Sparse Canonical Correlation Analysis Based on Rank-1 Matrix Approximation and its Application for fMRI Signals

Abdeldjalil Aissa El Bey, Abd-Krim Seghouane

SPARSE CANONICAL CORRELATION ANALYSIS BASED ON RANK-1 MATRIX APPROXIMATION AND ITS APPLICATION FOR FMRI SIGNALS

Abdeldjalil Aïssa-El-Bey

Institut Mines-Télécom; Télécom Bratgne
UMR CNRS 6285 Lab-STIC, Technopôle
Brest-Iroise CS 83818, 29238 Brest, France

Abd-Krim Seghouane

University of Melbourne, Department of
Electrical and Electronic Engineering
Parkville, Australia

ABSTRACT

Canonical correlation analysis (CCA) is a well-known technique used to characterize the relationship between two sets of multidimensional variables by finding linear combinations of variables with maximal correlation. Sparse CCA or regularized CCA are two widely used variants of CCA because of the improved interpretability of the former and the better performance of the later. So far the cross-matrix product of the two sets of multidimensional variables has been widely used for the derivation of these variants. In this paper a new algorithm for sparse CCA is proposed. This algorithm differs from the existing ones in their derivation which is based on penalized rank one matrix approximation and the orthogonal projectors onto the space spanned by the two sets of multidimensional variables instead of the simple cross-matrix product. The performance and effectiveness of the proposed algorithm are tested on simulated experiments. On these results it can be observed that they outperform the state of the art sparse CCA algorithms.

Index Terms— Canonical correlation analysis (CCA), Sparse representation, Rank-1 matrix approximation.

1. INTRODUCTION

Canonical correlation analysis (CCA) [1] is a multivariate analysis method, the aim of which is to identify and quantify the association between two sets of variables. The two sets of variables can be associated with a pair of linear transforms (projectors) such that the correlation between the projections of the variables in lower-dimensional space through these linear transforms are mutually maximized. The pair of canonical projectors are easily obtained by solving a simple generalized eigenvalue decomposition problem, which only involves the covariance and cross-covariance matrices of the considered random vectors. CCA has been widely applied in many important fields, for instance, facial expression recognition [2], detection of neural activity in functional magnetic resonance imaging (fMRI) [3, 4], machine learning [5, 6] and blind source separation [7, 8]. In the context of high-dimensional data, there are usually a large portion of features that are not informative in data analysis. When the canonical variables involve all features in the original space, the canonical projectors are in general not sparse. Therefore, it is not easy to interpret canonical variables in such high-dimensional data analysis. These problems may be tackled by selecting sparse subsets of variables, i.e. obtaining sparse canonical projectors in the linear combinations of variables of each data set [6, 9–11]. For example, in [10] the authors propose a new criterion for sparse CCA and applied a penalized matrix decomposition approach to solve the sparse CCA problem and in [9] the presented sparse CCA approach computes the canonical projectors from primal and dual representations.

In this paper, we adopt an alternative formulation of CCA problem which is based on rank-1 matrix approximation of the orthogonal projectors of data sets [12]. Based on this new formulation of CCA problem, we developed a new sparse CCA based on penalized rank-1 matrix approximation which aims to overcome the drawback of CCA in the context of high-dimensional data and improve interpretability. The proposed sparse CCA seeks to obtain iteratively a sparse pairwise of canonical projectors by solving a penalized rank-1 matrix approximation via sparse coding method. This proposed algorithm differs from the existing ones in their derivation which is based on penalized rank one matrix approximation and the orthogonal projectors onto the space spanned by the two sets of multidimensional variables instead of the simple cross-matrix product [6, 9–11].

2. CANONICAL CORRELATION ANALYSIS

In this section, we present briefly a review of CCA and its optimization problem. Let $\boldsymbol{x} \in \mathbb{R}^{d_x}$ and $\boldsymbol{y} \in \mathbb{R}^{d_y}$ be two random vectors and we assume, without loss of generality, that both $\boldsymbol{x}$ and $\boldsymbol{y}$ have zero mean, i.e. $\mathbb{E}[\boldsymbol{x}] = \boldsymbol{0}$ and $\mathbb{E}[\boldsymbol{y}] = \boldsymbol{0}$ where $\mathbb{E}[\cdot]$ is the expectation operator. CCA seeks a pair of linear transform $\boldsymbol{w}_x \in \mathbb{R}^{d_x}$ and $\boldsymbol{w}_y \in \mathbb{R}^{d_y}$, such that correlation between $\boldsymbol{w}_x^T\boldsymbol{x}$ and $\boldsymbol{w}_y^T\boldsymbol{y}$ is maximized. Mathematically, the objective function to be maximized is given by:

