\mathbf{W}_1^{-1}\right) - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\mathbf{I}\right)\psi_j^{(1)} - \frac{1}{2}\log\left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)\mathbf{E}_{\mathbf{S}}\psi_j^{(1)}. \tag{46}$$

Now we control the right hand side term by term. Recall that for any analytic function f over an open set in $\mathbb{R}$ that contains the spectrum of $\mathbf{W}_2\mathbf{W}_1^{-1}$, we can define $f(\mathbf{W}_2\mathbf{W}_1^{-1})$. Since $\mathbf{W}_2\mathbf{W}_1^{-1}$ and $\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}$ are similar, we have

$$f(\mathbf{W}_2\mathbf{W}_1^{-1}) = \mathbf{W}_1^{1/2} f\left(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}\right)\mathbf{W}_1^{-1/2}, \tag{47}$$

and hence

$$f\left(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}\right)\psi_j^{(1)} = \mathbf{W}_1^{-1/2} f(\mathbf{W}_2\mathbf{W}_1^{-1})\mathbf{W}_1^{1/2}\psi_j^{(1)} = \sqrt{\lambda_j^{(1)}}\mathbf{W}_1^{-1/2} f(\mathbf{W}_2\mathbf{W}_1^{-1})\psi_j^{(1)}. \tag{48}$$

Let μ_i and v_i denote the eigenvalues and eigenvectors of the matrix $\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}$, respectively, for $i = 1, \ldots, N$. Since $\{\psi_j^{(1)}\}_{j=1}^N$ and $\{v_j\}_{j=1}^N$ are both orthonormal bases of $\mathbb{R}^N$, we have $\psi_j^{(1)} = \sum_i \alpha_{ji} v_i$, where $\alpha_{ji} \in \mathbb{R}$ and $\sum_i \alpha_{ji}^2 = 1$ for all j. By (48), we have

$$\begin{aligned}&\frac{\sqrt{\lambda_j^{(1)}\lambda_j^{(2)}}}{2}\left(\log(\mathbf{W}_2\mathbf{W}_1^{-1}) - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\mathbf{I}\right)\psi_j^{(1)} \\ &= \frac{\sqrt{\lambda_j^{(2)}}}{2}\mathbf{W}_1^{1/2}\left(\log(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}) - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\mathbf{I}\right)\psi_j^{(1)} \\ &= \frac{\sqrt{\lambda_j^{(2)}}}{2}\mathbf{W}_1^{1/2}\sum_{i=1}^N\left(\log(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}) - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\mathbf{I}\right)\alpha_{ji}v_i. \end{aligned} \tag{49}$$

Using the fact that

$$\left(\log\left(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}\right) - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\mathbf{I}\right)v_i = \left(\log\mu_i - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\right)v_i, \tag{50}$$

yields

$$\begin{aligned}&\frac{\sqrt{\lambda_j^{(2)}}}{2}\mathbf{W}_1^{1/2}\sum_{i=1}^N\left(\log(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}) - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\mathbf{I}\right)\alpha_{ji}v_i \\ &= \frac{\sqrt{\lambda_j^{(2)}}}{2}\mathbf{W}_1^{1/2}\sum_{i=1}^N\left(\log\mu_i - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\right)\alpha_{ji}v_i. \end{aligned} \tag{51}$$

Therefore, the squared L^2 norm of the first term in the right hand side of (46) becomes

$$\left\| \frac{\sqrt{\lambda_j^{(1)}\lambda_j^{(2)}}}{2}\left(\log(\mathbf{W}_2\mathbf{W}_1^{-1}) - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\mathrm{I}\right)\psi_j^{(1)} \right\|_2^2$$

$$= \left\| \frac{\sqrt{\lambda_j^{(2)}}}{2}\mathbf{W}_1^{1/2}\sum_{i=1}^{N}\left(\log\mu_i - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\right)\alpha_{ji}v_i \right\|_2^2 \le \frac{\lambda_j^{(2)}}{4}\left\|\mathbf{W}_1^{1/2}\right\|^2 \left\| \sum_{i=1}^{N}\left(\log\mu_i - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\right)\alpha_{ji}v_i \right\|_2^2$$

$$= \frac{\lambda_j^{(2)}}{4}\sum_{i=1}^{N}\alpha_{ji}^2\left(\log\mu_i - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\right)^2, \tag{52}$$

where we use the fact that $\{v_j\}$ form an orthonormal basis and that the operator norm $\left\|\mathbf{W}_1^{1/2}\right\|^2 = 1$, since $\mathbf{W}_1$ is normalized.

Next, as in (43), we set

$$\mathbf{E}_{\mathbf{S}}\psi_j^{(1)} = -\left(\mathbf{S} - \sqrt{\lambda_j^{(1)}\lambda_j^{(2)}}\mathrm{I}\right)\psi_j^{(1)} = -\lambda_j^{(1)}\left((\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2} - \sqrt{\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}}\mathrm{I}\right)\psi_j^{(1)}. \tag{53}$$

Since $f(x) = \log(x)\sqrt{x}$ is analytic over an open set that contains the spectrum of $\mathbf{W}_1\mathbf{W}_2^{-1}$, by the same argument as that for (51), we have

$$\log\left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)\mathbf{E}_{\mathbf{S}}\psi_j^{(1)} = -\lambda_j^{(1)}\log\left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)\left((\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2} - \sqrt{\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}}\mathrm{I}\right)\psi_j^{(1)}$$

$$= -\sqrt{\lambda_j^{(1)}}\mathbf{W}_1^{1/2}\sum_{i=1}^{N}\alpha_{ji}\log(\mu_i)\left(\sqrt{\mu_i} - \sqrt{\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}}\right)v_i,$$

and hence the squared L^2 norm of the second term in the right hand side of (46) becomes

$$\left\|\log\left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)\mathbf{E}_{\mathbf{S}}\psi_j^{(1)}\right\|_2^2 \le \lambda_j^{(1)}\sum_{i=1}^{N}\alpha_{ji}^2(\log\mu_i)^2\left(\sqrt{\mu_i} - \sqrt{\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}}\right)^2, \tag{54}$$

where we again use the fact that $\{v_j\}$ form an orthonormal basis and that the operator norm $\left\|\mathbf{W}_1^{1/2}\right\|^2 = 1$. To finish the proof, we control α_{ji} and the relationship between μ_i and $\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}$ in (52) and (54) using matrix perturbation theory. By a direct expansion, we have

$$\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2} = \mathbf{U}_1\mathbf{L}_2\mathbf{L}_1^{-1}\mathbf{U}_1^\top + \epsilon\mathbf{B}, \tag{55}$$

where $\mathbf{B} = \mathbf{U}_1\mathbf{L}_1^{-1/2}\mathbf{U}_1^\top(\mathbf{A}\mathbf{L}_2\mathbf{U}_1^\top + \mathbf{U}_1\mathbf{L}_2\mathbf{A}^\top)\mathbf{U}_1\mathbf{L}_1^{-1/2}\mathbf{U}_1^\top + \epsilon\mathbf{U}_1\mathbf{L}_1^{-1/2}\mathbf{U}_1^\top\mathbf{A}\mathbf{L}_2\mathbf{A}^\top\mathbf{U}_1\mathbf{L}_1^{-1/2}\mathbf{U}_1^\top$. To simplify the notation, we assume that the diagonal entries of $\mathbf{L}_2\mathbf{L}_1^{-1} = \mathrm{diag}\left(\frac{\lambda_1^{(2)}}{\lambda_1^{(1)}}, \ldots, \frac{\lambda_N^{(2)}}{\lambda_N^{(1)}}\right) = \mathrm{diag}\left(\ell_1, \ldots, \ell_N\right)$ are all distinct, or the following argument could be carried out with eigenprojections. Thus, by a standard perturbation argument, when ϵ is sufficiently small, we have

$$v_i = \psi_i^{(1)} + \epsilon\sum_{k\neq i}\frac{\langle\psi_k^{(1)}, \mathbf{B}\psi_i^{(1)}\rangle}{\lambda_i^{(2)}/\lambda_i^{(1)} - \lambda_k^{(2)}/\lambda_k^{(1)}}\psi_k^{(1)} + O(\epsilon^2), \quad \mu_i = \frac{\lambda_i^{(2)}}{\lambda_i^{(1)}} + \epsilon\langle\psi_i^{(1)}, \mathbf{B}\psi_i^{(1)}\rangle + O(\epsilon^2) \tag{56}$$

for each $i = 1, \ldots, N$. Note that

$$I_N = \mathbf{U}_2^\top\mathbf{U}_2 = (\mathbf{U}_1 + \epsilon\mathbf{A})^\top(\mathbf{U}_1 + \epsilon\mathbf{A}) = I_N + \epsilon(\mathbf{A}^\top\mathbf{U}_1 + \mathbf{U}_1^\top\mathbf{A}) + \epsilon^2\mathbf{A}^\top\mathbf{A},$$

so we have

$$\mathbf{A}^\top\mathbf{U}_1 = -\mathbf{U}_1^\top\mathbf{A} - \epsilon\mathbf{A}^\top\mathbf{A}. \tag{57}$$

