A projection CCA method for effective fMRI data analysis

Muhammad Ali Qadar, Abd-Krim Seghouane, Senior Member, IEEE

Abstract—Objective: Canonical correlation analysis (CCA) is a data driven method that has been successfully used in functional magnetic resonance imaging (fMRI) data analysis. Standard CCA extracts meaningful information from a pair of data sets by seeking pairs of linear combinations from two sets of variables with maximum pairwise correlation. So far, however, this method has been used without incorporating prior information available for fMRI data. In this paper, we address this issue by proposing a new CCA method named pCCA (for projection CCA). Methods: The proposed method is obtained by projection onto a set of basis vectors that better characterize temporal information in the fMRI data set. A methodology is presented to describe the basis selection process using discrete cosine transform (DCT) basis functions. Employing DCT guides the estimated canonical variates, yielding a more computationally efficient CCA procedure. Results: The performance gain of the proposed pCCA algorithm over standard and regularized CCA is illustrated on both simulated and real fMRI datasets from resting state, block paradigm task-related and event-related experiments. The results show that pCCA can successfully extract the task as well resting-state latent components with increased specificity. Conclusion: In addition to offering a new CCA approach, when applied on fMRI data, the proposed algorithm adapts to variations of brain activity patterns and reveals the functionally connected brain regions. Significance: The proposed method can be seen as a regularized CCA method where regularization is introduced via basis expansion, which has the advantage of enforcing smoothness on canonical components.

Index Terms—Canonical correlation analysis, functional magnetic resonance imaging, basis expansion, regularization.

I. INTRODUCTION

There are two broad classes of methods used to analyse functional magnetic resonance imaging (fMRI) data. The first class of methods involve model based analysis, such as the general linear model (GLM) [1], that uses a priori information about the stimulus function to show the goodness-of-fit of the model and form inferences to study the brain function. Despite the wide acceptance and extensive usage of model-based methods in medical image analysis, their applicability is limited due to relying on a priori knowledge, which is unavailable in cases such as resting-state data or while studying naturalistic behaviour such as continuous listening or watching a movie [2].

The second class of methods are based on data-driven analysis, which has been successfully used for detection of functionally connected voxels [3]. These methods are suitable for fMRI data analysis as they explore the structure of data by identifying interesting components using minimal assumptions on the underlying characteristics of the data. Data-driven methods are applicable to both task-related and resting-state fMRI datasets. Under specific objectives, these methods decompose the observation matrix using a factor model [2]. Different objectives produce different data-driven techniques. Maximizing the variance leads to principal component analysis (PCA) [4], the independence of components leads to the spatial and temporal independent component analysis (sICA and tICA) [5], the sparsity of components leads to dictionary learning [6].

Maximum correlation constraints, which leads to CCA, has also been successfully employed to analyse fMRI data [7]–[17]. Generally, CCA analyses fMRI data from a temporal or spatial perspective. Temporal CCA (tCCA) aims to recover the time courses of interest by maximizing the correlation between two fMRI datasets X and Y [3]. The temporal decomposition of fMRI data facilitates the analysis of hemodynamic signals, which are critical to study brain function and in exploring the temporal properties of brain activity [18]. Whereas, spatial CCA (sCCA) seeks spatial activity patterns to determine the basis images with maximum spatial autocorrelation [3]. The present study focuses on temporal CCA analysis of fMRI data.

CCA is commonly used to find canonical eigenvectors corresponding to maximum spatially or temporally autocorrelated signals within a data set [19]. The performance of CCA can be significantly improved when it is used in conjunction with other data driven methods or constraints. In [20], [21], the specificity and sensitivity of dictionary learning methods were improved by including autocorrelation constraints in fMRI data. In [22], a sparse CCA method, derived via penalized rank one matrix approximation, was applied to resting-state fMRI data analysis. However, the presence of spatial structure and connectedness of the local neighbourhood voxels in fMRI data has not been exploited in sparsity based CCA methods. In [23], a sparse CCA approach is applied to the fusion of fMRI and structural MRI data. It is shown to preserve spatial structure by imposing non-negativity constraints. Recently in [24], CCA was used alongside ICA in order to preserve low frequency fMRI signals. In this study, the objective of projection CCA (pCCA) is also to exploit low frequency temporal information.