$$\rho(\boldsymbol{w}_x, \boldsymbol{w}_y) = \frac{\operatorname{cov}(\boldsymbol{w}_x^T\boldsymbol{x}, \boldsymbol{w}_y^T\boldsymbol{y})}{\sqrt{\operatorname{var}(\boldsymbol{w}_x^T\boldsymbol{x})\operatorname{var}(\boldsymbol{w}_y^T\boldsymbol{y})}}. \tag{1}$$

Then, the objective function ρ can be rewritten as:

$$\rho(\boldsymbol{w}_x, \boldsymbol{w}_y) = \frac{\boldsymbol{w}_x^T\boldsymbol{C}_{xy}\boldsymbol{w}_y}{\sqrt{(\boldsymbol{w}_x^T\boldsymbol{C}_{xx}\boldsymbol{w}_x)(\boldsymbol{w}_y^T\boldsymbol{C}_{yy}\boldsymbol{w}_y)}}, \tag{2}$$

where $\boldsymbol{C}_{xx} = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T]$, $\boldsymbol{C}_{yy} = \mathbb{E}[\boldsymbol{y}\boldsymbol{y}^T]$ and $\boldsymbol{C}_{xy} = \mathbb{E}[\boldsymbol{x}\boldsymbol{y}^T]$ are the covariance matrices. Since the value of $\rho(\boldsymbol{w}_x, \boldsymbol{w}_y)$ is invariant with the magnitude of the projection direction, we can turn to solve the following optimization problem

$$\begin{aligned} \underset{\boldsymbol{w}_x, \boldsymbol{w}_y}{\arg\max} \quad & \boldsymbol{w}_x^T\boldsymbol{C}_{xy}\boldsymbol{w}_y \\ \text{subject to} \quad & \boldsymbol{w}_x^T\boldsymbol{C}_{xx}\boldsymbol{w}_x = 1, \quad \boldsymbol{w}_y^T\boldsymbol{C}_{yy}\boldsymbol{w}_y = 1. \end{aligned}$$

One way to solve this problem is as proposed in [5] by assuming $\boldsymbol{C}_{yy}$ is invertible, we can write

$$\boldsymbol{w}_y = \frac{1}{\lambda}\boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{xy}^T\boldsymbol{w}_x, \tag{3}$$

where λ is the Lagrangian parameter and by substituting and assuming $\boldsymbol{C}_{xx}$ is invertible gives

$$\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy}\boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{xy}^{T}\boldsymbol{w}_x = \lambda^2\,\boldsymbol{w}_x. \tag{4}$$

It has been shown in [5] that we can choose the associated eigenvectors corresponding to the top eigenvalues of the generalized eigenvalue problem in (4) and then use (3) to find the corresponding $\boldsymbol{w}_y$. A number of existing methods for sparse CCA have used the description provided above of CCA and focused on the use of the cross matrix $\boldsymbol{C}_{xy}$ for the derivation of new CCA variant algorithms [6, 9–11]. For the derivation of the proposed CCA variants we adopt an alternative description of CCA which is based on the orthogonal projectors onto the space spanned by the two sets of multidimensional variables [12].

3. CANONICAL CORRELATION ANALYSIS BASED ON RANK-1 MATRIX APPROXIMATION

In practice, the covariance matrices $\boldsymbol{C}_{xx}$, $\boldsymbol{C}_{yy}$ and $\boldsymbol{C}_{xy}$ are usually not available. Instead, the estimated covariance matrices are constructed based on given sample data. Let $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N] \in \mathbb{R}^{d_x \times N}$ and $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N] \in \mathbb{R}^{d_y \times N}$ are two sets of instances of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. Then, the optimization problem for CCA based on estimated covariance matrices is given by

