Multimodal manifold analysis by simultaneous diagonalization of Laplacians

Davide Eynard, Artiom Kovnatsky, Michael M. Bronstein, *Senior Member, IEEE,* Klaus Glashoff, Alexander M. Bronstein, *Senior Member, IEEE*

Abstract—We construct an extension of spectral and diffusion geometry to multiple modalities through simultaneous diagonalization of Laplacian matrices. This naturally extends classical data analysis tools based on spectral geometry, such as diffusion maps and spectral clustering. We provide several synthetic and real examples of manifold learning, object classification, and clustering, showing that the joint spectral geometry better captures the inherent structure of multi-modal data. We also show the relation of many previous approaches for multimodal manifold analysis to our framework.

Index Terms—joint diagonalization, multimodal data, manifold alignment, manifold learning, Laplace-Beltrami operator, dimensionality reduction, diffusion distances, multimodal clustering

1 Introduction

The Laplacian operator and related constructions play a pivotal role in a wide range of applications in machine learning, pattern recognition, and computer vision. It has been shown that many problems in these fields boil down to finding a few smallest/largest eigenvectors and eigenvalues of a Laplacian constructed on some high-dimensional data. Important examples include *spectral clustering* [1] where clusters are determined by the first eigenvectors of the Laplacian; *eigenmaps* [2] and more generally *diffusion maps* [3], where one tries to find a low-dimensional manifold structure using the first smallest eigenvectors of the Laplacian; and *diffusion metrics* [4] measuring the "connectivity" of points on a manifold and expressed through the eigenvalues and eigenvectors of the Laplacian. Other applications heavily relying on the properties of the Laplacian include *spectral graph partitioning* [5], *spectral hashing* [6], spectral correspondence, image segmentation [7], and spectral shape analysis [8], [9], [10], [11]. Because of the intimate relation between the Laplacian operator, Riemannian geometry, and diffusion processes [12], it is common to encounter the umbrella term *spectral* or *diffusion geometry* in relation to the above problems.

These applications have been considered mostly in the context of uni-modal data, i.e., a single data space. However, many applications involve observations and measurements of data done using different modalities, such as multimedia documents [13], [14], [15], audio and video [16], [17], or medical imaging modalities like PET and CT [18]. Such problems of multimodal (or multi-view) data analysis have gained increasing interest in the computer vision and pattern recognition communities, however there have been only few attempts extending the powerful spectral methods to such settings.

In this paper, we propose a general framework allowing to extend different diffusion and spectral methods to the multimodal setting by finding a common eigenbasis of multiple Laplacians. Numerically, this problem is posed as simultaneous diagonalization. Such methods have received limited attention in the numerical mathematics community [19] and in blind source separation applications [20], [21], [22], [23].

In [24], [11], we showed the application of simultaneous diagonalization to the construction of compatible quasi-harmonic bases in shape analysis and manifold learning applications. The present paper is an extension of this line of works, focused on the application of simultaneous diagonalization of graph Laplacians for dimensionality reduction, manifold alignment, multimodal clustering, and object classification.

The paper is organized as follows: In Section 2, we overview the basic notions in spectral geometry of manifolds and graphs. In Section 3.1 we outline our framework and show two optimization problems (joint diagonalization, Section 3.2, and coupled diagonalization, Section 3.3) for simultaneous diagonalization of Laplacians. Besides providing a principled approach to data fusion, this approach gives a theoretical explanation to existing methods for multimodal data analysis. In Section 4, we show that many recent works on multi-view clustering [25], [26], [27], [28], [29] and manifold alignment [30], [31] can be considered as particular instances of our framework. Section 5 presents experimental results on synthetic and real datasets, and Section 6 concludes the paper.

D.E., A.K., M.B, and K.G. are with the Institute of Computational Science, Faculty of Informatics, University of Lugano (USI), Switzerland. A.B. is with the School of Electrical Engineering, Tel-Aviv University, Israel. A.B. and M.B. are also with the Perceptual Computing Group, Intel, Israel.
D.E., A.K., M.B., and K.G. are supported by the ERC Starting Grant No. 307047. A.B. is supported by the ERC Starting Grant No. 335491.

2 BACKGROUND

We assume that our data is represented as a k-dimensional compact manifold $X \subset \mathbb{R}^d$, embedded into a d-dimensional Euclidean space. In many applications d is very large while the intrinsic dimension of the data k is small, and one tries to study the structure of the manifold rather than its d-dimensional embedding. Such a structure can be characterized by means of the *Laplace-Beltrami operator* Δ, defined axiomatically through the Stokes identity[1] as

$$\int_X f \Delta h d\mu = \int_X \langle \nabla f, \nabla h \rangle d\mu, \tag{1}$$

where $f, h : X \to \mathbb{R}$ are smooth scalar fields on the manifold, $d\mu$ is a volume element, ∇ is the intrinsic gradient, and $\langle \cdot, \cdot \rangle$ is the Riemannian metric (inner product on the tangent space).

The eigenfunctions u_i of the Laplacian satisfying $\Delta u_i = \lambda_i u_i$ are often referred to as *manifold harmonics* and are analogous to the Fourier harmonics (which, in the 1D case, can be considered as a particular setting thereof being the eigenfunctions of the second-order derivative operator, $\frac{d^2}{dx} e^{-\iota\omega x} = -\omega^2 e^{-\iota\omega x}$). The eigenvalues play the role of "frequency" of the corresponding harmonics. Low-frequency harmonics (smallest eigenfunctions) capture the high-level structure of the manifold, while the high-frequency ones capture the "details". Manifold harmonics allow to extend standard harmonic analysis to manifolds. A real square-integrable function $f \in L^2(X)$ can be represented as the Fourier series

$$f = \sum_{i \geq 1} \langle f, u_i \rangle_{L^2(X)} u_i, \tag{2}$$

where $\langle f, g \rangle_{L^2(X)} = \int_X f g d\mu$ is the standard inner product on the space of real functions defined on the manifold.

In the discrete setting, the manifold is often represented by a weighted graph with vertices $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset X$ and edge weights $w_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ representing local connectivity using e.g. Gaussian kernel [32]. The Laplace-Beltrami operator can be discretized[2] as an $n \times n$ matrix $\mathbf{L} = \mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-1/2}$, where $\mathbf{W} = (w_{ij})$ and $\mathbf{D} = \operatorname{diag}(\sum_{j \neq i} w_{ij})$. Such a discretization is often referred to as *symmetric normalized Laplacian* and admits a unitary diagonalization $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, $\mathbf{U}^\top\mathbf{U} = \mathbf{I}$ where $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ is the matrix of column eigenvectors and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix of corresponding eigenvalues $\lambda_1 = 0 \leq \lambda_2 \leq \ldots \leq \lambda_n$. The eigenvectors $\mathbf{u}_i$ of the matrix $\mathbf{L}$ can be considered as a discretization of eigenfunctions u_i of the continuous operator Δ; it can be shown that under certain conditions on the discretization of the Laplacian, they converge to the continuous counterparts [33].

1. Note that we define the Laplacian as a positive-semidefinite operator unlike the convention in physics.

2. There exist many different constructions of the discrete Laplacian. For the sake of simplicity, we adopt the symmetric Laplacian. Our framework is applicable to other discretizations as well.

Geometric constructions associated with eigenvectors and eigenvalues of the Laplacian play an important role in manifold analysis, since several archetypical problems can be formulated in these terms. We list three classical examples below.

2.1 Eigenmaps

Non-linear dimensionality reduction methods try to capture the intrinsic low-dimensional structure of the manifold X. Belkin and Niyogi [2] showed that finding a neighborhood-preserving k-dimensional embedding of X can be posed as the *minimum eigenvalue problem*,

$$\min_{\mathbf{\Phi} \in \mathbb{R}^{n \times k}} \operatorname{tr}(\mathbf{\Phi}^\top \mathbf{L} \mathbf{\Phi}) \text{ s.t. } \mathbf{\Phi}^\top \mathbf{\Phi} = \mathbf{I}, \tag{3}$$

which has an analytic solution $\mathbf{\Phi} = (\mathbf{u_1}, \ldots, \mathbf{u_k})$ containing the first k eigenvectors of $\mathbf{L}$, thus effectively embedding the data by means of the eigenvectors of the Laplacian operator (the constant eigenvectors corresponding to the zero eigenvalues are usually discarded). Such an embedding is referred to as *Laplacian eigenmap* [2]. The neighborhood-preserving property of the eigenmaps is related to the fact the the smallest "low-frequency" eigenvectors of the Laplacian vary smoothly on the manifold.

More generally, a *diffusion map* is given as a mapping of the form $\mathbf{\Psi} = (K(\lambda_2)\mathbf{u}_2, \ldots, K(\lambda_k)\mathbf{u}_k)$, where $K(\lambda)$ is some transfer function acting as a "low-pass filter" on eigenvalues λ [4], [3].

2.2 Diffusion distances

Coifman et al. [4], [3], [12] related the eigenmaps to heat diffusion and random processes on manifolds and defined a family of *diffusion metrics* that in the most general setting can be written as

$$d^2(\mathbf{x}_i, \mathbf{x}_j) = \sum_l K(\lambda_l)(u_{il} - u_{jl})^2 = \|\mathbf{\Psi}(\mathbf{x}_i) - \mathbf{\Psi}(\mathbf{x}_j)\|_2^2. \tag{4}$$

Particular choice of $K(\lambda) = e^{-\lambda t}$ gives the *heat diffusion distance*, related to the connectivity of points $\mathbf{x}_i, \mathbf{x}_j$ on the manifold by means of diffusion process of length t. Such distances are intrinsic and thus invariant to manifold embedding and are robust to topological noise.

2.3 Spectral clustering

Ng et al. [1] showed a very efficient and robust clustering approach based on the observation that the multiplicity of the null eigenvalue of $\mathbf{L}$ is equal to the number of connected components of X. The corresponding eigenvectors act as indicator functions of these components. Embedding the data using the null eigenvectors and then applying some standard clustering algorithm such as K-means was shown to produce significantly better results than clustering the high-dimensional data directly. Spectral clustering can also be considered as a relaxation of the normalized cut method of Shi and Malik [7].

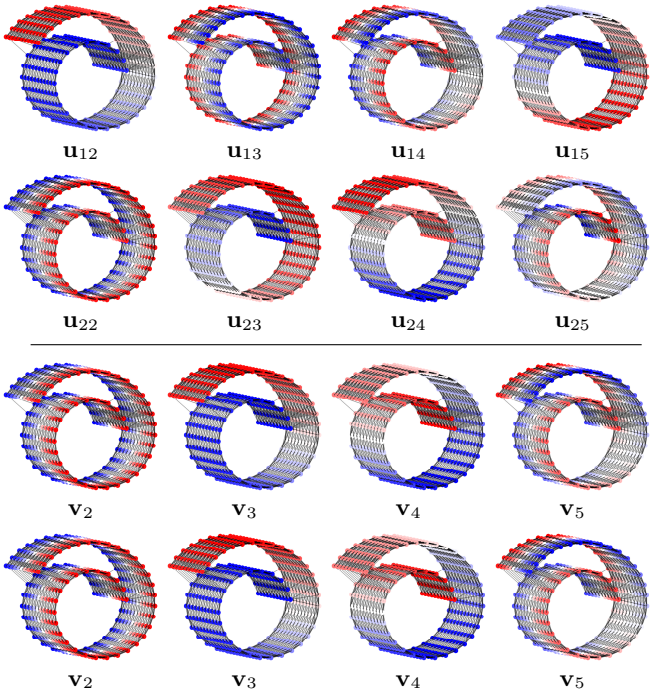

Fig. 1. Top: the first few Laplacian eigenvectors $\mathbf{u}_{1k}$ and $\mathbf{u}_{2k}$, $k = 2, \ldots, 5$ of two Swiss rolls (first and second rows) with slightly different connectivity (shown with lines). The difference in the connectivity results in different behavior of the eigenvectors (e.g. the third and the second eigenvectors are swapped). Bottom: joint approximate eigenvectors $\mathbf{v}_k$ computed on the same datasets using JADE behave in the same way. (Hot colors represent positive values; cold colors represent negative values).

