Published in final edited form as:
Inf Process Med Imaging. 2013 ; 23: 316–327.

Locality Preserving Non-negative Basis Learning with Graph Embedding

Yasser Ghanbari[1], **John Herrington**[2], **Ruben C. Gur**[3], **Robert T. Schultz**[2], and **Ragini Verma**[1,*]

Yasser Ghanbari: Yasser.Ghanbari@uphs.upenn.edu; John Herrington: herringtonj@mail.chop.edu; Ruben C. Gur: gur@mail.med.upenn.edu; Robert T. Schultz: schultzrt@mail.chop.edu; Ragini Verma: Ragini.Verma@uphs.upenn.edu

[1]Section of Biomedical Image Analysis, University of Pennsylvania, Philadelphia, PA

[2]Center for Autism Research, Children's Hospital of Philadelphia, Philadelphia, PA

[3]Brain Behavior Laboratory, Department of Psychiatry, University of Pennsylvania, Philadelphia, PA

Abstract

The high dimensionality of connectivity networks necessitates the development of methods identifying the connectivity building blocks that not only characterize the patterns of brain pathology but also reveal representative population patterns. In this paper, we present a non-negative component analysis framework for learning localized and sparse sub-network patterns of connectivity matrices by decomposing them into two sets of discriminative and reconstructive bases. In order to obtain components that are designed towards extracting population differences, we exploit the geometry of the population by using a graph-theoretical scheme that imposes locality-preserving properties as well as maintaining the underlying distance between distant nodes in the original and the projected space. The effectiveness of the proposed framework is demonstrated by applying it to two clinical studies using connectivity matrices derived from DTI to study a population of subjects with ASD, as well as a developmental study of structural brain connectivity that extracts gender differences.

Keywords

Connectivity analysis; non-negative matrix factorization; locality-preserving dimensionality reduction; graph embedding

1 Introduction

Computational techniques applied to neuroimaging data have helped unveil the underlying structural or functional differences between groups of interest, e.g. patients and healthy controls. Altered brain connectivity has recently gained a lot of attention in investigating the origin of many brain disorders such as autism spectrum disorder (ASD) [1] and in developmental studies [2]. Hence, advanced techniques of brain connectivity analysis are emerging as a powerful tool in pathological studies of brain disorders. Such tools quantify the connectivity between two regions of interest in DTI, fMRI, EEG, or MEG by calculating tractography [3], mutual information, or synchronization [1] measures.

*The authors acknowledge support for this work, from the following grants: NIH- MH092862(PI: R. Verma), MH089983 & MH089924 (PI: R. Gur), Pennsylvania Department of Health (SAP # 4100042728, SAP # 4100047863, PI: R. Schultz) and IDDRC (P30 HD026979, PI: M. Yudkoff).

A number of established analysis methods are available for studying the underlying brain structure via a succinct representation. A successful analysis methodology must possess a means of identifying relevant sub-networks providing an interpretable representation of the brain connectivity, while also facilitating the statistical analysis that describes how this representation is affected by disease. The traditional approaches, i.e. principal and independent components analysis (PCA and ICA) used for investigating brain networks [4] provide dimensionality reduction but may lack physiological interpretability.

Recently, non-negative matrix factorization (NMF) and its alternatives have received extensive attention and proven effective in providing an interpretable set of bases characterizing multivariate data. After being first introduced by Lee and Seung [5], NMF has been successfully employed in many applications such as signal processing, pattern recognition, and medical imaging [6–10]. This was later extended to enforce higher sparseness by adding certain regularization terms [11]. NMF's part-based representation of image data, as well as non-negativity constraints on both the bases and coefficients, facilitates interpretability, and its small size of the basis set categorizes NMF among the unsupervised dimensionality reduction techniques. Although the unsupervised methods are useful in interpretation due to their positivity and sparsity, they do not necessarily provide discriminative bases, only bases which best reconstruct the original data.

The approach taken here is a decomposition of connectivity matrices into interpretable basis components while enforcing positivity of both the components and coefficients. Such a decomposition maintains the interpretation of each component as a network connectivity matrix and the coefficients associated with these components as activations of those networks, while providing a succinct low dimensional representation of the population amenable to statistical analysis. We split the components into two sets of *discriminative* and *reconstructive* bases, which are learned during the optimization process by a graph embedding scheme.

The reconstructive basis set is modeled by minimizing the Frobenius norm of the reconstruction error matrix. To reach our discriminatory basis components, we create two graphs in the high dimensional space of connectivity elements (we call them high dimensional points here): A graph of nearest neighbors to maintain representatives of nearby high dimensional points as close as possible after dimensionality reduction, and a second graph connecting distant high dimensional points to maintain the long distance between their representatives in the projected space. Accounting for the geometrical information in the unsupervised projective NMF helps us categorize the non-negative basis set into discriminative (i.e. providing group differences) and reconstructive (i.e. providing low reconstruction error) components.