$$\begin{aligned} \underset{\boldsymbol{w}_x, \boldsymbol{w}_y}{\arg\max} \quad & \boldsymbol{w}_x^T \boldsymbol{X}\boldsymbol{Y}^T \boldsymbol{w}_y \\ \text{subject to} \quad & \boldsymbol{w}_x^T \boldsymbol{X}\boldsymbol{X}^T \boldsymbol{w}_x = 1, \quad \boldsymbol{w}_y^T \boldsymbol{Y}\boldsymbol{Y}^T \boldsymbol{w}_y = 1, \end{aligned} \tag{5}$$

and the generalized eigenvalue problem can be rewritten as

$$\boldsymbol{X}\boldsymbol{Y}^T\boldsymbol{w}_y = \lambda\,\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{w}_x \tag{6}$$

$$\boldsymbol{Y}\boldsymbol{X}^T\boldsymbol{w}_x = \lambda\,\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{w}_y. \tag{7}$$

Then, by multiplying the both side of equations (6) and (7) by $\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}$ and $\boldsymbol{Y}^T(\boldsymbol{Y}\boldsymbol{Y}^T)^{-1}$ respectively, we obtain:

$$\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\boldsymbol{Y}^T\boldsymbol{w}_y = \boldsymbol{P}_x\boldsymbol{Y}^T\boldsymbol{w}_y = \lambda\,\boldsymbol{X}^T\boldsymbol{w}_x \tag{8}$$

$$\boldsymbol{Y}^T(\boldsymbol{Y}\boldsymbol{Y}^T)^{-1}\boldsymbol{Y}\boldsymbol{X}^T\boldsymbol{w}_x = \boldsymbol{P}_y\boldsymbol{X}^T\boldsymbol{w}_x = \lambda\,\boldsymbol{Y}^T\boldsymbol{w}_y, \tag{9}$$

where $\boldsymbol{P}_x = \boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}$ and $\boldsymbol{P}_y = \boldsymbol{Y}^T(\boldsymbol{Y}\boldsymbol{Y}^T)^{-1}\boldsymbol{Y}$ are the orthogonal projectors onto the linear spans of the rows of $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively. So substituting $\boldsymbol{X}^T\boldsymbol{w}_x$ in equation (9) and $\boldsymbol{Y}^T\boldsymbol{w}_y$ in equation (8) gives

$$\begin{aligned} \boldsymbol{P}_x\boldsymbol{P}_y\boldsymbol{X}^T\boldsymbol{w}_x &= \lambda^2\,\boldsymbol{X}^T\boldsymbol{w}_x \\ \boldsymbol{P}_y\boldsymbol{P}_x\boldsymbol{Y}^T\boldsymbol{w}_y &= \lambda^2\,\boldsymbol{Y}^T\boldsymbol{w}_y, \end{aligned}$$

Therefore, the rank-1 matrix approximation of $\boldsymbol{K}_{xy} = \boldsymbol{P}_x\boldsymbol{P}_y$ can be formulated as solving the following optimization from:

$$\underset{\boldsymbol{w}_x, \boldsymbol{w}_y}{\arg\min} \|\boldsymbol{K}_{xy} - \boldsymbol{X}^T\boldsymbol{w}_x\boldsymbol{w}_y^T\boldsymbol{Y}\|_F^2 \tag{10}$$

where $\|\cdot\|_F^2$ is the squared Frobenius norm. Consequently, the projected data $\boldsymbol{w}_x^T\boldsymbol{X}$ and $\boldsymbol{w}_y^T\boldsymbol{Y}$ consist on the eigenvectors associated to the largest eigenvalue of the matrix $\boldsymbol{K}_{xy}$. Hence, for multiple projected data the solution consist on the associated eigenvectors corresponding to the top eigenvalues of the matrix $\boldsymbol{K}_{xy}$. One disadvantage of the above approach is the restriction that $\boldsymbol{X}\boldsymbol{X}^T$ and $\boldsymbol{Y}\boldsymbol{Y}^T$ must be nonsingular. In order to prevent overfitting and avoid the singularity of $\boldsymbol{X}\boldsymbol{X}^T$ and $\boldsymbol{Y}\boldsymbol{Y}^T$ [5], two regularization terms $\gamma_x\boldsymbol{I}_{d_x}$ and $\gamma_y\boldsymbol{I}_{d_y}$, with $\gamma_x > 0$, $\gamma_y > 0$ are added in (5). Therefore, the regularized version solves the generalized eigenvalue problem with $\boldsymbol{P}_x = \boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T + \gamma_x\boldsymbol{I}_{d_x})^{-1}\boldsymbol{X}$ and $\boldsymbol{P}_y = \boldsymbol{Y}^T(\boldsymbol{Y}\boldsymbol{Y}^T + \gamma_y\boldsymbol{I}_{d_y})^{-1}\boldsymbol{Y}$.

