

Recovering Hidden Components in Multimodal Data with Composite Diffusion Operators*

Tal Shnitzer[†], Mirela Ben-Chen[‡], Leonidas Guibas[§], Ronen Talmon[†], and Hau-Tieng Wu[¶]

Abstract. Finding appropriate low-dimensional representations of high-dimensional multimodal data can be challenging, since each modality embodies unique deformations and interferences. In this paper, we address the problem using manifold learning, where the data from each modality is assumed to lie on some manifold. In this context, the goal is to characterize the relations between the different modalities by studying their underlying manifolds. We propose two new diffusion operators that allow us to isolate, enhance, and attenuate the hidden components of multimodal data in a data-driven manner. Based on these new operators, efficient low-dimensional representations can be constructed for such data, which characterize the common structures and the differences between the manifolds underlying the different modalities. The capabilities of the proposed operators are demonstrated on 3D shapes and on a fetal heart rate monitoring application.

Key words. manifold learning, diffusion maps, multimodal data, sensor fusion, common variable, shape differences

AMS subject classifications. 62-07, 62M15, 57R40, 57M50

DOI. 10.1137/18M1218157

1. Introduction. Recent technological progress has led to highly heterogeneous datasets consisting of multimodal samples acquired by a multitude of sensors. Current research is plagued by the problem of finding the “appropriate,” often low-dimensional representation for such high-dimensional multimodal data. Indeed, obtaining meaningful representations from multimodal data is truly challenging, since such data comprise many latent sources of variability, each source embodying unique and possibly redundant information; while some of these sources are important, some are completely superfluous. This naturally leads to such problems as how to discover and isolate the different sources, how to identify and extract the relevant information, and how to merge data from different modalities.

Various studies have addressed multimodal data analysis problems [27]. A few examples

*Received by the editors October 1, 2018; accepted for publication (in revised form) July 10, 2019; published electronically September 24, 2019.

<https://doi.org/10.1137/18M1218157>

Funding: The work of the first and fourth authors was supported by the European Research Council (ERC starting grant 802735 DIFFOP) and by the Israel Science Foundation (grant 1490/16). The work of the second author was supported by the European Research Council (ERC starting grant 714776 OPREP) and by the Israel Science Foundation (grant 504/16).

[†]Andrew and Erna Viterbi Faculty of Electrical Engineering, Technion – Israel Institute of Technology, Technion City, Haifa, Israel 3200000 (shnitzer@campus.technion.ac.il, ronen@ee.technion.ac.il).

[‡]Department of Computer Science, Technion – Israel Institute of Technology, Technion City, Haifa, Israel 3200000 (mirela@cs.technion.ac.il).

[§]Department of Computer Science, Stanford University, Stanford, CA 94305 (guibas@cs.stanford.edu).

[¶]Department of Mathematics and Department of Statistical Science, Duke University, Durham, NC 27708 (hauwu@math.duke.edu).

include the classical Canonical Correlation Analysis (CCA) [22], which recovers highly correlated linear projections from two datasets, and recent CCA extensions which involve kernels to address nonlinearities [28, 3, 34]. Methods relying on kernels are of particular interest in the context of the present work. For example, methods for spectral clustering of multimodal data based on kernel manipulation are presented in [47, 24, 13, 32]. In [47], spectral clustering is performed on the multimodal data by solving the generalized eigenvalue problem of a new matrix, constructed based on a mixture of random walks defined on multiple graphs, each representing a different view. In [24], multimodal spectral clustering is learned by iteratively clustering each view separately and then modifying the graph structures accordingly. Two other methods [13, 32] combine affinity matrices of two graphs, representing two different views, by constructing a larger symmetric affinity matrix which is based on their multiplication. Other related work includes (i) the construction of a joint manifold by concatenating samples from several sensors, each represented by a separate manifold [12]; (ii) metric fusion obtained by combining similarity measures through kernel multiplication [45]; and (iii) a new representation of multiview data learned by jointly diagonalizing Laplacians of different views [15, 16]. In addition, [9] presents a method for mapping low-dimensional graph Laplacian representations of different views (or times) into a common latent space, allowing for the analysis of multimodal data in a low-dimensional intrinsic space.

Our specific focus here is on a manifold learning approach. Consider a single high-dimensional dataset assumed to live on a single manifold. Analyzing this dataset with typical manifold learning methods, such as Laplacian eigenmaps [5] or diffusion maps [7], simplifies to computing a kernel based on an affinity suitable for the dataset at hand. Then, by employing spectral analysis, the data are embedded in a new Euclidean space that captures their underlying manifold structure. A question then naturally arises: Are the required mathematical properties for spectral analysis transferable to settings comprising several datasets? If this can be achieved, the data analysis procedure could be naturally extended to analyze multiple datasets, deforming the intrinsic space in different ways.

Apparently, manifold learning techniques almost exclusively address only a single manifold structure. In a recent work [29], a data-driven method for recovering the common latent variable underlying multiple multimodal sensor data based on alternating products of diffusion operators was presented. This work was later extended in [44], showing that the alternating products of diffusion operators recover a common manifold structure. In addition, as proven in [44, 29], it ignores the components specific to each modality. However, the product of diffusion operators does not necessarily have a real spectrum. Other recent works [18, 33] propose to analyze dynamical systems based on products of diffusion operators, in a manner related to [29]. There, data from each time frame is modeled as samples from a manifold with a time-evolving metric, and by revealing the common latent variables of several time frames, they recover coherent sets (in [18]) or a representation of the common latent manifold in time (in [33]).

In this paper, we propose new diffusion operators defined on data arising from multiple sensors, allowing for a nonlinear efficient data-driven way to isolate, enhance, and attenuate various hidden components. More concretely, we propose two operators that reveal the common structures and the differences between manifolds. We show that these two operators have a meaningful spectral decomposition, which we leverage to construct an efficient low-dimensional representation.

The capabilities of the presented operators in extracting hidden components are demonstrated in simulations and on a real-world application for fetal heart rate (fHR) monitoring. fHR monitoring is widely used for the assessment of the fetus's health during both pregnancy and delivery. The most accurate method, relying on the placement of electrodes on the fetus's scalp, is invasive and therefore carries many risks. Consequently, noninvasive measurements are usually carried out by placing electrodes on the abdomen of the mother (see a comprehensive review in [39]). Naturally, the measured signal contains, in addition to the fetal electrocardiogram (ECG), the maternal ECG, masking the desired information. In order to suppress the maternal ECG and to extract the fetal ECG, common practice is to use another (reference) electrode, placed on the mother's thorax, for the purpose of measuring only the maternal ECG. Then the relation between the measured abdomen and thorax signals is extracted, using, for example, the adaptive least mean squares (LMS) algorithm [46]. In this work, we detect the fetal ECG from two abdomen signals, which is considered a challenging problem that currently has no definitive solution. We show that the proposed operators discover the relations between the signals acquired with multiple sensors in a data-driven manner, revealing their hidden components.

2. Problem formulation. Consider two diffeomorphic manifolds, $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$, with a diffeomorphism $\phi : \mathcal{M}^{(1)} \mapsto \mathcal{M}^{(2)}$, where each manifold $\mathcal{M}^{(\ell)}$ is a compact Riemannian manifold without a boundary of dimension d with a metric $g^{(\ell)}$. In this work, we will distinguish between the following two structures:

$$(2.1) \quad \Omega_\alpha = \left\{ x \in \mathcal{M}^{(1)} : \nabla \phi|_x = \alpha I \right\} \subset \mathcal{M}^{(1)},$$

$$(2.2) \quad \Omega_\alpha^c = \mathcal{M}^{(1)} \setminus \overset{\circ}{\Omega}_\alpha,$$

where $\alpha > 0$ is a scaling factor, I denotes a $d \times d$ identity matrix, $\nabla \phi|_x$ is represented by a pair of properly chosen orthonormal bases at $T_x \mathcal{M}^{(1)}$ and $T_{\phi(x)} \mathcal{M}^{(2)}$, and $\overset{\circ}{\Omega}_\alpha$ denotes the maximal open subset of the closed set Ω_α . Therefore, Ω_α denotes all structures which are similar, up to a scaling $\alpha > 0$, in the two manifolds $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$.

Our goal is to identify and isolate Ω_α and Ω_α^c in a data-driven manner, given pairs of observation samples (x, y) , such that $x \in \mathcal{M}^{(1)}$, $y \in \mathcal{M}^{(2)}$, and $y = \phi(x)$. We will show in what follows that the two structures Ω_α and Ω_α^c have great importance in data analysis problems.

For example, consider the two geometric shapes presented in Figure 1. Figure 1(a) depicts a 2-sphere, and Figure 1(b) depicts a scaled and deformed sphere, i.e., a scaled sphere with a “bump.” Denote these two shapes by $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$, respectively. The deformation and scaling of $\mathcal{M}^{(2)}$ can be represented by a diffeomorphism between the two shapes, $\phi : \mathcal{M}^{(1)} \rightarrow \mathcal{M}^{(2)}$. In this example, by definition the undeformed sphere structure (up to scaling) is represented by Ω_α and the “bump” is represented by Ω_α^c . Therefore, given the two shapes, our goal is to recover a separate representation for Ω_α and Ω_α^c .

This problem formulation, describing common structures of two manifolds, i.e., Ω_α , can be seen as analogous to recent work [18, 19]. There, a framework for recovering coherent sets in dynamical systems is proposed, where each time instance is represented by some underlying manifold, \mathcal{M} , and the system dynamics are represented by a diffeomorphism, ϕ . Since coherent



Figure 1. Two diffeomorphic geometric shapes: (a) a sphere, and (b) a scaled sphere with a deformation (a “bump”).

sets represent system behavior that changes slowly in time, they can be described by the common structures, i.e., Ω_α in our formulation.

3. Diffusion operators for multimodal data. In this section, we present the derivation of the proposed operators, starting from a single manifold setting in subsection 3.1, similarly to [7]. In subsection 3.2, we present an extension to two manifolds, as a variant of [44, 29], and finally, in subsection 3.3, we present the proposed new operators for revealing the common and difference structures of two manifolds.

3.1. Preliminaries—single manifold setting. Define the following symmetric kernel for a manifold \mathcal{M} , based on its distance function, denoted by d_g , corresponding to the metric g on \mathcal{M} :

$$(3.1) \quad k_\epsilon(x, x') = \exp\left(-\frac{d_g(x, x')^2}{\epsilon^2}\right),$$

where $x, x' \in \mathcal{M}$ and $\epsilon > 0$. The kernel is then normalized by

$$(3.2) \quad p_\epsilon(x, x') = \frac{k_\epsilon(x, x')}{d_\epsilon(x)},$$

where $d_\epsilon(x) = \int k_\epsilon(x, x') \mu(x') dV(x')$, V is the volume measure induced by g , and $\mu(x')$ is the density function of the points on \mathcal{M} . Similarly, define the following normalized kernel by

$$(3.3) \quad q_\epsilon(x, x') = \frac{k_\epsilon(x, x')}{d_\epsilon(x')}.$$

Based on $q_\epsilon(x, x')$ and $p_\epsilon(x, x')$, we define the following “backward” and “forward” diffusion operators:

$$(3.4) \quad P_\epsilon f(x) = \int p_\epsilon(x, x') f(x') \mu(x') dV(x'),$$

$$(3.5) \quad Q_\epsilon f(x) = \int q_\epsilon(x, x') f(x') \mu(x') dV(x')$$

for any $f \in C^\infty(\mathcal{M})$.

Proposition 3.1. Suppose $\mu, f \in C^4(\mathcal{M})$, where \mathcal{M} is a smooth Riemannian manifold with a metric g . The asymptotic expansion of the operators P_ϵ and Q_ϵ , when ϵ is sufficiently small, is given by

$$(3.6) \quad P_\epsilon f(x) = m_0 f(x) - m_2 \epsilon^2 \left(\Delta f + \frac{2\nabla f \cdot \nabla \mu}{\mu} \right)(x) + O(\epsilon^4),$$

$$(3.7) \quad Q_\epsilon f(x) = f(x) - \frac{m_2}{m_0} \epsilon^2 \left(\Delta f - \frac{f \Delta \mu}{\mu} \right)(x) + O(\epsilon^4),$$

where ∇ denotes the covariant derivative on the manifold, \mathcal{M} , Δ denotes the Laplace–Beltrami operator, and m_0, m_2 are two manifold related constants.

For the sake of readability, in the following asymptotic expansions we assume without loss of generality that the kernel function k_ϵ is scaled such that the constants m_0 and m_2 are equal to 1, similarly to [7, Appendix B] and [35, Appendix A], since these constants are moments of the Gaussian kernel. This derivation of the backward operator, P_ϵ , is shown in [7], and the derivation of the forward operator, Q_ϵ , is shown in the supplementary material, Appendix A.

The operator Q_ϵ is the *forward* operator, similarly defined in [35], which can be interpreted as an operator that propagates probability density functions on the manifold. Operator P_ϵ is the *backward* operator, which can be interpreted as propagating averages of functions on the manifold. These two operators are adjoint under the inner product with μ [35, 7].

From the spectral decomposition of the operators P_ϵ and Q_ϵ , a new low-dimensional representation for \mathcal{M} is typically obtained. In [7], it was shown that this low-dimensional representation embeds the data in a Euclidean space, where the Euclidean distance between the points in the new representation approximates the diffusion distance between data points, $x, x' \in \mathcal{M}$, defined as follows:

$$(3.8) \quad D_\epsilon(x, x') = \int_{\mathcal{M}} (p_\epsilon(x, u) - p_\epsilon(x', u))^2 \frac{\mu(u)}{\pi(u)} dV(u),$$

where $p_\epsilon(x, u)$, defined in (3.2), can be interpreted as the probability of transitioning from x to u in one propagation step of size ϵ and $\pi(u) = d_\epsilon(u) / \sum_{z \in \mathcal{M}} d_\epsilon(z)$ denotes the stationary distribution of the data on \mathcal{M} .

The backward operator P_ϵ was previously used in numerous applications to recover a meaningful representation of the data (e.g., [42, 43, 25]).