Thus, by a direct expansion with the definition of $\mathbf{B}$, we have

$$\langle\psi_k^{(1)}, \mathbf{B}\psi_i^{(1)}\rangle = (\lambda_i^{(1)}\lambda_k^{(1)})^{-1/2}e_k^\top\mathbf{U}_1^\top(\mathbf{A}\mathbf{L}_2\mathbf{U}_1^\top + \mathbf{U}_1\mathbf{L}_2\mathbf{A}^\top)\mathbf{U}_1e_i + O(\epsilon)$$

$$= (\lambda_i^{(1)}\lambda_k^{(1)})^{-1/2}(\lambda_i^{(2)}e_k^\top\mathbf{U}_1^\top\mathbf{A}e_i + \lambda_k^{(2)}e_k^\top\mathbf{A}^\top\mathbf{U}_1e_i) + O(\epsilon)$$

$$= (\lambda_i^{(1)}\lambda_k^{(1)})^{-1/2}\left((\lambda_i^{(2)} - \lambda_k^{(2)})e_k^\top \mathbf{U}_1^\top \mathbf{A}e_i - \epsilon\lambda_k^{(2)} e_k \mathbf{A}^\top \mathbf{A}e_i\right) + (\lambda_i^{(1)}\lambda_k^{(1)})^{-1/2}O(\epsilon)$$
$$= (\lambda_i^{(1)}\lambda_k^{(1)})^{-1/2}(\lambda_i^{(2)} - \lambda_k^{(2)})e_k^\top \mathbf{U}_1^\top \mathbf{A}e_i + O(\epsilon),$$

where the equality before last comes from (57), the last equality is due to $|e_k^\top \mathbf{A}^\top \mathbf{A}e_i| \leq 1$, since $\|A\| = 1$, and the constant in this derivation depends on $(\lambda_i^{(1)}\lambda_k^{(1)})^{-1/2}$. Thus, for the j-th eigenpair we are concerned with, we have $\mu_j = \frac{\lambda_j^{(2)}}{\lambda_j^{(1)}} + O(\epsilon^2)$, since $\langle \psi_j^{(1)}, \mathbf{B}\psi_j^{(1)}\rangle = O(\epsilon)$. For the eigenvector, when $k \neq i$, we have

$$\left|\frac{\langle \psi_k^{(1)}, \mathbf{B}\psi_i^{(1)}\rangle}{\lambda_i^{(2)}/\lambda_i^{(1)} - \lambda_k^{(2)}/\lambda_k^{(1)}}\right| \leq \frac{(\lambda_i^{(1)}\lambda_k^{(1)})^{-1/2}\left|\lambda_i^{(2)} - \lambda_k^{(2)}\right|}{\left|\lambda_i^{(2)}/\lambda_i^{(1)} - \lambda_k^{(2)}/\lambda_k^{(1)}\right|}\left|e_k^\top \mathbf{U}_1^\top \mathbf{A}e_i\right| + O(\epsilon)$$
$$\leq \frac{1}{\gamma_i}\frac{\left|(\lambda_i^{(2)} - \lambda_k^{(2)})\right|}{\sqrt{\lambda_i^{(1)}\lambda_k^{(1)}}}\left|e_k^\top \mathbf{U}_1^\top \mathbf{A}e_i\right| + O(\epsilon),$$

where the constant depends on $\frac{1}{\sqrt{\lambda_i^{(1)}\lambda_k^{(1)}}\gamma_i}$.

By (56) we have for $j \neq i$:

$$\alpha_{ji} = \langle \psi_j^{(1)}, v_i\rangle = \langle \psi_j^{(1)}, \psi_i^{(1)} + \epsilon\sum_{k\neq i}\frac{\langle \psi_k^{(1)}, \mathbf{B}\psi_i^{(1)}\rangle}{\lambda_i^{(2)}/\lambda_i^{(1)} - \lambda_k^{(2)}/\lambda_k^{(1)}}\psi_k^{(1)} + O(\epsilon^2)\rangle$$
$$= \epsilon\sum_{k\neq i}\frac{\langle \psi_k^{(1)}, \mathbf{B}\psi_i^{(1)}\rangle}{\lambda_i^{(2)}/\lambda_i^{(1)} - \lambda_k^{(2)}/\lambda_k^{(1)}}\langle \psi_j^{(1)}, \psi_k^{(1)}\rangle + O(\epsilon^2)$$
$$= \epsilon\frac{\langle \psi_j^{(1)}, \mathbf{B}\psi_i^{(1)}\rangle}{\lambda_i^{(2)}/\lambda_i^{(1)} - \lambda_j^{(2)}/\lambda_j^{(1)}} + O(\epsilon^2).$$

Combining this with the inequality above, we have for $j \neq i$:

$$\left|\alpha_{ji}\right| \leq \epsilon\frac{1}{\gamma_i}\frac{\left|(\lambda_i^{(2)} - \lambda_j^{(2)})\right|}{\sqrt{\lambda_i^{(1)}\lambda_j^{(1)}}}\left|e_j^\top \mathbf{U}_1^\top \mathbf{A}e_i\right| + O(\epsilon^2), \tag{58}$$

where the constant depends on $\frac{1}{\sqrt{\lambda_i^{(1)}\lambda_j^{(1)}}\gamma_i}$.

We thus have for (52):

$$\frac{\lambda_j^{(2)}}{4}\sum_{i=1}^N \alpha_{ji}^2\left(\log\mu_i - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\right)^2$$
$$= \frac{\lambda_j^{(2)}}{4}\alpha_{jj}^2\left(\log\mu_j - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\right)^2 + \frac{\lambda_j^{(2)}}{4}\sum_{i\neq j}\alpha_{ji}^2\left(\log\mu_i - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\right)^2 = O(\epsilon^2), \tag{59}$$

where the constant depends on $\frac{c(\ln c)^2}{\min_i\{\gamma_i^2\lambda_i^{(1)}\}}$.

This is because the first term is $O(\epsilon^2)$ by

$$\alpha_{jj}^2\left(\log\mu_j - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\right)^2 = \alpha_{jj}^2\left(\epsilon\langle \psi_j^{(1)}, \mathbf{B}\psi_j^{(1)}\rangle\frac{\lambda_j^{(1)}}{\lambda_j^{(2)}}\right)^2 + O(\epsilon^2) \tag{60}$$
$$= \epsilon^2\alpha_{jj}^2\left(\frac{\lambda_j^{(1)}}{\lambda_j^{(2)}}\right)^2(\lambda_j^{(1)}\lambda_j^{(1)})^{-1}\left((\lambda_j^{(2)} - \lambda_j^{(2)})e_j^\top \mathbf{U}_1^\top \mathbf{A}e_j + O(\epsilon)\right)^2 + O(\epsilon^2) \tag{61}$$

using Taylor expansion $\log(x + h) = \log(x) + h/x + O(h^2)$ for $x > 0$ and sufficiently small h. In addition, noting that $\left(\log\mu_i - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\right)^2 \leq 5(\ln c)^2$ we have that the second term is bounded by:

$$\frac{\lambda_j^{(2)}}{4}\sum_{i\neq j}\alpha_{ji}^2\left(\log\mu_i - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\right)^2 \leq \frac{\lambda_j^{(2)}}{4}5(\ln c)^2\sum_{i\neq j}\alpha_{ji}^2$$

$$\leq \frac{5\lambda_j^{(2)}}{4}(\ln c)^2 \sum_{i\neq j} \epsilon^2 \frac{1}{\gamma_i^2} \frac{(\lambda_i^{(2)}-\lambda_j^{(2)})^2}{\lambda_i^{(1)}\lambda_j^{(1)}} \left|e_j^\top \mathbf{U}_1^\top \mathbf{A} e_i\right|^2$$

$$\leq \frac{5}{4}\epsilon^2(\ln c)^2 \frac{\lambda_j^{(2)}}{\lambda_j^{(1)}} \sum_{i\neq j} \frac{1}{\gamma_i^2} \frac{(\lambda_i^{(2)}-\lambda_j^{(2)})^2}{\lambda_i^{(1)}} \left|e_j^\top \mathbf{U}_1^\top \mathbf{A} e_i\right|^2 . \tag{62}$$

Continuing with a few coarse steps:

$$\frac{\lambda_j^{(2)}}{4} \sum_{i\neq j} \alpha_{ji}^2 \left(\log \mu_i - \log\left(\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}\right)\right)^2 \leq \frac{5}{4}\epsilon^2 \frac{c(\ln c)^2}{\min_i\{\gamma_i^2\lambda_i^{(1)}\}} \sum_{i\neq j} \left|e_j^\top \mathbf{U}_1^\top \mathbf{A} e_i\right|^2$$

$$\leq \frac{5}{2}\epsilon^2 \frac{c(\ln c)^2}{\min_i\{\gamma_i^2\lambda_i^{(1)}\}} \tag{63}$$

where $\left|\lambda_i^{(2)}-\lambda_j^{(2)}\right|^2 \leq 1$ due to the normalization of $\mathbf{W}_1$ and $\mathbf{W}_2$ and $\sum_{i\neq j}\left|e_j^\top \mathbf{U}_1^\top \mathbf{A} e_i\right|^2 \leq 2\|\mathbf{A}e_j\|^2 \leq 2$ due to $\|\mathbf{A}\|=1$.