4. SPARSE CCA ALGORITHM BASED ON RANK-1 MATRIX APPROXIMATION

In this section, we will propose the sparse CCA method based on rank-1 matrix approximation by penalizing the optimization problem (10). Then, we propose an efficient iterative algorithm to solve the sparse solution of the proposed criterion. In general cases, the canonical projectors $\boldsymbol{w}_x$ and $\boldsymbol{w}_y$ solutions of equation (10) are not sparse, i.e., the entries of both $\boldsymbol{w}_x$ and $\boldsymbol{w}_y$ are nonzeros. To obtain the sparse solution, we adopt the similar trick used in [6, 10, 11] by imposing penalty functions on the optimization problem (10). Therefore, we can write the new optimization problem as:

$$\begin{aligned} \underset{\boldsymbol{w}_x, \boldsymbol{w}_y}{\arg\min} \quad & \|\boldsymbol{K}_{xy} - \boldsymbol{X}^T\boldsymbol{w}_x\boldsymbol{w}_y^T\boldsymbol{Y}\|_F^2 \\ \text{subject to} \quad & \mathcal{F}_x(\boldsymbol{w}_x) \le \beta_x \quad \text{and} \quad \mathcal{F}_y(\boldsymbol{w}_y) \le \beta_y, \end{aligned} \tag{11}$$

where $\mathcal{F}_x(\cdot)$ and $\mathcal{F}_y(\cdot)$ are penalty functions, which can take on a variety of forms. Useful examples are: ℓ_0-quasi-norm $\mathcal{F}(\boldsymbol{z}) = \|\boldsymbol{z}\|_0$ which count the nonzero entries of a vector; Lasso penalty with ℓ_1-norm $\mathcal{F}(\boldsymbol{z}) = \|\boldsymbol{z}\|_1$ and so on. The optimization problem (11) can be alternatively solved by optimizing $\boldsymbol{w}_x$ and $\boldsymbol{w}_y$. Specifically, we first fix $\boldsymbol{w}_y$ and solve $\boldsymbol{w}_x$ by minimizing (11). Then, we fix $\boldsymbol{w}_x$ and minimize (11) to obtain $\boldsymbol{w}_y$. The above two procedures are repeated until convergence. The straightforward approach to solve this problem is to formulate it as an ordinary sparse coding task. Then, for a fix $\boldsymbol{w}_y$ the problem (11) is equivalent to much simpler sparse coding problem

$$\underset{\boldsymbol{w}_x}{\arg\min} \|\boldsymbol{K}_{xy}\boldsymbol{Y}^T\boldsymbol{w}_y - \boldsymbol{X}^T\boldsymbol{w}_x\|_2^2 \quad \text{subject to} \quad \mathcal{F}_x(\boldsymbol{w}_x) \le \beta_x$$

which can be solved by using any sparse approximation method. In the same way, we can solve the problem (11) regarding $\boldsymbol{w}_y$ for a fix $\boldsymbol{w}_x$ by minimizing the following criterion:

$$\underset{\boldsymbol{w}_y}{\arg\min} \|\boldsymbol{K}_{xy}^T\boldsymbol{X}^T\boldsymbol{w}_x - \boldsymbol{Y}^T\boldsymbol{w}_y\|_2^2 \quad \text{subject to} \quad \mathcal{F}_y(\boldsymbol{w}_y) \le \beta_y$$

Based on the above description, we can obtain the first pairwise of sparse projectors $\boldsymbol{w}_x$ and $\boldsymbol{w}_y$. For multiple projection vectors, we propose to use a deflation procedure where the second pairwise of sparse projectors are defined by using the corresponding residual matrices $\boldsymbol{K}_{xy} - \boldsymbol{w}_x^T\boldsymbol{X}\boldsymbol{K}_{xy}\boldsymbol{Y}^T\boldsymbol{w}_y\boldsymbol{X}^T\boldsymbol{w}_x\boldsymbol{w}_y^T\boldsymbol{Y}$. Using the same way, we can define the other pairwise of sparse projectors. Then, we summarized the method of solving the entire Sparse rank-1 matrix approximation CCA in Algorithm 1