3.2. Modified alternating diffusion in a two manifold setting. Given two manifolds, denoted by $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$, consider the following $C^\infty(\mathcal{M}^{(1)}) \rightarrow C^\infty(\mathcal{M}^{(1)})$ operators:

$$(3.9) \quad G_{\epsilon_1, \epsilon_2} f(x) = \phi^* P_{\epsilon_2}^{(2)} (\phi^*)^{-1} Q_{\epsilon_1}^{(1)} f(x),$$

$$(3.10) \quad H_{\epsilon_1, \epsilon_2} f(x) = P_{\epsilon_1}^{(1)} \phi^* Q_{\epsilon_2}^{(2)} (\phi^*)^{-1} f(x)$$

for any function $f \in C^\infty(\mathcal{M}^{(1)})$, where $\epsilon_1, \epsilon_2 > 0$, $\phi^* : C^\infty(\mathcal{M}^{(2)}) \rightarrow C^\infty(\mathcal{M}^{(1)})$ denotes the operator corresponding to the pullback from $\mathcal{M}^{(2)}$ to $\mathcal{M}^{(1)}$, i.e., $(\phi^* g)(x) = (g)(\phi(x))$ for $x \in \mathcal{M}^{(1)}$, $g \in C^\infty(\mathcal{M}^{(2)})$, and $(\phi^*)^{-1}$ denotes the pushforward from $\mathcal{M}^{(2)}$ to $\mathcal{M}^{(1)}$, i.e., the inverse of ϕ^* .

Note that for such a composition of operators, the interpretation of the forward operator as propagating probability density functions does not extend to the operators G_ϵ and H_ϵ . In the following proposition, we present an analysis for the new operators, $G_{\epsilon_1, \epsilon_2}$ and $H_{\epsilon_1, \epsilon_2}$, which are the composition of $P_{\epsilon_\ell}^{(\ell)}$ and $Q_{\epsilon_\ell}^{(\ell)}$, $\ell = 1, 2$, based on their asymptotic expansions.

Proposition 3.2. *When $\epsilon_1, \epsilon_2 > 0$ are sufficiently small and $\mu^{(1)}$ is smooth enough, the asymptotic expansions of the operators $G_{\epsilon_1, \epsilon_2}$ and $H_{\epsilon_1, \epsilon_2}$ are given by*

$$(3.11) \quad G_{\epsilon_1, \epsilon_2} f(x) = f(x) - \epsilon_1^2 \Delta^{(1)} f(x) - \epsilon_2^2 \phi^* \Delta^{(2)}(\phi^*)^{-1} f(x)$$

$$(3.12) \quad - \epsilon_2^2 \phi^* \frac{2\nabla^{(2)}(\phi^*)^{-1} f \cdot \nabla^{(2)} \mu^{(2)}}{\mu^{(2)}}(x) + \epsilon_1^2 \frac{f \Delta^{(1)} \mu^{(1)}}{\mu^{(1)}}(x) + O(\epsilon_1^4 + \epsilon_2^4),$$

$$(3.13) \quad H_{\epsilon_1, \epsilon_2} f(x) = f(x) - \epsilon_1^2 \Delta^{(1)} f(x) - \epsilon_2^2 \phi^* \Delta^{(2)}(\phi^*)^{-1} f(x)$$

$$(3.14) \quad - \epsilon_1^2 \frac{2\nabla^{(1)} f \cdot \nabla^{(1)} \mu^{(1)}}{\mu^{(1)}}(x) + \epsilon_2^2 \phi^* \frac{\Delta^{(2)} \mu^{(2)}}{\mu^{(2)}}(x) + O(\epsilon_1^4 + \epsilon_2^4).$$

The derivations for both operators appear in the supplementary material, Appendix B.

Note that the asymptotic expansion of these operators can be described by a term which depends on the geometry, the Laplace–Beltrami operators $\Delta^{(1)}$ and $\Delta^{(2)}$ in both (3.11) and (3.13), and a term which depends on both the geometry and the densities, $\mu^{(1)}$, $\mu^{(2)}$, in both (3.12) and (3.14).

In [29, 44], alternating diffusion operators are defined in a related manner. In [29], the operator $\phi^* Q_{\epsilon_2}^{(2)}(\phi^*)^{-1} Q_{\epsilon_1}^{(1)}$ was introduced, and in [44] the operator $\phi^* P_{\epsilon_2}^{(2)}(\phi^*)^{-1} P_{\epsilon_1}^{(1)}$ was studied. Both variants are compositions of two operators, each corresponding to a different manifold. It was shown there that these operators reveal the common structure of the two manifolds. Note that the alternating diffusion operators are different from the operators proposed here, due to the use of two backward or forward operators in alternating diffusion, instead of one backward and one forward operator, as proposed here. We will show that the modification considered here not only is semantic but also leads to a different asymptotic behavior than the one described in [44]. The difference between the asymptotic expansions in (3.12) and (3.14) and the corresponding asymptotic expansion of the alternating diffusion operator is described in detail in the supplementary material, Appendix D.

3.3. Composite operators in a two manifold setting. The operators in subsection 3.2 and in [44, 29] suffer from several shortcomings. First, as presented in [44], the alternating diffusion operator highly depends on the order of the kernel multiplication (in a realistic discrete setting). Note that this is also true for operators $G_{\epsilon_1, \epsilon_2}$ and $H_{\epsilon_1, \epsilon_2}$, which depend on the kernel order even in the continuous setting, as portrayed by their asymptotic expansions. Second, these operators are neither self-adjoint nor normal (see the supplementary material, Appendix D), and therefore the spectral theorem does not hold. In this subsection, we address these problems and propose two new operators: $S_{\epsilon_1, \epsilon_2}$, which will be shown to reveal *common structures*, and $A_{\epsilon_1, \epsilon_2}$, which will be shown to reveal *differences*.

Define

$$(3.15) \quad S_{\epsilon_1, \epsilon_2} f(x) = \frac{1}{2} (G_{\epsilon_1, \epsilon_2} f(x) + H_{\epsilon_1, \epsilon_2} f(x)),$$

$$(3.16) \quad A_{\epsilon_1, \epsilon_2} f(x) = \frac{1}{2} (G_{\epsilon_1, \epsilon_2} f(x) - H_{\epsilon_1, \epsilon_2} f(x)).$$

Proposition 3.3. When $\epsilon_1, \epsilon_2 > 0$ are sufficiently small and $\mu^{(1)}$ is smooth enough, the asymptotic expansions of the operators $S_{\epsilon_1, \epsilon_2}$ and $A_{\epsilon_1, \epsilon_2}$ are given by

$$(3.17) \quad S_{\epsilon_1, \epsilon_2} f(x) = f(x) - \epsilon_1^2 \Delta^{(1)} f(x) - \epsilon_2^2 \phi^* \Delta^{(2)}(\phi^*)^{-1} f(x)$$

$$(3.18) \quad - \frac{\epsilon_2^2}{2} \left(\phi^* \frac{2\nabla^{(2)}(\phi^*)^{-1} f \cdot \nabla^{(2)} \mu^{(2)}}{\mu^{(2)}}(x) - f \phi^* \frac{\Delta^{(2)} \mu^{(2)}}{\mu^{(2)}}(x) \right)$$

$$(3.19) \quad - \frac{\epsilon_1^2}{2} \left(\frac{2\nabla^{(1)} f \cdot \nabla^{(1)} \mu^{(1)}}{\mu^{(1)}}(x) - f \Delta^{(1)} \mu^{(1)}(x) \right) + O(\epsilon_1^4 + \epsilon_2^4),$$

$$(3.20) \quad A_{\epsilon_1, \epsilon_2} f(x) = \frac{\epsilon_1^2}{2} \left(\frac{2\nabla^{(1)} f \cdot \nabla^{(1)} \mu^{(1)}}{\mu^{(1)}}(x) + f \Delta^{(1)} \mu^{(1)}(x) \right)$$

$$(3.21) \quad - \frac{\epsilon_2^2}{2} \left(\phi^* \frac{2\nabla^{(2)}(\phi^*)^{-1} f \cdot \nabla^{(2)} \mu^{(2)}}{\mu^{(2)}}(x) + f \phi^* \frac{\Delta^{(2)} \mu^{(2)}}{\mu^{(2)}}(x) \right) + O(\epsilon_1^4 + \epsilon_2^4).$$

The derivations for both operators appear in the supplementary material, Appendix C.

Note that since ϕ is a diffeomorphism from $\mathcal{M}^{(1)}$ to $\mathcal{M}^{(2)}$, the probability density function of the manifold $\mathcal{M}^{(2)}$, denoted by $\mu^{(2)}$, can be written as a function of $\mu^{(1)}$ and ϕ :

$$(3.22) \quad \mu^{(2)}(y) = |\det(\nabla\phi^{-1}(y))| \mu^{(1)}(\phi^{-1}(y)),$$

where $y \in \mathcal{M}^{(2)}$ and $\det()$ denotes the determinant.

The asymptotic expansion of $S_{\epsilon_1, \epsilon_2}$ includes a summation of two Laplace–Beltrami operators (the right term in (3.17)), corresponding to the two considered manifolds, $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$. This term relates to the dynamic Laplacian, defined in [18, 19], which was shown to be equivalent to the summation of two Laplace–Beltrami operators, from two different time-instances, when assuming a uniform density. The dynamic Laplacian reveals coherent sets in dynamical systems, representing common system behavior in different time-instances. Therefore, this similarity strengthens the claim that the operator $S_{\epsilon_1, \epsilon_2}$ reveals the common structure of the two manifolds. Conversely, the asymptotic expansion of $A_{\epsilon_1, \epsilon_2}$ is composed of the subtraction between the term (3.20), which is based on $\mathcal{M}^{(1)}$, and the term (3.21), which is based on $\mathcal{M}^{(2)}$. These two terms are functions of the probability densities, $\mu^{(\ell)}$, and the diffeomorphism, ϕ . Importantly, in the asymptotic expansion of $A_{\epsilon_1, \epsilon_2}$, the two Laplace–Beltrami operators of the two manifolds that are applied to f in (3.17) are absent (see the supplementary material, Appendix C). Clearly, when ϕ is the identity function, i.e., the two manifolds are identical, then the expansion (3.17)–(3.19) of $S_{\epsilon_1, \epsilon_2}$ recovers the result in [7] and $A_{\epsilon_1, \epsilon_2}$ is zero.

In the following we will show that $A_{\epsilon_1, \epsilon_2}$ characterizes the difference between the manifolds based on differences in their density functions. In addition, we will show that the eigenfunctions of $A_{\epsilon_1, \epsilon_2}$ are supported on Ω_α^c , the regions containing these differences. To complement

the analysis, in section 4 we will support these claims in a discrete setting, and in sections 5 and 6 we will demonstrate them using both synthetic and real applications.

Consider a special case in which the density $\mu^{(1)}$ of manifold $\mathcal{M}^{(1)}$ is uniform. In this case, the asymptotic expansions in Proposition 3.3 reduce to

$$(3.23) \quad S_{\epsilon_1, \epsilon_2} f(x) = f(x) - \epsilon_1^2 \Delta^{(1)} f(x) - \epsilon_2 \phi^* \Delta^{(2)}(\phi^*)^{-1} f(x)$$

$$(3.24) \quad - \frac{\epsilon_2^2}{2} \left(\phi^* \frac{2\nabla^{(2)}(\phi^*)^{-1} f \cdot \nabla^{(2)} \mu^{(2)}}{\mu^{(2)}}(x) - f \phi^* \frac{\Delta^{(2)} \mu^{(2)}}{\mu^{(2)}}(x) \right) + O(\epsilon_1^4 + \epsilon_2^4),$$

$$(3.25) \quad A_{\epsilon_1, \epsilon_2} f(x) = -\frac{\epsilon_2^2}{2} \left(\phi^* \frac{2\nabla^{(2)}(\phi^*)^{-1} f \cdot \nabla^{(2)} \mu^{(2)}}{\mu^{(2)}}(x) + f \phi^* \frac{\Delta^{(2)} \mu^{(2)}}{\mu^{(2)}}(x) \right) + O(\epsilon_1^4 + \epsilon_2^4),$$

where $\mu^{(2)}(x) = |\det(\nabla\phi^{-1}(x))|$.

In addition, when considering a volume-preserving diffeomorphism, similarly to [18], $\mu^{(2)}(x)$ is uniform as well. In such a case, the asymptotic expansion of the operator $S_{\epsilon_1, \epsilon_2}$ is reduced to the addition of the two Laplace–Beltrami operators in (3.23). Moreover, the second order terms in the asymptotic expansion of $A_{\epsilon_1, \epsilon_2}$ vanish. This special case emphasizes that the operator $S_{\epsilon_1, \epsilon_2}$ depends mostly on the geometry of the two manifolds, whereas $A_{\epsilon_1, \epsilon_2}$ depends on the diffeomorphism and the probability density functions of the two manifolds.

Proposition 3.4. Denote $\epsilon_1 = \epsilon$ and suppose $\epsilon_2 = \alpha\epsilon$ for some $\alpha > 0$. The operators $A_\alpha : C^\infty(\mathcal{M}^{(1)}) \rightarrow C^\infty(\mathcal{M}^{(1)})$ and $S_\alpha : C^\infty(\mathcal{M}^{(1)}) \rightarrow C^\infty(\mathcal{M}^{(1)})$ are anti-self-adjoint and self-adjoint, respectively, where $A_\alpha = \lim_{\epsilon \rightarrow 0} A_{\epsilon_1, \epsilon_2}/\epsilon^2$ and $S_\alpha = \lim_{\epsilon \rightarrow 0} S_{\epsilon_1, \epsilon_2}/\epsilon^2$.

The proof is given in the supplementary material, Appendix E.

As presented in this section, the proposed operators, $S_{\epsilon_1, \epsilon_2}$ and $A_{\epsilon_1, \epsilon_2}$, solve the two main shortcomings of the alternating diffusion operator. First, from their asymptotic expansions, it can be seen that there is no dependency on the order of the kernels (this will be revisited in the discrete setting in section 4). Second, based on Proposition 3.4, they are self-adjoint and anti-self-adjoint, respectively, and therefore, the spectral theorem holds for these operators.

Based on the latter property, we strengthen the claim that A_α represents the differences between the two manifolds by showing that the eigenfunctions of $A_\alpha = \lim_{\epsilon \rightarrow 0} A_{\epsilon_1, \epsilon_2}/\epsilon^2$, $\epsilon_2 = \alpha\epsilon_1 = \alpha\epsilon$, are supported on Ω_α^c .

Proposition 3.5. Given $f \in C^\infty(\mathcal{M}^{(1)})$, if $\text{supp } f \subset \mathring{\Omega}_\alpha$, then $A_\alpha f(x) = 0$.

The proof is given in the supplementary material, Appendix F. A direct consequence of this proposition is that if $A_\alpha f = \lambda f$, $f \neq 0$, then $\text{supp } f \subset \mathcal{M}^{(1)} \setminus \mathring{\Omega}_\alpha = \Omega_\alpha^c$. Therefore, the eigenfunctions of the difference operator A_α (when $\epsilon_1, \epsilon_2 \rightarrow 0$) are nonzero only in regions where there are differences between the two manifolds, $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$. Note that this proposition does not guarantee the behavior of f on Ω_α^c .