3 Multimodal manifold analysis by simultaneous diagonalization

Recently, we witness increasing popularity of attempts to analyze different "views" or modalities of data. Such data can be modeled as m different manifolds $X^1 \subset \mathbb{R}^{d_1}, \ldots, X^m \subset \mathbb{R}^{d_m}$, which can have embeddings of different dimensionality $(d_1, \ldots, d_m)$ and sometimes different structure. Taking as example the multimedia retrieval application where one tries to find images matching to text tags or vice versa, different structure of the image and text tags manifolds can stem from ambiguities: e.g. "Cayenne" can refer to a city (capital of French Guiana), a plant (Cayenne pepper), or a car (Porsche Cayenne). We are interested in analyzing these manifolds simultaneously in order to extract their joint intrinsic structure.

We assume that we are given n_i samples $\{(\mathbf{x}_1^i, \ldots, \mathbf{x}_{n_i}^i) \subset \mathbb{R}^{d_i}\}_{i=1}^m$ on the manifolds and can construct the Laplacian matrices $\{\mathbf{L}_i \in \mathbb{R}^{n_i \times n_i}\}_{i=1}^m$ as described in Section 2. Trying to use the eigenbases $\mathbf{U}_1, \ldots, \mathbf{U}_m$ of the Laplacian matrices $\mathbf{L}_1, \ldots, \mathbf{L}_m$ to represent the data from different modalities in a common space is problematic since they do not "speak the same language": even if the manifolds are isometric (have the same intrinsic structure) and have simple spectrum (the Laplacian eigenvalues have no multiplicity), the corresponding eigenvectors may differ up to a sign. For an eigenvalue of multiplicity p, any basis spanning the p-dimensional subspace of the eigenspace constitutes a valid set of eigenvectors. More generally, for non-isometric manifolds (which is usually the case in real applications), the Laplacian eigenvectors can differ dramatically (Figure 1, top), making the data from different modalities mutually incomparable.

3.1 Main idea

The key idea of our paper in addressing this problem is to try to find the eigenbases of the Laplacians *simultaneously*. If the Laplacians $\mathbf{L}_1, \ldots, \mathbf{L}_m$ are of equal size $n_i = n$ and commute ($\mathbf{L}_i\mathbf{L}_j = \mathbf{L}_j\mathbf{L}_i$ for $i, j = 1, \ldots, m$), they are *jointly diagonalizable* in the sense that there exists a single set of orthonormal vectors $\mathbf{V}$ (*joint eigenvectors*) such that $\mathbf{V}^\top \mathbf{L}_i \mathbf{V} = \mathbf{\Lambda}_i = \mathrm{diag}(\lambda_{i1}, \ldots, \lambda_{i,n_i})$ are diagonal matrices of the eigenvalues of $\mathbf{L}_i$. Joint diagonalization allows to remove the ambiguities and incompatibilities between different modalities (Figure 1, bottom). It also allows us to naturally extend the spectral geometric methods discussed in Section 2 (eigenmaps, diffusion distances, spectral clustering, etc.) to the multimodal setting by simply replacing the eigenvalues and eigenvectors of a single Laplacian by the joint ones obtained from multiple Laplacians (in the following, we denote by $\mathbf{U}$ the eigenvectors and by $\mathbf{V}$ the joint eigenvectors, respectively).

In practice, however, due to differences between the modalities, the presence of noise, and possibly different number of samples, the Laplacian matrices $\mathbf{L}_1, \ldots, \mathbf{L}_m$ are not jointly diagonalizable. We thus need to look for bases $\mathbf{V}_1, \ldots, \mathbf{V}_m$ *approximately* satisfying the following properties:

Diagonalization: the basis $\mathbf{V}_i$ diagonalizes the Laplacian $\mathbf{L}_i$ for $i = 1, \ldots, m$, i.e., $\mathbf{V}_i^\top \mathbf{L}_i \mathbf{V}_i = \mathrm{diag}(\lambda_{i1}, \ldots, \lambda_{i,n_i})$. In this case, the eigenvalues $\lambda_{i1}, \ldots, \lambda_{i,n_i}$ can be regarded as "frequencies" and the columns of $\mathbf{V}_i = (\mathbf{v}_{i1}, \ldots, \mathbf{v}_{i,n_i})$ as an analogy of the harmonic (Fourier) basis. Using the first k (low frequency) eigenvectors of $\mathbf{L}_i$ to embed the samples of X^i ensures that the embedding preserves well the neighborhood structures [2].

Orthogonality: $\mathbf{V}_i^\top \mathbf{V}_i = \mathbf{I}$. This ensures that the dimensions of the embedding are uncorrelated and thus the embedding is "efficient".

Coupling: all the bases $\mathbf{V}_i$ behave consistently. In the most general setting, this consistency can be defined as follows: given a set of vectors $\mathbf{F}_i = (\mathbf{f}_{i1}, \ldots, \mathbf{f}_{iq})$ on X^i and a set of corresponding vectors $\mathbf{F}_j = (\mathbf{f}_{j1}, \ldots, \mathbf{f}_{jq})$ on X^j, their Fourier coefficients in the respective basis must coincide, $\mathbf{F}_i^\top \mathbf{V}_i = \mathbf{F}_j^\top \mathbf{V}_j$. We refer to this formulation as *Fourier coupling*. For example, the columns of $\mathbf{F}_i$ and $\mathbf{F}_j$ can be indicator functions of subsets ("blobs") on X^i and X^j.

A particular setting of the above is when the blobs are of single vertex size: given a set of q corresponding samples $k_{i1},\dots,k_{iq}$ and $k_{j1},\dots,k_{jq}$ in modalities i and j, respectively, the column $\mathbf{f}_{il}$ of $\mathbf{F}_i$ contains one at the index k_{il} and zeros elsewhere; the corresponding column $\mathbf{f}_{jl}$ of $\mathbf{F}_j$ contains one at the index k_{jl}. This setting is referred to as *point-wise* or *sparse coupling*.

Finally, the simplest case of coupling is when the manifolds are sampled at equal number of points and the samples are ordered in the same way in all modalities: $n_i = n$, $q = n$, and $\mathbf{F}_i = \mathbf{I}$ for all $i = 1,\dots,m$. In this case, referred to as *full coupling*, we simply have a single basis $\mathbf{V}_i = \mathbf{V}$ for all modalities.

In this section, we present several approaches for an approximate solution of the simultaneous diagonalization problem. We start with the problem of *joint approximate diagonalization*, in which full coupling is assumed and the optimization is performed over a single approximate eigenbasis $\mathbf{V}$ for all the Laplacians. Then, we present a more generic problem of *coupled diagonalization*, in which we are looking for m approximate eigenbases $\mathbf{V}_1,\dots,\mathbf{V}_m$ with Fourier or point-wise sparse coupling.

3.2 Joint diagonalization with generalized Jacobi method

Recall that the eigendecomposition problem (3) can be formulated as the minimization of the off-diagonal elements of the matrix $\mathbf{U}^\top\mathbf{L}\mathbf{U}$ over the space of orthonormal matrices $\mathbf{U}$,

$$\min_{\mathbf{U}^\top\mathbf{U}=\mathbf{I}} \operatorname{off}(\mathbf{U}^\top\mathbf{L}\mathbf{U}) \tag{5}$$

where $\operatorname{off}(\mathbf{X})$ is some off-diagonality criterion, e.g. the sum of squared off-diagonal elements, $\operatorname{off}(\mathbf{X}) = \|\mathbf{X} - \operatorname{Diag}(\mathbf{X})\|_{\mathrm{F}}^2$, where $\operatorname{Diag}(\mathbf{X})$ denotes a diagonal matrix containing only the diagonal values of $\mathbf{X}$. For a symmetric matrix $\mathbf{L}$, optimization (5) achieves the minimum value of zero, with a minimizer $\mathbf{U}$ being the eigenvectors of $\mathbf{L}$.

This type of optimization lies in the heart of a class of eigensolvers based on the *Jacobi iteration* [34]. One can observe that if $\mathbf{R}_\theta$ is a rotation matrix by angle θ in the plane ij, then $\mathbf{L}' = \mathbf{R}_\theta^\top\mathbf{L}\mathbf{R}_\theta$ is similar to $\mathbf{L}$ (i.e., both matrices have the same eigenvalues) and $\|\mathbf{L}'\|_{\mathrm{F}} = \|\mathbf{L}\|_{\mathrm{F}}$. However, we can choose $\theta = \tan^{-1}(\frac{2l_{ij}}{l_{jj}-l_{ii}})$ which results in $\mathbf{L}'_{ij} = 0$. This way, we find the angle θ minimizing $\operatorname{off}(\mathbf{R}_\theta^\top\mathbf{L}\mathbf{R}_\theta)$ which means that by applying such rotation $\mathbf{R}_\theta$ we reduce the off-diagonal elements of $\mathbf{L}$. Also note that since $\mathbf{L}'$ has only the ith and jth rows and columns different from $\mathbf{L}$, the rotation can be applied "in-place" and does not require matrix multiplication.

The idea of the Jacobi method for eigenvalue calculation is to construct $\mathbf{U}$ as a sequence of plane rotations $\mathbf{U} = \cdots\mathbf{R}_2\mathbf{R}_1$ in order to sequentially minimize the off-diagonal elements in (5). Being a product of rotations, the matrix $\mathbf{U}$ is orthonormal by construction.

JADE method. The same idea can be extended to finding the approximate joint eigenvectors of a few matrices in the full coupling setting [19], [20], [21], by solving the optimization problem

$$\min_{\mathbf{V}^\top\mathbf{V}=\mathbf{I}} \sum_{i=1}^{m} \operatorname{off}(\mathbf{V}^\top\mathbf{L}_i\mathbf{V}) \tag{6}$$

In this case, the minimum of zero is achieved iff $\mathbf{L}_1,\dots,\mathbf{L}_m$ commute. The *generalized Jacobi method* (referred to as JADE [21]) follows the standard Jacobi method, with the exception that in JADE the rotations are applied to reduce the off-diagonality criterion $\sum_{i=1}^m \operatorname{off}(\mathbf{R}_\theta^\top\mathbf{L}_i\mathbf{R}_\theta)$ in each step rather than $\operatorname{off}(\mathbf{R}_\theta^\top\mathbf{L}\mathbf{R}_\theta)$ used in the standard Jacobi iteration. Cardoso and Soloumiac [21] show that for a rotation matrix $\mathbf{R}_\theta$ the rotation angle θ minimizing $\sum_{i=1}^m \operatorname{off}(\mathbf{R}_\theta^\top\mathbf{L}_i\mathbf{R}_\theta)$ can be computed as follows: Let $\mathbf{G} = \sum_{i=1}^m h(\mathbf{L}_i)h(\mathbf{L}_i)^\top$, where $h(\mathbf{L}) = (l_{ii} - l_{jj}, l_{ij} + l_{ji})^\top$, and let $\alpha = g_{11} - g_{22}$ and $\beta = g_{12} + g_{22}$. Then, $\theta = \frac{1}{2}\tan^{-1}(\beta/(\alpha + \sqrt{\alpha^2+\beta^2}))$. The complexity of JADE is akin to that of the standard Jacobi iteration.