Similarly, it can be shown for (54) that

$$\sum_{i=1}^{N} \alpha_{ji}^2 (\log \mu_i)^2 \left(\sqrt{\mu_i} - \sqrt{\frac{\lambda_j^{(2)}}{\lambda_j^{(1)}}}\right)^2 = O(\epsilon^2),$$

and the proof is concluded. □

Remark 1. Note that the implied constant $\frac{\sqrt{c}\ln c}{\min_i\left(\gamma_i\sqrt{\lambda_i^{(1)}}\right)}$ might be large, narrowing the scope of Theorem 4. Particularly in our context, the matrix $\mathbf{W}_1$ (and $\mathbf{W}_2$) tends to be close to low rank, for which $\min_i\left(\gamma_i\sqrt{\lambda_i^{(1)}}\right)$ is small.

Remark 2. Empirically, we observe that $\psi_j^{(1)} = \sum_i \alpha_{ji} v_i \simeq \sum_{i\sim j} \alpha_{ji} v_i$, i.e., only a small number of expansion coefficients α_{ji} are non-negligible, for which $\lambda_i^{(1)}$ is close to $\lambda_j^{(1)}$. Therefore, in practice, the implied constant depends on $1/\min_{i\sim j}\left(\gamma_i\sqrt{\lambda_i^{(1)}}\right)$. Since we are usually interested in principal components $\psi_j^{(1)}$ (i.e., with large $\lambda_j^{(1)}$), the implied constant is typically sufficiently large.

7. Conclusions

In this work, we introduce a new multi-resolution analysis of temporal high-dimensional data with an underlying time-varying manifold structure. Our analysis is based on the definition of two composite operators that represent the relation of two aligned datasets jointly sampled from two diffeomorphic manifolds in terms of their spectral components. Specifically, we showed that these operators not only recover but also distinguish different types of common spectral components of the underlying manifolds and that each operator emphasizes different properties. One operator was shown to emphasize common components that are similarly expressed in the two manifolds, and the other operator was shown to emphasize the common components that are expressed with significantly different eigenvalues. In the context of spatiotemporal data analysis, the application of the operators is analogous to low-pass and high-pass filters. Therefore, by applying them in multiple resolutions, we devise a wavelet-like analysis framework. We demonstrated this framework on a dynamical system describing a transitory double-gyre flow, showing that such a framework can be used for the analysis of non-stationary multivariate time-series.

In addition to spatiotemporal analysis, we showed that the defined composite operators may be useful for multimodal data analysis as well. Specifically, we showed application to remote sensing, demonstrating the recovery of meaningful properties expressed by different sensing modalities.

In the future, we plan to extend the definition of the operators **S** and **F** from two to more time frames (datasets). In addition, since our analysis results in a large number of vectors representing the common components at different scales and time-points, we plan to develop compact representations of these components, which may lead to improved, more conclusive results for highly non-stationary time-series.

Finally, we remark that in our model, we represent each sample by an undirected weighted graph, and then, analyze the temporal sequence of graphs. Another interesting future work would be to investigate our Riemannian composite operators in the context of graph neural networks (GNNs) and graph convolutional networks (GCNs) [13,44,66].

Data availability

Data will be made available on request.

Acknowledgments

We would like to thank the Associate Editor and the anonymous reviewers for their helpful comments and suggestions. The work of T.S. and R.T. was supported by the European Union's Horizon2020 research and innovation programme under grant agreement No. 802735-ERC-DIFFOP. R.T. also acknowledges the support of the Schmidt Career Advancement Chair.

Appendix A. Additional statements

The following equivalent forms of operators $\mathbf{S}$ and $\mathbf{F}$ are used in proofs of the theorems.

Proposition 1 *(Equivalent Forms of the Operators S and F). We have*

$$\mathbf{S} = \mathbf{W}_1^{1/2}\left(\mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2}\right)^{1/2}\mathbf{W}_1^{1/2} = \left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)^{1/2}\mathbf{W}_1$$

and

$$\mathbf{F} = \mathbf{S}^{1/2}\log\left(\mathbf{S}^{-1/2}\mathbf{W}_1\mathbf{S}^{-1/2}\right)\mathbf{S}^{1/2} = \log\left(\mathbf{W}_1\mathbf{S}^{-1}\right)\mathbf{S}.$$

Proof. To show the claim for $\mathbf{S}$, define the following:

$$\mathbf{M} = \mathbf{W}_1^{-1/2}\mathbf{W}_2\mathbf{W}_1^{-1/2} \quad \text{and} \quad \tilde{\mathbf{M}} = \mathbf{W}_2\mathbf{W}_1^{-1}. \tag{64}$$

Since $\mathbf{W}_1$ and $\mathbf{W}_2$ are positive definite, the matrix $\mathbf{M}$ is positive definite, and hence $\mathbf{M}$ and $\tilde{\mathbf{M}}$ are similar via $\mathbf{M} = \mathbf{W}_1^{-1/2}\tilde{\mathbf{M}}\mathbf{W}_1^{1/2}$. Denote the eigenvalue and eigenvector matrices of $\mathbf{M}$ by $\mathbf{\Lambda}^{(\mathbf{M})}$ and $\mathbf{V}^{(\mathbf{M})}$ respectively. Therefore, the eigenvalue matrix of $\tilde{\mathbf{M}}$ is $\mathbf{\Lambda}^{(\tilde{\mathbf{M}})} = \mathbf{\Lambda}^{(\mathbf{M})}$, and the right and left eigenvectors are $\mathbf{V}_R^{(\tilde{\mathbf{M}})} = \mathbf{W}_1^{1/2}\mathbf{V}^{(\mathbf{M})}$ and $\mathbf{V}_L^{(\tilde{\mathbf{M}})} = \mathbf{W}_1^{-1/2}\mathbf{V}^{(\mathbf{M})}$ respectively. Thus, we have

$$\begin{aligned}\mathbf{M}^{1/2} &= \mathbf{V}^{(\mathbf{M})}\left(\mathbf{\Lambda}^{(\mathbf{M})}\right)^{1/2}\left(\mathbf{V}^{(\mathbf{M})}\right)^T \\ \tilde{\mathbf{M}}^{1/2} &= \mathbf{W}_1^{1/2}\mathbf{V}^{(\mathbf{M})}\left(\mathbf{\Lambda}^{(\mathbf{M})}\right)^{1/2}\left(\mathbf{V}^{(\mathbf{M})}\right)^T\mathbf{W}_1^{-1/2}\end{aligned} \tag{65}$$

and hence $\tilde{\mathbf{M}}^{1/2} = \mathbf{W}_1^{1/2}\mathbf{M}^{1/2}\mathbf{W}_1^{-1/2}$. As a result, we have

$$\mathbf{S} = \mathbf{W}_1^{1/2}\mathbf{M}^{1/2}\mathbf{W}_1^{1/2} = \mathbf{W}_1^{1/2}\mathbf{M}^{1/2}\mathbf{W}_1^{-1/2}\mathbf{W}_1 = \tilde{\mathbf{M}}^{1/2}\mathbf{W}_1,$$

which shows the claim.

The proof for the claim for the operator $\mathbf{F}$ is similar. Define

$$\mathbf{N} := \mathbf{S}^{-1/2}\mathbf{W}_1\mathbf{S}^{-1/2} \quad \text{and} \quad \tilde{\mathbf{N}} := \mathbf{W}_1\mathbf{S}^{-1}. \tag{66}$$

Since $\mathbf{S}$ is positive definite, $\mathbf{N}$ and $\tilde{\mathbf{N}}$ are similar and their eigenvalues and eigenvectors are related by $\mathbf{\Lambda}^{(\tilde{\mathbf{N}})} = \mathbf{\Lambda}^{(\mathbf{N})}$, $\mathbf{V}_R^{(\tilde{\mathbf{N}})} = \mathbf{S}^{1/2}\mathbf{V}^{(\mathbf{N})}$ and $\mathbf{V}_L^{(\tilde{\mathbf{N}})} = \mathbf{S}^{-1/2}\mathbf{V}^{(\mathbf{N})}$. The matrix logarithm of $\mathbf{N}$ and $\tilde{\mathbf{N}}$ can then be expressed by:

$$\begin{aligned}\log(\mathbf{N}) &= \mathbf{V}^{(\mathbf{N})}\log\left(\mathbf{\Lambda}^{(\mathbf{N})}\right)\left(\mathbf{V}^{(\mathbf{N})}\right)^T \\ \log(\tilde{\mathbf{N}}) &= \mathbf{S}^{1/2}\mathbf{V}^{(\mathbf{N})}\log\left(\mathbf{\Lambda}^{(\mathbf{N})}\right)\left(\mathbf{V}^{(\mathbf{N})}\right)^T\mathbf{S}^{-1/2}.\end{aligned} \tag{67}$$

Based on these expressions the relationship between the logarithm of the two matrices is $\log(\tilde{\mathbf{N}}) = \mathbf{S}^{1/2}\log(\mathbf{N})\mathbf{S}^{-1/2}$. The proof can now be concluded by

$$\mathbf{F} = \mathbf{S}^{1/2}\log(\mathbf{N})\mathbf{S}^{1/2} = \mathbf{S}^{1/2}\log(\mathbf{N})\mathbf{S}^{-1/2}\mathbf{S} = \log(\tilde{\mathbf{N}})\mathbf{S}. \quad \square$$

Next, for completeness, we explicitly show that the equivalence between definitions 2 and 3 of the pseudo-spectrum in Definition 1 includes a shared pseudo-eigenvector.