4. Discrete setting for data analysis. We now present our proposed method in the discrete setting. We begin by introducing the discrete counterparts of the operators presented in section 3. In subsection 4.1, we present a discrete analysis of the operator $A_{\epsilon_1, \epsilon_2}$, and in subsection 4.2, we present our construction of a new coordinate system for the data based on these discrete operators.

Let $\{x_i\}_{i=1}^N$, $x_i \in \mathbb{R}^{p_x}$, and $\{y_i\}_{i=1}^N$, $y_i \in \mathbb{R}^{p_y}$, be two datasets of N samples from $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$, respectively, such that $y_i = \phi(x_i)$. Assume that the data are embedded in two different high-dimensional ambient spaces, each corresponding to some measurement of $\mathcal{M}^{(1)}$ or $\mathcal{M}^{(2)}$. Since we only have access to the ambient space, there is no direct access to the geometric structure of $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$, and identifying Ω_α and Ω_α^c is nontrivial.

The following are the discrete counterparts of the operators presented in section 3.

Let $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$ be two $N \times N$ affinity (kernel) matrices defined by

$$(4.1) \quad W_{i,j}^{(1)} = k_{\epsilon_1}^{(1)}(x_i, x_j) = \exp(-d_X(x_i, x_j)^2/\epsilon_1^2),$$

$$(4.2) \quad W_{i,j}^{(2)} = k_{\epsilon_2}^{(2)}(y_i, y_j) = \exp(-d_Y(y_i, y_j)^2/\epsilon_2^2),$$

where $\epsilon_1, \epsilon_2 > 0$ may be different, and let $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$ be two $N \times N$ diagonal matrices, with diagonal elements given by

$$(4.3) \quad D_{i,i}^{(1)} = \sum_{j=1}^N k_{\epsilon_1}^{(1)}(x_i, x_j),$$

$$D_{i,i}^{(2)} = \sum_{j=1}^N k_{\epsilon_2}^{(2)}(y_i, y_j).$$

In the construction of the kernels, $k_{\epsilon_1}^{(1)}(x_i, x_j)$ and $k_{\epsilon_2}^{(2)}(y_i, y_j)$, the distance functions, d_X and d_Y , can be chosen as any metric induced distance that best describes the datasets, e.g., Euclidean distance, geodesic distance, and Wasserstein distance [36, 37].

Note that a common choice for ϵ_1 and ϵ_2 in the construction of the kernel used in (4.1), (4.2), and (4.3) is some scalar multiplication of the median of the distances between the dataset samples, i.e., $\epsilon_1 = c \text{ median}\{d_X(x_i, x_j)\}$ and $\epsilon_2 = c \text{ median}\{d_Y(y_i, y_j)\}$, where $c > 0$ is some scalar. By constructing the kernels in this manner, if the distance functions $d_X(x_i, x_j)$ and $d_Y(y_i, y_j)$ are induced by a metric, the resulting operators are scale invariant. Therefore, in the discrete setting, Ω_α is defined similarly to (2.1), by $\Omega_\alpha = \{x \in \mathcal{M}^{(1)} : \nabla \phi|_x = \alpha \mathbf{I}\}$, where $\alpha > 0$ denotes the scaling and \mathbf{I} is the identity matrix.

Let $\mathbf{P}^{(\ell)}, \mathbf{Q}^{(\ell)} \in \mathbb{R}^{N \times N}$ be the discrete counterparts of the operators $P_{\epsilon_\ell}^{(\ell)}$ and $Q_{\epsilon_\ell}^{(\ell)}$ given by

$$(4.4) \quad \mathbf{P}^{(\ell)} = (\mathbf{D}^{(\ell)})^{-1} \mathbf{W}^{(\ell)},$$

$$(4.5) \quad \mathbf{Q}^{(\ell)} = \mathbf{W}^{(\ell)} (\mathbf{D}^{(\ell)})^{-1}$$

for $\ell = 1, 2$. It is clear that $(\mathbf{P}^{(\ell)})^T = \mathbf{Q}^{(\ell)}$, where $(\cdot)^T$ denotes the transpose operator. Note that $\mathbf{Q}^{(\ell)}$ is a column stochastic matrix and therefore can be interpreted as a Markov transition matrix, defined on the data, which propagates probabilities, analogously to the continuous-time forward operator $Q_{\epsilon_\ell}^{(\ell)}$.

For any $f \in C^\infty(\mathcal{M}^{(1)})$, define $\mathbf{v} \in \mathbb{R}^N$ by $v(j) = f(x_j)$. Our formulations are based on the assumption that the discrete matrix and kernel operations approximate the continuous

operators, i.e.,

$$(4.6) \quad P_{\epsilon_\ell}^{(\ell)} f(x_j) \approx \left(\mathbf{P}^{(\ell)} \mathbf{v} \right) (j),$$

$$(4.7) \quad Q_{\epsilon_\ell}^{(\ell)} f(x_j) \approx \left(\mathbf{Q}^{(\ell)} \mathbf{v} \right) (j).$$

This approximation can be justified by a standard large deviation argument, similarly to [40], which we omit for brevity.

Accordingly, the discrete counterparts of the operators $G_{\epsilon_1, \epsilon_2}$ and $H_{\epsilon_1, \epsilon_2}$ are

$$(4.8) \quad \mathbf{G} = \mathbf{P}^{(2)} \mathbf{Q}^{(1)},$$

$$(4.9) \quad \mathbf{H} = \mathbf{P}^{(1)} \mathbf{Q}^{(2)},$$

and those of the operators S_ϵ and A_ϵ are

$$(4.10) \quad \mathbf{S} = \mathbf{G} + \mathbf{H},$$

$$(4.11) \quad \mathbf{A} = \mathbf{G} - \mathbf{H}.$$

Note that in this construction of the discrete operators, the probability density function of each manifold, $\mu^{(\ell)}$, is reflected in the sampling of the points in the dataset. In addition, we assume that the diffeomorphism ϕ , which appears explicitly in the continuous operators, $S_{\epsilon_1, \epsilon_2}$ and $A_{\epsilon_1, \epsilon_2}$, is implicitly contained in the discrete operators \mathbf{P} and \mathbf{Q} . However, if the two datasets differ significantly in their densities or metrics, this assumption may be inaccurate. We discuss this issue and propose two possible solutions in Appendix G.

Proposition 4.1. \mathbf{S} is symmetric and \mathbf{A} is antisymmetric.

Based on the definitions of \mathbf{S} and \mathbf{A} above, it is easy to show that $\mathbf{S}^T = \mathbf{S}$ and that $\mathbf{A}^T = -\mathbf{A}$. Specifically, $(j\mathbf{A})^H = j\mathbf{A}$, where $(\cdot)^H$ denotes conjugate transpose and $j = \sqrt{-1}$.

Note that both the discrete alternating diffusion operator [29] and the operators \mathbf{G} and \mathbf{H} are not Hermitian, and therefore there is no spectral decomposition for them. Moreover, by their definition, they depend on the order of the matrix multiplication, e.g., whether we define $\mathbf{G} = \mathbf{P}^{(2)} \mathbf{Q}^{(1)}$ or $\mathbf{G} = \mathbf{P}^{(1)} \mathbf{Q}^{(2)}$.

The use of the symmetric and antisymmetric parts of an operator in the context of constructing a new representation was also presented in [17], where representations for directed graphs were obtained based on the symmetric and antisymmetric parts of the nonsymmetric weight matrix of the graph.

4.1. Discrete analysis of the operator \mathbf{A} . In this subsection we present an analysis for the discrete operator \mathbf{A} , showing that it is supported on the locations of the differences between the datasets, similarly to the continuous operator A_ϵ .

Consider two datasets $\{x_i\}_{i=1}^N$ and $\{y_i\}_{i=1}^N$, each consisting of N points, which are samples of $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$, respectively. The affinity matrices for the datasets $\{x_i\}$ and $\{y_i\}$ are constructed according to (4.1) and (4.2) and are denoted by $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$, respectively. Define $V_{\Omega_\alpha} = \{i \in V \mid W_{i,j}^{(1)} = W_{i,j}^{(2)} \forall j \in V\}$, where $V = \{1, \dots, N\}$, and $V_{\Omega_\alpha^c} = V \setminus V_{\Omega_\alpha}$. Assume that the correspondence between pairs of points in $\{x_i\}$ and $\{y_i\}$ is given and that the

datasets differ in the affinities between the points, i.e., $W_{i,j}^{(1)} \neq W_{i,j}^{(2)}$ if and only if $i \in V_{\Omega_\alpha^c}$ and $j \in V_{\Omega_\alpha^c}$. Note that since $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ are not isometric, the global distance profiles differ between the two manifolds, i.e., for any pair in the undeformed structures, (x_i, y_i) , $i \in V_{\Omega_\alpha}$, we can find a pair (x_j, y_j) , $j \in V_{\Omega_\alpha^c}$, such that $d_X(x_i, x_j) \neq d_Y(y_i, y_j)$. However, due to the construction of the affinity matrices using the Gaussian kernel, only the local distance profiles are captured in $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$, where the “locality” is induced by the kernel scales ϵ_1 and ϵ_2 in (4.1) and (4.2). Therefore, for a proper choice of the kernel scales, the local distance profiles of any pair (x_i, y_i) , $i \in V_{\Omega_\alpha}$, will be similar for the two manifolds, i.e., $W_{i,j}^{(1)} = W_{i,j}^{(2)}$ $\forall j \in V$. Conversely, for any pair belonging to the difference structures, (x_ℓ, y_ℓ) , $\ell \in V_{\Omega_\alpha^c}$, the local distance profiles will not be similar at any kernel scale, i.e., $\exists j \in V$ s.t. $W_{\ell,j}^{(1)} \neq W_{\ell,j}^{(2)}$ for any choice of ϵ_1 and ϵ_2 (assuming the local neighborhoods induced by ϵ_1 and ϵ_2 include at least 2 points). This desired property can be obtained by choosing small kernel scales, such that the effect of the differences between the two manifolds is localized around them. In addition, setting ϵ_1 and ϵ_2 to be some multiple of the *median* of the distances is important for guaranteeing that $W_{i,j}^{(1)} = W_{i,j}^{(2)}$ for $i \in V_{\Omega_\alpha}$. This indicates that the choice of ϵ is important in the construction of \mathbf{A} and should be some small multiple of the median of the distances in the data.

Proposition 4.2. Suppose $|V_{\Omega_\alpha^c}| = m \leq N/2$. The discrete operator \mathbf{A} has the following properties:

1. $A_{i,j} \neq 0$ only when $i \in V_{\Omega_\alpha^c}$ or $j \in V_{\Omega_\alpha^c}$.
2. The rank of \mathbf{A} is bounded by $2|V_{\Omega_\alpha^c}| = 2m$.

This proposition states that the discrete operator \mathbf{A} is nonzero only in regions where the two datasets differ and that its rank is related to the dimensionality of the differences. A direct consequence of this proposition is that the eigenvectors of \mathbf{A} encode information related to the location of the nontrivial diffeomorphism, and hence \mathbf{A} can be utilized for representing the differences between the two datasets.

Proof. Based on the definition of the datasets and the assumptions stated above, the difference between the affinity matrices $\mathbf{W}^{(1)}, \mathbf{W}^{(2)} \in \mathbb{R}^{N \times N}$ can be represented by

$$(4.12) \quad \mathbf{W}^{(2)} = \mathbf{W}^{(1)} + \mathbf{B}^T \mathbf{B},$$

where $\mathbf{B} \in \mathbb{R}^{N \times N}$, $\mathbf{B}e_i = 0$ if $i \in V_{\Omega_\alpha}$ and e_i are vectors which contain 1 at index i and 0 elsewhere, i.e., \mathbf{B} is a matrix in which column i contains only 0 $\forall i \in V_{\Omega_\alpha}$. Note that $(\mathbf{B}^T \mathbf{B})_{i,j} = 0$ if $i \in V_{\Omega_\alpha}$ or if $j \in V_{\Omega_\alpha}$.

Then, based on the definition of $\mathbf{D}^{(\ell)}$ in (4.3), $\mathbf{D}^{(2)} = \mathbf{D}^{(1)} + \text{diag}(\mathbf{B}^T \mathbf{B} \mathbf{1})$, where $\mathbf{1} \in \mathbb{R}^N$ is a vector containing only 1's and $\text{diag}(z)$ denotes a diagonal matrix with the elements of z on its diagonal. Note that $\text{diag}(\mathbf{B}^T \mathbf{B} \mathbf{1})$ is a diagonal matrix with nonzero diagonal entries only for $i \in V_{\Omega_\alpha^c}$.

The operator $\mathbf{P}^{(2)}$ is then given by

$$(4.13) \quad \mathbf{P}^{(2)} = (\mathbf{D}^{(2)})^{-1} \mathbf{W}^{(2)} = (\mathbf{D}^{(1)} + \text{diag}(\mathbf{B}^T \mathbf{B} \mathbf{1}))^{-1} (\mathbf{W}^{(1)} + \mathbf{B}^T \mathbf{B}).$$

Denote the inverse of the sum $(\mathbf{D}^{(1)} + \text{diag}(\mathbf{B}^T \mathbf{B} \mathbf{1}))$ by $((\mathbf{D}^{(1)})^{-1} - \mathbf{J})$, where

$$(4.14) \quad \mathbf{J} = (\mathbf{D}^{(1)} + \text{diag}(\mathbf{B}^T \mathbf{B} \mathbf{1}))^{-1} \text{diag}(\mathbf{B}^T \mathbf{B} \mathbf{1}) (\mathbf{D}^{(1)})^{-1}.$$

Based on the expression in (4.14), \mathbf{J} is a diagonal matrix with nonzero values only for $i \in V_{\Omega_\alpha^c}$, i.e., $J_{i,i} = 0 \quad \forall i \notin V_{\Omega_\alpha^c}$.