Perturbation analysis of JADE. The joint approximate eigenvectors obtained as the solution of problem (6) are related to the eigenvectors of the Laplacian matrices by the following

Theorem 3.1: Let $\mathbf{L}_1 = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ have a simple τ-separated spectrum $|\lambda_i - \lambda_j| \geq \tau$ for all $i \neq j$, and let $\mathbf{L}_2 = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top + \epsilon\mathbf{R}$ be a perturbation of $\mathbf{L}_1$. Then (ignoring permutation of eigenfunctions and sign flips), the joint approximate eigenbasis can be written as the first-order perturbation

$$\mathbf{v}_i = \mathbf{u}_i + \epsilon\sum_{j\neq i}\alpha_{ij}\mathbf{u}_j + \mathcal{O}(\epsilon^2), \tag{7}$$

where $\alpha_{ij} = \mathbf{u}_i^\top\mathbf{R}\mathbf{u}_j/2(\lambda_j - \lambda_i)$. For proof, we refer the reader to [35], [11].

3.3 Coupled diagonalization

The notable drawback of the joint diagonalization problem (6) is the full coupling assumption, requiring an equal number of data points in all the modalities and bijective correspondence between them. In many settings, this assumption could be too restrictive: for example, in multimedia retrieval applications, establishing the correspondence between images and annotations requires some human intelligence (tagging the images).

Coupling and decoupling. In a more general setting of the problem to which we refer as *coupled diagonalization* (CD), we use sparse point-wise (or, more generally, Fourier) coupling by means of a set of matrices $\{\mathbf{F}_i \in \mathbb{R}^{n_i\times q}\}_{i=1}^m$ containing as columns corresponding vectors in the respective modalities. We are looking for a set of *coupled bases* $\{\mathbf{V}_i \in \mathbb{R}^{n_i\times n_i} : \mathbf{V}_i^\top\mathbf{V}_i = \mathbf{I}\}_{i=1}^m$ such that $\mathbf{V}_i^\top\mathbf{L}_i\mathbf{V}_i$ are approximately diagonal for $i = 1,\dots,m$. To ensure that the bases $\mathbf{V}_1,\dots,\mathbf{V}_m$ behave consistently, we introduce *coupling constraints*: given a vector $\mathbf{f}^i$ on manifold X^i and a corresponding vector $\mathbf{f}^j$ on manifold X^j, we require that their Fourier coefficients in the respective bases coincide, $\mathbf{V}_i\mathbf{f}^i = \mathbf{V}_j\mathbf{f}^j$. Note that such coupling does not necessarily require the knowledge of

corresponding *points*, but rather of corresponding *vectors* $\mathbf{f}^i, \mathbf{f}^j$ [11]. Similarly, we can introduce a *decoupling term* on different vectors $\mathbf{g}^i$, $\mathbf{g}^j$, requiring their respective Fourier coefficients to be as different as possible. We can write the *coupled diagonalization* problem as

$$\min_{\mathbf{V}_i^\top \mathbf{V}_i = \mathbf{I}} \sum_{i=1}^m \mathrm{off}(\mathbf{V}_i^\top \mathbf{L}_i \mathbf{V}_i) + \mu_c \sum_{i,j=1}^m \|\mathbf{F}_i^\top \mathbf{V}_i - \mathbf{F}_j^\top \mathbf{V}_j\|_{\mathrm{F}}^2 - \mu_d \sum_{i,j=1}^m \|\mathbf{G}_i^\top \mathbf{V}_i - \mathbf{G}_j^\top \mathbf{V}_j\|_{\mathrm{F}}^2 \quad (8)$$

Because we have to look for a basis for each modality rather than for a single common basis, the number of variables in problem (8) is $\sum_{i=1}^m n_i^2$, compared to n^2 variables in problem (6). In the case $n_i = n = q$, $\mathbf{F}_i = \mathbf{I}$, $\mu_c \to \infty$, and $\mu_d = 0$, the coupled diagonalization problem (8) boils down to the joint diagonalization problem (6). Corresponding vectors can be delta functions (representing sparse point-wise correspondence between the manifolds), blobs (e.g. if one has information about corresponding sets of points), distance functions, etc.

Finding first k joint eigenvectors. One drawback of the JADE problem (6) is that it looks for *all* the joint approximate eigenvectors of the Laplacians. Since in most applications we do not need the $n_i \times n_i$ full basis but rather the first k coupled eigenvectors, we can solve a smaller problem over the matrices $\bar{\mathbf{V}}_i = (\mathbf{v}_{i1}, \ldots, \mathbf{v}_{ik})$ of size $n_i \times k$. Note that since *any* subset of columns of $\mathbf{V}_i$ will approximately diagonalize $\mathbf{L}_i$, when using the off penalty like in (8), *any* k approximate joint eigenvectors would be a solution, not necessarily the *first* ones. For this reason, in order to find the smallest approximate joint eigenvectors, we resort to a different off-diagonality penalty similar to one used in [22],

$$\min_{\bar{\mathbf{V}}_i^\top \bar{\mathbf{V}}_i = \mathbf{I}} \sum_{i=1}^m \|\bar{\mathbf{V}}_i^\top \mathbf{L}_i \bar{\mathbf{V}}_i - \bar{\mathbf{\Lambda}}_i\|_{\mathrm{F}}^2 + \mu_c \sum_{i,j=1}^m \|\mathbf{F}_i^\top \bar{\mathbf{V}}_i - \mathbf{F}_j^\top \bar{\mathbf{V}}_j\|_{\mathrm{F}}^2 - \mu_d \sum_{i,j=1}^m \|\mathbf{G}_i^\top \bar{\mathbf{V}}_i - \mathbf{G}_j^\top \bar{\mathbf{V}}_j\|_{\mathrm{F}}^2 \quad (9)$$

where $\bar{\mathbf{\Lambda}}_i = \mathrm{diag}(\lambda_{i1}, \ldots, \lambda_{ik})$ denotes the diagonal matrix containing the first k eigenvalues of $\mathbf{L}_i$. The number of variables in problem (9) is $\sum_{i=1}^m n_i k$.

Subspace parametrization. By virtue of Theorem 3.1, we have that for approximately jointly diagonalizable Laplacians, $\mathrm{span}\{\mathbf{v}_{i1}, \ldots, \mathbf{v}_{ik}\} \approx \mathrm{span}\{\mathbf{u}_{i1}, \ldots, \mathbf{u}_{ik}\}$. We can thus approximate the first k vectors of the coupled bases as a linear combination of the first $k' \geq k$ eigenvectors of $\mathbf{L}_i$, denoted by $\bar{\mathbf{U}}_i = (\mathbf{u}_{i1}, \ldots, \mathbf{u}_{ik'})$. We parametrize the coupled basis of modality i as $\bar{\mathbf{V}}_i = \bar{\mathbf{U}}_i \mathbf{A}_i$, where $\mathbf{A}_i$ is the $k' \times k$ matrix of linear combination coefficients. From the orthogonality of $\bar{\mathbf{V}}_i$, it follows that $\mathbf{A}_i^\top \mathbf{A}_i = \mathbf{I}$ [11]. Plugging this subspace parametrization into (8) and observing that $\bar{\mathbf{V}}_i^\top \mathbf{L}_i \bar{\mathbf{V}}_i = \mathbf{A}_i^\top \bar{\mathbf{\Lambda}}_i \mathbf{A}_i$, where $\bar{\mathbf{\Lambda}}_i = \mathrm{diag}(\lambda_{i,1}, \ldots, \lambda_{i,k'})$ is the diagonal matrix containing the first k' eigenvalues of $\mathbf{L}_i$, we get a problem with mkk' variables,

$$\min_{\mathbf{A}_i^\top \mathbf{A}_i = \mathbf{I}} \sum_{i=1}^m \mathrm{off}(\mathbf{A}_i^\top \bar{\mathbf{\Lambda}}_i \mathbf{A}_i) + \mu_c \sum_{i,j=1}^m \|\mathbf{F}_i^\top \bar{\mathbf{U}}_i \mathbf{A}_i - \mathbf{F}_j^\top \bar{\mathbf{U}}_j \mathbf{A}_j\|_{\mathrm{F}}^2 - \mu_d \sum_{i,j=1}^m \|\mathbf{G}_i^\top \bar{\mathbf{U}}_i \mathbf{A}_i - \mathbf{G}_j^\top \bar{\mathbf{U}}_j \mathbf{A}_j\|_{\mathrm{F}}^2 \quad (10)$$

The use of subspace parametrization offers several advantages. First, unlike problems (6) and (8), problem (10) is of fixed size mkk' *independent* of n_i. This potentially allows to deal with very large datasets. [3] Since typical values are $n \sim 10^3 - 10^4$, while $k', k \sim 10 - 100$, the reduction of the number of variables can be of several orders of magnitude. Second, as we represent our coupled basis vectors as linear combinations of the first k' low-frequency eigenvectors, the coupled basis vectors have guaranteed smooth behavior which ensures neighborhood-preservation property typical of Laplacian embeddings. Finally, differently from JADE, in (10) the Laplacians are *not used explicitly*. This key difference makes the problem agnostic to the specific discretization of the Laplacian.

Optimization. The solution of problem (10) can be carried out using standard constrained optimization techniques such as `fmincon` in MATLAB which require the gradients of the cost function and the constraints. The gradient of the off-diagonal penalty is given by

$$\nabla_{\mathbf{A}_i} \|\mathbf{A}_i^\top \bar{\mathbf{\Lambda}}_i \mathbf{A}_i - \bar{\mathbf{\Lambda}}_i\|_{\mathrm{F}}^2 = 4(\bar{\mathbf{\Lambda}}_i \mathbf{A}_i \mathbf{A}_i^\top \bar{\mathbf{\Lambda}}_i \mathbf{A}_i - \bar{\mathbf{\Lambda}}_i \mathbf{A}_i \bar{\mathbf{\Lambda}}_i).$$

The gradient of the coupling/decoupling term is

$$\nabla_{\mathbf{A}_i} \|\mathbf{F}_i^\top \bar{\mathbf{U}}_i \mathbf{A}_i - \mathbf{F}_j^\top \bar{\mathbf{U}}_j \mathbf{A}_j\|_{\mathrm{F}}^2 = 2\bar{\mathbf{U}}_i^\top \mathbf{F}_i (\mathbf{F}_i^\top \bar{\mathbf{U}}_i \mathbf{A}_i - \mathbf{F}_j^\top \bar{\mathbf{U}}_j \mathbf{A}_j).$$

Alternatively, recent techniques [36], [37] for optimization on *Stiefel manifold* $\mathbb{V}_k(\mathbb{R}^{k'}) = \{\mathbf{A} \in \mathbb{R}^{k' \times k} : \mathbf{A}^\top \mathbf{A} = \mathbf{I}\}$ can be employed. In this approach, we solve an unconstrained problem having the orthonormality constraint built into the optimization method in the form of projected descent. Technically, optimization is applied in a block-coordinate manner w.r.t. to $\mathbf{A}_i$ fixing all other $\mathbf{A}_{j \neq i}$,

$$\min_{\mathbf{A}_i \in \mathbb{V}_k(\mathbb{R}^{k'})} \|\mathbf{A}_i^\top \bar{\mathbf{\Lambda}}_i \mathbf{A}_i - \bar{\mathbf{\Lambda}}_i\|_{\mathrm{F}}^2 + \mu_c \sum_{j=1}^m \|\mathbf{F}_i^\top \bar{\mathbf{U}}_i \mathbf{A}_i - \mathbf{F}_j^\top \bar{\mathbf{U}}_j \mathbf{A}_j\|_{\mathrm{F}}^2 - \mu_d \sum_{j=1}^m \|\mathbf{G}_i^\top \bar{\mathbf{U}}_i \mathbf{A}_i - \mathbf{G}_j^\top \bar{\mathbf{U}}_j \mathbf{A}_j\|_{\mathrm{F}}^2$$

alternatingly for all $i = 1, \ldots, m$.