Proposition 2. *Consider $\mathbf{M} \in \mathbb{R}^{N\times N}$ and a small $\epsilon > 0$. If $v \in \mathbb{R}^N$ with $\|v\|_2 = 1$ s.t. $\|(\mathbf{M} - \lambda\mathbf{I})v\|_2 \le \epsilon$ for $\lambda \in \mathbb{R}$, then there exists $\mathbf{E} \in \mathbb{R}^{N\times N}$ with $\|\mathbf{E}\| \le \epsilon$ s.t. $(\mathbf{M} + \mathbf{E})v = \lambda v$.*

Proof. Define the following rank one operator:

$$\mathbf{B}u = -\langle u, v\rangle(\mathbf{M} - \lambda\mathbf{I})v,$$

where $u \in \mathbf{R}^N$. Then, we have that $\|\mathbf{B}\| \le \epsilon$ and $(\mathbf{M} + \mathbf{B})v = \lambda v$. $\square$

A.1. Additional motivation for the operator **F**

We provide additional motivation for the operator $\mathbf{F}$ by showing that it captures differences between the diffusion operators $\mathbf{W}_1$ and $\mathbf{W}_2$, through small perturbations of these matrices.

Proposition 3. *Assume* $\mathbf{W}_2 = \mathbf{W}_1 + \mathbf{E}$ *s.t.* $\|\mathbf{E}\mathbf{W}_1^{-1}\| < 1$, *then* $\mathbf{F} \approx -\frac{1}{2}\left(\mathbf{W}_2 - \mathbf{W}_1\right)\left(\mathbf{W}_1^{-1}\mathbf{W}_2\right)^{1/2}$.

Proof.

$$\begin{aligned}
\mathbf{F} &= \log\left(\mathbf{W}_1\mathbf{S}^{-1}\right)\mathbf{S} \\
&= \log\left(\mathbf{W}_1\left((\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2}\mathbf{W}_1\right)^{-1}\right)\left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)^{1/2}\mathbf{W}_1 \\
&= -\tfrac{1}{2}\log\left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)\left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)^{1/2}\mathbf{W}_1 \\
&= -\tfrac{1}{2}\log\left(\mathbf{I} + \mathbf{E}\mathbf{W}_1^{-1}\right)\left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)^{1/2}\mathbf{W}_1 \\
&\approx -\tfrac{1}{2}\mathbf{E}\mathbf{W}_1^{-1}\left(\mathbf{W}_2\mathbf{W}_1^{-1}\right)^{1/2}\mathbf{W}_1 \\
&= -\tfrac{1}{2}\mathbf{E}\left(\mathbf{W}_1^{-1}\mathbf{W}_2\right)^{1/2} \\
&= -\tfrac{1}{2}\left(\mathbf{W}_2 - \mathbf{W}_1\right)\left(\mathbf{W}_1^{-1}\mathbf{W}_2\right)^{1/2}
\end{aligned} \tag{68}$$

where the approximation is based on a power series of the matrix logarithm, and the transition before last is due to $\mathbf{W}_2\mathbf{W}_1^{-1}$ being similar to a symmetric matrix, and therefore, diagonalizable, and due to $\mathbf{X}^{-1}\mathbf{A}^{1/2}\mathbf{X} = \left(\mathbf{X}^{-1}\mathbf{A}\mathbf{X}\right)^{1/2}$ for a diagonalizable matrix $\mathbf{A}$ and an invertible matrix $\mathbf{X}$. □

This result shows that the operator $\mathbf{F}$ is related to the differences between the diffusion operators, factored by a term related to the Riemannian metric of the space of SPD matrices. More intuitively, we remark that mapping a point from the manifold to the tangent space (via the logarithmic map) is the Riemannian counterpart of subtraction in a linear space.

Note that a similar conclusion can be obtained by noting that $\mathbf{E}\mathbf{W}_1^{-1}$ and $(\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2}$ commute. This leads to $\mathbf{F} \approx -\frac{1}{2}(\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2}(\mathbf{W}_2 - \mathbf{W}_1)$. This result is equivalent to the equation in (68), since $\mathbf{F}$, $\mathbf{W}_1$ and $\mathbf{W}_2$ are symmetric matrices, and therefore:

$$\begin{aligned}
\mathbf{F} = \mathbf{F}^T &\approx \left(-\frac{1}{2}(\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2}(\mathbf{W}_2 - \mathbf{W}_1)\right)^T \\
&= -\frac{1}{2}(\mathbf{W}_2 - \mathbf{W}_1)\left((\mathbf{W}_2\mathbf{W}_1^{-1})^{1/2}\right)^T \\
&= -\frac{1}{2}(\mathbf{W}_2 - \mathbf{W}_1)(\mathbf{W}_1^{-1}\mathbf{W}_2)^{1/2}
\end{aligned}$$

A follow up paper, [17], demonstrated additional uses for this difference operator, where $\mathbf{F}$ was applied to feature selection tasks.

Appendix B. 3-dimensional tori example

In this subsection we consider datasets of samples from two 3-dimensional tori. Using these datasets we demonstrate that the operators $\mathbf{S}$ and $\mathbf{F}$ indeed recover the similarly expressed common components and the differently expressed common components, respectively. In addition, in Subsection B.2, we demonstrate that these operators still recover the common components, even when modality-specific unique structures exist.

B.1. With common components only

Consider two 3-dimensional tori in two observation spaces, denoted by $\mathcal{O}_1$ and $\mathcal{O}_2$. Both tori are obtained by sampling the product of three $\mathcal{S}^1$ manifolds that differ in scaling, which are embedded into two ambient spaces:

$$\mathcal{O}_1 = f_1^{(1)}\left(\mathcal{S}^1\right) \times f_1^{(2)}\left(\mathcal{S}^1\right) \times \mathcal{S}^1 \mapsto \mathcal{X}_1 \tag{69}$$

$$\mathcal{O}_2 = f_2^{(1)}\left(\mathcal{S}^1\right) \times f_2^{(2)}\left(\mathcal{S}^1\right) \times \mathcal{S}^1 \mapsto \mathcal{X}_2 \tag{70}$$

where $f_\ell^{(k)} : \mathcal{S}^1 \to \mathcal{S}^1$, $\nabla f_\ell^{(k)}|_x = \alpha_\ell^{(k)}\ \forall x \in \mathcal{S}^1$ and $\alpha_1^{(1)} = \alpha_2^{(2)}$, $\alpha_1^{(2)} = \alpha_2^{(1)}$. We assume that the embedding into the ambient space preserves the order of the $\mathcal{S}^1$ manifolds and therefore, these scale differences can be represented by switching two of the main angles in the parameterization of the two tori. The samples in the ambient spaces can be explicitly described by the following embedding in 4D:

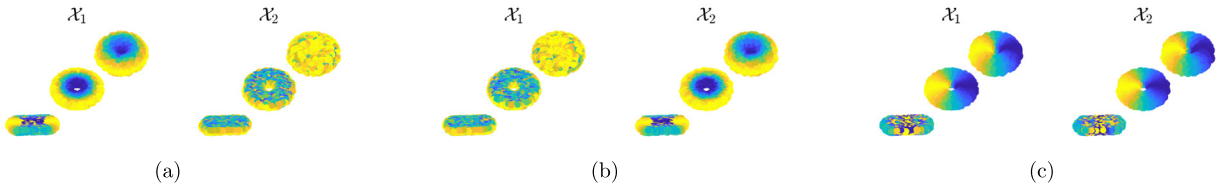


Fig. 10. Two 3D tori with angles θ_1, θ_2, θ_3. Angles θ_1 and θ_2 are switched between the tori and angles θ_3 is similar. Both tori are colored according to (a) $\cos(\theta_1)$, (b) $\cos(\theta_2)$ and (c) $\cos(\theta_3)$.