The two capabilities of low reconstruction error and good discrimination are unified into minimizing one objectives function by using a gradient descent approach with guaranteeing the positivity of bases and their coefficients.

While the method is generalizable to any type of non-negative connectivity matrix, in this work we apply it to structural connectivity networks computed from Diffusion Tensor Imaging (DTI) data from two different populations of subjects with autism spectrum disorder (ASD) and a developmental study. When NMF is combined with the geometrical information, we show that the discriminative bases are grouped in the discriminative set and dominant bases gather in the reconstructive set.

2 Methods

We hypothesize that each connectivity matrix obtained from the brain connectivity network of a subject, is a linear combination of several fundamental connectivity matrices called connectivity components. Due to the symmetry of connectivity matrices, a vector of all elements of the upper triangular part of any connectivity matrix is considered as the representative of that matrix, and is used as an observation vector $\boldsymbol{x}_i$ for the corresponding subject i. To compute the connectivity components whose mixture approximately constructs the observed connectivity matrices, a matrix factorization model is used as $\boldsymbol{X} = \boldsymbol{W}\boldsymbol{\Phi} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ represents the residual error matrix, columns of $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{m\times n}$, i.e. $\boldsymbol{x}_i$ ($1 \le i \le n$), are the connectivity matrix representatives, and columns of $\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_p] \in \mathbb{R}^{m\times p}$, i.e. $\boldsymbol{w}_j$ ($1 \le j \le p$), are representative of the normalized basis connectivity components, i.e. the upper triangular elements of the matrix of the corresponding connectivity component. These components ($\boldsymbol{w}_j$) are then mixed by the elements of each column of the loading matrix $\boldsymbol{\phi} = [\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_n] \in \mathbb{R}^{p\times n}$ to approximate the corresponding column of $\boldsymbol{X}$ [7, 10], i.e. $\boldsymbol{x}_i \approx \sum_{j=1}^{p} \Phi_{ji}\boldsymbol{w}_j$; $1 \le i \le n$.

2.1 Unsupervised Learning of Projective Non-negative Bases

Inspired by [12], we assume that $\boldsymbol{\Phi}$ is the projection of $\boldsymbol{X}$ onto $\boldsymbol{W}$, i.e. $\boldsymbol{\Phi} = \boldsymbol{W}^T\boldsymbol{X}$, the non-negativity constraint on the elements of $\boldsymbol{W}$ and $\boldsymbol{\Phi}$ makes our non-negative component analysis an optimization problem of minimizing the cost function $F(\boldsymbol{W}) = \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\|^2$ with respect to $\boldsymbol{W}$, where $\|.\|$ represents the matrix norm. Considering the Frobenius norm, the minimization problem can be denoted by

$$\min_{\boldsymbol{W}\ge 0} F_1(\boldsymbol{W}) = \min_{\boldsymbol{W}\ge 0} trace\left\{(\boldsymbol{X} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X})(\boldsymbol{X} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X})^T\right\}. \quad (1)$$

2.2 Locality Preserving Bases with Graph Embedding

In order to impose locality preserving properties, we split the set of projective bases into two sets of $\boldsymbol{W} = [\hat{\boldsymbol{W}}, \tilde{\boldsymbol{W}}]$ in which $\hat{\boldsymbol{W}} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_q] \in \mathbb{R}^{m\times q}$ ($q < p$) captures the *discriminative* basis components while $\tilde{\boldsymbol{W}} = [\mathbf{w}_{q+1}, \ldots, \mathbf{w}_p] \in \mathbb{R}^{m\times(p-q)}$ is the complimentary space containing the *reconstructive* basis components which minimizes the reconstruction error together with $\tilde{\boldsymbol{W}}$. Thus, the coefficient matrix is also split into $\boldsymbol{\Phi} = \begin{bmatrix} \hat{\boldsymbol{\Phi}} \\ \tilde{\boldsymbol{\Phi}} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{W}}^T \\ \tilde{\boldsymbol{W}}^T \end{bmatrix} \boldsymbol{X}$. A proper modeling of such intent would provide at most q of those bases which are likeliest to provide discrimination to belong to $\hat{\boldsymbol{W}}$.

Assuming that the two-group multivariate m-dimensional points, e.g. connectivity elements of patients and controls, lie on a manifold, we would like the basis components captured in $\hat{\boldsymbol{W}}$ to be discriminative, meaning that they should group the p-dimensional coefficients corresponding to the similar (i.e. nearby) m-dimensional points close to each other while keeping the p-dimensional coefficients corresponding to dissimilar (i.e. far) m-dimensional points as far as possible after dimensionality reduction. In this section, we propose a model to satisfy such goals.