Complexity. Assuming the number of modalities m is small and that the $q \times k'$ Fourier coefficients matrices $\mathbf{F}_i^\top \bar{\mathbf{U}}_i$ are pre-computed, the cost function and its gradient computation (and consequently, the cost of a single optimization iteration) has complexity $\mathcal{O}(qk'k)$.

3. Obviously, in order to construct the subspace parametrization, one has to calculate the first k' harmonics of each of the manifolds, an operation dependent on the data size. However, such a computation would be required anyway if a single-modality spectral method were applied to each modality separately. Our parametrization allows to couple exact harmonics by solving an additional modestly-sized optimization problem.

4 RELATION TO PREVIOUS WORKS

In this section, we overview other spectral methods for dealing with multimodal data and show their relations to the proposed approach. In our analysis, we distinguish between two broad groups of approaches: those assuming the *full coupling* setting (i.e., equal number of vertices $n_i = n$ and known bijective correspondence between them) and *sparse coupling* setting (each modality might have a different number of vertices, and the correspondence is known only between a few of them).

4.1 Full coupling setting

Laplacian averaging. Assuming the full coupling setting and that the first k eigenvalues of the Laplacians are zero, we want to find $\mathbf{V} \in \mathbb{R}^{n\times k}$ such that $\mathbf{L}_i\mathbf{V} = 0$ for all $i = 1, \ldots, m$ and $\mathbf{V}^\top\mathbf{V} = \mathbf{I}$ by reformulating (6) as

$$\min_{\mathbf{V}^\top\mathbf{V}=\mathbf{I}} \sum_{i=1}^{m} \|\mathbf{L}_i\mathbf{V}\|_{\mathrm{F}}^2 \tag{11}$$

Since $\sum_{i=1}^m \|\mathbf{L}_i\mathbf{V}\|_{\mathrm{F}}^2 = \mathrm{tr}\,(\mathbf{V}^\top(\sum_{i=1}^m \mathbf{L}_i^\top\mathbf{L}_i)\mathbf{V})$, the problem can be equivalently recast as finding the null eigenvectors of the "average" Laplacian matrix $\bar{\mathbf{L}} = \sum_{i=1}^m \mathbf{L}_i^\top\mathbf{L}_i$. One can also consider other averaging operators, such as arithmetic mean $\bar{\mathbf{L}} = \frac{1}{m}\sum_{i=1}^m \mathbf{L}_i$ or harmonic mean $\bar{\mathbf{L}} = (\sum_{i=1}^m \mathbf{L}_i^{-1})^{-1}$. In what follows, we will show that many approaches for multimodal manifold alignment boil down to simple Laplacian averaging in their limit cases. Laplacian averaging methods seem to be the most 'naïve' way of producing multimodal spectral geometry and have been used in several applications such as clustering [26]. [4]

Matrix factorization. Tang et al. [27] proposed graph clustering through low-rank factorization of the weight matrix, trying to find a common factor $\mathbf{U}$ such that $\mathbf{W}_i \approx \mathbf{U}\mathbf{\Lambda}_i\mathbf{U}^\top$ by solving

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k}, \mathbf{\Lambda}_i\in\mathbb{R}^{n\times n}} \sum_{i=1}^{m} \|\mathbf{W}_i - \mathbf{U}\mathbf{\Lambda}_i\mathbf{U}^\top\|_{\mathrm{F}}^2, \tag{12}$$

using the quasi-Newton method. Besides the fact that the factorization is applied to the weight matrix (it can be equivalently applied to the Laplacian), we see here a (non-orthogonal) joint diagonalization problem with an off-diagonality criterion considered by Yeredor [22].

MVSC. Cai et al. [28] proposed a method for *multi-view spectral clustering* (MVSC) by solving[5]

$$\min_{\mathbf{V}^\top\mathbf{V}=\mathbf{I}} \sum_{i=1}^{m} \mathrm{tr}\,(\mathbf{V}_i^\top\mathbf{L}_i\mathbf{V}_i) + \mu\|\mathbf{V}_i - \mathbf{V}\|_{\mathrm{F}}^2 \tag{13}$$

4. We refer to reader to [38] for a recent attempt to generalize Laplacian averaging methods to a more general setting of sparse coupling.

5. Cai et al. [28] also impose a non-negativity constraint on the matrix $\mathbf{V}$ in order to obtain cluster indicators directly and bypass the K-means clustering stage. We ignore this additional constraint for the simplicity of discussion; such a constraint can be added to all the problems discussed in this paper.

and show that this problem can be equivalently posed as

$$\max_{\mathbf{V}^\top\mathbf{V}=\mathbf{I}} \mathrm{tr}\left(\mathbf{V}^\top \textstyle\sum_{i=1}^m (\mathbf{L}_i + \mu\mathbf{I})^{-1}\,\mathbf{V}\right) \tag{14}$$

We observe that problem (13) consists of m minimum-eigenvalue problems w.r.t. bases $\mathbf{V}_i$, with the addition of a coupling term, encouraging $\mathbf{V}_i$ as close as possible to some common basis $\mathbf{V}$ (note that the authors do not impose orthogonality constraints $\mathbf{V}_i^\top\mathbf{V} = \mathbf{I}$, but for $\mu \to \infty$, the proximity to orthogonal $\mathbf{V}$ makes $\mathbf{V}_i$ approximately orthogonal). Thus, it is possible to interpret (13) as a kind of joint diagonalization criterion similar to manifold alignment discussed in Section 4.2.

Problem (14) can be rewritten as a minimum eigenvalue problem

$$\min_{\mathbf{V}^\top\mathbf{V}=\mathbf{I}} \mathrm{tr}\left(\mathbf{V}^\top \left(\textstyle\sum_{i=1}^m (\mathbf{L}_i + \mu\mathbf{I})^{-1}\right)^{-1}\mathbf{V}\right) \tag{15}$$

whose solution is given by the first k eigenvectors of the matrix $\left(\sum_{i=1}^m (\mathbf{L}_i + \mu\mathbf{I})^{-1}\right)^{-1}$. For $\mu = 0$, this is simply the harmonic mean of the Laplacians. In order to obtain the limit case $\mu \to \infty$, observe that

$$\begin{aligned}\textstyle\sum_{i=1}^m (\mathbf{L}_i + \mu\mathbf{I})^{-1} &= \tfrac{1}{\mu}\sum_{i=1}^m (\tfrac{1}{\mu}\mathbf{L}_i + \mathbf{I})^{-1} \approx \tfrac{1}{\mu}\sum_{i=1}^m \mathbf{I} - \tfrac{1}{\mu}\mathbf{L}_i \\ &= \tfrac{m}{\mu}\mathbf{I} - \tfrac{1}{\mu}\sum_{i=1}^m \mathbf{L}_i. \end{aligned} \tag{16}$$

Plugged into (14) and normalized, in the limit $\mu \to \infty$ expression (16) becomes

$$\min_{\mathbf{V}^\top\mathbf{V}=\mathbf{I}} \mathrm{tr}\left(\mathbf{V}^\top \textstyle\sum_{i=1}^m \mathbf{L}_i\mathbf{V}\right) \tag{17}$$

thus essentially boiling down to arithmetic mean of the Laplacians. [6]

Co-regularization. Kumar et al. [29] proposed the *centroid co-regularization* approach for multimodal clustering based on the minimization of

$$\min_{\mathbf{V}_i^\top\mathbf{V}_i=\mathbf{I};\ \mathbf{V}^\top\mathbf{V}=\mathbf{I}} \sum_{i=1}^{m} \mathrm{tr}\,(\mathbf{V}_i^\top\mathbf{L}_i\mathbf{V}_i) - \mu\mathrm{tr}\,(\mathbf{V}_i\mathbf{V}_i^\top\mathbf{V}\mathbf{V}^\top) \tag{18}$$

This function is alternatingly minimized, first with respect to the $n \times k$ matrices $\mathbf{V}_i$, then with respect to $\mathbf{V}$. The term $-\mathrm{tr}\,(\mathbf{V}_i\mathbf{V}_i^\top\mathbf{V}\mathbf{V}^\top) = \|\mathbf{V}_i\mathbf{V}_i^\top - \mathbf{V}\mathbf{V}^\top\|_{\mathrm{F}}^2 - k$ measures the Grassmanian distance between the column subspaces $\mathrm{span}\{\mathbf{v}_{i1}, \ldots, \mathbf{v}_{ik}\}$ and $\mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ [39] and has an effect similar to our coupling term $\|\mathbf{V}_i - \mathbf{V}\|_{\mathrm{F}}^2$.

SC-ML. Dong et al. [39] proposed an approach for *spectral clustering on multi-layer graphs* (SC-ML) similar to co-regularization, trying to find an $n \times k$ matrix $\mathbf{V}$ which minimizes that Laplacian quadratic form and is closest

6. The same result can be obtained by analyzing (13) and noticing that for $\mu \to \infty$ we have $\mathbf{V}_i = \mathbf{V}$ and the problem becomes

$$\min_{\mathbf{V}} \mathrm{tr}\,\left(\mathbf{V}^\top \textstyle\sum_{i=1}^m \mathbf{L}_i\mathbf{V}\right)\ \text{ s.t. }\ \mathbf{V}^\top\mathbf{V} = \mathbf{I}.$$

to the subspaces spanned by the first k eigenvectors $\bar{\mathbf{U}}_i$ of the Laplacians $\mathbf{L}_i$,

$$\min_{\mathbf{V}^\top\mathbf{V}=\mathbf{I}} \sum_{i=1}^{m} \mathrm{tr}\,(\mathbf{V}^\top\mathbf{L}_i\mathbf{V}) - \mu\mathrm{tr}\,(\bar{\mathbf{U}}_i\bar{\mathbf{U}}_i^\top\mathbf{V}\mathbf{V}^\top). \tag{19}$$

Rewriting (19) as

$$\min_{\mathbf{V}^\top\mathbf{V}=\mathbf{I}} \mathrm{tr}\left(\mathbf{V}^\top\left(\sum_{i=1}^{m}\mathbf{L}_i - \mu\bar{\mathbf{U}}_i\bar{\mathbf{U}}_i^\top\right)\mathbf{V}\right), \tag{20}$$

one obtains a closed-form solution to $\mathbf{V}$ by finding the first k eigenvectors of the matrix $\sum_{i=1}^{m}\mathbf{L}_i - \mu\bar{\mathbf{U}}_i\bar{\mathbf{U}}_i^\top$.