$$\mathcal{X}_1 \ni \mathbf{x}_1[i] = \begin{Bmatrix} x_\ell[i] &=& (\tilde{R} + (R + r\cos(\theta_2[i]))\cos(\theta_1[i]))\cos(\theta_3[i]) \\ y_\ell[i] &=& (\tilde{R} + (R + r\cos(\theta_2[i]))\cos(\theta_1[i]))\sin(\theta_3[i]) \\ z_\ell[i] &=& (R + r\cos(\theta_2[i]))\sin(\theta_1[i]) \\ w_\ell[i] &=& r\sin(\theta_2[i]) \end{Bmatrix} \tag{71}$$

$$\mathcal{X}_2 \ni \mathbf{x}_2[i] = \begin{Bmatrix} x_\ell[i] &=& (\tilde{R} + (R + r\cos(\theta_1[i]))\cos(\theta_2[i]))\cos(\theta_3[i]) \\ y_\ell[i] &=& (\tilde{R} + (R + r\cos(\theta_1[i]))\cos(\theta_2[i]))\sin(\theta_3[i]) \\ z_\ell[i] &=& (R + r\cos(\theta_1[i]))\sin(\theta_2[i]) \\ w_\ell[i] &=& r\sin(\theta_1[i]) \end{Bmatrix} \tag{72}$$

where $\theta_1[i], \theta_2[i], \theta_3[i] \in [0, 2\pi]$, $r = 2$, $R = 7$ and $\tilde{R} = 15$. In this setting, the radii R and r are related to the scale parameters $\alpha_1^{(1)} = \alpha_2^{(2)}$ and $\alpha_1^{(2)} = \alpha_2^{(1)}$ that define the diffeomorphisms $f_\ell^{(k)}$ in (69) and (70).

We sample $N = 2000$ points from each torus, $\{\mathbf{x}_\ell[i]\}_{i=1}^N$, $\mathbf{x}_\ell[i] \in \mathbb{R}^4$, $\ell = 1, 2$, with point correspondence between the two tori, which is obtained through correspondence of the samples of θ_1, θ_2 and θ_3.

The data is visualized in Fig. 10 where projections of the two tori are colored according to the 3 different angles, $\cos(\theta_1)$ in (a), $\cos(\theta_2)$ in (b) and $\cos(\theta_3)$ in (c). This figure presents projections of the tori to the following 3-dimensional spaces: 'xyz', 'xyw' and 'yzw'. In this example, the two modalities, represented by $\mathcal{X}_1$ and $\mathcal{X}_2$, includes only common components, some of which are differently expressed and some are similarly expressed. Specifically, θ_3 is related to the similarly expressed common components, as depicted by Fig. 10 (c), and θ_1 and θ_2 are related to the differently expressed common components, as depicted by Fig. 10 (a) and (b).

We apply Algorithm 3 to the samples of the two 3-dimensional tori with $\sigma_\ell = \text{median}(d(\mathbf{x}_\ell[i], \mathbf{x}_\ell[j]))$, $i, j = 1, ..., N$, $\ell = 1, 2$, and compute the eigenvectors of the kernels $\mathbf{W}_1$, $\mathbf{W}_2$, as well as the operators $\mathbf{S}$ and $\mathbf{F}$ and their eigenvectors. These eigenvectors compose the embedding of each operator. The eigenvectors are denoted in the following figures by $\psi_n^{(1)}$, $\psi_n^{(2)}$, $\psi_n^{(\mathbf{S})}$ and $\psi_n^{(\mathbf{F})}$, respectively, and are ordered according to a decreasing magnitude of their corresponding eigenvalues, denoted by $\lambda_n^{(1)}$, $\lambda_n^{(2)}$, $\lambda_n^{(\mathbf{S})}$ and $\lambda_n^{(\mathbf{F})}$, respectively.

To demonstrate that the operator $\mathbf{S}$ indeed emphasizes the similarly expressed common components and that $\mathbf{F}$ emphasizes the differently expressed common components, we compare their eigenvectors with the eigenvectors of $\mathbf{W}_1$ and $\mathbf{W}_2$ and with the angles θ_1, θ_2 and θ_3 parameterizing the tori. Note that the sign of the eigenvalues of $\mathbf{F}$ is meaningful and provides information on the source of the dominant difference components, as a demonstrated by Theorem 2. Therefore, in order to clearly distinguish between the largest positive and smallest negative eigenvalues, an eigenvector of $\mathbf{F}$ that correspond to the kth largest (positive) eigenvalue is denoted by $\psi_k^{(\mathbf{F})}$ and an eigenvector that corresponds to the ℓth smallest negative eigenvalue is denoted by $\psi_{-\ell \bmod N}^{(\mathbf{A})}$, throughout this section.

Fig. 11 presents the two tori, top and bottom in Fig. 11 (a) and in Fig. 11 (b), colored according to these eigenvectors and the torus angles. The vector that each torus was colored by is denoted in the title of each plot. In Fig. 11 (a), the tori are colored according to the two leading eigenvectors of $\mathbf{S}$, the leading eigenvector of $\mathbf{W}_1$ (top) and $\mathbf{W}_2$ (bottom) and the cosine of θ_3, from left to right. In Fig. 11 (b), the tori are colored according to two leading eigenvectors of $\mathbf{F}$ (corresponding to positive and negative eigenvalues), eigenvector number 7 of $\mathbf{W}_1$ (top) and $\mathbf{W}_2$ (bottom), the cosine of θ_1 and the cosine of θ_2, from left to right.

Fig. 11 (a) depicts that the operator $\mathbf{S}$ captures the similarly expressed common component, i.e. θ_3, the angle that is related to the $\mathcal{S}^1$ manifold that does not undergo any transformation between the two observation spaces. This is indicated by similarity of the torus colors (up to some rotation), when comparing the coloring according to the eigenvectors of $\mathbf{S}$ with the coloring according to $\cos(\theta_3)$. In addition, the torus colors are also very similar when colored according to the second eigenvectors of $\mathbf{W}_1$ and $\mathbf{W}_2$. This implies that θ_3 is highly expressed by both kernels $\mathbf{W}_1$ and $\mathbf{W}_2$, meaning that it is a dominant component in both tori.

Note that the first eigenvectors of $\mathbf{W}_1$ and $\mathbf{W}_2$ (and in this example also of $\mathbf{S}$) are related to the point density on each torus, which is of less interest in this example, and were omitted therefor.

In contrast, Fig. 11 (b) depicts that the operator $\mathbf{F}$ captures the differently expressed common components, i.e. θ_1 and θ_2, the angles that are related to $\mathcal{S}^1$ manifolds that undergo transformations between the two observation spaces. This is indicated by the similarity in the torus colors, when comparing the coloring according to the eigenvectors of $\mathbf{F}$ with the coloring according to $\cos(\theta_1)$ and $\cos(\theta_2)$. Note that this figure also nicely demonstrates the connection between the sign of the eigenvalues of $\mathbf{F}$ and the source of the dominant component. For example, the angle θ_1 is a more dominant component in torus $\mathcal{X}_1$, i.e. its corresponding $\mathcal{S}^1$ manifold has a larger radius compared with $\mathcal{X}_2$, and this angle is captured by an eigenvector of $\mathbf{F}$ that corresponds to a positive eigenvalue. In

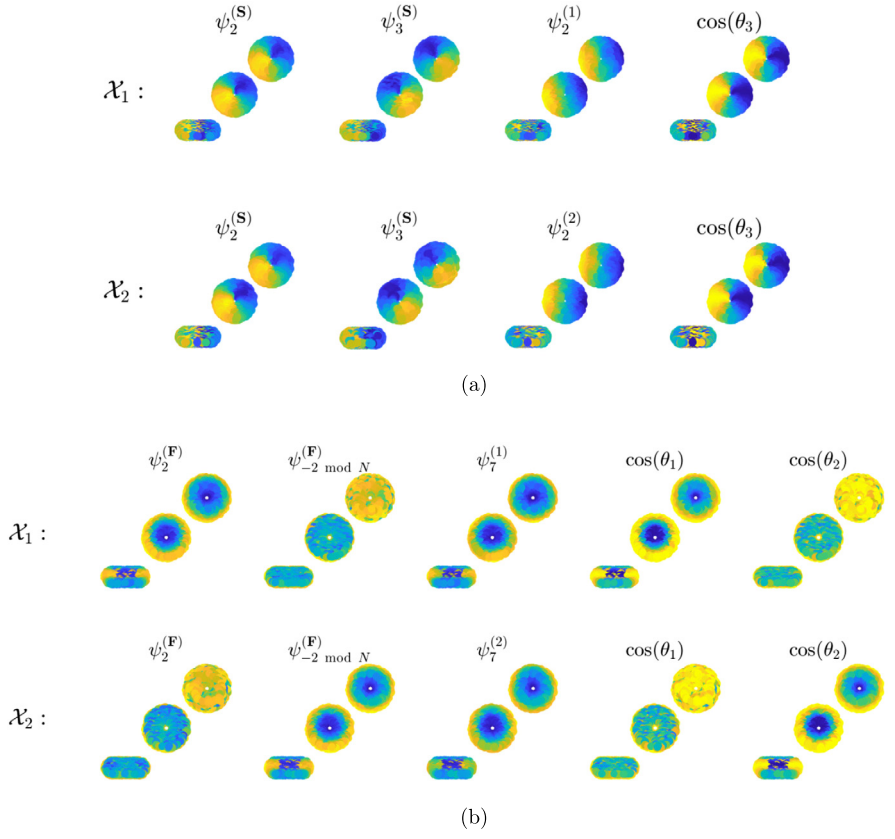


Fig. 11. The two tori colored according to eigenvectors of the operators and kernels and according to the torus angles. (a) Eigenvectors and angle that are captured by the operator **S**, (b) eigenvectors and angles that are captured by the operator **F**.

contrast, the eigenvector that corresponds to a negative eigenvalue of **F** captures the angle θ_2, which is a more dominant component in torus $\mathcal{X}_2$.