To clarify the general idea behind our mathematical modeling given later in this section, suppose that the m-dimensional points of two groups lie on a manifold, as illustrated in Fig. 1(a), and are to be projected into a p=2 dimensional space with q = 1 discriminative and $p - q$ = 1 reconstructive basis. Therefore, between the two orthogonal bases $\vec{\boldsymbol{x}}$ and $\vec{\boldsymbol{y}}$, it is

desirable for $\hat{\boldsymbol{W}}$ to include $\vec{\boldsymbol{y}}$ (which is the most discriminative), while $\tilde{\boldsymbol{W}}$ is to include $\vec{\boldsymbol{x}}$ (which would have the best reconstruction together with $\hat{\boldsymbol{W}}$. Since $\hat{\boldsymbol{W}}$ defines our discriminatory space, it is supposed to provide clustering of coefficients. In order to get the 1-dimensional projections (i.e. coefficients) of *m*-dimensional points clustered in the 1-dimensional space, we need to keep projections of nearby *m*-dimensional points as close as possible. This can be obtained by taking advantage of an intrinsic *k*-nearest-neighbor graph [6]. However, due to the unsupervised nature of our approach, $\hat{\boldsymbol{W}}$ may pick up $\vec{\boldsymbol{x}}$ which keeps points closer to each other than $\vec{\boldsymbol{y}}$ and therefore unfavorably merge the two present groups together. To inhibit this, we incorporate a second graph to keep the projections of the distant *m*-dimensional points as far as possible. Then, we impose the reconstructive basis $\tilde{\boldsymbol{W}}$ to keep the projections of the distant *m*-dimensional points close to each other, i.e. imposing $\tilde{\boldsymbol{W}}$ to pick up $\vec{\boldsymbol{x}}$. This will prevent the discriminatory basis from picking up $\vec{\boldsymbol{x}}$ because it is already picked by $\tilde{\boldsymbol{W}}$ and also because there is some degrees of inherent orthogonality [12] in the bases for the best reconstruction. As a result, the discriminatory basis picks up hyperplanes which maintain coefficients of nearby *m*-dimensional points together but keeps coefficients of distant points as far as possible, i.e. in this example $\hat{\boldsymbol{W}}$ will be $\vec{\boldsymbol{y}}$.

There are a variety of approaches that can characterize separability of multivariate data-points. Most of such techniques can be unified in the framework of graph embedding [6]. Let $G = \{\boldsymbol{X}, \boldsymbol{S}\}$ be an undirected weighted graph of *n* vertices, i.e. data points $\boldsymbol{x}_i$, with a symmetric similarity matrix $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ with non-negative elements corresponding to the edge weight of the graph ($\boldsymbol{S}$ has zero diagonal elements). The Laplacian matrix $\boldsymbol{L}$ of the graph is then defined by

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{S}, \quad D_{ii} = \sum_{j=1}^{n} S_{ij}, \ \forall i. \quad (2)$$

In order for the bases in $\hat{\boldsymbol{W}}$ to provide discriminatory information, we would like the resulting coefficients of nearby $\boldsymbol{x}_i$ points to stay close to each other when projected into $\hat{\boldsymbol{W}}$. To satisfy such intent, we first construct a *k*-nearest-neighbor graph $\hat{G} = \{\boldsymbol{X}, \hat{\boldsymbol{S}}\}$ of the *m*-dimensional points $\boldsymbol{x}_i$, as illustrated in Fig. 1(b), in which the edge weight of neighbor points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is defined by

$$\hat{S}_{ij} = e^{-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{\hat{\sigma}^2}}, \quad (3)$$

where $\hat{\sigma}$ is a scaling parameter. In this scheme, $\hat{S}_{ij}$ is non-zero, if and only if $\boldsymbol{x}_i$ is among the *k*-nearest-neighbors of $\boldsymbol{x}_j$ or vice versa, i.e. $\hat{\boldsymbol{S}}$ is sparse and symmetric. Hence, minimizing the following cost function will preserve coefficients (φ_i) of the nearby points as close as possible.

$$\min_{\boldsymbol{W} \geq 0} F_2(\boldsymbol{W}) = \min_{\hat{\boldsymbol{W}} \geq 0} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\hat{\boldsymbol{\phi}}_i - \hat{\boldsymbol{\phi}}_j\|^2 \hat{S}_{ij} = \min_{\hat{\boldsymbol{W}} \geq 0} trace \left\{ \hat{\boldsymbol{\Phi}} \hat{\boldsymbol{L}} \hat{\boldsymbol{\Phi}}^T \right\}. \quad (4)$$

According to the equation (4), if data-points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are close, their graph edge weight S_{ij} will be large, and therefore, the cost function $F_2(\hat{\boldsymbol{W}})$ gets minimized only if the corresponding coefficients φ_i and φ_j remain close.