Substituting these derivations into the definition of \mathbf{A} , we get

$$(4.15) \quad \mathbf{A} = \mathbf{P}^{(1)} (\mathbf{P}^{(2)})^T - \mathbf{P}^{(2)} (\mathbf{P}^{(1)})^T$$

$$(4.16) \quad = \mathbf{P}^{(1)} \left(\mathbf{B}^T \mathbf{B} (\mathbf{D}^{(1)})^{-1} - \mathbf{W}^{(1)} \mathbf{J} - \mathbf{B}^T \mathbf{B} \mathbf{J} \right)$$

$$(4.17) \quad - \left((\mathbf{D}^{(1)})^{-1} \mathbf{B}^T \mathbf{B} - \mathbf{J} \mathbf{W}^{(1)} - \mathbf{J} \mathbf{B}^T \mathbf{B} \right) (\mathbf{P}^{(1)})^T.$$

Since all the elements in this expression are multiplied by either \mathbf{J} or $\mathbf{B}^T \mathbf{B}$, which contain nonzero values only for rows and columns corresponding to $i \in V_{\Omega_\alpha^c}$, the value of the discrete operator are $A_{i,j} \neq 0$ only when $i \in V_{\Omega_\alpha^c}$ or $j \in V_{\Omega_\alpha^c}$. In addition, this indicates that the rank of \mathbf{A} is bounded by $2|V_{\Omega_\alpha^c}| = 2m$, i.e., twice the number of elements in $V_{\Omega_\alpha^c}$. ■

4.2. New representations of the data based on \mathbf{S} and \mathbf{A} . Our goal is to obtain new representations for multimodal data based on the operators \mathbf{S} and \mathbf{A} , analogous to the diffusion maps coordinates [7] that represent the diffusion distances in the data. Specifically, we seek nonlinear mappings of the data to new coordinate systems, which describe the common structures or the differences between the modalities (manifolds). In addition, to obtain a compact representation, we require the constructed coordinates to be orthogonal. In this subsection, we present one option for obtaining such representations.

Since \mathbf{S} is a symmetric matrix, it has real eigenvalues and eigenvectors. The eigenvectors are orthogonal, and hence we can construct a new low-dimensional representation for the common structures in the datasets based on \mathbf{S} by taking its eigenvectors corresponding to the largest eigenvalues. Moreover, if the matrices $\mathbf{P}^{(\ell)}$ and $\mathbf{Q}^{(\ell)}$, $\ell = 1, 2$, are constructed such that they are doubly stochastic [8, 41], an analogous diffusion distance can be defined based on the operator \mathbf{S} and its eigenvectors.

The operator \mathbf{A} is antisymmetric and therefore has purely imaginary eigenvalues, in conjugate pairs, and complex eigenvectors. We propose to construct a new low-dimensional representation of the differences between the datasets based on the eigenvectors of \mathbf{A} . Consider the following embedding:

$$(4.18) \quad \tilde{\Psi}^{(\mathbf{A})}(i) = [\mu_1^{(\mathbf{A})} \psi_1^{(\mathbf{A})}(i), \dots, \mu_N^{(\mathbf{A})} \psi_N^{(\mathbf{A})}(i)],$$

where $\mathbf{A} \psi_k^{(\mathbf{A})} = \mu_k^{(\mathbf{A})} \psi_k^{(\mathbf{A})}$, $k = 1, \dots, N$ and $i = 1, \dots, N$. Assume that the eigenvalues, $\mu_k^{(\mathbf{A})}$, and corresponding eigenvectors, $\psi_k^{(\mathbf{A})}$, are ordered such that $|\mu_1^{(\mathbf{A})}| \geq |\mu_2^{(\mathbf{A})}| \geq \dots \geq |\mu_N^{(\mathbf{A})}|$. The low-dimensional representation can then be defined by taking the elements corresponding to the $M < N$ largest eigenvalues (in absolute value) as follows:

$$(4.19) \quad \Psi^{(\mathbf{A})}(i) = [\mu_1^{(\mathbf{A})} \psi_1^{(\mathbf{A})}(i), \dots, \mu_M^{(\mathbf{A})} \psi_M^{(\mathbf{A})}(i)].$$

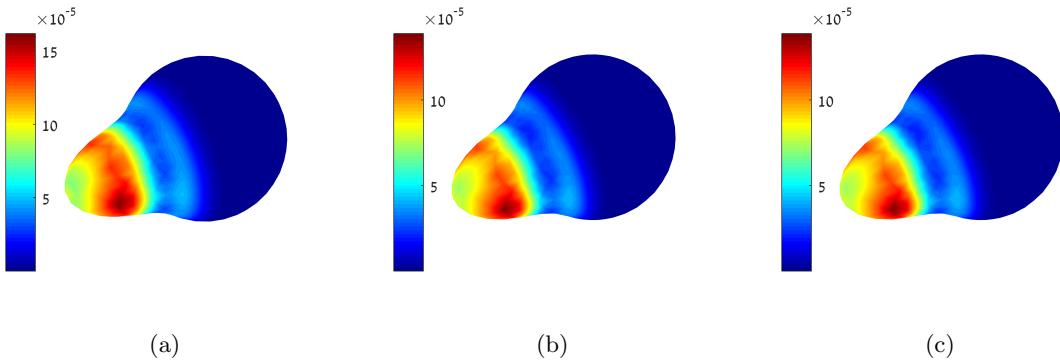


Figure 2. The deformed sphere colored according to (a) $\|A_{i,:}\|_2^2$, (b) $\|\Psi^{(A)}(i)\|_2^2$ with $M = 6$, and (c) $2\|\Psi_{nc}^{(A)}(i)\|_2^2$ with $M = 6$.

To explain the motivation for this choice, consider the following sum of squares of a row in \mathbf{A} :

$$(4.20) \quad \begin{aligned} \|A_{i,:}\|_2^2 &= \sum_{r=1}^N \sum_{k=1}^N \left| \mu_k^{(A)} \right|^2 \left| \psi_k^{(A)}(r) \right|^2 \left| \psi_k^{(A)}(i) \right|^2 \\ &= \sum_{k=1}^N \left| \mu_k^{(A)} \right|^2 \left| \psi_k^{(A)}(i) \right|^2 = \left\| \Psi^{(A)}(i) \right\|_2^2 + O\left(\left| \mu_{M+1}^{(A)} \right|^2\right), \end{aligned}$$

where we assume that the eigenvectors are normalized such that $\sum_{r=1}^N |\psi_k^{(A)}(r)|^2 = 1 \forall k$. Note that since the eigenvalues typically decay fast, the error term in (4.20) is small.

In subsection 4.1, we proved that $A_{i,j} \neq 0$ if and only if $x_i \in \Omega_\alpha^c$ or $x_j \in \Omega_\alpha^c$. Therefore, the squared row sum of the operator \mathbf{A} is related to the differences between the datasets, where $\|A_{i,:}\|_2^2$ is higher for $x_i \in \Omega_\alpha^c$. Furthermore, based on the relationship in (4.20), the low-dimensional embedding of the points in (4.19) establishes a new space, where the Euclidean norm provides a measure for the difference between the datasets at each point i . This result shows that the eigenvectors of \mathbf{A} can be used as a new representation of the differences between the datasets and that the eigenvectors corresponding to smaller eigenvalues in absolute value can be disregarded.

An illustration of this relationship between the sum of squares of the rows of \mathbf{A} and the location of the differences is presented in Figure 2, where the operator \mathbf{A} was constructed based on the sphere and the deformed sphere presented in Figure 1. Plots (a) and (b) in Figure 2 present the deformed sphere, colored according to (a) $\|A_{i,:}\|_2^2$ and (b) $\|\Psi^{(A)}(i)\|_2^2$ with $M = 6$. This figure depicts that $\|A_{i,:}\|_2^2$ is roughly supported on the location of the differences and that the first six eigenvectors provide a good approximation of $\|A_{i,:}\|_2^2$.

Note that the proposed embedding in (4.19) is still lacking, since it ignores the complex nature of the eigenvectors. Therefore, in the following, we propose a modified embedding, which takes into account the complex values by taking the real and imaginary parts of *nonconjugate*

eigenvalues and concatenating them as follows:

$$(4.21) \quad \{x_i, y_i\} \mapsto \Psi_{nc}^{(\mathbf{A})}(i) = \{\text{real}\{\psi_k^{(\mathbf{A})}(i)\}; \text{imag}\{\psi_k^{(\mathbf{A})}(i)\}\}_{k=2r}, \quad r = 1, \dots, M/2.$$

In the remainder of this subsection, we will explain the motivation for this choice by showing that (i) the eigenvectors corresponding to conjugate eigenvalues—i.e., $\psi_k^{(\mathbf{A})}$ and $\psi_\ell^{(\mathbf{A})}$ s.t. $\mu_k^{(\mathbf{A})} = \bar{\mu}_\ell^{(\mathbf{A})}$, where \bar{a} denotes the complex conjugate of a —contain redundant information, and therefore it is sufficient to take only the eigenvectors corresponding to nonconjugate eigenvalues as described in (4.21); (ii) the real and imaginary parts of these eigenvectors form a set of orthogonal vectors. Figure 2(c) demonstrates this redundancy of the eigenvectors corresponding to conjugate eigenvalues. This plot presents the deformed sphere, colored according to $2\|\Psi_{nc}^{(\mathbf{A})}(i)\|_2^2$ with $M = 6$, i.e., the Euclidean norm of the first 3 complex eigenvectors, corresponding to the largest nonconjugate eigenvalues (multiplied by 2). The norm is multiplied by 2 since we use here only 3 eigenvectors (of nonconjugate eigenvalues) and compare their norm to the norm calculated based on 6 eigenvectors (of conjugate eigenvalues). Note that plot (c) in this figure is identical to plot (b), which presents the Euclidean norm of the first 6 complex eigenvectors corresponding to the largest eigenvalues (including the conjugate pairs).

The spectral decomposition of a real antisymmetric matrix is given by

$$(4.22) \quad \mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^T,$$

where \mathbf{V} is a matrix containing the eigenvectors of \mathbf{A} in its columns and Λ is a diagonal matrix, containing the eigenvalues in conjugate pairs, i.e.,

$$(4.23) \quad \begin{bmatrix} j\lambda_1 & 0 & 0 & \dots \\ 0 & -j\lambda_1 & 0 & \dots \\ 0 & 0 & j\lambda_2 & \dots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots \end{bmatrix},$$

where λ_k , $k = 1, \dots, [N/2]$, are real and positive and j denotes $\sqrt{-1}$. Note that when N is odd, $\lambda_{[N/2]} = 0$. In relation to the notation in (4.19), $\mu_r^{(\mathbf{A})} = (-1)^{r+1}j\lambda_{\lfloor(r+1)/2\rfloor}$, $r = 1, \dots, N$.

This spectral decomposition is related to a real orthogonal decomposition of the form

$$(4.24) \quad \mathbf{A} = \mathbf{U}\Sigma\mathbf{U}^T,$$

where \mathbf{U} is orthogonal and real, and Σ is a block diagonal matrix, with k th 2×2 diagonal blocks of the form

$$(4.25) \quad \begin{bmatrix} 0 & \lambda_k \\ -\lambda_k & 0 \end{bmatrix},$$

where $k = 1, \dots, [N/2]$ [20]. By comparing this form to the spectral decomposition of the antisymmetric matrix, \mathbf{A} , it can be shown that the real and imaginary parts of eigenvectors corresponding to nonconjugate nonzero eigenvalues of \mathbf{A} are equal to different orthogonal

vectors in U , i.e., $\text{real}\{\psi_\ell^{(\mathbf{A})}\} = u_k$, $\text{imag}\{\psi_\ell^{(\mathbf{A})}\} = u_n$, where $\psi_\ell^{(\mathbf{A})}$ is the ℓ th eigenvector of \mathbf{A} and u_k and u_n are the k th and n th columns of \mathbf{U} ($n \neq k$). From the orthogonality of \mathbf{U} , we obtain $\langle \text{real}\{\psi_\ell^{(\mathbf{A})}\}, \text{imag}\{\psi_r^{(\mathbf{A})}\} \rangle = 0 \quad \forall \lambda_\ell \neq \lambda_r$ as well as for the real and imaginary parts of the same eigenvector ($\ell = r$). The real and imaginary parts of these eigenvectors can then be used for the construction of a new orthogonal representation for the differences between the datasets.

Algorithm 4.1 summarizes the procedure for obtaining the new representations for the data based on \mathbf{S} and \mathbf{A} .

Algorithm 4.1. Representation of the common structures and the differences between datasets.

1. Construct the affinity matrices for the two datasets

$$(4.26) \quad W_{i,j}^{(1)} = \exp\left(-\frac{d_X(x_i, x_j)^2}{\epsilon_1^2}\right), \quad W_{i,j}^{(2)} = \exp\left(-\frac{d_Y(y_i, y_j)^2}{\epsilon_2^2}\right),$$

where d_X, d_Y are some suitable notion of distance, defined on the data (e.g., the Euclidean distance if the data are in an ambient Euclidean space), and ϵ_1, ϵ_2 are the kernel scales, commonly taken as some multiple of the median of the distances.

2. Create the row stochastic and column stochastic matrices

$$(4.27) \quad \begin{aligned} \mathbf{P}^{(1)} &= \left(\mathbf{D}^{(1)}\right)^{-1} \mathbf{W}^{(1)}, & \mathbf{Q}^{(1)} &= \mathbf{W}^{(1)} \left(\mathbf{D}^{(1)}\right)^{-1}, \\ \mathbf{P}^{(2)} &= \left(\mathbf{D}^{(2)}\right)^{-1} \mathbf{W}^{(2)}, & \mathbf{Q}^{(2)} &= \mathbf{W}^{(2)} \left(\mathbf{D}^{(2)}\right)^{-1}, \end{aligned}$$

where $\mathbf{D}^{(\ell)}$ is a diagonal matrix with $D_{i,i}^{(\ell)} = \sum_{j=1}^N W_{i,j}^{(\ell)}$ and $\ell = 1, 2$.

3. Construct the symmetric and antisymmetric matrices

$$(4.28) \quad \mathbf{S} = \mathbf{P}^{(2)} \mathbf{Q}^{(1)} + \mathbf{P}^{(1)} \mathbf{Q}^{(2)}, \quad \mathbf{A} = \mathbf{P}^{(2)} \mathbf{Q}^{(1)} - \mathbf{P}^{(1)} \mathbf{Q}^{(2)}.$$

4. To obtain a new representation of dimension M for the common structures in the two datasets, calculate the eigenvalue decomposition of \mathbf{S} , $\mathbf{S} \psi_k^{(\mathbf{S})} = \mu_k^{(\mathbf{S})} \psi_k^{(\mathbf{S})}$, and take the first M eigenvectors, corresponding to the largest eigenvalues, $\{x_i, y_i\} \mapsto \{\psi_k^{(\mathbf{S})}(i)\}_{k=1}^M$.
 5. To obtain a representation for the differences between the datasets, calculate the eigenvalue decomposition of \mathbf{A} , $\mathbf{A} \psi_k^{(\mathbf{A})} = \mu_k^{(\mathbf{A})} \psi_k^{(\mathbf{A})}$ and take the real and imaginary parts of the first $M/2$ eigenvectors, corresponding to the largest (in absolute value) *nonconjugate* eigenvalues, $\{x_i, y_i\} \mapsto \{\text{real}\{\psi_k^{(\mathbf{A})}(i)\}; \text{imag}\{\psi_k^{(\mathbf{A})}(i)\}\}_{k=1}^{M/2}$.
-

5. Nonisometric shapes analysis. In this section, we demonstrate the properties of the proposed operators \mathbf{S} and \mathbf{A} using a toy example composed of two manifolds with a nonisometric diffeomorphism. We show that operator \mathbf{S} recovers the common manifold, whereas operator \mathbf{A} captures the “difference” between the two manifolds.