Note that in kernels $\mathbf{W}_1$ and $\mathbf{W}_2$ the angles θ_1 and θ_2 are expressed only in less dominant eigenvectors (corresponding to much smaller eigenvalues). Specifically, θ_1 and θ_2 are first captured by $\mathbf{W}_1$ only in eigenvector number 6 (and 7 as presented in the figure) and eigenvector number 17, respectively, and their location in $\mathbf{W}_2$ is similar but reversed. Therefore, this demonstrates that the operator **F** significantly emphasized these differently expressed common components. We note in addition, that the eigenvectors of **F** that correspond to the first (largest) positive and negative eigenvalues were highly related to θ_1 and θ_2 as well, specifically, they captured $\sin(\theta_1)$ and $\sin(\theta_2)$, and were omitted for brevity.

B.2. With unique components

Consider now a slightly different setting of the two 3D tori, in which the third S^1 manifold undergoes two different diffeomorphisms, which lead to a unique structure in each observation space. The formulation for this setting is given by:

$$\mathcal{O}_1 = f_1^{(1)}\left(S^1\right) \times f_1^{(2)}\left(S^1\right) \times \mathcal{N}_1 \mapsto \mathcal{X}_1 \tag{73}$$

$$\mathcal{O}_2 = f_2^{(1)}\left(S^1\right) \times f_2^{(2)}\left(S^1\right) \times \mathcal{N}_2 \mapsto \mathcal{X}_2 \tag{74}$$

where $\mathcal{N}_1$ and $\mathcal{N}_2$ denote two different unique structures, $f_\ell^{(k)} : S^1 \to S^1$, $\nabla f_\ell^{(k)}|_x = \alpha_\ell^{(k)}\ \forall x \in S^1$ and $\alpha_1^{(1)} = \alpha_2^{(2)}$, $\alpha_1^{(2)} = \alpha_2^{(1)}$. We assume that the embedding into the ambient spaces express these unique structures as different permutations of the angel corresponding to this S^1 manifold. In addition, similarly to Subsection B.1, we assume that the scale differences between the first two manifolds can be represented by switching two of the main angles in the parameterization of the two tori. In this case, the samples in the ambient spaces can be explicitly described by the following embedding in 4D:

$$\mathcal{X}_1 \ni \mathbf{x}_1[i] = \left\{ \begin{array}{lcl} x_\ell[i] & = & (\tilde{R} + (R + r\cos(\theta_2[i]))\cos(\theta_1[i]))\cos(\theta_3[i]) \\ y_\ell[i] & = & (\tilde{R} + (R + r\cos(\theta_2[i]))\cos(\theta_1[i]))\sin(\theta_3[i]) \\ z_\ell[i] & = & (R + r\cos(\theta_2[i]))\sin(\theta_1[i]) \\ w_\ell[i] & = & r\sin(\theta_2[i]) \end{array} \right\} \tag{75}$$

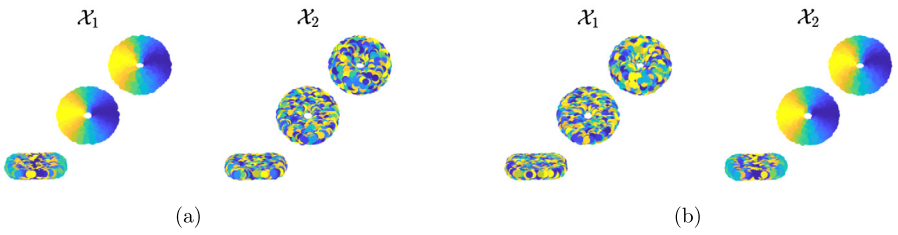


Fig. 12. Two 3D tori colored according to (a) $\cos(\theta_3)$ and (b) $\cos(\theta_4)$. Angles θ_3 and θ_4 are unique to each torus.

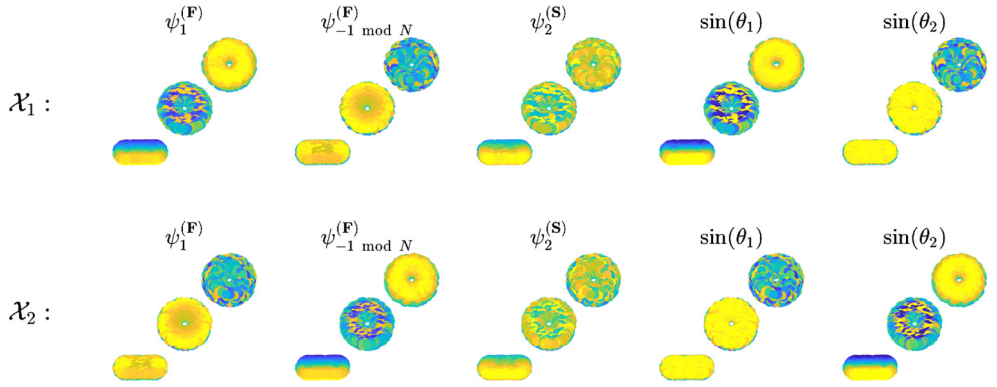


Fig. 13. The two tori (top and bottom) colored according to the eigenvectors of $\mathbf{F}$ and $\mathbf{S}$ and according to $\sin(\theta_1)$ and $\sin(\theta_2)$.

$$\mathcal{X}_2 \ni \mathbf{x}_2[i] = \left\{ \begin{array}{lcl} x_\ell[i] & = & (\tilde{R} + (R + r\cos(\theta_1[i]))\cos(\theta_2[i]))\cos(\theta_4[i]) \\ y_\ell[i] & = & (\tilde{R} + (R + r\cos(\theta_1[i]))\cos(\theta_2[i]))\sin(\theta_4[i]) \\ z_\ell[i] & = & (R + r\cos(\theta_1[i]))\sin(\theta_2[i]) \\ w_\ell[i] & = & r\sin(\theta_1[i]) \end{array} \right\} \tag{76}$$

where $\theta_1[i], \theta_2[i], \theta_3[i], \theta_4[i] \in [0, 2\pi]$, $r = 2$, $R = 7$ and $\tilde{R} = 15$. Note that the difference between the parameterization in this setting compared with the setting in Subsection B.1 is the angle θ_4, which differs from θ_3. These two angles are visualized in Fig. 12, which presents projections of the tori to the following 3-dimensional spaces: 'xyz', 'xyw' and 'yzw'. This figure demonstrate that θ_3 is unique to $\mathcal{X}_1$ and that θ_4 is unique to $\mathcal{X}_2$, since coloring $\mathcal{X}_2$ according to θ_3 and $\mathcal{X}_1$ according to θ_4 leads to random coloring, as presented in the right plot of Fig. 12 (a) and the left plot of Fig. 12 (b), respectively.

In summary, the two modalities, represented by $\mathcal{X}_1$ and $\mathcal{X}_2$, include common components that are differently expressed, which are related to θ_1 and θ_2 (similarly to the previous setting), and a unique component, which is related to θ_3 and θ_4.

We sample $N = 2000$ points from each torus similarly to Subsection B.1 and apply Algorithm 3 with $\sigma_\ell = \text{median}(d(\mathbf{x}_\ell[i], \mathbf{x}_\ell[j]))$, $i, j = 1, \ldots, N$, $\ell = 1, 2$. We then compute the eigenvectors of the operators $\mathbf{S}$ and $\mathbf{F}$.

Fig. 13 presents the two tori (top and bottom) colored according to the two leading eigenvectors of $\mathbf{F}$ (corresponding to positive and negative eigenvalues), the leading eigenvector of $\mathbf{S}$, the sine of θ_1 and the sine of θ_2, from left to right. Note that this figure presents the sine of the angles compared with a different eigenvector of $\mathbf{F}$ as a complementary example to Fig. 11, since the two leading eigenvectors of $\mathbf{F}$ (with either negative or positive eigenvalues) capture the sine and cosine of the differently expressed angles.

Fig. 13 depicts that the operator $\mathbf{F}$ captures the differently expressed common components, i.e. θ_1 and θ_2. This is indicated by the similarity in the torus colors, when comparing the coloring according to the eigenvectors of $\mathbf{F}$ with the coloring according to $\sin(\theta_1)$ and $\sin(\theta_2)$. Note the connection between the sign of the eigenvalues of $\mathbf{F}$ and the source of the dominant component, as depicted by the two left-most and two right-most plots in this figure.