CCO. Using the relation between joint diagonalizability and commutativity [40], [41], Bronstein et al. [42] considered a class of problems referred to as *closest commuting operators* (CCO), where one seeks the smallest perturbation $\tilde{\mathbf{L}}_1, \tilde{\mathbf{L}}_2$ of $\mathbf{L}_1, \mathbf{L}_2$ such that $\tilde{\mathbf{L}}_1, \tilde{\mathbf{L}}_2$ commute,

$$\min_{\tilde{\mathbf{L}}_k\in\mathcal{M}} \sum_{k=1}^{2} \|\tilde{\mathbf{L}}_k - \mathbf{L}_k\|_{\mathrm{F}}^2 \;\text{ s.t. }\; \tilde{\mathbf{L}}_1\tilde{\mathbf{L}}_2 = \tilde{\mathbf{L}}_2\tilde{\mathbf{L}}_1, \tag{21}$$

(where $\mathcal{M} \subseteq \mathbb{R}^{n\times n}$ is some class of matrices). It was shown that for $\mathcal{M} = \mathbb{R}^{n\times n}$, problem (21) is equivalent to JADE (6), in the following sense: since the minimizers of (21) are commuting matrices, they are jointly diagonalizable, and their joint eigenbasis is the minimizer of (6) [42]. Using as $\mathcal{M}$ the set of Laplacian matrices, one can impose a specific sparse structure on $\tilde{\mathbf{L}}_1, \tilde{\mathbf{L}}_2$. The solution of (21) is carried out by parametrizing $\tilde{\mathbf{L}}_i$ through the non-zero elements of the adjacency matrix $\mathbf{W}_i$. The complexity of the problem depends both on the *size* and the *structure* of $\mathbf{W}_i$: assuming that each row of the adjacency matrix has at most s non-zero elements, computing the cost function and the constraints and their gradients requires $\mathcal{O}(sn^2)$ operations.

4.2 Sparse coupling setting

Manifold alignment. Ham et al. [30] introduced *manifold alignment* as a way to construct embeddings that are consistent in two different modalities. Let us be given two weighted adjacency graphs with n vertices in the spaces X^1 and X^2, and let us assume w.l.o.g. that the points are ordered such that the first l points in the two modalities correspond. The main idea of manifold alignment is to construct a big graph with $2n$ vertices where the edges connecting corresponding points in different modalities have some weight μ. The joint Laplacian of such a graph is a $2n\times 2n$ matrix of the form

$$\hat{\mathbf{L}} = \begin{pmatrix} \mathbf{L}_1 + \mu\mathbf{Q} & -\mu\mathbf{Q} \\ -\mu\mathbf{Q} & \mathbf{L}_2 + \mu\mathbf{Q} \end{pmatrix} \tag{22}$$

where $\mathbf{Q}$ is an $n\times n$ diagonal matrix with first l diagonal elements equal to one and the rest to zero. Ham et al. then compute the eigenmap of the joint Laplacian

$$\min_{\mathbf{Z}\in\mathbb{R}^{2n\times k}} \mathrm{tr}\,(\mathbf{Z}^\top\hat{\mathbf{L}}\mathbf{Z}) \;\text{ s.t. }\; \mathbf{Z}^\top\mathbf{Z} = \mathbf{I}, \tag{23}$$

and use the rows $1,\dots,n$ and $n+1,\dots,2n$ of $\mathbf{Z}$ as the k-dimensional embeddings of manifolds X^1 and X^2, respectively. Larger values of μ ensure that the embedding coordinates of the corresponding points coincide.

Denoting $\mathbf{Z} = (\mathbf{V}_1; \mathbf{V}_2)$ we can rewrite (23) as

$$\min_{\mathbf{V}_i\in\mathbb{R}^{n\times k}} \mathrm{tr}\,(\mathbf{V}_1^\top\mathbf{L}_1\mathbf{V}_1) + \mathrm{tr}\,(\mathbf{V}_2^\top\mathbf{L}_2\mathbf{V}_2) + \mu\|\mathbf{Q}\mathbf{V}_1 - \mathbf{Q}\mathbf{V}_2\|_{\mathrm{F}}^2 \;\text{ s.t. }\; \mathbf{V}_1^\top\mathbf{V}_1 + \mathbf{V}_2^\top\mathbf{V}_2 = \mathbf{I}, \tag{24}$$

We recognize in the first terms the cost used in the classical eigenmap (3). In the case $\mu = 0$, $\hat{\mathbf{L}}$ becomes a block-diagonal matrix and $\mathbf{V}_i = \mathbf{U}_i$ the eigenvectors of the Laplacians $\mathbf{L}_i$. For $\mu > 0$, the third term serves as a coupling similar to our coupled diagonalization problem (8). An important difference, however, is that $\mathbf{V}_i$ in this formulation are not orthonormal (the orthogonality constraint is on their sum). Furthermore, the trace terms ignore the off-diagonal elements and do not promote diagonalization of the Laplacians. Thus, the resulting bases are not quasi-harmonic.

Finally, observe that in the limit case $\mu\to\infty$ and $\mathbf{Q} = \mathbf{I}$, problem (23) becomes (up to scaling) the minimum eigenvalue problem

$$\min_{\mathbf{V}^\top\mathbf{V}=\mathbf{I}} \mathrm{tr}\,(\mathbf{V}^\top(\mathbf{L}_1 + \mathbf{L}_2)\mathbf{V}) \tag{25}$$

for the matrix $\mathbf{L}_1 + \mathbf{L}_2$. Thus, in this case manifold alignment boils down to arithmetic mean of the Laplacians.

Procrustes analysis. Wang and Mahadevan proposed to align the low-dimensional embeddings of two modalities by solving an orthogonal Procrustes problem [31]. Note that this problem is a particular setting of our coupled joint diagonalization problem (10) for $m = 2, k = k'$ and $\mu\to\infty$: in this case, we can ignore the off-diagonality penalty and obtain

$$\min_{\mathbf{A}_1,\mathbf{A}_2\in\mathbb{R}^{k\times k'}} \|\mathbf{F}_1^\top\bar{\mathbf{U}}_1\mathbf{A}_1 - \mathbf{F}_2^\top\bar{\mathbf{U}}_2\mathbf{A}_2\|_{\mathrm{F}}^2 \;\text{ s.t. }\; \mathbf{A}_i^\top\mathbf{A}_i = \mathbf{I}. \tag{26}$$

Using the invariance of the Frobenius norm under orthogonal transformation, we can rewrite problem (26) as an orthogonal Procrustes problem

$$\min_{\mathbf{\Omega}\in\mathbb{R}^{k\times k}} \|\mathbf{F}_1^\top\bar{\mathbf{U}}_1 - \mathbf{F}_2^\top\bar{\mathbf{U}}_2\mathbf{\Omega}\|_{\mathrm{F}}^2 \;\text{ s.t. }\; \mathbf{\Omega}^\top\mathbf{\Omega} = \mathbf{I}. \tag{27}$$

where $\mathbf{\Omega} = \mathbf{A}_2\mathbf{A}_1^\top$. The problem has an analytic solution $\mathbf{\Omega} = \mathbf{S}\mathbf{R}^\top$, where $\bar{\mathbf{U}}_1^\top\mathbf{F}_1\mathbf{F}_2^\top\bar{\mathbf{U}}_2 = \mathbf{S}\mathbf{\Sigma}\mathbf{R}^\top$ is the singular value decomposition of the matrix $\bar{\mathbf{U}}_1^\top\mathbf{F}_1\mathbf{F}_2^\top\bar{\mathbf{U}}_2$ with left- and right singular vectors $\mathbf{S}, \mathbf{R}$ [43]. Then, $\mathbf{A}_1 = \mathbf{S}$ and $\mathbf{A}_2 = \mathbf{R}$.

Geometrically, the matrices $\mathbf{A}_1, \mathbf{A}_2$ can be interpreted as rotations aligning the eigenvectors $\bar{\mathbf{U}}_1, \bar{\mathbf{U}}_2$ to coincide in the best way at the corresponding points. This solution is ambiguous up to a rotation $\mathbf{A}$, i.e., $\mathbf{A}_2\mathbf{A}, \mathbf{A}^\top\mathbf{A}_1$ is also a solution. More importantly, the resulting coupled bases are not guaranteed to be quasi-harmonic since the off-diagonality penalty is not used.

Functional correspondence. Ovsjanikov et al. [10] proposed a framework for finding functional correspondence between X^1 and X^2. Let $\mathbf{f}^1$ and $\mathbf{f}^2$ be corresponding functions on X^1 and X^2, respectively. One can define

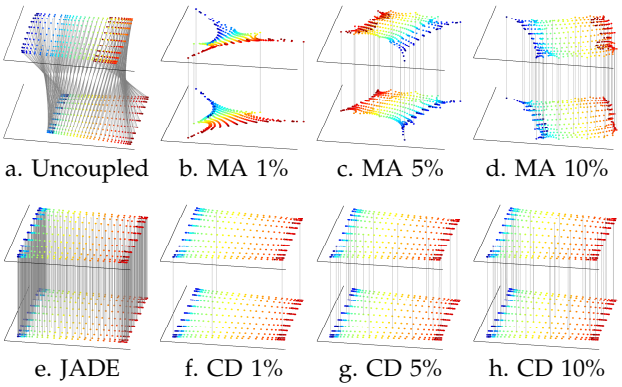

Fig. 2. Simultaneous two-dimensional embedding of two Swiss rolls with slightly different connectivity using JADE, coupled diagonalization (CD) and manifold alignment (MA). Ideally, the embedding should 'unroll' the rolls into rectangular regions, and the embeddings of the two modalities should coincide. Using the same sparse coupling (from 10% to 1% of the points, shown in gray lines), CD produces a significantly better alignment than MA.

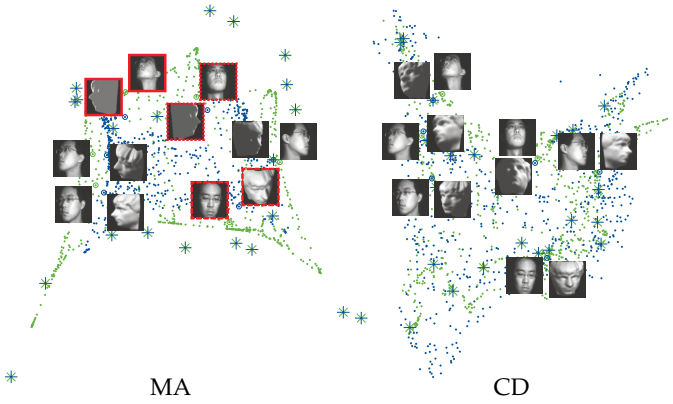

Fig. 3. Alignment of face (green) and statue (blue) manifolds. Each point represents an image in the respective dataset; circles represent corresponding poses of the statue and face images shown. Crosses denote the data points used for coupling. Note some significant misalignment between the manifolds in the MA results (marked in red).

the functional map as an $n_2 \times n_1$ matrix $\mathbf{T}$, such that $\mathbf{f}^2 = \mathbf{T}\mathbf{f}^1$. The functional map can be approximated using k first Laplacian eigenvectors, $\mathbf{T} \approx \bar{\mathbf{U}}_2\mathbf{C}^\top\bar{\mathbf{U}}_1^\top$, where the $k \times k$ matrix $\mathbf{C}$ translates Fourier coefficients from basis $\bar{\mathbf{U}}_1$ to basis $\bar{\mathbf{U}}_2$. Given a set of corresponding vectors $\mathbf{F}_i$ of size $n_i \times q$, $i = 1, 2$, one can find $\mathbf{C}$ by solving the system of qk linear equations in k^2 variables,

$$\mathbf{F}_2^\top\bar{\mathbf{U}}_2 = \mathbf{F}_1^\top\bar{\mathbf{U}}_1\mathbf{C}. \tag{28}$$

One needs $q \geq k$ corresponding vectors in order to make the system (28) determined. If X^1, X^2 are isometric, $\mathbf{C}$ is orthonormal.