Consider two manifolds, $\mathcal{M}^{(1)}$, which is a sphere, and $\mathcal{M}^{(2)}$, which is a sphere with scaling

and a nonisometric deformation, which we will refer to as a “bump.” The two manifolds (shapes) are depicted in Figure 1. In this example, Ω_α , defined in (2.1), represents the part of the sphere that does not undergo deformation, and Ω_α^c , defined in (2.2), represents the deformed part. In order to construct the operators \mathbf{S} and \mathbf{A} , we first construct the two diffusion operators, $\mathbf{P}^{(\ell)}$ and $\mathbf{Q}^{(\ell)}$, for each manifold $\ell = 1, 2$, as described in (4.27). Here, d_X and d_Y are the geodesic distances on the shapes, calculated based on the heat flow constructed for the triangulated meshes [10]. Note that the geodesic distance captures the geometry of the shape better than the Euclidean distance. However, in this simple example, using the Euclidean distances between the 3D points leads to comparable results. We then construct the symmetric and antisymmetric discrete operators, \mathbf{S} and \mathbf{A} , respectively, according to (4.10) and (4.11). Finally, the eigenvalue decompositions of \mathbf{S} and \mathbf{A} are calculated, and the eigenvectors are sorted according to the imaginary part of the eigenvalues in descending order. The kernel scales, ϵ_1 and ϵ_2 in (4.26), were set to be the median of the distances, $d_X(x_i, x_j)$ and $d_Y(y_i, y_j)$, respectively, divided by some scalar. In the construction of \mathbf{S} , ϵ_1 was set to $\text{median}(d_X(x_i, x_j))/2$ and ϵ_2 to $\text{median}(d_Y(y_i, y_j))/2$, and in the construction of \mathbf{A} , ϵ_1 was set to $\text{median}(d_X(x_i, x_j))/5$ and ϵ_2 to $\text{median}(d_Y(y_i, y_j))/5$. In this example and in the examples in section 6, the values of the kernel scales, ϵ_1 and ϵ_2 , were set to $\text{median}(d_X(x_i, x_j))$ and $\text{median}(d_Y(y_i, y_j))$, respectively, times a scale factor $c > 0$, which was chosen by trial and error. On the one hand, the choice of c was motivated by the discrete analysis presented in subsection 4.1, which showed that for small kernel scales, i.e., a small factor c , the operator \mathbf{A} will be supported on $V_{\Omega_\alpha^c}$. On the other hand, if the kernel scales are too small, the obtained embedding may contain significant outliers or degenerate to a line. Therefore, in our simulations, we decreased the parameter c up to the point in which the embedding started to degenerate.

Figure 3 presents the sphere (top plots) and the bump (bottom plots), colored by the eigenvectors of the operators \mathbf{S} and \mathbf{A} . Plots (a) and (d) are colored by the first 4 eigenvectors of \mathbf{S} . Plots (b) and (e) are colored by the real part of the first 4 eigenvectors of \mathbf{A} , and plots (c) and (f) are colored by the imaginary part of the first 4 eigenvectors of \mathbf{A} . Note that in both \mathbf{S} and \mathbf{A} , the eigenvector corresponding to the largest eigenvalue (top plot) separates between the location of the deformation and the similar parts of the sphere. The other 3 eigenvectors of \mathbf{S} and \mathbf{A} exhibit different properties. The eigenvectors of \mathbf{S} are supported on the entire sphere and take the form of standard spherical harmonics. Conversely, the eigenvectors of \mathbf{A} (both real and imaginary parts) are supported on the deformed part Ω_α^c and take the form of local standard spherical harmonics there. Namely, the eigenvectors of \mathbf{A} are supported on the regions where the diffeomorphism is nonisometric, and within their support, the “standard” (yet local) harmonic oscillations are obtained.

6. Fetal ECG. In this section, we demonstrate the properties of the proposed operators in a fetal heart activity identification problem from two transabdominal maternal ECG (ta-mECG) contacts. This problem consists of two oscillatory signals: one is the undesired maternal ECG signal, and the other is the desired fetal ECG signal. The two signals are observed by two ECG contacts located on the maternal abdomen. In each contact, a mixture of the two oscillatory signals is captured. Based on the physiological properties, we assume that both observations capture the same view of the maternal ECG signal, since the source

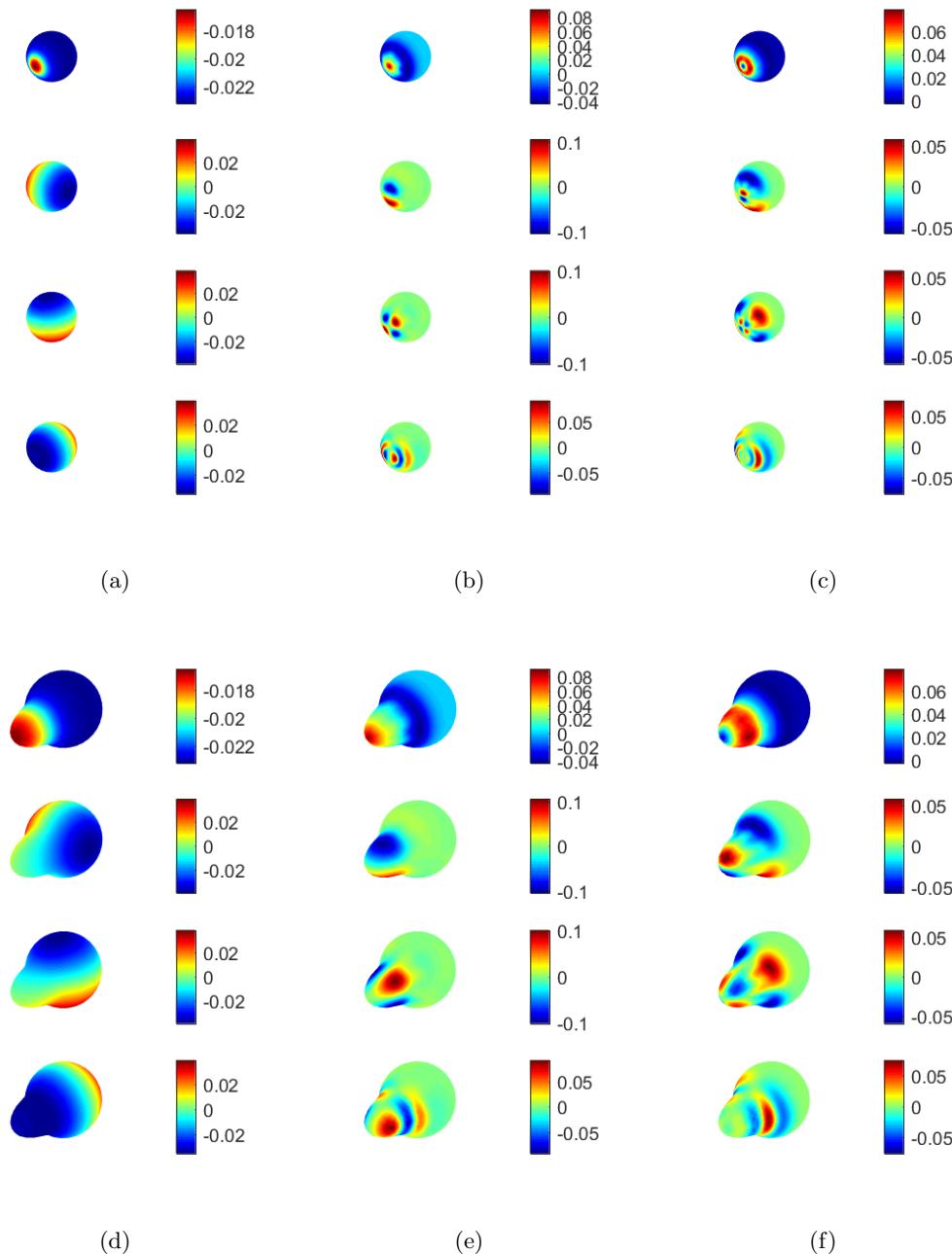


Figure 3. Application of operators \mathbf{S} and \mathbf{A} to two diffeomorphic manifolds: a sphere (plots (a)–(c)) and a deformed sphere (plots (d)–(f)). The plots in the first row depict the sphere colored by (a) the eigenvectors of the symmetric operator \mathbf{S} , (b) the real part of the eigenvectors of the antisymmetric operator \mathbf{A} , and (c) the imaginary part of the eigenvectors of the antisymmetric operator. The coloring of the plots in the second row, (d)–(f), corresponds to the coloring of plots (a)–(c).

of the maternal ECG is located remotely from the two abdominal contacts. Conversely, we assume that the two observations capture different views of the fetal ECG signal, since its source is located close to each of the contacts.

Fetal heart rate (fHR) provides significant information about fetal health. For example, fetal distress monitoring can be obtained through fHR analysis [23]. In recent years, analyzing how fHR fluctuates has attracted increasing attention due to its potential to enhance our understanding of the dynamics of various physiological systems, as well as to contribute to clinical procedures, e.g., inflammation detection [14]. Obtaining intrapartum fHR noninvasively is not an easy task. Traditionally, cardiotocogram is the standard tool to obtain the fHR. However, it is well known that the cardiotocogram does not have a sufficiently high sampling rate for fHR fluctuation analysis. In the past decades, studies have focused on obtaining the fHR through the ta-mECG, due to the high sampling rate of the ECG. See, for example, [1, 30] and references therein. However, to date, while many algorithms and products based on multiple channels (more than 4) have been proposed, there is no gold standard that works in all situations when there are only one or two channels. While we do not presume to provide a state-of-the-art algorithm, in this section, we show the potential of the operator \mathbf{A} in extracting the fHR from two ta-mECG signals.

The section is structured as follows. In subsection 6.1, we present our basic geometric model of the problem to justify the application of the operator \mathbf{A} . Results on simulation data are presented in subsection 6.2 and on real measured data in subsection 6.3.

6.1. Model. Let $s^{(\ell)}(t)$, $\ell = 1, 2$, be the measured signal at the first and second ta-mECG leads, given by

$$s^{(\ell)}(t) = m^{(\ell)}(t) + f^{(\ell)}(t),$$

where $f^{(\ell)}(t)$ and $m^{(\ell)}(t)$ denote the fetal and maternal ECG signals, respectively. The signal $f^{(\ell)}(t)$ (resp., $m^{(\ell)}(t)$) consists of a (quasi-)periodic oscillation representing the fetal (resp., the maternal) heart beat. “Quasi” here indicates that the heart rate and ECG morphology change occasionally. To simplify the discussion, we assume that the relationship between the two (separate) cardio systems entails that the maternal and fetal ECG signals are approximately perpendicular in short time periods, i.e.,

$$(6.1) \quad \int_I m^{(\ell)}(t) f^{(\ell)}(t) dt \approx 0,$$

for all time intervals I of length 1 second. Note that this is an oversimplified model motivated by the fact that the maternal heart rate is about 1Hz and the fetal and maternal heart beats are not synchronized. Indeed, when the QRS complexes of the maternal and fetal ECG overlap, this assumption may not hold.

Using lag map embedding, the measured signals can be written as

$$\mathbf{s}^{(\ell)}(t) = \mathbf{m}^{(\ell)}(t) + \mathbf{f}^{(\ell)}(t) \in \mathbb{R}^p,$$

where $\mathbf{s}^{(\ell)}(t) = [s^{(\ell)}(t), T s^{(\ell)}(t), \dots, T^{p-1} s^{(\ell)}(t)]$, T denotes an operator that propagates $s^{(\ell)}(t)$ one time step forward, and p is the number of time steps in the lag map embedding of each time interval I .

Let $\mathcal{E}^{(1)} \subset \mathbb{R}^p$ and $\mathcal{E}^{(2)} \subset \mathbb{R}^p$ be the embedding of $\mathbf{s}^{(1)}(t)$ and $\mathbf{s}^{(2)}(t)$ in I , respectively. By assumption (6.1), we can write

$$\mathcal{E}^{(1)} = \mathcal{M}^{(1)} \oplus \mathcal{F}^{(1)},$$

where $\mathcal{F}^{(1)}$ and $\mathcal{M}^{(1)}$ are the manifolds underlying $\mathbf{f}^{(1)}(t)$ and $\mathbf{m}^{(1)}(t)$ in I , respectively. Similarly, let $\mathcal{F}^{(2)}$ and $\mathcal{M}^{(2)}$ be the manifolds underlying $\mathbf{f}^{(2)}(t)$ and $\mathbf{m}^{(2)}(t)$ in I , respectively.

As described at the beginning of section 6, the locations of the two abdominal leads entail that $\mathbf{m}^{(1)}(t) \approx \mathbf{m}^{(2)}(t)$, while $\mathbf{f}^{(1)}(t)$ and $\mathbf{f}^{(2)}(t)$ are different. As a result, $\mathcal{M}^{(1)} \approx \mathcal{M}^{(2)}$ and the diffeomorphism between $\mathcal{E}^{(1)}$ and $\mathcal{E}^{(2)}$ can be modeled as

$$\mathcal{E}^{(2)} = \phi(\mathcal{E}^{(1)}) = \phi(\mathcal{M}^{(1)} \oplus \mathcal{F}^{(1)}) = \mathcal{M}^{(1)} \oplus \tilde{\phi}(\mathcal{F}^{(1)}),$$

where $\tilde{\phi} : \mathcal{F}^{(1)} \rightarrow \mathcal{F}^{(2)}$ is a smooth diffeomorphism.

Define $\mu^{(\ell)}$ as the probability density on $\mathcal{E}^{(\ell)}$, $\mu_m^{(\ell)}$ as the marginal density of $\mu^{(\ell)}$ on $\mathcal{M}^{(\ell)}$, and $\mu_f^{(\ell)}$ as the marginal density of $\mu^{(\ell)}$ on $\mathcal{F}^{(\ell)}$.