In addition, in this setting, the operator $\mathbf{S}$ recovers the same common components, θ_1 and θ_2, since there are no similarly expressed common components. However, the coloring of the middle plot in Fig. 13 depicts that the leading eigenvector of $\mathbf{S}$ is not identical to either one of the angles. This is due to rotations of eigenvector subspaces of $\mathbf{S}$, which stem from the symmetry of the two tori problem. Due to the symmetry, the angles θ_1 and θ_2 are expressed by the eigenvectors of both kernels, $\mathbf{W}_1$ and $\mathbf{W}_2$, but in a different complementary order. For example, if θ_1 appears in eigenvector number 6 of $\mathbf{W}_1$ and eigenvector number 18 of $\mathbf{W}_2$, then θ_2 will appear in eigenvector number 18 of $\mathbf{W}_1$ and eigenvector number 6 of $\mathbf{W}_2$, and with similar complementary eigenvalues. Since according to Theorem 1 the eigenvalues of $\mathbf{S}$ are given by $\sqrt{\lambda^{(1)}\lambda^{(2)}}$, where $\lambda^{(1)}$ and $\lambda^{(2)}$ correspond to common eigenvectors, the eigenvector of $\mathbf{S}$ that captures θ_1 and the eigenvector of $\mathbf{S}$ that captures θ_2 have the same eigenvalue, with a multiplicity of 2. This may lead to rotation of the eigenvectors in the eigenspace of eigenvalue $\sqrt{\lambda^{(1)}\lambda^{(2)}}$. As a result, the computed eigenvectors correspond

to a combination of θ_1 and θ_2, as depicted by the middle plot in Fig. 13. Moreover, note that this explains why the middle-top and middle-bottom plots appear to be colored similarly.

Finally, note that in this example, the leading eigenvectors of $\mathbf{W}_1$ and $\mathbf{W}_2$ capture angles θ_3 and θ_4, respectively, and that angles θ_1 and θ_2 first appear only in eigenvectors number 6 and 18 of $\mathbf{W}_1$, respectively, and similarly for $\mathbf{W}_2$ but in reversed order. This demonstrates that even in the presence of unique structures in each torus, the operators $\mathbf{S}$ and $\mathbf{F}$ successfully recovers the common components, and that these common components are significantly enhanced compared with the other unique components.

References

[1] V. Arsigny, P. Fillard, X. Pennec, N. Ayache, Geometric means in a novel vector space structure on symmetric positive-definite matrices, SIAM J. Matrix Anal. Appl. 29 (1) (2007) 328–347.
[2] R. Banisch, P. Koltai, Understanding the geometry of transport: diffusion maps for Lagrangian trajectory data unravel coherent sets, Chaos, Interdiscip. J. Nonlinear Sci. 27 (3) (2017) 035804.
[3] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Classification of covariance matrices using a Riemannian-based kernel for bci applications, Neurocomputing 112 (2013) 172–178.
[4] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (2003) 1373–1396.
[5] M. Belkin, P. Niyogi, Convergence of Laplacian eigenmaps, in: Advances in Neural Information Processing Systems, 2007, pp. 129–136.
[6] P. Bérard, G. Besson, S. Gallot, Embedding Riemannian manifolds by their heat kernel, Geom. Funct. Anal. 4 (4) (1994) 373–398.
[7] R. Bergmann, J.H. Fitschen, J. Persch, G. Steidl, Priors with coupled first and second order differences for manifold-valued image processing, J. Math. Imaging Vis. 60 (9) (2018) 1459–1481.
[8] T. Berry, D. Giannakis, J. Harlim, Nonparametric forecasting of low-dimensional dynamical systems, Phys. Rev. E 91 (3) (2015) 032915.
[9] T. Berry, J. Harlim, Variable bandwidth diffusion kernels, Appl. Comput. Harmon. Anal. 40 (1) (2016) 68–96.
[10] R. Bhatia, Positive Definite Matrices, vol. 24, Princeton University Press, 2009.
[11] R. Bhatia, T. Jain, Y. Lim, On the Bures–Wasserstein distance between positive definite matrices, Expo. Math. 37 (2) (2019) 165–191.
[12] S. Bonnabel, R. Sepulchre, Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank, SIAM J. Matrix Anal. Appl. 31 (3) (2010) 1055–1070.
[13] M.M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, P. Vandergheynst, Geometric deep learning: going beyond Euclidean data, IEEE Signal Process. Mag. 34 (4) (2017) 18–42.
[14] M. Budišić, R. Mohr, I. Mezić, Applied koopmanism, Chaos, Interdiscip. J. Nonlinear Sci. 22 (4) (2012) 047510.
[15] Z. Chebbi, M. Moakher, Means of Hermitian positive-definite matrices based on the log-determinant α-divergence function, Linear Algebra Appl. 436 (7) (2012) 1872–1889.
[16] X. Cheng, H.-T. Wu, Convergence of graph Laplacian with kNN self-tuned kernels, Inf. Inference 11 (3) (2022) 889–957.
[17] D. Cohen, T. Shnitzer, Y. Kluger, R. Talmon, Few-sample feature selection via feature manifold learning, in: Proceedings of the 40th International Conference on Machine Learning, PMLR, 2023, pp. 6296–6319.
[18] R. Coifman, S. Lafon, Diffusion maps, Appl. Comput. Harmon. Anal. 21 (Jul. 2006) 5–30.
[19] R. Coifman, M. Maggioni, Diffusion wavelets, Appl. Comput. Harmon. Anal. 21 (1) (2006) 53–94.
[20] R.R. Coifman, N.F. Marshall, S. Steinerberger, A common variable minimax theorem for graphs, Found. Comput. Math. (2022) 1–25.
[21] W. Czaja, B. Manning, L. McLean, J.M. Murphy, Fusion of aerial gamma-ray survey and remote sensing data for a deeper understanding of radionuclide fate after radiological incidents: examples from the Fukushima Dai-Ichi response, J. Radioanal. Nucl. Chem. 307 (3) (2016) 2397–2401.
[22] I. Daubechies, Ten Lectures on Wavelets, SIAM, 1992.
[23] V.R. De Sa, Spectral clustering with two views, in: ICML Workshop on Learning with Multiple Views, 2005, pp. 20–27.
[24] X. Ding, H.-T. Wu, How do kernel-based sensor fusion algorithms behave under high dimensional noise?, arXiv preprint, arXiv:2111.10940, 2021.
[25] X. Ding, H.-T. Wu, Impact of Signal-to-Noise Ratio and Bandwidth on Graph Laplacian Spectrum from High-Dimensional Noisy Point Cloud, 2021.
[26] X. Ding, H.-T. Wu, Impact of signal-to-noise ratio and bandwidth on graph Laplacian spectrum from high-dimensional noisy point cloud, IEEE Trans. Inf. Theory 69 (3) (2022) 1899–1931.
[27] C.J. Dsilva, R. Talmon, C.W. Gear, R.R. Coifman, I.G. Kevrekidis, Data-driven reduction for a class of multiscale fast-slow stochastic dynamical systems, SIAM J. Appl. Dyn. Syst. 15 (3) (2016) 1327–1351.
[28] D.B. Dunson, H.-T. Wu, N. Wu, Spectral convergence of graph Laplacian and heat kernel reconstruction in l^∞ from random samples, Appl. Comput. Harmon. Anal. 55 (2021) 282–336.
[29] N. El Karoui, On information plus noise kernel random matrices, Ann. Stat. 38 (5) (2010) 3191–3216.
[30] N. El Karoui, H.-T. Wu, Graph connection Laplacian methods can be made robust to noise, Ann. Stat. 44 (1) (2016) 346–372.
[31] D. Eynard, A. Kovnatsky, M.M. Bronstein, K. Glashoff, A.M. Bronstein, Multimodal manifold analysis by simultaneous diagonalization of Laplacians, IEEE Trans. Pattern Anal. Mach. Intell. 37 (12) (2015) 2505–2517.
[32] G. Froyland, Dynamic isoperimetry and the geometry of Lagrangian coherent structures, Nonlinearity 28 (10) (2015) 3587.
[33] G. Froyland, E. Kwok, A dynamic Laplacian for identifying Lagrangian coherent structures on weighted Riemannian manifolds, J. Nonlinear Sci. 30 (5) (2020) 1889–1971.
[34] G. Froyland, K. Padberg-Gehle, Almost-invariant and finite-time coherent sets: directionality, duration, and diffusion, in: Ergodic Theory, Open Dynamics, and Coherent Structures, Springer, 2014, pp. 171–216.
[35] G. Froyland, K. Padberg-Gehle, A rough-and-ready cluster-based approach for extracting finite-time coherent sets from sparse and incomplete trajectory data, Chaos, Interdiscip. J. Nonlinear Sci. 25 (8) (2015) 087406.
[36] D. Giannakis, Dynamics-adapted cone kernels, SIAM J. Appl. Dyn. Syst. 14 (2) (2015) 556–608.
[37] V. Gómez, Multivariate Time Series with Linear State Space Structure, Springer, 2016.
[38] D. Hammond, P. Vandergheynst, R. Gribonval, Wavelets on graphs via spectral graph theory, Appl. Comput. Harmon. Anal. 30 (2) (2011) 129–150.
[39] M.T. Harandi, M. Salzmann, R. Hartley, From manifold to manifold: geometry-aware dimensionality reduction for spd matrices, in: European Conference on Computer Vision, Springer, 2014, pp. 17–32.
[40] M. Hein, Uniform convergence of adaptive graph-based regularization, in: International Conference on Computational Learning Theory, Springer, 2006, pp. 50–64.
[41] P.W. Jones, M. Maggioni, R. Schul, Manifold parametrizations by eigenfunctions of the Laplacian and heat kernels, Proc. Natl. Acad. Sci. USA 105 (6) (2008) 1803–1808.
[42] O. Katz, R.R. Lederman, R. Talmon, Spectral flow on the manifold of spd matrices for multimodal data processing, arXiv preprint, arXiv:2009.08062, 2020.
[43] O. Katz, R. Talmon, Y.-L. Lo, H.-T. Wu, Alternating diffusion maps for multimodal data fusion, Inf. Fusion 45 (2019) 346–360.
[44] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, arXiv preprint, arXiv:1609.02907, 2016.