Solving (28) in the least-squares sense in this setting is equivalent to the orthogonal Procrustes problem (27), which we showed to be a limit case of our coupled diagonalization problem. More generally, the use of the off-diagonal penalty in (10) imposes a diagonal structure on the matrix $\mathbf{C}$, which serves as a regularization allowing to reduce the amount of data required to make the problem determined.

5 Results

In this section, we show several examples of the application of our simultaneous diagonalization approach. The leitmotif of all the experiments in, given several graphs ('modalities') representing same or related data in slightly different ways, to take advantage of this multimodal information. The experiments are structured as follows: In Section 5.1, we show examples of dimensionality reduction by embedding the multimodal data into low-dimensional spaces using joint Laplacian eigenvectors. We show that the eigenmaps of different modalities are well aligned in this way. These results are mostly qualitative. In Section 5.2, we compare different state-of-the-art multimodal clustering algorithms on standard datasets. In Section 5.3 we compute diffusion distances in the joint Laplacian eigenspaces and use them to classify objects. Section 5.4 shows further example of applying diffusion distances for meaningful subsampling of the datasets using farthest point sampling (FPS) technique [44]. Finally, in Section 5.5 we analyze the complexity of our approach.

5.1 Dimensionality reduction

Swiss rolls. In this experiment, we used two Swiss roll surfaces with slightly different embedding as two different data modalities. The rolls were constructed in such a way that in each modality there is topological noise (connectivity "across" the roll loops) at different points. The rolls contained $n = 451$ points. Laplacians were constructed as in [2], using 8-neighbor connectivity and Gaussian weights with scale parameter $t = 5$. Figure 1 shows the first few eigenvectors computed using each Laplacian individually (top) and jointly with JADE (bottom). Using individual modalities, the difference in the connectivity produces different eigenvectors, with respect to both their order and their behavior (values across two different faces are closer where there are links). When using joint eigenvectors, instead, we are able to correctly capture the intrinsic structure of the data (i.e. links across faces are not influent anymore), and the eigenvectors behave the same way. This effect is evident in Figure 2, where the second and third uncoupled (a) and joint (e) eigenvectors of the same rolls are plotted.

Sparsely-coupled Swiss rolls. Next, we repeat the same experiment using correspondence between a small subset of vertices (sparse coupling) rather than all the points. The corresponding sparse points were sampled using farthest point sampling. Since the rolls are (up

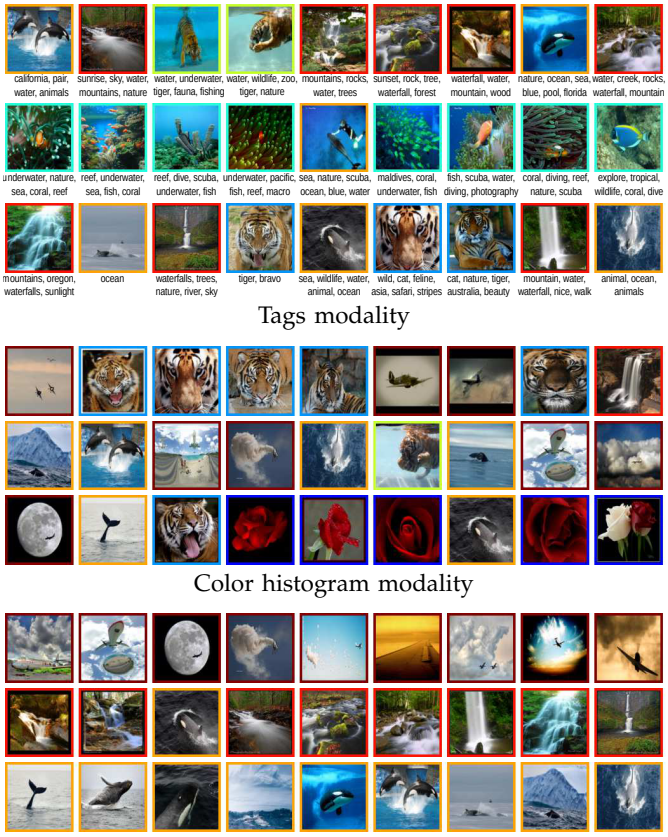

Fig. 4. Spectral clustering of the NUS dataset. Shown are a few images (randomly sampled) attributed to a single cluster by spectral clustering using the Tags modality only (top), the Color modality only (middle) and the Tags+Color multimodal clustering using JADE (bottom). Groundtruth clusters are shown in different colors. Note the ambiguities in the Tag-based clustering (e.g. swimming tigers and underwater scenes) and Color-base clustering (e.g. yellowish tigers and autumn scenes).

Fig. 5. Spectral clustering of the Caltech-7 dataset. Shown are a few images (randomly sampled) attributed to a single cluster by spectral clustering using the Bio-inspired modality only (top), the PHOW modality only (middle) and the multimodal clustering using JADE (bottom). Groundtruth clusters are shown in different colors. Ideally, a cluster should contain images from a single class only.

to topological noise) isometric to a plane, their ideal embeddings should be rectangular patches. Figure 2 (f-h) shows the result of joint embedding using our CD with sparse point-wise correspondence. With as little as 1% correspondences, we obtain results similar to JADE (which uses full coupling). Figure 2 (b-d) shows the result of manifold alignment (MA) [30] with the same sparse correspondences. It is evident that MA requires many more points to achieve results similar to CD.

Alignment of visual manifolds. As an additional comparison of CD and MA, we reproduce the problem of alignment of two visual manifolds using the data of [30]: 831 120×100 images of a face and 698 64×64 images of a statue. The datasets were coupled sampling 25 points from the statue dataset with FPS and then manually matching them with corresponding images in the faces dataset. Figure 3 shows the result of the alignment of face (green) and statue (blue) manifolds. As an example, we took 6 face pictures in different poses (green circles) and showed them, for both methods, next to their closest counterparts on the statue manifold (blue circles). We observe that, with the same number of correspondences, the alignment of the two manifolds is significantly better using CD compared to MA: as a consequence, pictures in the statue dataset tend to be closer to pictures of faces in the same pose.

5.2 Multimodal clustering

We performed multimodal spectral clustering on six different multimodal datasets. *Circles* and *Text* are two synthetic datasets purposely built to be noisy in each modality (overlapping clusters) and to have modalities that disambiguate each other (clusters which are close in one modality are far apart in the other one, see Figure 6). *NUS* is a subset of the NUS-WIDE dataset [45] containing images (represented by 64-dimensional color histograms) and their text annotations (represented by 1000-dimensional bags of words). Images were purposely selected to have ambiguous content and annotations (e.g., swimming tigers are also tagged as "water" making them confuse e.g. with whales). *Caltech* is a subset of the Caltech-101 dataset with the same 7 image

			Accuracy/NMI (%)					
		Method	Circles	Text	Caltech [28], [24]	NUS [45], [24]	Digits [46], [47]	Reuters [48], [47]
		#points	800	800	105	145	2000	600
		*Uncoupled**	53.0/39.5	60.4/50.9	77.1/75.0	84.8/81.9	83.2/82.2	52.3/41.1
		Harmonic Mean	95.6/90.1	97.2/91.0	84.8/79.2	89.0/83.8	87.0/86.3	52.3/40.9
		Arithmetic Mean	96.5/91.2	96.9/89.6	88.6/83.1	**95.2**/**92.1**	85.2/84.8	52.2/**41.4**
		Comraf [49]	40.8/16.9	60.8/41.7	–[7]	86.9/84.3	81.6/77.0	53.2/30.7
		MVSC [28]	95.6/90.1	97.2/91.2	85.7/80.8	89.0/83.8	83.0/84.8	52.3/40.9
		MultiNMF [47]	41.1/14.2	50.5/23.2	–[7]	77.4/79.3	87.2/79.3	53.1/40.9
		SC-ML [39]	98.2/94.6	97.8/93.1	88.6/81.6	94.5/90.7	87.8/85.3	52.8/38.4
		JADE [24]	**100**/**100**	**98.4**/**94.1**	86.7/80.6	93.1/87.5	85.1/85.1	52.3/40.9
CD*	pos	10%	52.5/26.0	54.5/26.2	78.7/75.3	78.6/77.9	94.2/87.8	53.7/34.4
		20%	61.3/40.2	60.0/41.9	80.8/76.0	82.9/78.2	94.1/87.4	54.2/33.7
		60%	93.7/85.4	86.5/69.7	87.0/80.0	87.2/78.9	93.9/87.1	54.7/36.5
		100%	98.9/95.5	96.8/89.4	**89.5**/83.3	94.5/90.6	93.9/87.1	54.8/36.9
	pos+neg	10%	67.3/46.5	63.6/42.1	86.5/80.9	92.7/86.2	**94.9**/**88.9**	**59.0**/37.7
		20%	69.6/50.2	67.8/50.0	87.9/81.2	93.3/87.0	94.8/88.7	57.6/37.1
		60%	95.2/87.9	87.0/68.5	89.2/**84.0**	94.5/88.5	94.8/88.7	57.0/38.8

TABLE 1
Performance of different multimodal clustering methods of different datasets (accuracy / normalized mutual information in %, the higher the better). References provide additional details about the datasets, experiments, and methods. *Best performing modality is shown.