5. EXPERIMENTS

In this section, we present several computer simulations to demonstrate the effectiveness of the proposed algorithm. We compare the performance of the proposed algorithm with existing state of the art sparse CCA methods:

- The sparse CCA presented in [10], relying on a penalized matrix decomposition denoted PMD. An R package implementing this algorithm, called PMA, is available at `http://cran.r-project.org/web/packages/PMA/index.html`. Sparsity parameters are selected using the permutation approach presented in [13] of which the code is provided in PMA package.
- The sparse CCA presented in [6] where the CCA is reformulated as a least-squares problem denoted LS CCA.

Algorithm 1 Sparse rank-1 matrix approximation CCA algorithm

Input: Training data $\boldsymbol{X} \in \mathbb{R}^{d_x \times N}$ and $\boldsymbol{Y} \in \mathbb{R}^{d_y \times N}$.

Output: The r pairs of canonical projector $\boldsymbol{W}_x \in \mathbb{R}^{d_x \times r}$ and $\boldsymbol{W}_y \in \mathbb{R}^{d_y \times r}$.

Compute $\boldsymbol{P}_x = \boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T + \gamma_x \boldsymbol{I}_{d_x})^{-1}\boldsymbol{X}$, $\boldsymbol{P}_y = \boldsymbol{Y}^T(\boldsymbol{Y}\boldsymbol{Y}^T + \gamma_y \boldsymbol{I}_{d_y})^{-1}\boldsymbol{Y}$ and $\boldsymbol{K}_{xy} = \boldsymbol{P}_x\boldsymbol{P}_y$;
for $i = 1, 2, \ldots, r$ **do**
  Perform the SVD of $\boldsymbol{K}_{xy}$: $\boldsymbol{K}_{xy} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T$;
  Initialize $\widetilde{\boldsymbol{u}} = \boldsymbol{u}_1$ and $\widetilde{\boldsymbol{v}} = \boldsymbol{v}_1$;
  repeat
    Update the i-th column of $\boldsymbol{W}_x$: $\boldsymbol{W}_x(:,i) = \underset{\boldsymbol{W}_x(:,i)}{\arg\min} \|\boldsymbol{K}_{xy}\widetilde{\boldsymbol{v}} - \boldsymbol{X}^T\boldsymbol{W}_x(:,i)\|_2^2$ subject to $\mathcal{F}_x(\boldsymbol{W}_x(:,i)) \leq \beta_x$;
    Update $\widetilde{\boldsymbol{u}} = \frac{\boldsymbol{X}^T\boldsymbol{W}_x(:,i)}{\|\boldsymbol{X}^T\boldsymbol{W}_x(:,i)\|_2}$;
    Update the i-th column of $\boldsymbol{W}_y$: $\boldsymbol{W}_y(:,i) = \underset{\boldsymbol{W}_y(:,i)}{\arg\min} \|\boldsymbol{K}_{xy}^T\widetilde{\boldsymbol{u}} - \boldsymbol{Y}^T\boldsymbol{W}_y(:,i)\|_2^2$ subject to $\mathcal{F}_y(\boldsymbol{W}_y(:,i)) \leq \beta_y$;
    Update $\widetilde{\boldsymbol{v}} = \frac{\boldsymbol{Y}^T\boldsymbol{W}_y(:,i)}{\|\boldsymbol{Y}^T\boldsymbol{W}_y(:,i)\|_2}$;
  until convergence
  Update $\boldsymbol{K}_{xy}$: $\boldsymbol{K}_{xy} \leftarrow \boldsymbol{K}_{xy} - \widetilde{\boldsymbol{u}}^T\boldsymbol{K}_{xy}\widetilde{\boldsymbol{v}}\widetilde{\boldsymbol{u}}\widetilde{\boldsymbol{v}}^T$;
end for

A Matlab package implementing this algorithm is available at `http://www.public.asu.edu/~jye02/Software/CCA/`.

- The sparse CCA presented in [11] where the sparse canonical projectors are computed by solving two ℓ_1-minimization problems by using Linearized Bregman iterative method [14]. This algorithm is denoted CCA LB (Linearized Bregman). We re-implemented the sparse CCA algorithm proposed [11] in Matlab.