Corollary 6.1. Define $A_\alpha = \lim_{\epsilon \rightarrow 0} A_{\epsilon_1, \epsilon_2} / \epsilon^2$, where $\epsilon_2 = \alpha \epsilon$ and $\epsilon_1 = \epsilon$, $\alpha > 0$. For all $g \in C^\infty(\mathcal{E}^{(1)})$, if $\text{supp } g \subset \mathcal{M}^{(1)} \oplus \tilde{\Omega}_{f,\alpha}$, then $A_\alpha g = 0$. Hence, if $A_\alpha g = \lambda g$, $g \neq 0$, then $\text{supp } g \subset \mathcal{M}^{(1)} \oplus \Omega_{f,\alpha}^c$.

According to this corollary, the eigenfunctions of the operator A_α are supported on the differences. We assume that the differences in the measured fetal ECG signals are manifested mainly during heart activity, i.e., depolarization (QRS complex and P wave) and repolarization (T wave). Therefore, based on the model presented in this subsection, the eigenfunctions of A_α can serve as indicators for fetal heart activity. In addition, the common component in this model, i.e., the maternal heart activity, $\mathbf{m}^{(1)}(t)$ and $\mathbf{m}^{(2)}(t)$, represents an almost periodic oscillation. By the Takens embedding theorem, the manifolds underlying such signals can be well recovered, up to a diffeomorphism, by a 1D manifold that is diffeomorphic to S^1 . Therefore, we expect that the eigenfunctions of the operator S_α will represent S^1 .

6.2. Fetal heart rate detection—synthetic example. In this subsection, we begin with a synthetic problem setting of fetal ECG detection to demonstrate the main properties of our composite operators for such data.

Following the model described in subsection 6.1, we create synthetic data of two ta-mECG leads from three different ECG recordings, denoted by $z_i^{(1)}$, $z_i^{(2)}$, and $z_i^{(3)}$, where $i = 1, \dots, N$ and $N = 4 \times 10^4$ is the number of samples. These recordings are taken from the QT database in Physionet [26, 21], which contains annotated 2-lead ECG recordings, sampled at 250Hz. The signals $z_i^{(2)}$ and $z_i^{(3)}$ are taken from the same recording, i.e., taken from two corresponding ECG leads which were recorded simultaneously. These recordings were filtered by a notch filter to remove the 60Hz net noise and by a median filter (with a window size of 100 samples) to remove the baseline drift. In order to obtain more samples per heart cycle, we increased the number of samples in these recordings using interpolation. One recording, $z_i^{(1)}$, represents the maternal ECG and is upsampled by a factor of 4. The other two recordings, $z_i^{(2)}$ and $z_i^{(3)}$, represent the fetal ECG, which commonly has a higher heart rate, and therefore they

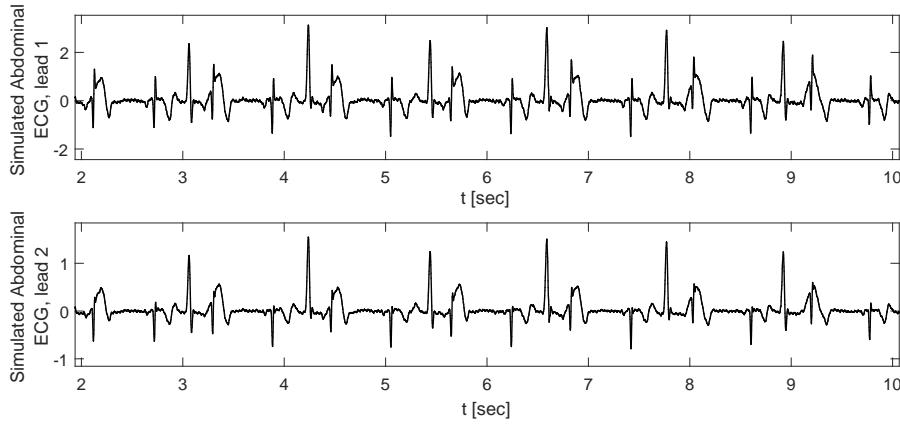


Figure 4. Two simulated ECG leads representing two ta-mECG recordings.

are upsampled by a factor of 2. The simulated ta-mECG signals $s_i^{(1)}$ and $s_i^{(2)}$ are generated according to

$$(6.2) \quad s_i^{(1)} = 2z_i^{(1)} - z_i^{(2)},$$

$$(6.3) \quad s_i^{(2)} = z_i^{(1)} - 0.5z_i^{(3)},$$

where the common maternal ECG $z_i^{(1)}$ is identical up to a scaling factor. In these simulated signals, $z_i^{(1)}$, which is denoted by $m^{(\ell)}(t)$ in subsection 6.1, is assumed to be part of the common structure, whereas the fetal ECG signals $z_i^{(2)}$ and $z_i^{(3)}$ are captured differently by the two abdominal leads. With regard to the model described in subsection 6.1, the fetal ECG signals $z_i^{(2)}$ and $z_i^{(3)}$, denoted there by $f^{(\ell)}(t)$, undergo a diffeomorphism, which mainly distorts the higher values in the signal—the QRS complexes. Therefore, in this example, $\Omega_{f,\alpha}^c$ describes these QRS complexes, and we expect the eigenvectors of operator \mathbf{A} to be supported there. Figure 4 presents an example for the resulting simulated ta-mECG leads.

Using the simulated signals described above, we illustrate some of the properties of operators \mathbf{S} and \mathbf{A} . We construct these operators according to Algorithm 4.1. First, a lag map is constructed from each signal, $s_i^{(\ell)}$, $\ell = 1, 2$, in windows of 12 samples and with an overlap of 6 samples, in order to obtain a better representation of the data. Denote the lag map of signal ℓ by $s_{i,lag}^{(\ell)}$. Second, an affinity matrix is constructed for each signal according to (4.26) by treating each time frame (lag) as one sample, denoted by x_i or y_i in (4.26). The affinity matrices were constructed using the Euclidean distances between the samples, i.e., $d(x_i, x_j) = \|x_i - x_j\|_2$, and the kernel scales, ϵ_1, ϵ_2 , were set to be the median of the distances, which is common practice. Third, operators $\mathbf{Q}^{(\ell)}$ and $\mathbf{P}^{(\ell)}$, $\ell = 1, 2$, are constructed for both $s_{i,lag}^{(1)}$ and $s_{i,lag}^{(2)}$ according to (4.27). Finally, the operators \mathbf{S} and \mathbf{A} are constructed as in (4.28).

In Figure 5, scatter plots of the second and third eigenvectors of operators \mathbf{S} and \mathbf{A} are presented and compared to the eigenvectors of diffusion maps applied to each channel separately. Note that the choice to present the second and third eigenvectors of \mathbf{S} and \mathbf{A} is

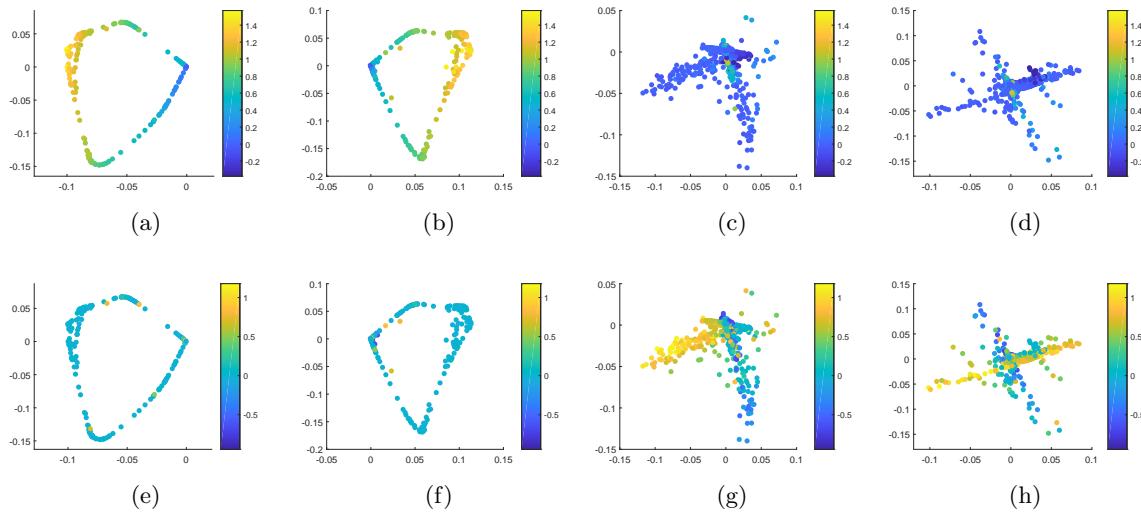


Figure 5. Synthetic fHR detection example. Presenting the second and third eigenvectors of operator \mathbf{S} (plots (b) and (f)), the imaginary parts of the second and third eigenvectors of operator \mathbf{A} (plots (c) and (g)), and the real parts of the second and third eigenvectors of operator \mathbf{A} (plots (d) and (h)), compared with the first and second (nontrivial) eigenvectors of diffusion maps for ECG lead 1 (plots (a) and (e)). The plots are colored according to the maternal ECG in the top row, and according to the fetal ECG in the bottom row.

motivated by the result in section 5, where the respective first eigenvectors of \mathbf{S} and \mathbf{A} are similar and only represent the support of the nonisometric parts between the two manifolds. In this figure, plots (a) and (e) depict 2 eigenvectors (corresponding to the largest nontrivial eigenvalues) of diffusion maps, constructed based on the ECG lead $s_{i, \text{lag}}^{(1)}$. Plots (b) and (f) depict the second and third eigenvectors of the operator \mathbf{S} , plots (c) and (g) depict the imaginary part of the second and third eigenvectors of the operator \mathbf{A} , and plots (d) and (h) depict the real part of the second and third eigenvectors of the operator \mathbf{A} . The plots in the first row (plots (a), (b), (c), and (d)) are colored according to the maternal ECG $z_i^{(1)}$, and the plots in the second row (plots (e), (f), (g), and (h)) are colored according to one of the fetal ECG signals $z_i^{(3)}$.

These plots show that in the eigenvectors of \mathbf{A} the fetal ECG is significantly emphasized, compared with the eigenvectors of \mathbf{S} and the diffusion maps embedding of the two channels. Furthermore, both the ECG lead $s_{i, \text{lag}}^{(1)}$ and the operator \mathbf{S} , which mainly describe the (more dominant) maternal ECG signal, lead to an embedding that corresponds to an embedding of \mathcal{S}^1 , as can be seen in plots (a), (b), (e), and (f). This strengthens the model described in subsection 6.1, in which the underlying manifolds $\mathcal{E}_m^{(1)}$ and $\mathcal{E}_m^{(2)}$ are diffeomorphic to \mathcal{S}^1 . In contrast, the eigenvectors of \mathbf{A} describe a different structure, since the difference between the ECG leads, $\Omega_{f,\alpha}^c$, is only a subset of $\mathcal{E}_f^{(1)}$ and $\mathcal{E}_f^{(2)}$. We note that the second ECG lead, $s_{i, \text{lag}}^{(2)}$, led to plots that are similar to plots (a) and (e) in Figure 5 depicting the first ECG lead and were omitted for brevity.

Figure 6 presents a short simulated ta-mECG segment from lead $s_i^{(1)}$, containing both fetal and maternal components. Plots (a) and (b) are colored by an index vector, containing ones

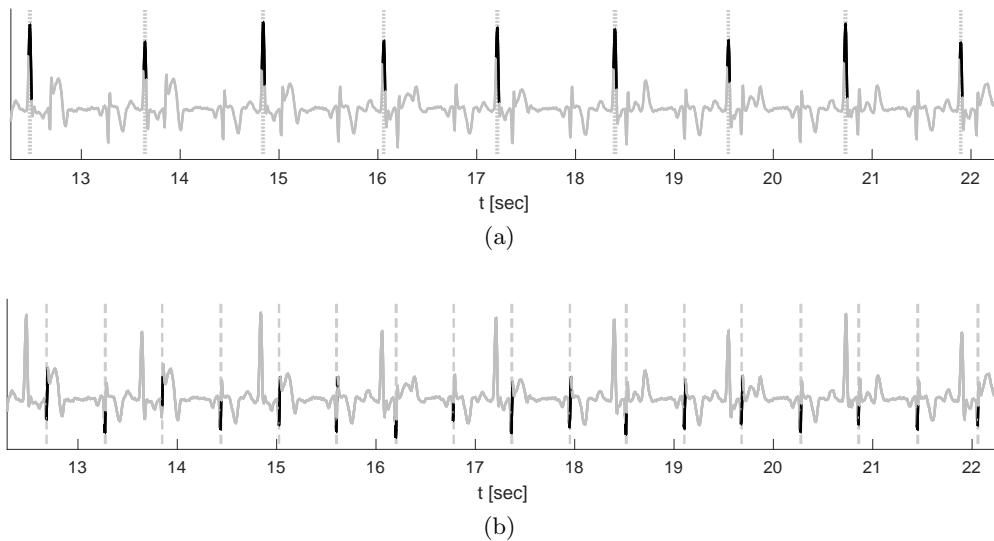


Figure 6. *ta-mECG segment from lead $s_i^{(1)}$, colored by (a) the second eigenvector of operator \mathbf{S} , (b) the imaginary part of the second eigenvector of operator \mathbf{A} . The vertical dotted lines in plot (a) mark the locations of the true maternal heart beats and the vertical dashed lines in plot (b) mark the locations of the true fetal heart beats.*

(colored in black) where the absolute value of the considered eigenvector exceeds a certain threshold and zeros (colored in gray) elsewhere. In plot (a), the segment is colored according to the second eigenvector of \mathbf{S} with a threshold of 10^{-2} ; i.e., locations in which the eigenvector exceeds the threshold are colored in black. In plot (b), the segment is colored according to the imaginary part of the second eigenvector of \mathbf{A} with a threshold of 3×10^{-2} . The dotted vertical gray lines in plot (a) mark the locations of the true maternal beats, and the dashed vertical gray lines in plot (b) mark the locations of the true fetal beats. These plots further demonstrate that \mathbf{A} reveals the fetal beat locations, as the fetal heart beat morphologies are captured differently by the two synthetic leads. In addition, the eigenvectors of \mathbf{A} are supported mainly on the fetal QRS complexes, as assumed in the model presented in subsection 6.1.

The MATLAB code of this synthetic example is available on [GitHub](#).

6.3. Fetal heart rate detection—real data. Following the synthetic example in subsection 6.2, we address fHR detection from real ta-mECG recordings and propose to extract the fHR by constructing the operator \mathbf{A} based on two ta-mECG leads. Similarly to the synthetic example, we expect that the operator \mathbf{A} will provide a new representation of the signals that emphasizes the fetal beats.