As explained earlier, to avoid the issue of merging the two groups by $\hat{\boldsymbol{W}}$, we introduce the graph of *k*-farthest points $\tilde{G} = \{\boldsymbol{X}, \tilde{\boldsymbol{S}}\}$ where, as illustrated in Fig. 1(c), each point $\boldsymbol{x}_i$ is

connected by a non-zero weighted edge to its *k* most distant points in the *m*-dimensional space by

$$\tilde{S}_{ij}=e^{-\frac{\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2}{\tilde{\sigma}^2}}, \quad (5)$$

where $\tilde{\sigma}$ is a scaling parameter. $\tilde{S}_{ij}$ is non-zero, if and only if $\boldsymbol{x}_i$ is among the *k*- farthest points of $\boldsymbol{x}_j$ or vice versa, i.e. $\tilde{\boldsymbol{S}}$ is sparse and symmetric. With a similar rationale as in the nearest-neighbor graph $\hat{G}$, we exploit this graph $\tilde{G}$ to impose $\boldsymbol{W}$ to keep the representative coefficients (φ_i) of the the farthest points as close as possible in the lower dimensional space. This is performed by minimizing

$$\min_{\boldsymbol{W}\geq 0} F_3(\boldsymbol{W})=\min_{\widehat{\boldsymbol{W}}\geq 0}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\hat{\boldsymbol{\phi}}_i-\hat{\boldsymbol{\phi}}_j\|^2\hat{S}_{ij}=\min_{\widehat{\boldsymbol{W}}\geq 0} trace\left\{\tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{L}}\tilde{\boldsymbol{\Phi}}^T\right\}. \quad (6)$$

This will lead to resolve the aforementioned issue, and thereby $\hat{\boldsymbol{W}}$ will gain discriminatory properties by picking up the projection direction than maintains the original nearby points as close as possible while keeping the distant points as far as possible.

2.3 Objective Function

To achieve the above three objectives of (1), (4), and (6), the final objective function is to minimize $F(\boldsymbol{W}) = F_1(\boldsymbol{W}) + \lambda (F_2(\boldsymbol{W}) + F_3(\boldsymbol{W}))$, and according to the projective properties of the model, i.e. $\boldsymbol{\Phi} = \boldsymbol{W}^T\boldsymbol{X}$, the final objective function can be rewritten as follows

$$F(\boldsymbol{W})=trace\left\{(\boldsymbol{X}-\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X})(\boldsymbol{X}-\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X})^T\right\}+\lambda\left(trace\left\{\widehat{\boldsymbol{W}}^T\boldsymbol{X}\hat{\boldsymbol{L}}\boldsymbol{X}^T\widehat{\boldsymbol{W}}\right\}+trace\left\{\tilde{\boldsymbol{W}}^T\boldsymbol{X}\tilde{\boldsymbol{L}}\boldsymbol{X}^T\tilde{\boldsymbol{W}}\right\}\right), \quad (7)$$

where λ is a tunable parameter to balance the two terms of reconstruction error norm and graph embedding.

2.4 Optimization Solution

Minimizing the objective function of the equation (7) with non-negativity constraints of $\boldsymbol{W}$ yields the optimal projective bases among which *q* likeliest discriminative ones are obtained in $\hat{\boldsymbol{W}}$. To minimize our objective function, we use a gradient descent approach, i.e. updating $W_{ij}=W_{ij}-\eta_{ij}\frac{\partial F}{\partial W_{ij}}$ with a positive step-size η_{ij}, where

$$\frac{\partial F}{\partial \boldsymbol{W}}=-4(\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W})+2(\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W})+2(\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W})+\left[2\lambda\boldsymbol{X}\hat{\boldsymbol{L}}\boldsymbol{X}^T\widehat{\boldsymbol{W}}, 2\lambda\boldsymbol{X}\tilde{\boldsymbol{L}}\boldsymbol{X}^T\tilde{\boldsymbol{W}}\right]. \quad (8)$$

Regarding that $\boldsymbol{L}\hat{=}\boldsymbol{D}\hat{-}\hat{\boldsymbol{S}}$ and $\boldsymbol{L}\tilde{=}\boldsymbol{D}\tilde{-}\boldsymbol{S}$, and the fact that both $\boldsymbol{D}$ and $\boldsymbol{S}$ matrices have non-negative elements, our non-negativity constraint is guaranteed by positive initialization of $\boldsymbol{W}$ and applying the step-size as follows:

$$\eta_{i,j}=\frac{\frac{1}{2}W_{ij}}{(\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W})_{ij}+(\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W})_{ij}+\left[\lambda\boldsymbol{X}\widehat{\boldsymbol{D}}\boldsymbol{X}^T\widehat{\boldsymbol{W}}, \lambda\boldsymbol{X}\tilde{\boldsymbol{D}}\boldsymbol{X}^T\tilde{\boldsymbol{W}}\right]_{ij}}. \quad (9)$$