For proposed sparse CCA algorithm, we have used $\mathcal{F}_x(\boldsymbol{z}) = \mathcal{F}_y(\boldsymbol{z}) = \|\boldsymbol{z}\|_0$ as penalty functions. We solve the sparse coding problem by using orthogonal matching pursuit algorithm [15, 16].

5.1. Synthetic data

This simulation setup is inspired from [17]. The synthetic data $\boldsymbol{X}$ and $\boldsymbol{Y}$ were generating according to multivariate normal distribution, with covariance matrices described in Table 1. The number of simulations with each configuration was $N_r = 1000$. We compare the performance of our algorithm to methods of the state of the art by estimating the precision accuracy of the space spanned by estimated canonical projectors. Then, we compute for each simulation run r the angle $\theta^r(\hat{\boldsymbol{W}}_x^r, \boldsymbol{W}_x)$ between the subspace spanned by the estimated canonical projectors contained in the columns of $\hat{\boldsymbol{W}}^r$ and the subspace spanned by the true canonical projectors contained in the columns of $\boldsymbol{W}_x$ solution of the eigenproblem (4). The same criterion is used for the canonical projectors $\boldsymbol{W}_y$. The average angles are estimated over N_r Monte-Carlo run such that:

$$\theta_x = \frac{1}{N_r}\sum_{r=1}^{N_r}\theta^r(\hat{\boldsymbol{W}}_x^r, \boldsymbol{W}_x) \quad \text{and} \quad \theta_y = \frac{1}{N_r}\sum_{r=1}^{N_r}\theta^r(\hat{\boldsymbol{W}}_y^r, \boldsymbol{W}_y)$$

For each algorithm, we used the following parameters; LS CCA algorithm with $\lambda_x = \lambda_y = 0.5$, CCA LB algorithm with $\mu_x = \mu_y = 2$ and Algorithm 1 with $\beta_x = \beta_y = 3$. The simulation performance on the estimated angle between the subspace spanned by the true canonical projectors and the estimated one by the different methods are reported in Table 2. We can observe that the simulation accuracy

Table 2. Simulation results with synthetic data

	θ_x	θ_y	θ_x	θ_y	θ_x	θ_y
Method	$N = 50$		$N = 100$		$N = 200$	
Scen 1:						
CCA	0.539	0.503	0.346	0.347	0.227	0.238
LS CCA	0.416	0.369	0.264	0.265	0.178	0.187
CCA LB	0.517	0.515	0.331	0.334	0.225	0.222
PMD	0.220	0.242	0.090	0.050	0.020	0.017
Algo 1	0.201	0.219	0.049	0.027	0.004	0.005
Scen 2:						
CCA	0.509	0.668	0.310	0.412	0.208	0.277
LS CCA	0.348	0.508	0.228	0.324	0.160	0.218
CCA LB	0.300	0.376	0.022	0.022	0.001	0.001
PMD	0.2061	0.306	0.023	0.070	0.004	0.044
Algo 1	0.116	0.150	0.001	0.001	10^{-4}	10^{-4}
Scen 3:						
CCA	0.812	0.995	0.560	0.667	0.339	0.448
LS CCA	0.527	0.730	0.355	0.471	0.241	0.344
CCA LB	0.760	0.920	0.278	0.516	0.014	0.315
PMD	0.611	0.827	0.203	0.461	0.039	0.337
Algo 1	0.399	0.685	0.017	0.323	10^{-4}	0.303
Scen 4:						
CCA	1.379	1.376	0.887	0.874	0.470	0.472
LS CCA	0.853	0.829	0.523	0.518	0.337	0.337
CCA LB	1.368	1.365	0.726	0.734	0.047	0.041
PMD	1.397	1.354	1.131	1.034	0.408	0.382
Algo 1	1.118	1.098	0.027	0.027	10^{-4}	10^{-4}