We validate our approach using the publicly available database of ta-mECG signals, [2013 PhysioNet/Computing in Cardiology Challenge](#), abbreviated as CinC2013. We focus on the set A, which consists of 75 recordings, each of length 1 minute with R peak annotation and with reference to a ground-truth fetal ECG signal, acquired from an invasive fetal scalp electrode. Each recording includes four noninvasive ta-mECG channels recorded from multiple positions using different electrodes (with possibly different configuration). The recordings are resampled at 1000Hz. The lead placements on the maternal abdomen and the fetal/maternal

health status are unknown. We disregard recording number 54 since it was excluded by the Challenge organizers [2]. In addition, we disregard recordings 33, 38, 47, 52, 71, and 74, since they contain inaccurate reference fetal annotations, as identified by [4]. We focus on the remaining 68 recordings.

We first perform a preprocessing stage for each signal, which includes a low-pass filter (below 100Hz), trend removal (median filtering with a window size of 101 samples), and computation of a lag map with a window of 8 samples and a 7-sample overlap. After the preprocessing step, in the first stage of the proposed algorithm, we construct the forward and backward diffusion operators, $\mathbf{P}^{(\ell)}, \mathbf{Q}^{(\ell)}$, $\ell = 1, 2$, from the lag map of the two ta-mECG leads, and we compute the operator \mathbf{A} based on (4.11). In this application, the two kernel scales, ϵ_1 and ϵ_2 , were set to 3 times the medians of the distances, where the factor $c = 3$ was chosen by trial and error. This example required larger kernel scales due to noise and due to the structural properties of the ECG signals. Taking smaller kernel scales in this example led to significant outliers for some of the subjects in the database.

The eigenvectors of the operator \mathbf{A} are computed and sorted as described in subsection 6.2. The deshape Short Time Fourier Transform (dsSTFT) [31] is then applied to the real and imaginary parts of each of the first 20 eigenvectors of \mathbf{A} , resulting in 40 spectrograms, depicting the dominant frequencies in each eigenvector. The median (pixelwise) over all of the dsSTFT spectrograms is taken as a new spectrogram for each subject, depicting both the fetal and maternal instantaneous heart rates. This can be viewed as a variation of the recently developed generalized multitaper approach for time-frequency analysis, called concentration of frequency and time [11]. Here, we use the eigenvectors of \mathbf{A} , which capture the oscillatory behavior of the signal, instead of the multiple windows in [11]. An example for such a spectrogram is presented in Figure 7. In plots (a) and (b), the dsSTFT of the two ta-mECG leads are presented. The thick black line in these two plots represents the maternal heart rate. In plot (c), the median spectrogram of the eigenvectors of \mathbf{A} is presented. In this plot, the red arrow marks the location of the maternal heart rate line and the blue arrow marks the location of the fHR line. Plot (d) depicts the same spectrogram as plot (c), along with the ground truth of the fetal heart rate, marked by a dotted blue line. Plots (c) and (d) demonstrate that the operator \mathbf{A} leads to a result which significantly emphasizes the true fHR, compared with the original ta-mECG signals.

In the next stage of the algorithm, the fHR is extracted from the spectrogram, presented in Figure 7(c). This is performed by first obtaining the maternal heart rate from the dsSTFT of the original ta-mECG signals (plots (a) and (b)) and removing its curve from the spectrogram of the operator \mathbf{A} . Second, the most dominant curve in the remaining spectrogram is extracted using the algorithm described in [30]. This curve is assumed to represent the fHR. In order to extract the fetal ECG and the beat locations, we continue the analysis as described in [30], after the dsSTFT stage.

The algorithm we applied to the ta-mECG leads is summarized in Algorithm 6.1.

For performance evaluation we consider the F1 score, which is the harmonic mean of the sensitivity (SE) and the positive predictive value (PPV), similarly to [30]. The true positive (TP), false positive (FP), and false negative (FN) measures, used in the calculation of SE and PPV, were defined using a window of 50ms; i.e., a true positive classification means that an estimated beat is located within a window of 50ms around a true beat from the

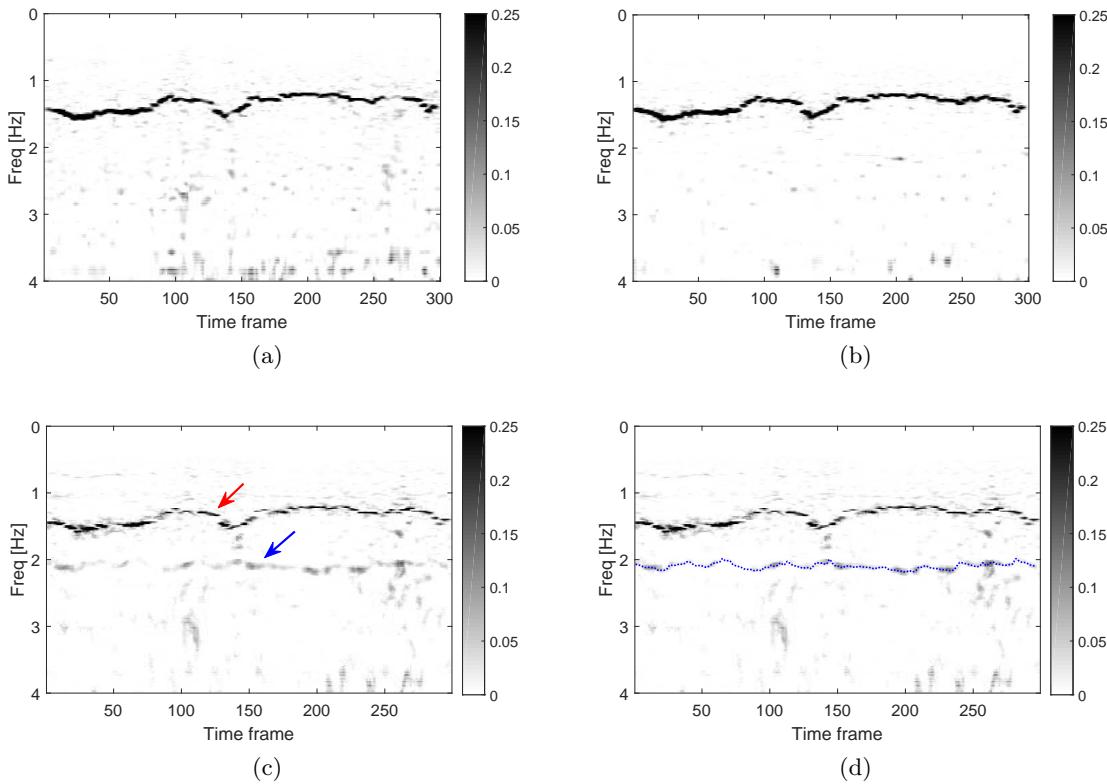


Figure 7. Plots (a) and (b) present the spectrograms (after dsSTFT) of the ECG signal from two abdomen leads. The visually dominant frequency is the maternal ECG. Plot (c) presents the spectrogram extracted from the antisymmetric diffusion operator \mathbf{A} (applied to the two abdomen signals). In this plot, the hidden fetal ECG (marked by a blue arrow) is significantly enhanced. Plot (d) depicts the same spectrogram as plot (c), as well as the ground-truth fHR (marked by a dotted blue line).

provided annotations. We report the results of ta-mECG lead pair 1 and 4, which provided the best performance, out of the 6 possible pairs, for all algorithms. In addition, in the above performance measures, to avoid the boundary effect, the first and last 2 seconds in every recording are not evaluated.

Table 1 presents the F1 results obtained by the operators \mathbf{A} and \mathbf{S} using Algorithm 6.1, as well as reference results obtained based on the ta-mECG leads by applying the filtering in the preprocessing stage and principal component analysis (PCA). The application of PCA includes the extraction of the first two principal components, i.e., obtaining a rotation of the channels, denoted by $s_{PCA}^{(1)}$ and $s_{PCA}^{(2)}$. The state-of-the-art results of [30] are also presented in Table 1 for comparison. The results of [30] are reported for the same ta-mECG lead pair (1 and 4) and the same 68 recordings used in our experiments. To obtain the PCA-based reference results, the dsSTFT was applied to $s_{PCA}^{(1)}$ and $s_{PCA}^{(2)}$ and the algorithm continued as described in Algorithm 6.1, step 3. For the operator \mathbf{A} , Algorithm 6.1 is applied as is, whereas for operator \mathbf{S} , the only modification is the use of operator \mathbf{S} instead of \mathbf{A} . The

Algorithm 6.1. Fetal ECG extraction using operator **A**.

1. **Preprocessing** – Low-pass filtering below 100Hz, trend removal (median filtering), and computation of a lag map with a window of 8 samples (7-sample overlap) for each ta-mECG lead. Denote the resulting signals by $s_{i, \text{lag}}^{(\ell)}$, $\ell = 1, 2$, where $i = 1, \dots, N$ and $N = 6 \times 10^4$ is the number of samples in each ta-mECG recording.
2. **Fetal Instantaneous Heart Rate Detection**
 - (a) Construct the operator **A** from the two ta-mECG leads, $s_{i, \text{lag}}^{(1)}$ and $s_{i, \text{lag}}^{(2)}$, and compute its eigenvectors $\psi_k^{(\mathbf{A})}(i)$ (sorted as in subsection 6.2).
 - (b) Apply deshape Short Time Fourier Transform (dsSTFT) [31] to the real and imaginary parts of $\{\psi_k^{(\mathbf{A})}(i)\}_{k=1}^{20}$, separately.
 - (c) Take the median over all the resulting spectrograms of the dsSTFT of $\{\psi_k^{(\mathbf{A})}(i)\}_{k=1}^{20}$.
3. **Maternal ECG Removal and Fetal ECG Estimation** – Continue similarly to the algorithm described in [30], after the dsSTFT stage, using the spectrogram obtained in step 2c above.

Table 1

*F1 measure for the CinC2013 dataset calculated using Algorithm 6.1 with the operators **A** and **S**, compared to PCA applied directly to the ta-mECG signals and to the state-of-the-art results of [30]. Note that here we report the results of [30] for the same 68 recordings used in our experiments [30, footnote 6 after Table 3].*

	Mean(F1)%	STD(F1)%	median(F1)%	IQR(F1)%
A	82.74	28.37	98.41	12.7
S	78.7	28.66	97.86	46.31
$s_{PCA}^{(1)}$	73.01	29.95	94.13	57.99
$s_{PCA}^{(2)}$	78.59	27.84	97.02	49.09
[30]	89.81	20.84	98.41	5.1

mean, standard deviation (STD), median and interquartile range (IQR) over the F1 values of the 68 subjects are presented. This table depicts that the operator **A** extracts significant information related to the fHR from the ta-mECG leads. It improves the results obtained by using the ta-mECG signals after applying PCA, $s_{PCA}^{(1)}$ and $s_{PCA}^{(2)}$.

Note that the comparison of the embedding obtained by **A** to the PCA of the channels shows that a simple decomposition of the channels to their principal components does not capture the fetal ECG and demonstrates the role of the operator **A**.

We further note that the state-of-the-art results, obtained by [30], outperform our results as presented in Table 1. In [30], the maternal ECG is first removed and then only the remaining fetal ECG is processed. For comparison, we applied the proposed operators, **A** and **S**, after first removing the maternal ECG in a similar manner. This led to improved results, which are closer to the state-of-the-art results. For **A**, the median and IQR of the F1 measure in this case were 98.5% and 6.7%, respectively, and for **S**, they were 98.5% and 6.2%, respectively. The mean and STD of the F1 measure were $87.3\% \pm 23.8\%$ for **A** and $87.1\% \pm 24.4\%$ for **S**. In this setting, the performance of operator **S** is comparable to operator **A**, whereas in Table 1,

the operator \mathbf{A} leads to significantly better performance. These results further demonstrate the properties of the proposed operators. In the latter case, removing the maternal ECG causes the fetal ECG to become the dominant *common* component in the two signals, which leads to its identification by the operator \mathbf{S} . In addition, since each ECG lead captures a different view of the fetal ECG it is still revealed by operator \mathbf{A} as well.

While the reported performance does not outperform the state-of-the-art method based on two channels reported in [30], these results support the potential of the antisymmetric operator in fHR extraction, which was demonstrated by the synthetic example in subsection 6.2 as well. In addition, one advantage of our method over [30] is that our method can be applied directly to the ta-mECG channels, prior to the removal of the maternal ECG waveform, which might distort the remaining signal and introduce additional noise.

One of the reasons for the degraded performance of the operator \mathbf{A} in the real application, compared with the synthetic example in subsection 6.2, is that the database is composed of heterogeneous signals—it is recorded using different machines, includes pregnant women of different gestational ages, uses different lead placements, contains different noise levels, etc. (all of which are unknown to us). In addition, the presence of significant noise in some ECG leads hampers the performance, since the noise is a part of the difference component (different between the two ECG leads) and therefore is captured by \mathbf{A} . Indeed, Table 1 demonstrates our method's sensitivity to noise and data heterogeneity, as the median F1 score of \mathbf{A} is similar to the state of the art [30], but the mean F1, STD, and IQR are higher. Therefore, based on these results and on the synthetic example in subsection 6.2, we believe that an additional extensive noise suppression stage may improve the performance of the proposed algorithm and lead to results which are closer to the state of the art.

7. Other related operators. The proposed operators are related to recent work, most of which concern the recovery of common structures from different views, i.e., acquired by different modalities, or from different time frames, similarly to the symmetric operator \mathbf{S} . Such methods include the previously mentioned alternating diffusion [29, 44], the dynamic Laplacian [18, 19], cross-diffusion [45], and the minimizing-disagreement algorithm [13]. One related work addressing the recovery of differences between shapes, similarly to \mathbf{A} , is presented in [38]. In this section, we present a short overview of some of these related operators and discuss their connection to our work.