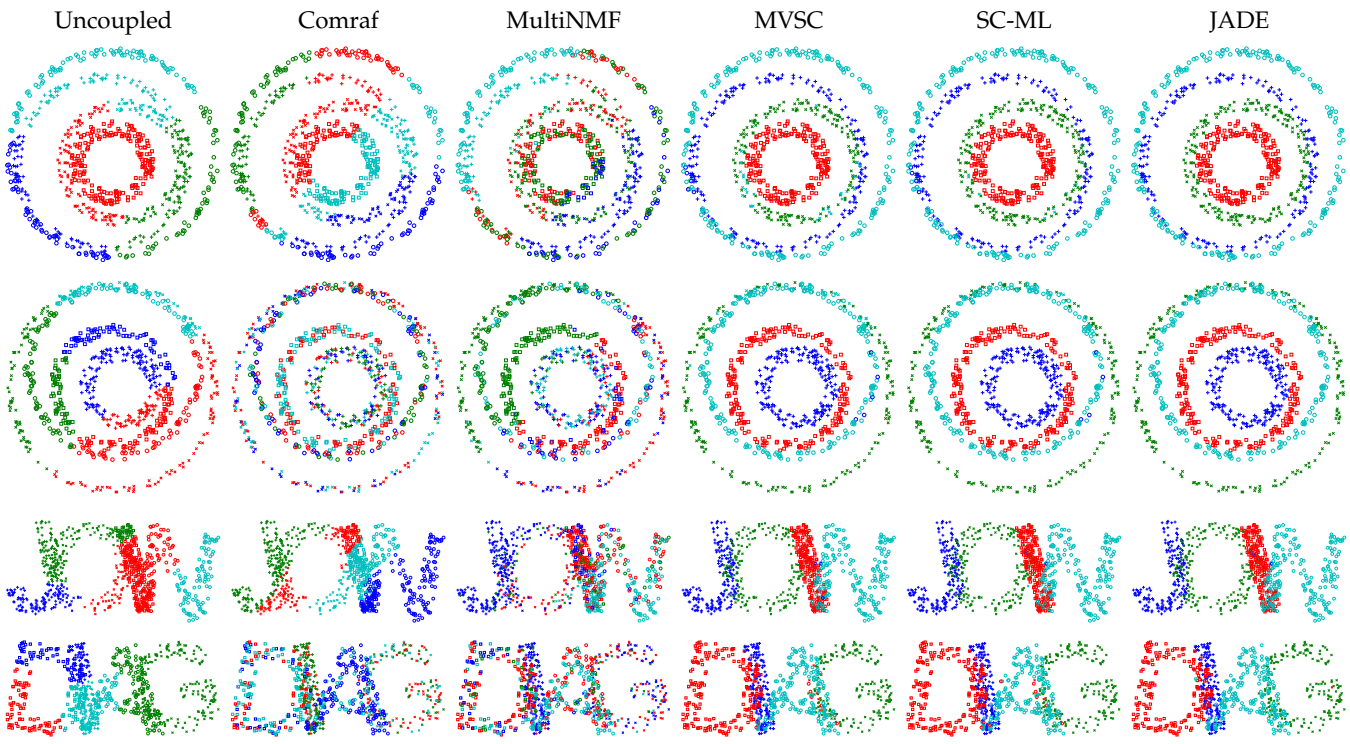

Fig. 6. Clustering of synthetic multimodal datasets *Circles* (two modalities shown in first and second rows) and *Text* (third and fourth rows). Marker shape represents ground truth clusters; marker color represents the clustering results produced by different methods (ideally, all markers of one type should have only one color).

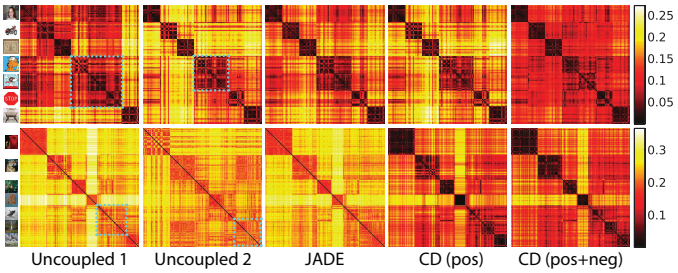

Fig. 7. Diffusion distances between objects from the Caltech (top) and NUS (bottom) datasets using separate modalities (first and second columns), JADE (third column) and CD with coupling (fourth column) and coupling+decoupling (fifth column) terms. Note the ambiguities between different classes of objects (marked in cyan) when using a single modality.

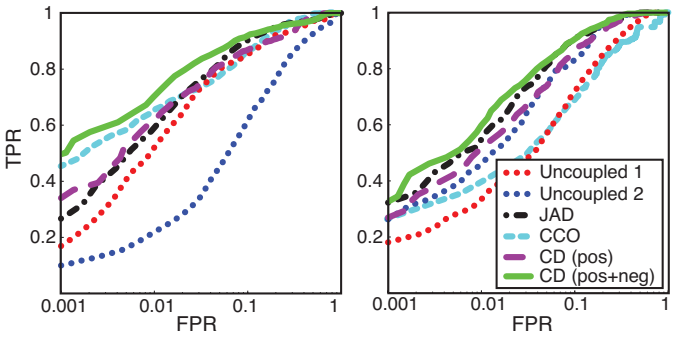

Fig. 8. Object classification performance on Caltech (left) and NUS (right) datasets using diffusion distances computed in each modality separately (Uncoupled), a joint eigenspace (JADE), coupled eigenspaces produced by CD with coupling (pos) and coupling+decoupling (pos+neg) terms, and the joint eigenspace of the closest commuting Laplacians (CCO). Note that CD (pos+neg) performs better than each modality on its own and outperforms the other methods.

classes as in Cai et al. [28]. For each image, kernels arising from different visual descriptors were given [50]: we chose the ht_bio_105728 bio-inspired features and 4x4 pyramid histogram of visual words (PHOW) as different modalities. *Digits* is the UCI Handwritten Digits dataset [46], [47], represented using 76 Fourier coefficients and the 240 pixel averages in 2×3 windows. *Reuters* is a subset of the Reuters multilingual text collection [48], [47] using the English and French languages as two different modalities.

Laplacians were constructed using the Gaussian weight selected with a self-tuning scale [51]. Spectral clustering was performed independently on each modality (Uncoupled), on the joint eigenspace calculated with JADE [21], and on the coupled bases calculated using CD with coupling only (pos, $\mu_d = 0$) as well as decoupling (pos+neg) terms. Sparse sets of corresponding points for coupling were generated using FPS on each cluster with random initial point. The results were averaged over ten runs with different sampling. Negatives were generated by choosing blobs of points belonging to ambiguous sets (e.g. clusters 5, 6, and 7 for NUS, and clusters 4 and 5 for Caltech). For reference, we show the performance of the following state-of-the-art multiview clustering methods: Comraf [49], MVSC [28], MultiNMF [47], and SC-ML [39].[7] We further compare with the two Laplacian averaging methods (harmonic mean and arithmetic mean). Clustering quality was measured using two standard criteria used in the evaluation of clustering algorithms: the *micro-averaged accuracy* [49] and the *normalized mutual information* (NMI) [52].

Figures 4 and 5 visualizes the results of unimodal spectral clustering with its multimodal extension (calculated using JADE) on the NUS and Caltech dataset. One can easily see the advantage of simultaneously using information from both modalities: images which are ambiguous in either modality (e.g. due to their colors, tags, or other visual features) are made unambiguous in the multimodal case.

Figure 6 visualizes the behavior of different multimodal clustering algorithms (all assuming the full coupling setting) on the synthetic datasets *Circles* and *Text*: due to the non-globular shapes of the clusters and their overlap, both the unimodal approach and the non-spectral multimodal ones perform poorly on these datasets. Clusters found by multimodal spectral clustering methods, instead, are all quite accurate, and JADE performs the best among them.

Finally, Table 1 summarizes the quantitative evaluation of different clustering methods. In the full coupling setting, we observe that multimodal spectral methods perform consistently better on non-globular clusters and very noisy datasets. In particular, methods that might look naive such as harmonic and arithmetic mean provide surprisingly good results, competing with other much more elaborate approaches. In the sparse coupling setting (using correspondence between 10%-100% points), CD is able to obtain performances close to (and often better than) the ones of full coupling methods with just a fraction of the data they need.

5.3 Object classification

In this experiment, we used the diffusion distances computed using Laplacian eigenvectors (individual and joint). The distances were computed with the first 100 eigenvectors according to (4) using heat diffusion kernel $K(\lambda) = e^{-5\lambda}$. Figure 7 shows the distance matrices between the objects in the *Caltech* (top) and *NUS* (bottom) datasets. Ideally, the distance matrix should contain zero blocks on the diagonal (objects of the same class) and non-zero elsewhere (objects from different classes). Thresholding these distances at a set of levels and measuring the false positives/true positive rates (FPR/TPR),

7. Since Comraf and Multi-NMF methods require explicit coordinates of the data points, while *Caltech* data is represented implicitly as kernels, we could not measure performance on this dataset.

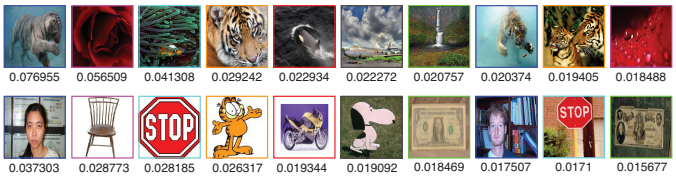

Fig. 9. Farthest point sampling of NUS (top) and Caltech-7 (bottom) datasets using the diffusion distance in the joint eigenspace computed by JADE. First point is on the left. Numbers indicate the sampling radius. Note that in both cases, the first seven samples cover all the image classes, providing a meaningful subsampling of the respective datasets.

we produce the ROC curves that clearly indicate the advantage of using multiple modalities (see Figure 8).

5.4 Manifold subsampling

Next, we used the same diffusion distances to progressively sample the *Caltech* (top) and *NUS* (bottom) datasets using the farthest point sampling strategy: starting with some point, pick up the second one as most distant from the first; then the third as the most distant from the first and second, and so on. Such sampling is almost-optimal [44] and is known to produce a progressively refined r-covering of the dataset. Figure 9 shows that the first seven samples produced in this way cover all the classes present in the dataset, providing thus a meaningful subsampling. This is an indication of the presence of data clusters in the coupled eigenspace which are cohesive (points in the same class are close to each other) and at the same time well separated (points in different classes are far from each other).

5.5 Complexity

In our final experiment, we studied the complexity of the CD approach and compare it to the CCO. The synthetic dataset using in this experiment was similar the *Circles* dataset, and contained four concentric circles with different connectivity in two modalities; the number of points n ranged between 400 and 2000. The Laplacians were constructed using s=5, 10, and 20 nearest neighbors. Since CCO assumes full coupling (bijective correspondence between the modalities), in order to make the comparison fair, we used the full coupling setting in the CD problem ($q = n$), thus making its complexity depend linearly on n.

Figure 10 shows the computational complexity comparison between CD and CCO, computed as average computation time of the cost functions (10) and (21) and their respective gradients. The CD problem complexity in this setting scales linearly with n, independently of s. The CCO problem complexity scales quadratically in n and linearly in s.

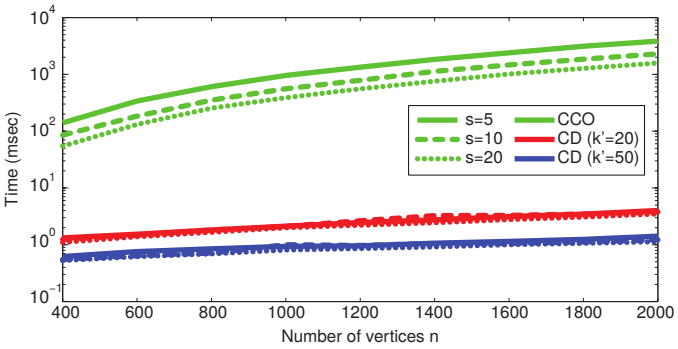

Fig. 10. Comparison of CCO and CD computational complexity, measured the mean time (in msec) per iteration for different number of vertices n and number of nearest neighbors s used in the definition of the Laplacian. CD is shown in two settings: $k' = 20$ (blue) and $k' = 50$ (red); CCO is shown in green. For a fair comparison, in CD the coupling is performed using all the points, hence the complexity grows linearly with n (the complexity of CCO grows quadratically with n). Also, note that CCO depends on the adjacency structure (number of nearest neighbors s) while CD does not.

6 Conclusions

We presented a framework for multi-modal data analysis using approximate joint diagonalization of Laplacian matrices, naturally extending the classical construction of diffusion geometry to the multi-modal setting. This construction allowed an almost straightforward extension of various diffusion-geometric data analysis tools such as spectral clustering and manifold learning based on diffusion maps.

Our starting point was the generalized Jacobi method (JADE) for joint diagonalization of matrices developed in the signal processing community for source separation problems. Though conceptually easy to understand, this method was developed for small full matrices while we have large sparse ones. Furthermore, JADE computes the full set of eigenvectors while most manifold analysis applications require only a few largest or smallest eigenvectors. As an alternative, we showed a method working in the subspace of the eigenvectors of the Laplacians. It is also easily extendable to the more generic setting of coupled diagonalization, in which only partial correspondence between the different modalities is known.