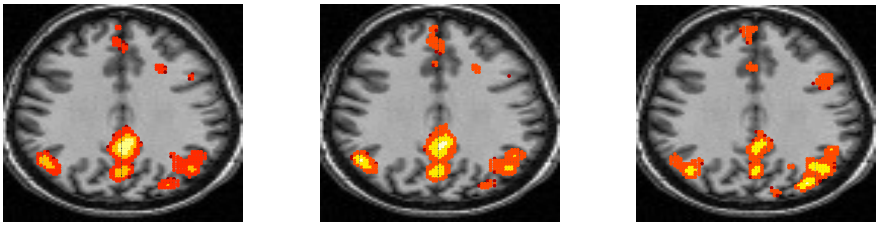
of the proposed sparse CCA method is significantly better compared to other CCA methods, especially for a large number of observations N. In the case of low number of observations the proposed sparse CCA method is still doing well and where the performance gain increases with increasing number of observations. This demonstrates the robustness of our sparse CCA method with respect to the number of available observations and the benefit of using our sparse CCA method in the context of a relatively low number of observations

5.2. Blind source separation for fMRI signals

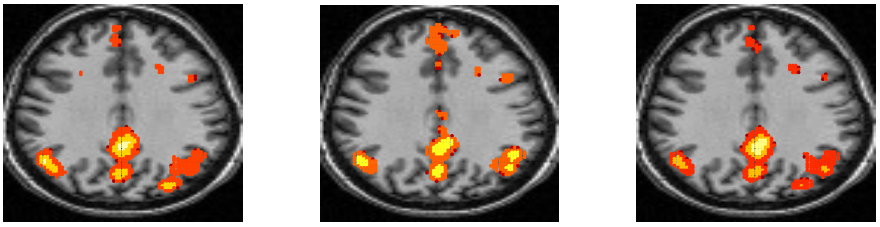
In this section we evaluate the performance of the proposed CCA variant algorithms on a problem of functional magnetic resonance imaging (fMRI) resting state experiment. In this case we are interested in functional connectivity and recovering a resting state network; i.e., the default mode network from a data matrix $\boldsymbol{Y}$ formed by vectorizing each time series observed in every voxel creating a matrix $n \times N$ where n is the number of time points and N the number of voxels ($\approx 10,000 - 100,000$) [18]. To estimate functionally connected brain voxels, response signal strength known as coefficient matrix estimated as $\boldsymbol{X}$ is considered [19]. According to the neural dynamics of interest, coefficient rows can be converted to z-scores to obtain sparsely distributed and clustered origin of the dynamics. The neural dynamics of interest can be obtained by correlating the modulation profile with the time-series representing average

Table 1. Simulation settings

Parameters	d_x	d_y	N	$\boldsymbol{C}_{xx}$	$\boldsymbol{C}_{yy}$	$\boldsymbol{C}_{xy}$
Scenario 1	4	4	{50, 100, 200}	$\boldsymbol{I}_4$	$\boldsymbol{I}_4$	$\begin{bmatrix} \frac{9}{10} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Scenario 2	4	6	{50, 100, 200}	$\boldsymbol{I}_4$	$\boldsymbol{I}_6$	$\begin{bmatrix} \frac{3}{5} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
Scenario 3	6	10	{50, 100, 200}	$\boldsymbol{I}_6$	$\begin{bmatrix} \boldsymbol{M} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_7 \end{bmatrix}$ with $M(i,j) = 0.3^{\lvert i-j \rvert}$	$\frac{1}{2}\begin{bmatrix} \boldsymbol{I}_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$
Scenario 4	20	20	{50, 100, 200}	$\boldsymbol{I}_{20}$	$\boldsymbol{I}_{20}$	$\frac{7}{10}\begin{bmatrix} \boldsymbol{I}_{10} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$



(a) Reference (b) CCA (c) LS CCA, $\lambda = 0.5$



(d) CCA LB, $\mu = 10$ (e) PMD (f) Algorithm 1, $\beta = 4$

Fig. 1. The functional connectivity results of a single subject for default mode network (DMN) using different CCA variant algorithms.

neural dynamics for regions of interest (ROIs). The representative time-series for cortical, subcortical, and cerebellum regions in the brain were obtained by parcellating the whole brain into 116 ROIs using automated anatomical labelling [20]. Only the first functional run from the first subject was used for functional connectivity analysis of a default mode network (DMN). The functional connectivity results of a single subject for DMN using eight different CCA variant algorithms are shown in Figure 1. To obtain these results the modulation profile that was most correlated with posterior cingulate cortex (PCC) representative time-series is used. Using the different CCA variant algorithms, the connected regions obtained for DMN are mostly PCC, medial pre-frontal cortex (MFC), and right inferior parietal lobe (IPL), where as for rest of the proposed algorithms. As there is no gold standard reference for DMN connectivity available, therefore, we relied on the similarity of temporal dynamics of DMN based modulation profile with PCC representative time-series. The similarity measure used was correlation and estimated as > 0.9 for all the algorithms.