This results in the a multiplicative updating solution as

$$W_{ij}=W_{ij}\frac{\left(2\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}+\lambda\left[\boldsymbol{X}\hat{\boldsymbol{S}}\boldsymbol{X}^T\widehat{\boldsymbol{W}},\boldsymbol{X}\tilde{\boldsymbol{S}}\boldsymbol{X}^T\tilde{\boldsymbol{W}}\right]\right)_{ij}}{\left(\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}+\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W}+\lambda\left[\boldsymbol{X}\widehat{\boldsymbol{D}}\boldsymbol{X}^T\widehat{\boldsymbol{W}},\boldsymbol{X}\tilde{\boldsymbol{D}}\boldsymbol{X}^T\tilde{\boldsymbol{W}}\right]\right)_{ij}}. \quad (10)$$

For stability of the convergence, at each iteration, each column of $\boldsymbol{W}$ is normalized by $\boldsymbol{w}_i=\frac{\boldsymbol{w}_i}{\|\boldsymbol{w}_i\|_2}$. Starting with initial random positive elements on $\boldsymbol{W}$, the iterative procedure will converge to the desired $\boldsymbol{W}\geq 0$, whose first q columns are likeliest discriminative bases and the rest are the reconstructive ones.

2.5 Scaling Parameter of Graph Edge Weights

The parameters $\hat{\sigma}$ and $\tilde{\sigma}$ in equations (3) and (5) are a scaling measure of similarity between two points. Such scaling parameters are commonly set by trial and error, but this approach requires manual intervention and is time-consuming [13]. We propose to set the scaling parameter of the graph $\hat{G}$ by

$$\hat{\sigma}=\frac{1}{n}\sum_{i=1}^{n}\hat{\delta}_i \;;\quad \hat{\delta}_i=\|\boldsymbol{x}_i-\hat{\hat{\boldsymbol{x}}}_{i,k}\|_2, \quad (11)$$

where $\hat{\hat{\boldsymbol{x}}}_{i,k}$ is the most distant point among the k-nearest neighbors of $\boldsymbol{x}_i$.

This results in a suitable scaling measure because $\hat{\sigma_i}$ becomes large for the outliers and small for the points near the center of each distribution in the high dimensional space. The average of the $\hat{\sigma_i}$s is dominated by the edges of the points around the center of population distributions, because the number of points around the distribution center exceeds the number of outliers. With the same rationale, the scaling parameter of the graph G is set by

$$\tilde{\sigma}=\frac{1}{n}\sum_{i=1}^{n}\tilde{\delta}_i \;;\quad \tilde{\delta}_i=\|\boldsymbol{x}_i-\tilde{\tilde{\boldsymbol{x}}}_{i,k}\|_2, \quad (12)$$

where $\tilde{\tilde{\boldsymbol{x}}}_{i,k}$ is the least distant point among the k-farthest points to $\boldsymbol{x}_i$.

2.6 Group Analysis Model

As stated by equation $\boldsymbol{X}\approx\boldsymbol{WPhi;}$, the n connectivity observations, i.e. $\boldsymbol{x}_{i:}$ $1\leq i\leq n$, in the matrix $\boldsymbol{X}$ are approximated by

$$[\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n]\approx[\boldsymbol{w}_1,\boldsymbol{w}_2,\ldots,\boldsymbol{w}_p]\begin{bmatrix}\Phi_{11} & \Phi_{12} & \ldots & \Phi_{1n}\\ \vdots & \vdots & & \vdots\\ \Phi_{p1} & \Phi_{p2} & \ldots & \Phi_{pn}\end{bmatrix}. \quad (13)$$

Each observation vector per subject i is thus, approximately reconstructed by

$$\boldsymbol{x}_i\approx\sum_{j=1}^{p}\Phi_{ji}\boldsymbol{w}_j=\sum_{j=1}^{p}\left(\boldsymbol{w}_j^T\boldsymbol{x}_i\right)\boldsymbol{w}_j \;;\; 1\leq i\leq n. \quad (14)$$

Thereby, the presence of each component $\boldsymbol{w}_j$ in the corresponding connectivity vector of a subject $\boldsymbol{x}_i$, is characterized by the corresponding coefficients ϕ_{ji}. Let us suppose, with no

loss of generality, that the first n_1 elements are from the first group (e.g. population of patients) and the remaining $n_2 = n - n_1$ from the second group (e.g. controls). Therefore, the statistical significance between the set of $\{\Phi_{ji:}\ 1 \le i \le n_1\}$ and $\{\Phi_{ji:}\ n_1 + 1 \le i \le n\}$ describes the importance of the corresponding connectivity basis $\boldsymbol{w}_j$ in differentiating the two groups, which can be verified by a two-sample t-test.