Surprisingly, it appears that many prior works on multi-modal data analysis can be considered as particular instances of our framework. In particular, previously proposed approaches to multi-modal spectral clustering are nearly equivalent and try to solve some version of the joint approximate diagonalization problem. Manifold alignment methods are also instances of the coupled diagonalization. We believe that the presented construction makes the need of such a tool central enough to deserve the interest of the entire machine learning community.

REFERENCES

[1] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. NIPS*, 2001.

[2] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, pp. 1373–1396, 2002.

[3] R. Coifman and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 21, pp. 5–30, 2006.

[4] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps," *PNAS*, vol. 102, no. 21, pp. 7426–7431, 2005.

[5] C. Ding, X. He, H. Zha, M. Gu, and H. Simon, "A min-max cut algorithm for graph partitioning and data clustering," in *Proc. Conf. Data Mining*, 2001.

[6] Y. Weiss, A. Torralba, and R. Fergus, "Spectral hashing," in *Proc. NIPS*, 2008.

[7] J. Shi and J. Malik, "Normalized cuts and image segmentation," *Trans. PAMI*, vol. 22, pp. 888–905, 1997.

[8] B. Levy, "Laplace-Beltrami eigenfunctions towards an algorithm that "understands" geometry," in *Proc. SMI*, 2006.

[9] G. Rong, Y. Cao, and X. Guo, "Spectral mesh deformation," *Visual Computer*, vol. 24, no. 7, pp. 787–796, 2008.

[10] M. Ovsjanikov, M. Ben-Chen, J. Solomon, A. Butscher, and L. Guibas, "Functional maps: A flexible representation of maps between shapes," *TOG*, vol. 31, no. 4, 2012.

[11] A. Kovnatsky, M. M. Bronstein, A. M. Bronstein, K. Glashoff, and R. Kimmel, "Coupled quasi-harmonic bases," *Computer Graphics Forum*, vol. 32, no. 2, pp. 439–448, 2013.

[12] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis, "Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators," in *Proc. NIPS*, 2005.

[13] J. Weston, S. Bengio, and N. Usunier, "Large scale image annotation: learning to rank with joint word-image embeddings," *Machine Learning*, vol. 81, no. 1, pp. 21–35, 2010.

[14] N. Rasiwasia, J. Costa Pereira, E. Coviello, G. Doyle, G. Lanckriet, R. Levy, and N. Vasconcelos, "A new approach to cross-modal multimedia retrieval," in *Proc. ICM*, 2010.

[15] B. McFee and G. R. G. Lanckriet, "Learning multi-modal similarity," *JMLR*, vol. 12, pp. 491–523, 2011.

[16] E. Kidron, Y. Y. Schechner, and M. Elad, "Pixels that sound," in *Proc. CVPR*, 2005.

[17] X. Alameda-Pineda, V. Khalidov, R. Horaud, and F. Forbes, "Finding audio-visual events in informal social gatherings," in *Proc. ICMI*, 2011.

[18] M. M. Bronstein, A. M. Bronstein, F. Michel, and N. Paragios, "Data fusion through cross-modality metric learning using similarity-sensitive hashing," in *Proc. CVPR*, 2010.

[19] A. Bunse-Gerstner, R. Byers, and V. Mehrmann, "Numerical methods for simultaneous diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 14, no. 4, pp. 927–949, 1993.

[20] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *Radar and Signal Processing*, vol. 140, no. 6, pp. 362–370, 1993.

[21] J. F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Matrix Anal. Appl*, vol. 17, pp. 161–164, 1996.

[22] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *Trans. Signal Proc.*, vol. 50, no. 7, pp. 1545 –1553, 2002.

[23] A. Ziehe, *Blind Source Separation based on Joint Diagonalization of Matrices with Applications in Biomedical Signal Processing*, U. o. L. Dissertation, Ed. Dissertation, University of Potsdam, 2005.

[24] D. Eynard, K. Glashoff, M. Bronstein, and A. Bronstein, "Multimodal diffusion geometry by joint diagonalization of laplacians," *arXiv:1209.2295*, 2012.

[25] V. de Sa, "Spectral clustering with two views," in *Proc. ICML Workshop on learning with multiple views*, 2005.

[26] C. Ma and C.-H. Lee, "Unsupervised anchor shot detection using multi-modal spectral clustering," in *Proc. ICASSP*, 2008.

[27] W. Tang, Z. Lu, and I. Dhillon, "Clustering with multiple graphs," in *Proc. Data Mining*, 2009.

[28] X. Cai, F. Nie, H. Huang, and F. Kamangar, "Heterogeneous image feature integration via multi-modal spectral clustering," in *Proc. CVPR*, 2011.

[29] A. Kumar, P. Rai, and H. Daumé III, "Co-regularized multi-view spectral clustering," in *Proc. NIPS*, 2011.

[30] J. Ham, D. Lee, and L. Saul, "Semisupervised alignment of manifolds," in *Proc. Annual Conference on Uncertainty in Artificial Intelligence*, 2005.

[31] C. Wang and S. Mahadevan, "Manifold alignment using procrustes analysis," in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08. New York, NY, USA: ACM, 2008, pp. 1120–1127.

[32] U. von Luxburg, "A tutorial on spectral clustering," 2007.

[33] M. Wardetzky, "Convergence of the cotangent formula: An overview," in *Discrete Differential Geometry*, ser. Oberwolfach Seminars, A. I. Bobenko, J. M. Sullivan, P. Schröder, and G. M. Ziegler, Eds. Birkhäuser Basel, 2008, vol. 38, pp. 275–286.

[34] C. Jacobi, "Über ein leichtes Verfahren, die in der Theorie der Säkularstörungen vorkommenden Gleichungen numerisch aufzulösen," *Journal für reine und angewandte Mathematik*, vol. 30, pp. 51–95, 1846.

[35] J. F. Cardoso, "Perturbation of joint diagonalizers," 1995.

[36] K. Rahbar and J. P. Reilly, "Geometric optimization methods for blind source separation of signals," in *Proc. ICA*, 2000.

[37] Z. Wen and W. Yin, "A feasible method for optimization with orthogonality constraints," *Mathematical Programming*, vol. 142, no. 1-2, pp. 397–434, 2013.

[38] M. M. Bronstein and K. Glashoff, "Heat kernel coupling for multiple graph analysis," *arXiv:1312.3035*, 2013.

[39] X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov, "Clustering on multi-layer graphs via subspace analysis on grassmann manifolds," *arXiv:1303.2221*, 2013.

[40] H. Lin, "Almost commuting selfadjoint matrices and applications." vol. 13, pp. 193–233, 1997.

[41] K. Glashoff and M. M. Bronstein, "Matrix commutators: their asymptotic metric properties and relation to approximate joint diagonalization," *Linear Algebra and its Applications*, vol. 439, no. 8, pp. 2503–2513, 2013.

[42] M. M. Bronstein, K. Glashoff, and T. A. Loring, "Making laplacians commute," *arXiv:1307.6549*, 2013.

[43] P. Schönemann, "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, vol. 31, no. 1, pp. 1–10, 1966.

[44] D. S. Hochbaum and D. B. Shmoys, "A best possible heuristic for the k-center problem," *Mathematics of operations research*, pp. 180–184, 1985.

[45] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y.-T. Zheng, "Nus-wide: A real-world web image database from national university of singapore," in *Proc. CIVR*, 2009.

[46] E. Alpaydin and C. Kaynak, "Cascading classifiers," *Kybernetika*, vol. 34, no. 4, pp. 369–374, 1998.

[47] J. Liu, C. Wang, J. Gao, and J. Han, "Multi-view clustering via joint nonnegative matrix factorization," in *Proc. SDM*, 2013.

[48] M. R. Amini, N. Usunier, and C. Goutte, "Learning from multiple partially observed views-an application to multilingual text categorization," in *Proc. NIPS*, 2009.

[49] R. Bekkerman and J. Jeon, "Multi-modal clustering for multimedia collections," in *Proc. CVPR*, 2007.

[50] N. Pinto, "UCSD-MIT Caltech-101-MKL Dataset," 2009. [Online]. Available: http://mkl.ucsd.edu/dataset/ucsd-mit-caltech-101-mkl-dataset

[51] P. Perona and L. Zelnik-Manor, "Self-tuning spectral clustering," in *Proc. NIPS*, 2004.

[52] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

Davide Eynard is a researcher at the Institute of Computational Science, Faculty of Informatics, University of Lugano (USI). He obtained both MS (2005) and PhD (2009) in computer engineering from Politecnico di Milano, Italy, and he has worked as a postdoc both at Politecnico di Milano and at USI. His research activity has always dealt with knowledge at different levels: how it can be formalized and collectively produced (Description Logics, Social Semantic Web), how it can be extracted from unstructured or semi-structured data (Social Web mining, ontology extraction from text), and how it can be exploited to make sense of vast amounts of information (machine learning, pattern recognition). His current work focuses on the concept of similarity between objects, its extension to different modalities/views and its application to problems in shape analysis, image processing, and machine learning.

Artiom Kovnatsky received the B.Sc. and M.Sc. degrees in Applied Mathematics at the Technion, Israel. He is now pursuing a Ph.D. degree in Computational Science in the Faculty of Informatics at the University of Lugano (USI), Switzerland. His research interests are in the fields of computer vision, computer graphics, data mining, numerical optimisation, an in particular, spectral methods for the analysis of heterogeneous manifolds.

Michael M. Bronstein is a professor in the Faculty of Informatics at the University of Lugano (USI), Switzerland and a Research Scientist at the Perceptual Computing group, Intel, Israel. He held visiting appointments at Politecnico di Milano (2008), Stanford university (2009), and University of Verona (2010, 2014). Michael got his B.Sc. in Electrical Engineering (2002) and Ph.D. in Computer Science (2007), both from the Technion, Israel. His main research interests are theoretical and computational methods in spectral and metric geometry and their application to problems in computer vision, pattern recognition, shape analysis, computer graphics, image processing, and machine learning. His research appeared in international media and was recognized by numerous awards. In 2012, Michael received the highly-competitive European Research Council (ERC) starting grant. In 2014, he was invited as a Young Scientist to the World Economic Forum New Champions meeting in China, an honor bestowed on forty world's leading scientists under the age of 40. Besides academic work, Prof. Bronstein is actively involved in the industry. He was the co-founder of the Silicon Valley start-up company Novafora, where he served as Vice President of technology (2006-2009), responsible for the development of algorithms for large-scale video analysis. He was one of the principle inventors and technologists at Invision, an Israeli startup developing 3D sensing technology acquired by Intel in 2012 and released under the RealSense brand.

Klaus Glashoff is a retired professor of Applied Mathematics, University of Hamburg, Germany. He is specialized in optimization algorithms. He also worked in the field of the history of logic (European and Indian logic). Since 2012, Prof. Glashoff is affiliated with the Institute of Computational Science, University of Lugano (USI), Switzerland. His current research interests are in the field of computational methods for data analysis.

Alexander M. Bronstein is a professor in the School of Electrical Engineering at Tel Aviv University. His main research interests are theoretical and computational methods in metric geometry and their application to problems in computer vision, pattern recognition, shape analysis, computer graphics, image processing, and machine learning. In addition to his academic activities, Alex Bronstein is an active technologist. His technology has been in the foundation of several successful startup companies.