6. CONCLUSION

In this paper, we have developed a new variant of CCA; more specifically we have introduced new algorithm for sparse CCA. The proposed algorithm is based on penalized rank one approximation and differs from existing ones in the matrices they use for their derivation. Indeed instead of focusing on the cross-matrix product of the two sets of multidimensional variables we have used the product of the orthogonal projectors onto the space spanned by the two sets of multidimensional variables. Using this approach the sparse CCA algorithm proposed differs only the penalty used in the penalized rank one matrix approximation. Simulation results illustrating the effectiveness of the proposed CCA variant algorithm are provided where we can observe that proposed sparse CCA outperforms state of the art methods.

7. REFERENCES

[1] H. Hotelling, "Relations between two sets of variables," *Biometrika*, vol. 28, no. 3-4, pp. 321–377, 1936.

[2] W. Zheng, X. Zhou, C. Zou, and L. Zhao, "Facial expression recognition using kernel canonical correlation analysis (KCCA)," *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 233–238, January 2006.

[3] O. Friman, J. Carlsson, P. Lundberg, M. Borga, and H. Knutsson, "Detection of neural activity in functional MRI using canonical correlation analysis," *Magnetic Resonance in Medicine*, vol. 45, no. 2, pp. 323–330, February 2001.

[4] D. R. Hardoon, J. Mourao-Miranda, M. Brammer, and J. Shawe-Taylor, "Unsupervised analysis of fmri data using kernel canonical correlation," *NeuroImage*, vol. 37, no. 4, pp. 1250 – 1259, 2007.

[5] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, December 2004.

[6] L. Sun, S. Ji, and J. Ye, "Canonical correlation analysis for multilabel classification: A least-squares formulation, extensions, and analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 194–200, January 2011.

[7] Wei Liu, D.P. Mandic, and A. Cichocki, "Analysis and online realization of the cca approach for blind source separation," *IEEE Transactions on Neural Networks*, vol. 18, no. 5, pp. 1505–1510, September 2007.

[8] Yi-Ou Li, T. Adali, Wei Wang, and V.D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, October 2009.

[9] D. R. Hardoon and J. Shawe-Taylor, "Sparse canonical correlation analysis," *Machine Learning*, vol. 83, no. 3, pp. 331–353, June 2011.

[10] D. M. Witten, R. Tibshirani, and T. Hastie, "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis," *Biostatistics*, vol. 10, no. 3, pp. 515–534, July 2009.

[11] D. Chu, L.Z. Liao, M.K. Ng, and X. Zhang, "Sparse canonical correlation analysis: New formulation and algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 3050–3065, December 2013.

[12] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, Academic Press, 1979.

[13] S. Gross, B. Narasimhan, R. Tibshirani, and D. Witten, "Correlate: Sparse canonical correlation analysis for the integrative analysis of genomic data," Tech. Rep. User guide and technical document, Stanford University, 2011.

[14] J.F. Cai, S. Osher, and Z. Shen, "Convergence of the linearized bregman iteration for ℓ_1-norm minimization," Tech. Rep. CAM Report 08-52, University of California Los Angeles, 2008.

[15] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Asilomar Conference on Signals, Systems and Computers*, November 1993, vol. 1, pp. 40–44.

[16] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approximation*, vol. 13, no. 1, pp. 57–98, 1997.

[17] J.A. Branco, C. Croux, P. Filzmoser, and M.R. Oliveira, "Robust canonical correlations: A comparative study," *Computational Statistics*, vol. 20, no. 2, pp. 203–229, 2005.

[18] N. A. Lazar, *The Statistical Analysis of Functional MRI Data*, Springer, 2008.

[19] N. M. Correa, T. Adali andY. Li, and V. D. Calhoun, "Canonical correlation analysis for data fusion and group inference," *Signal Processing Magazine*, vol. 20, pp. 39–50, 2010.

[20] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the mni mri single-subject brain," *NeuroImage*, vol. 15, pp. 273–289, 2002.