3 Results

The proposed method above provides a linear framework of dimensionality reduction yielding two sets of discriminative and reconstructive network components. The reconstructive basis set is expected to show the primary sub-networks of the overall connectivity which are dominant based on their magnitude of coefficients representing their average activation within the population.

The discriminatory set of basis components are expected to show localized sparse sub-networks which represent population clustering and differentiate the two groups but do not contribute considerably in reconstruction of the original connectivity matrices. These two basis sets help us understand the primary global dominant networks as well as pattern-based discriminatory sub-networks characterizing population differences.

In order to show the effectiveness of our method, we examine it over two separate datasets of DTI connectivity matrices. We will obtain the reconstructive bases as well as pattern-based discriminative bases, and examine the differences between the two groups pooled. Similar to the feature extraction problem, the number of bases (i.e. p and q) is population dependent; however, we show that with relatively small numbers for p and q, we obtain stable group differences.

Dataset demographics

Our first dataset consisted of 83 children, 24 ASD and 59 typically developing (TD), all male, aged 6–18 years (mean=12.9, SD=3.0 in ASD, and mean=11.6, SD=3.2 in TD). A standard t-test showed that the age difference between the two groups was not statistically significant. Our second dataset is a developmental study consisting of 595 subjects with 262 males and 333 females. Subjects are aged 8–22 years (mean=14.9, SD=3.2 in males, and mean=15.4, SD=3.3 in females). A standard t-test showed that there was no significant age difference between the genders.

Establishing structural connectivity

DTI data was acquired for each subject and brain extraction was performed on each diffusion-weighted volume and used in computing a fractional anisotropy (FA) volume. Cortical parcellation and sub-cortical segmentation was carried out for each subject using Freesurfer. For the developmental study, a total of 95 ROIs from the Desikan atlas [14] were extracted to represent the nodes of the structural network, comprising 68 cortical regions and 27 sub-cortical structures. In the ASD study, 79 ROIs were extracted comprising 68 cortical regions and 11 sub-cortical structures.

The seed region was limited to the GM-WM boundary of each ROI for reliable tracking. The seed regions were then transferred to the diffusion space via intra-subject affine coregistration between T1 and FA volumes, to act as node labels. Probabilistic tractography [3] was performed on all the subjects with 5000 streamline fibers sampled per voxel. The result was a 79×79 (in ASD) or 95×95 (in developmental) matrix of weighted connectivity values, where each (i, j) element represents the conditional probability of a pathway between regions i and j, normalized by the active surface area of the ROI i. This matrix was treated as a weighted, undirected symmetric network of each subject.

3.1 Connectivity analysis in ASD

The 79×79 connectivity matrix of each subject was vectorized to their $m = 3081$ upper triangular elements. In order to compute the bases and their linearly projected coefficients, we used a three-nearest-neighbor graph for $\hat{G}$ as well as a 3-farthest-point graph for G (i.e. $k = 3$) and correspondingly calculated their graph edge weights and node strengths, i.e. $\boldsymbol{S}$ and $\boldsymbol{D}$. Also, the tuning parameter was set to $\lambda = 1$. We used $p = 5$ for the number of bases with $q = 2$. The iterative procedure of equation (10) was performed and the two discriminative as well as three reconstructive bases learned are shown in Fig. 2.

A statistical group analysis, as described in Sect. 2.6, was performed over the resulting projective coefficients of each $\boldsymbol{w}_i$ basis, i.e. Φ_{ji}. The two-sample t-test is applied to the coefficients of each basis from the pooled population of AST–TD subjects and the p–values and t–values are given in Table 1, as well as the average of the coefficients in the entire population as a ranking criterion of their activation magnitude in the population. The average values here are scaled down similarly for all bases due to their large values.

It is observed that the discriminatory bases are quite sparse with localized patterns, as expected. The large average of the reconstructive coefficients shows that those bases are playing the main role in reconstructing the overall connectivity while the small coefficients in the discriminative bases confirm that they do not play a significant role in the reconstruction but are important in distinguishing the two groups. The statistical significance of the discriminative basis (a) in Fig. 2 shows distinct connectivity deficiencies in inter-hemisphere subcortical and parietal connections in children with ASD, as well as in short-range frontal connections with other nearby cortical and subcortical regions. The significance of the reconstructive basis (c) also shows that the brain is slightly deficient in children with autism in the overall short-range intra-hemisphere connectivity.

3.2 Connectivity analysis in the developmental study

The $m = 4465$-length vectors of the upper triangular part of the 95 × 95 connectivity matrices were used for basis learning. Similar to the ASD dataset, we used a three-nearest-neighbor graph for $\hat{G}$, as well as a 3-farthest-point graph for G (i.e. $k = 3$) and correspondingly calculated their graph matrices, $\boldsymbol{S}$ and $\boldsymbol{D}$. Also, the tuning parameter was similarly set to $\lambda = 1$. We applied our method with $p = 10$ basis components with the first $q = 6$ bases forming the set of discriminatory ones. The resulting bases are shown in Fig. 3.