fMRI datasets exhibit temporal structure that incorporate physiological fluctuations of the human brain. These fluctua-
tions, which generally are referred to as low frequency fluctuations (LFFs) arise due to blood oxygenation or blood flow [25], [26]. To observe low frequency signals, a prolonged change of brain state with higher oxygenation is required. It has been reported that low frequency information of less than 0.08 Hz exhibits high functional connectivity in motor areas of the brain, which has been proved to be an effective criterion to study brain disorders (e.g. Alzheimer disease) in resting-state fMRI [27]. Furthermore, the amplitude of low frequency components are used to reflect spontaneous regional neural activity and to overcome the problem of physiological noise in fMRI data [28]. Generally, these frequencies are extracted by preprocessing fMRI data using high pass filtering [28]. However, filtering original fMRI signals does not guarantee the removal of noise embedded in low frequency signals, therefore, data-driven methods are used to remove noise signals [24].

In this study, we propose to exploit the presence of temporal structure mentioned above when using CCA. This structure, which can be converted to a guided constraint of smoothness on the canonical variates, has been overlooked in classical CCA methods for fMRI data analysis. With fMRI datasets, regularizing the canonical variates to encourage smoothness of the dataset in the appropriate direction may therefore be of interest. Regularization can be obtained by adding a penalty to the CCA optimization problem, however, this approach ignores the low frequency temporal structure of fMRI data. Instead, we propose a CCA method that accounts for such a priori information about temporal structure by enforcing an appropriate smoothness on the canonical variates. Temporal structure can be imposed by using a set of basis functions to project the fMRI data onto a lower dimensional subspace [29], [30]. The projection subspace will have the advantage of incorporating a priori information available from the fMRI data. The proposed CCA method can be seen as a regularized CCA where regularization is introduced via basis expansion. The regularization is obtained by shrinking the canonical variates to a certain subspace defined by the basis functions, which are selected based on the a priori low frequency temporal information extracted from the fMRI dataset.

fMRI datasets are structured with notions of temporal smoothness and low frequencies. Classical CCA algorithms that overlook these structures can result in lower performance. Based on the motivations and description above, the specific contributions of this study is to propose a CCA learning method for regularized canonical variates through basis expansion and present the associated computationally efficient algorithm. We also present a paradigm to select the appropriate subspace using discrete cosine transform (DCT). Note that the spatial correlation is not taken into account in the proposed method as in [8], [31]–[33], the time series in each voxel are assumed independent.

As a comparison benchmark, the performance of our proposed method is presented alongside standard and ridge regularized (rCCA) [34] for simulated and real fMRI data. The rest of the paper is organized in the following manner: In section II, we provide a brief overview of CCA. In section III, we present an alternative CCA formulation by describing the computation and utilization of prior information in the CCA objective function. In section IV, we describe a method for selecting the basis vectors from fMRI data. In section V, we evaluate the proposed CCA method on simulated and experimental fMRI datasets. In section VI, we describe a statistical test to analyse the significance of activation detection. Finally, some concluding points are summarized in section VII.

II. BACKGROUND

This section briefly describes CCA in the form that is commonly used in the context of fMRI data analysis.

A. Canonical correlation analysis

Canonical correlation analysis is a standard technique for studying the relationship between two sets of multivariate variables. In fMRI data analysis, CCA is used to unmix the structured components from the underlying mixture by autocorrelation maximization [8]. CCA requires two datasets. Let \( \{x_p \in \mathbb{R}^N, p = 1, \ldots, P\} \) be a collection of \( P \) fMRI voxel time series of length \( N \), resulting in the original data matrix \( X \) of size \( N \times P \). Construct a second data matrix \( Y \in \mathbb{R}^{(N−1)\times P} \) from \( X \) by delaying it by one step along time [3]. Assuming both \( X \) and \( Y \) are centred and scaled, CCA finds two directions of projection, \( w_x^1 \in \mathbb{R}^N \) and \( w_y^1 \in \mathbb{R}^{N−1} \) by maximizing the correlation between their covariates such that

\[
\arg\max_{w_x, w_y} w_x^T C_{xx} w_x \ s.t. \ w_x^T C_{xx} w_x = w_y^T C_{yy} w_y = 1, \ w_x^T w_y = 0
\]