Most methods that address the recovery of common structures rely on operator composition. For example, the dynamic Laplacian [18] focuses on recovering coherent sets in dynamical system, which can be modeled as the common structures in a set of manifolds, each representing a different time frame. The dynamic Laplacian operator is constructed from the composition of an operator with its adjoint, $\mathcal{L}_\epsilon^* \mathcal{L}_\epsilon$, where $\mathcal{L}_\epsilon = P_{2,\epsilon} R P_{1,\epsilon}$, and $P_{1,\epsilon}$ is a smoothing (diffusion) operator of the manifold corresponding to the first time frame, $P_{2,\epsilon}$ is a smoothing (diffusion) operator of the second time frame, and R is the Perron–Frobenius operator representing the system dynamics. In the context of our work, the operators $P_{1,\epsilon}$ and $P_{2,\epsilon}$ are analogous to the operator $P_\epsilon^{(\ell)}$. In [18] it is shown that this operator has a spectrum and converges to the sum of the Laplace–Beltrami operators of the manifolds representing the two time frames, similarly to S_ϵ as noted in subsection 3.3. In the cross-diffusion algorithm [45] two diffusion operators, $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$, and their transpose, $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$, are

constructed (as in (4.4) and (4.5), respectively) based on two different metrics. A fusion of these metrics is then obtained by $[\mathbf{P}_{t+1}^{(1)} + \mathbf{P}_{t+1}^{(2)}]/2$, where $t > 0$ and $\mathbf{P}_{t+1}^{(1)} := \mathbf{P}^{(1)}\mathbf{P}_t^{(2)}\mathbf{Q}^{(1)}$ and $\mathbf{P}_{t+1}^{(2)} := \mathbf{P}^{(2)}\mathbf{P}_t^{(1)}\mathbf{Q}^{(2)}$ are constructed iteratively. Note that, similarly to \mathbf{S} and \mathbf{A} , this construction is also based on the composition of forward and backward operators, i.e., $\mathbf{Q}^{(\ell)}$ and $\mathbf{P}^{(\ell)}$, respectively. Two other papers [15, 32] proposed methods for constructing a low-dimensional representation of multiple views, based on kernel multiplication and construction of a symmetric operator in [32] and based on joint diagonalization of the Laplacians of different views in [15]. Similarly to our method, these papers rely on spectral decompositions to obtain the new embedding. The proposed operator \mathbf{S} is also closely related to the work presented in [6], where the common variable is recovered from multimodal data, where each modality contains different i.i.d. noise. In our setting, this i.i.d. noise can be modeled as the sensor specific components. In contrast to the operator in [6], the operator \mathbf{S} recovers a *new representation* of the common variable (instead of the common variable itself), but we show that this can be achieved in more complicated scenarios, i.e., including various sensor-specific components rather than only i.i.d. noise.

All of the above methods address the recovery of the common components only and ignore the differences. Therefore, compared with these operators, the novelty in the current work is the introduction of the difference revealing operator \mathbf{A} . A similar notion of difference characterization between manifolds was previously presented in [38]. There, a new linear operator for comparison of shape deformations was proposed, which provides a mapping between the shapes, and was shown to distort functions on the shapes only in areas where the shapes differ. This operator was constructed as a composition of operators representing the two shapes, with one of the operators inverted, e.g., $(\mathbf{H}^{(1)})^{-1}\mathbf{F}^T\mathbf{H}^{(2)}\mathbf{F}$, where $\mathbf{H}^{(\ell)}$ denotes a matrix representing the inner product on shape ℓ and \mathbf{F} is the functional map between the two shapes. In the context of our work, the operator \mathbf{F} is analogous to the diffeomorphism, ϕ , and the operator $\mathbf{H}^{(\ell)}$ is analogous to the operator $P_\epsilon^{(\ell)}$. In contrast to the proposed self-adjoint operator \mathbf{A} , this shape difference operator does not necessarily have a spectrum and depends on the order of the operator composition. In addition, in the discrete setting, it requires the inverse (or pseudoinverse) of a possibly large matrix.

Other operators for recovering differences between manifolds can be considered. For example, $\hat{\mathbf{A}} = (\mathbf{P}^{(1)} - \mathbf{P}^{(2)}) (\mathbf{P}^{(1)} - \mathbf{P}^{(2)})^T$ is a symmetric operator which obtains comparable results in the experimental results in sections 5 and 6. However, this operator was not considered in the current paper since in the asymptotic expansion of this operator, in contrast to A_ϵ , the second order terms, of order $O(\epsilon^2)$, cancel out and only fourth order terms and above remain. In the future, we plan to extend this work and explore such additional operators for recovering hidden components of multimodal data, and to create a “library” of operators. We plan to use this library of operators and construct a framework for characterizing the common and the difference structures in multimodal data or in data which lie on a time-evolving manifold. Moreover, we plan to devise a multiresolution analysis framework for time-varying manifolds based on such a library of operators, which can be seen as analogous to the wavelet analysis under the manifold setting.

Note the assumption hidden in both the composite operators \mathbf{S} and \mathbf{A} and in the presented alternative operator, $\hat{\mathbf{A}}$. The addition and subtraction operations in the composition imply

that the operators lie in a linear Euclidean space, which may violate the Riemannian structure of the operators. In future work, we plan to address this issue and investigate different ways of composing such operators using non-Euclidean settings.

REFERENCES

- [1] F. ANDREOTTI, J. BEHAR, S. ZAUNSEDER, J. OSTER, AND G. D. CLIFFORD, *An open-source framework for stress-testing non-invasive foetal ECG extraction algorithms*, Physiol. Meas., 37 (2016), pp. 627–648.
- [2] F. ANDREOTTI, M. RIEDL, T. HIMMELSBACH, D. WEDEKIND, N. WESSEL, H. STEPAN, C. SCHMIEDER, A. JANK, H. MALBERG, AND S. ZAUNSEDER, *Robust fetal ECG extraction and detection from abdominal leads*, Physiol. Meas., 35 (2014), pp. 1551–1567.
- [3] G. ANDREW, R. ARORA, J. BILMES, AND K. LIVESCU, *Deep canonical correlation analysis*, in International Conference on Machine Learning, 2013, pp. 1247–1255.
- [4] J. BEHAR, J. OSTER, AND G. D. CLIFFORD, *Combining and benchmarking methods of foetal ECG extraction without maternal or scalp electrode data*, Physiol. Meas., 35 (2014), pp. 1569–1589.
- [5] M. BELKIN AND P. NIYOGI, *Laplacian eigenmaps for dimensionality reduction and data representation*, Neural Comput., 15 (2003), pp. 1373–1396.
- [6] B. BOOTS AND G. GORDON, *Two-manifold problems with applications to nonlinear system identification*, in Proceedings of the 29th International Conference on Machine Learning (ICML-12), 2012, pp. 33–40.
- [7] R. COIFMAN AND S. LAFON, *Diffusion maps*, Appl. Comput. Harmon. Anal., 21 (2006), pp. 5–30.
- [8] R. R. COIFMAN AND M. J. HIRN, *Bi-stochastic kernels via asymmetric affinity functions*, Appl. Comput. Harmon. Anal., 35 (2013), pp. 177–180.
- [9] R. R. COIFMAN AND M. J. HIRN, *Diffusion maps for changing data*, Appl. Comput. Harmon. Anal., 36 (2014), pp. 79–107.
- [10] K. CRANE, C. WEISCHEDEL, AND M. WARDETZKY, *Geodesics in heat: A new approach to computing distance based on heat flow*, ACM Trans. Graphics, 32 (2013), 152.
- [11] I. DAUBECHIES, Y. G. WANG, AND H.-T. WU, *ConceFT: Concentration of frequency and time via a multitapered synchrosqueezed transform*, Philos. Trans. Roy. Soc. A, 374 (2016), 20150193.
- [12] M. A. DAVENPORT, C. HEGDE, M. F. DUARTE, AND R. G. BARANIUK, *Joint manifolds for data fusion*, IEEE Trans. Image Process., 19 (2010), pp. 2580–2594.
- [13] V. R. DE SA, *Spectral clustering with two views*, in ICML 2005 Workshop on Learning with Multiple Views, 2005, pp. 20–27.
- [14] L. D. DUROSIER, C. L. HERRY, M. CORTES, M. CAO, P. BURNS, A. DESROCHERS, G. FECTEAU, A. J. SEELY, AND M. G. FRASCH, *Does heart rate variability reflect the systemic inflammatory response in a fetal sheep model of lipopolysaccharide-induced sepsis?*, Physiol. Meas., 36 (2015), pp. 2089–2102.
- [15] D. EYNARD, K. GLASHOFF, M. M. BRONSTEIN, AND A. M. BRONSTEIN, *Multimodal Diffusion Geometry by Joint Diagonalization of Laplacians*, preprint, <https://arxiv.org/abs/1209.2295>, 2012.
- [16] D. EYNARD, A. KOVNATSKY, M. M. BRONSTEIN, K. GLASHOFF, AND A. M. BRONSTEIN, *Multimodal manifold analysis by simultaneous diagonalization of Laplacians*, IEEE Trans. Pattern Anal. Mach. Intell., 37 (2015), pp. 2505–2517.
- [17] M. FANUEL, C. M. ALAÍZ, Á. FERNÁNDEZ, AND J. A. SUYKENS, *Magnetic eigenmaps for the visualization of directed networks*, Appl. Comput. Harmon. Anal., 44 (2018), pp. 189–199.
- [18] G. FROYLAND, *Dynamic isoperimetry and the geometry of Lagrangian coherent structures*, Nonlinearity, 28 (2015), pp. 3587–3622.
- [19] G. FROYLAND AND E. KWOK, *A dynamic Laplacian for identifying Lagrangian coherent structures on weighted Riemannian manifolds*, J. Nonlinear Sci., (2017), pp. 1–83.
- [20] F. R. GANTMAKHER, *The Theory of Matrices*, AMS, 1998.
- [21] A. L. GOLDBERGER, L. A. N. AMARAL, L. GLASS, J. M. HAUSDORFF, P. C. IVANOV, R. G. MARK, J. E. MIETUS, G. B. MOODY, C.-K. PENG, AND H. E. STANLEY, *PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals*, Circulation, 101 (2000), pp. e215–e220, <https://doi.org/10.1161/01.CIR.101.23.e215>.

- [22] H. HOTELLING, *Relations between two sets of variates*, Biometrika, 28 (1936), pp. 321–377.
- [23] H. JENKINS, *Thirty years of electronic intrapartum fetal heart rate monitoring: Discussion paper*, J. Roy. Soc. Med., 82 (1989), pp. 210–214.
- [24] A. KUMAR AND H. DAUMÉ, *A co-training approach for multi-view spectral clustering*, in Proceedings of the 28th International Conference on Machine Learning (ICML-11), 2011, pp. 393–400.
- [25] S. LAFON, Y. KELLER, AND R. R. COIFMAN, *Data fusion and multicue data matching by diffusion maps*, IEEE Trans. Pattern Anal. Mach. Intell., 28 (2006), pp. 1784–1797.
- [26] P. LAGUNA, R. G. MARK, A. GOLDBERG, AND G. B. MOODY, *A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG*, in Computers in Cardiology 1997, Lund, Sweden, 1997, pp. 673–676.
- [27] D. LAHAT, T. ADALI, AND C. JUTTEN, *Multimodal data fusion: An overview of methods, challenges, and prospects*, Proc. IEEE, 103 (2015), pp. 1449–1477.
- [28] P. L. LAI AND C. FYFE, *Kernel and nonlinear canonical correlation analysis*, Internat. J. Neural Syst., 10 (2000), pp. 365–377.
- [29] R. R. LEDERMAN AND R. TALMON, *Learning the geometry of common latent variables using alternating-diffusion*, Appl. Comput. Harmon. Anal., 44 (2018), pp. 509–536.
- [30] R. LI, M. G. FRASCH, AND H.-T. WU, *Efficient fetal-maternal ECG signal separation from two channel maternal abdominal ECG via diffusion-based channel selection*, Frontiers Physiol., 8 (2017), 277.
- [31] C.-Y. LIN, L. SU, AND H.-T. WU, *Wave-shape function analysis*, J. Fourier Anal. Appl., 24 (2018), pp. 451–505.
- [32] O. LINDENBAUM, A. YEREDOR, M. SALHOV, AND A. AVERBUCH, *Multiview Diffusion Maps*, preprint, <https://arxiv.org/abs/1508.05550>, 2015.
- [33] N. F. MARSHALL AND M. J. HIRN, *Time coupled diffusion maps*, Appl. Comput. Harmon. Anal., 45 (2018), pp. 709–728.
- [34] T. MICHAELI, W. WANG, AND K. LIVESCU, *Nonparametric canonical correlation analysis*, in International Conference on Machine Learning, 2016, pp. 1967–1976.
- [35] B. NADLER, S. LAFON, R. R. COIFMAN, AND I. G. KEVREKIDIS, *Diffusion maps, spectral clustering and reaction coordinates of dynamical systems*, Appl. Comput. Harmon. Anal., 21 (2006), pp. 113–127.
- [36] Y. RUBNER, C. TOMASI, AND L. J. GUIBAS, *The earth mover’s distance as a metric for image retrieval*, Internat. J. Comput. Vis., 40 (2000), pp. 99–121.
- [37] L. RÜSCHENDORF, *The Wasserstein distance and approximation theorems*, Probab. Theory Related Fields, 70 (1985), pp. 117–129.
- [38] R. M. RUSTAMOV, M. OVSJANIKOV, O. AZENCOT, M. BEN-CHEN, F. CHAZAL, AND L. GUIBAS, *Map-based exploration of intrinsic shape differences and variability*, ACM Trans. Graphics, 32 (2013), 72.
- [39] R. SAMENI AND G. D. CLIFFORD, *A review of fetal ECG signal processing: Issues and promising directions*, Open Pacing Electrophysiol. Ther. J., 3 (2010), pp. 4–20.
- [40] A. SINGER AND H.-T. WU, *Spectral convergence of the connection Laplacian from random samples*, Inf. Inference, 6 (2016), pp. 58–123.
- [41] R. SINKHORN AND P. KNOPP, *Concerning nonnegative matrices and doubly stochastic matrices*, Pacific J. Math., 21 (1967), pp. 343–348.
- [42] R. TALMON, I. COHEN, S. GANNOT, AND 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 (2013), pp. 75–86.
- [43] R. TALMON AND R. R. COIFMAN, *Empirical intrinsic geometry for nonlinear modeling and time series filtering*, Proc. Natl. Acad. Sci., 110 (2013), pp. 12535–12540.
- [44] R. TALMON AND H.-T. WU, *Latent common manifold learning with alternating diffusion: Analysis and applications*, Appl. Comput. Harmon. Anal., (2018).
- [45] B. WANG, J. JIANG, W. WANG, Z.-H. ZHOU, AND Z. TU, *Unsupervised metric fusion by cross diffusion*, in Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012, pp. 2997–3004.
- [46] B. WIDROW, J. M. MCCOOL, M. G. LARIMORE, AND C. R. JOHNSON, *Stationary and nonstationary learning characteristics of the LMS adaptive filter*, Proc. IEEE, 64 (1976), pp. 1151–1162.
- [47] D. ZHOU AND C. J. BURGES, *Spectral clustering and transductive learning with multiple views*, in Proceedings of the 24th International Conference on Machine Learning, ACM, 2007, pp. 1159–1166.