The statistical group analysis results of the resulting coefficients of each basis are given in Table 2, as well as the average of the coefficients in the entire population. The average values here are scaled down similarly for all bases.

According to the statistics of the coefficients, it is seen from the discriminative set of bases in the top row of Fig. 3 that the discriminative bases are sub-networks of connectivity which mostly show strong group differences in both males and females. Males show stronger intra-hemisphere inter-cortical connectivity structure (see Fig. 3(a), (b), (d)) whereas women are distinguished by their stronger inter-hemisphere connectivity specially in the frontal regions (e.g. see Fig. 3(c)). Also, the four reconstructive bases at the bottom row of this figure show that the overall dominant brain connectivity is a collection of short-range connections which are stronger towards the male.

4 Conclusion

We have presented a novel technique for extracting the discriminative subnetworks of a population via graph embedding. This maps the connectivity patterns of the population onto a lower dimensional space to ease subsequent population statistics. Our method consists of

basis learning which simultaneously minimizes the reconstruction error, as well as provides discriminative bases which identify group differences in the presence of non-negativity constraints over both bases and coefficients. The method was evaluated on two datasets of connectivity analysis, one in a study of ASD which revealed significant inter-hemisphere connectivity deficiencies in a set of interpretable connections as part of a discriminatory basis. The second developmental dataset also showed dominant intra-hemisphere connectivity in males while the inter-hemisphere frontal connectivity appeared stronger in females. The presented technique represents a framework, that is in principle capable of handling other types of non-negative functional or structural connectivity networks from any modality for statistical group analysis. Its linearity makes our technique attractive for the unsupervised feature selection applications as well.

References

1. Vissers M, Cohen M, Geurts H. Brain connectivity and high functioning autism: a promising path of research that needs refined models, methodological convergence, and stronger behavioral links. Neurosci Biobehav Rev. 2012; 36(1):604–25. [PubMed: 21963441]
2. Dennis EL, Jahanshad N, et al. Development of brain structural connectivity between ages 12 and 30: A 4-tesla diffusion imaging study in 439 adolescents and adults. Neuroimage. Jan.2013 64:671–684. [PubMed: 22982357]
3. Behrens TEJ, Johansen-Berg H, et al. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nat Neurosci. Jul; 2003 6(7):750–757. [PubMed: 12808459]
4. Calhoun V, Kiehl K, Pearlson G. Modulation of temporally coherent brain networks estimated using ica at rest and during cognitive tasks. Hum Brain Mapp. 2008; 29(7):828–838. [PubMed: 18438867]
5. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature. Oct; 1999 401(6755):788–791. [PubMed: 10548103]
6. Yan S, Xu D, Zhang B, Zhang HJ, Yang Q, Lin S. Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans Pattern Anal Mach Intell. Jan; 2007 29(1):40–51. [PubMed: 17108382]
7. Batmanghelich N, Taskar B, Davatzikos C. Generative-discriminative basis learning for medical imaging. IEEE Trans Med Imaging. 2011; 31(1):51–69. [PubMed: 21791408]
8. Wang C, Yan S, Zhang L, Zhang H. Non-negative semi-supervised learning. Journal of Machine Learning Research. 2009; 5:575–582.
9. Berry MW, Browne M, Langville AN, Pauca VP, Plemmons RJ. Algorithms and applications for approximate nonnegative matrix factorization. Computational Statistics and Data Analysis. 2007; 52:155–173.
10. Yang Z, Oja E. Linear and nonlinear projective nonnegative matrix factorization. IEEE Trans Neural Netw. 2010; 21(5):1734–749.
11. Hoyer PO. Non-negative matrix factorization with sparseness constraints. The Journal of Machine Learning Research. 2004; 5:1457–1469.
12. Ghanbari Y, Bloy L, et al. Dominant component analysis of electrophysiologi-cal connectivity networks. Medical Image Computing and Computer-Assisted Intervention MICCAI 2012. 2012; 7512:231–238.
13. Manor LZ, Perona P. Self-tuning spectral clustering. Advances in Neural Information Processing Systems. 2005:1601–1608.
14. Desikan R, Sgonne F, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. NeuroImage. 2006; 31:968–980. [PubMed: 16530430]

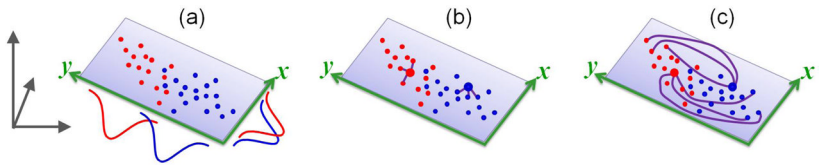

Fig. 1.