[45] J.N. Kutz, S.L. Brunton, B.W. Brunton, J.L. Proctor, Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems, SIAM, 2016.
[46] S. Lafon, Y. Keller, R.R. Coifman, Data fusion and multicue data matching by diffusion maps, IEEE Trans. Pattern Anal. Mach. Intell. 28 (11) (Nov. 2006) 1784–1797.
[47] R.R. Lederman, R. Talmon, Learning the geometry of common latent variables using alternating-diffusion, Appl. Comput. Harmon. Anal. (2015).
[48] R. Li, M.G. Frasch, H.-T. Wu, Efficient fetal-maternal ecg signal separation from two channel maternal abdominal ecg via diffusion-based channel selection, Front. Physiol. 8 (2017) 277.
[49] Z. Lin, Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition, SIAM J. Matrix Anal. Appl. 40 (4) (2019) 1353–1370.
[50] O. Lindenbaum, M. Salhov, A. Yeredor, A. Averbuch, Gaussian bandwidth selection for manifold learning and classification, Data Min. Knowl. Discov. 34 (2020) 1676–1712.
[51] O. Lindenbaum, A. Yeredor, M. Salhov, A. Averbuch, Multi-view diffusion maps, Inf. Fusion 55 (2020) 127–149.
[52] L. Malagò, L. Montrucchio, G. Pistone, Wasserstein Riemannian geometry of Gaussian densities, Inf. Geom. 1 (2) (2018) 137–179.
[53] S. Mallat, A Wavelet Tour of Signal Processing, Elsevier, 1999.
[54] N.F. Marshall, M.J. Hirn, Time coupled diffusion maps, Appl. Comput. Harmon. Anal. 45 (3) (2018) 709–728.
[55] T. Michaeli, W. Wang, K. Livescu, Nonparametric canonical correlation analysis, in: International Conference on Machine Learning, PMLR, 2016, pp. 1967–1976.
[56] M. Moakher, A differential geometric approach to the geometric mean of symmetric positive-definite matrices, SIAM J. Matrix Anal. Appl. 26 (3) (2005) 735–747.
[57] B.A. Mosovsky, J.D. Meiss, Transport in transitory dynamical systems, SIAM J. Appl. Dyn. Syst. 10 (1) (2011) 35–65.
[58] J.M. Murphy, M. Maggioni, Diffusion geometric methods for fusion of remotely sensed data, in: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIV, vol. 10644, International Society for Optics and Photonics, 2018, 106440I.
[59] A. Ortega, P. Frossard, J. Kovačević, J.M. Moura, P. Vandergheynst, Graph signal processing: overview, challenges, and applications, Proc. IEEE 106 (5) (2018) 808–828.
[60] X. Pennec, P. Fillard, N. Ayache, A Riemannian framework for tensor computing, Int. J. Comput. Vis. 66 (1) (2006) 41–66.
[61] M.H. Quang, M. San Biagio, V. Murino, Log-Hilbert-Schmidt metric between positive definite operators on Hilbert spaces, in: Advances in Neural Information Processing Systems, 2014, pp. 388–396.
[62] I. Ram, M. Elad, I. Cohen, Generalized tree-based wavelet transform, IEEE Trans. Signal Process. 59 (9) (2011) 4199–4209.
[63] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 260 (2000) 2323–2326.
[64] A. Sandryhaila, J.M. Moura, Discrete signal processing on graphs, IEEE Trans. Signal Process. 61 (7) (2013) 1644–1656.
[65] A. Sandryhaila, J.M. Moura, Discrete signal processing on graphs: frequency analysis, IEEE Trans. Signal Process. 62 (12) (2014) 3042–3054.
[66] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, IEEE Trans. Neural Netw. 20 (1) (2008) 61–80.
[67] P.J. Schmid, Dynamic mode decomposition of numerical and experimental data, J. Fluid Mech. 656 (2010) 5–28.
[68] T. Shnitzer, M. Ben-Chen, L. Guibas, R. Talmon, H.-T. Wu, Recovering hidden components in multimodal data with composite diffusion operators, SIAM J. Math. Data Sci. 1 (3) (2019) 588–616.
[69] T. Shnitzer, R. Talmon, J.-J. Slotine, Manifold learning with contracting observers for data-driven time-series analysis, IEEE Trans. Signal Process. 65 (4) (2016) 904–918.
[70] D.I. Shuman, S.K. Narang, P. Frossard, A. Ortega, P. Vandergheynst, The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains, IEEE Signal Process. Mag. 30 (3) (2013) 83–98.
[71] A. Singer, From graph to manifold Laplacian: the convergence rate, Appl. Comput. Harmon. Anal. 21 (1) (2006) 128–134.
[72] A. Singer, R.R. Coifman, Non-linear independent component analysis with diffusion maps, Appl. Comput. Harmon. Anal. 25 (2) (2008) 226–239.
[73] A. Singer, R. Erban, I.G. Kevrekidis, R.R. Coifman, Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps, Proc. Natl. Acad. Sci. USA 106 (38) (2009) 16090–16095.
[74] S. Sra, A new metric on the manifold of kernel matrices with application to matrix geometric means, Adv. Neural Inf. Process. Syst. 25 (2012) 144–152.
[75] A. Swatantran, R. Dubayah, D. Roberts, M. Hofton, J.B. Blair, Mapping biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion, Remote Sens. Environ. 115 (11) (2011) 2917–2930.
[76] R. Talmon, I. Cohen, S. Gannot, R.R. Coifman, Diffusion maps for signal processing: a deeper look at manifold-learning techniques based on kernels and graphs, IEEE Signal Process. Mag. 30 (4) (2013) 75–86.
[77] R. Talmon, R.R. Coifman, Empirical intrinsic geometry for nonlinear modeling and time series filtering, Proc. Natl. Acad. Sci. USA 110 (31) (2013) 12535–12540.
[78] R. Talmon, H.-T. Wu, Latent common manifold learning with alternating diffusion: analysis and applications, Appl. Comput. Harmon. Anal. 47 (3) (2019) 848–892.
[79] J.B. Tenenbaum, V. De Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
[80] L.N. Trefethen, M. Embree, Spectra and Pseudospectra: the Behavior of Nonnormal Matrices and Operators, Princeton University Press, 2005.
[81] R.S. Tsay, Multivariate Time Series Analysis: with R and Financial Applications, John Wiley & Sons, 2013.
[82] J.H. Tu, C.W. Rowley, D.M. Luchtenburg, S.L. Brunton, J.N. Kutz, On dynamic mode decomposition: theory and applications, J. Comput. Dyn. 1 (2) (2014) 391.
[83] O. Tuzel, F. Porikli, P. Meer, Pedestrian detection via classification on Riemannian manifolds, IEEE Trans. Pattern Anal. Mach. Intell. 30 (10) (2008) 1713–1727.
[84] H.-t. Wu, R. Talmon, Y.-L. Lo, Assess sleep stage by modern signal processing techniques, IEEE Trans. Biomed. Eng. 62 (4) (2014) 1159–1168.
[85] O. Yair, M. Ben-Chen, R. Talmon, Parallel transport on the cone manifold of spd matrices for domain adaptation, IEEE Trans. Signal Process. 67 (7) (2019) 1797–1811.
[86] L. Zelnik-Manor, P. Perona, Self-tuning spectral clustering, Adv. Neural Inf. Process. Syst. 17 (2004).
[87] Z. Zhao, D. Giannakis, Analog forecasting with dynamics-adapted kernels, Nonlinearity 29 (9) (2016) 2888.