Illustration of a two-group multivariate point distribution on a manifold in the *m*-dimensional space. (a) The point distribution when projected into the direction $\vec{x}$ or $\vec{y}$. (b) The 3-nearest-neighbor graph of two selected magnified points. (c) The 3-farthest-point graph of the same two selected points as in b.

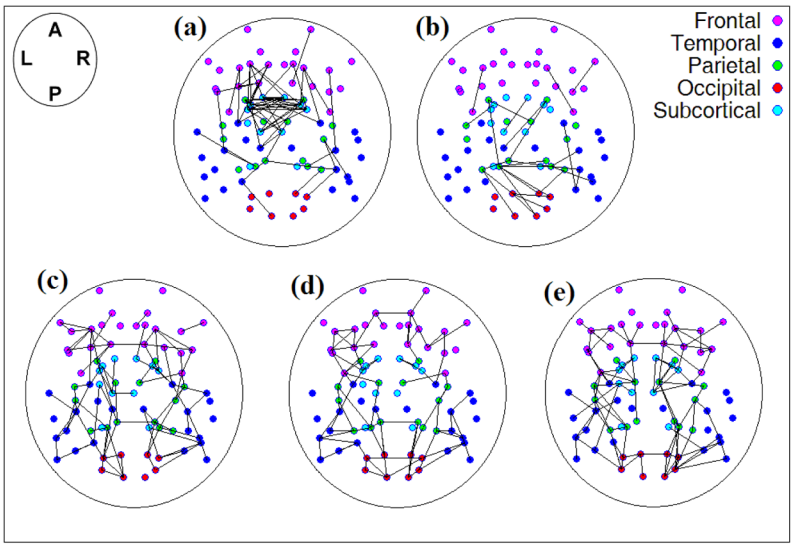

Fig. 2.

The $p = 5$ connectivity bases (ASD study) learned by the proposed method. (a)–(b) are the $q = 2$ discriminatory, and (c)–(e) are the three reconstructive bases.

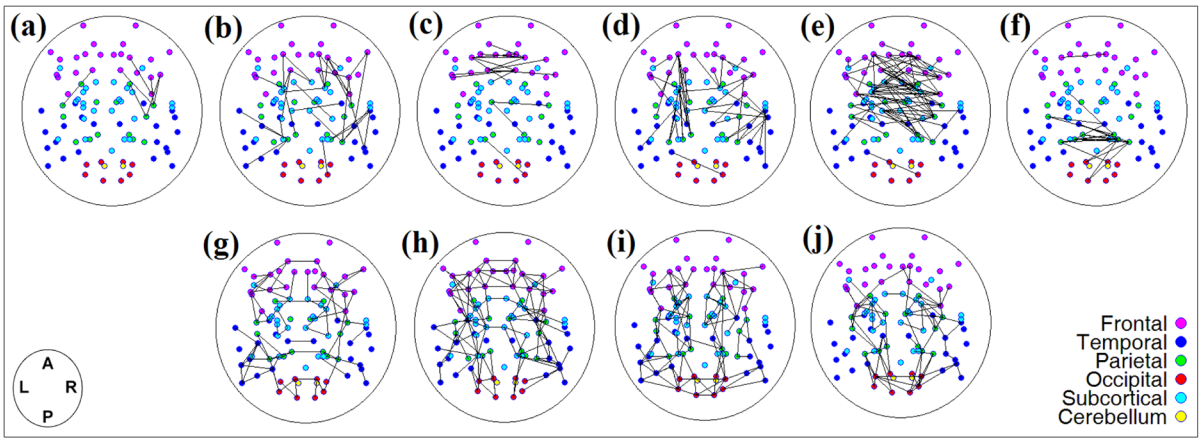

Fig. 3.

The $p = 10$ connectivity bases (developmental study) learned by the proposed method. (a)–(f) are the $q = 6$ discriminatory, and (g)–(j) are the 4 reconstructive bases.

Table 1

Statistical group analysis of the coefficients of the ASD-TD bases

Component label	Pooled ASD-TD average coefficients	ASD-TD group *p* – value	ASD-TD group *t*–value
a	0.6	0.002	−3.3
b	0.4	0.96	−0.1
c	5.5	0.02	−2.5
d	5.8	0.44	−0.8
e	5.7	0.16	−1.4

Table 2

Statistical group analysis of the developmental basis coefficients

Component label	Pooled Male-Female average coefficients	Male-Female group *p*–value	Male-Female group *t*–value
a	0.15	9.8e–12	+7.0
b	0.17	1.5e–6	+4.9
c	0.10	1.6e–3	−3.2
d	0.13	1.2e–2	+2.5
e	0.17	0.56	−0.59
f	0.21	0.69	+0.39
g	4.1	6.1e–14	+7.7
h	4.2	2.5e–9	+6.1
i	4.4	4.2e–15	+8.1
j	3.6	3.0e–12	+7.1