where \( C_{xx} = 1/P \sum_{p=1}^P x_p x_p^T \), \( C_{yy} = 1/P \sum_{p=1}^P y_p y_p^T \), \( C_{xy} = 1/P \sum_{p=1}^P x_p y_p^T \), and \( C_{yx} = C_{xy}^T \). The matrices \( C_{xx} \) or \( C_{yy} \) and \( C_{xy} \) are the covariance and cross-covariance matrices, respectively. The vectors \( w_x^1 \) and \( w_y^1 \) correspond to the first pair of the canonical vectors where \( (w_x^1, w_y^1) \) are the first pair of canonical variates and \( r_1 = w_x^1^T C_{xx} w_x^1 \) is first canonical correlation. There are \( r \leq \min(\text{rank}(X, Y)) \) number of pairs that can be found in uncorrelated directions, with \( r \) such canonical vectors \( W_x \in \mathbb{R}^{N \times r} \) and \( W_y \in \mathbb{R}^{(N−1) \times r} \) are obtained by employing a deflationary approach with following constraints

\[
W_x^T C_{xx} W_x = W_y^T C_{yy} W_y = I_r, \quad W_x^T C_{xy} W_y = \text{diag}(\rho_1, \ldots, \rho_r)
\]

In order to obtain closed form solutions for \( W_x \) and \( W_y \), constrained optimization problem (1) is solved using Lagrange multipliers for first \( r \) eigenvectors of \( C_{xx}^{-1} C_{xy} C_{yy}^{-1} \) and \( C_{yy}^{-1} C_{yx} C_{xx}^{-1} \), respectively. Standard CCA suffers from the singularity of \( C_{xx} \) and \( C_{yy} \), therefore, it is necessary to add a scaled identity matrix to apply regularization [34], [35]. The regularized correlation matrices are expressed as

\[
\tilde{C}_{xx} = C_{xx} + \gamma_x I_N \quad \text{and} \quad \tilde{C}_{yy} = C_{yy} + \gamma_y I_{(N−1)}
\]
where $\gamma_x > 0$ and $\gamma_y > 0$. Then, rCCA objective can be obtained by updating (1) i.e.
\[
\arg \max_{w_x^1, w_y^1} w_x^1^T C_{xx} w_x^1 \text{ s.t. } w_x^1^T C_{xz} w_z^1 = w_y^1^T C_{yx} w_y^1 = 1
\]  
(4)
The rCCA method is accompanied with important disadvantages, it is computationally expensive, requires regularization parameter selection and leads to inaccurate solutions by filtering out important information [36]. To ease these drawbacks and perform effective analysis of fMRI data using CCA, we regularize CCA by incorporating prior information about the low frequency fMRI signals by building a basis that better spans the subspace of fMRI signals. The proposed projection-based regularization enforces smoothness on canonical component, which makes the decomposition of data more efficient and increase the specificity of activated voxels.

III. PROPOSED METHOD

A. The projection CCA

The objective function defined in (1) is invariant to scaling of the canonical loadings $w_x$ and $w_y$, therefore, (1) can be reformulated as
\[
u = C_{zy}^1 w_y, \quad w_x^T C_{xx} w_x = u^T u = 1
\] and
\[
u = C_{xy}^1 w_y, \quad w_y^T C_{yy} w_y = v^T v = 1
\] Then, CCA optimization (1) becomes
\[
\max_{u,v} u^T C_{xx}^1 C_{xy}^1 v \text{ subject to } u^T u = v^T v = 1
\]
The objective is to find a linear representation of the matrix $C = C_{xx}^1 C_{xy} C_{yz}^1$ by minimizing the following objective function
\[
\mathcal{L} = \|T - suv^T\|_F^2 = \|T - s B_N \theta \psi^T B_{(N-1)}^T\|_F^2
\]  
(5)
where $u = B_N \theta$ and $v = B_{(N-1)} \psi$. $\theta$, $\psi$, $\psi$ are vectors of size $K \times 1$ to be estimated, $u$ and $v$ are left and right singular vectors of size $N \times 1$ and $(N-1) \times 1$, respectively, $s$ corresponds to singular values and $\| \cdot \|_F$ is the Frobenius norm. The subspace matrix $B_N$ of size $N \times K$ contains $K$ basis functions (for $K \ll N$), whereas, $B_{(N-1)}$ of size $(N-1) \times K$ is created by delaying $B_N$ by one step along time. Due to the presence of large number of samples, delaying $B_N$ by one row does not affect the CCA estimation procedure largely. Therefore, the subspace matrix for both $u$ and $v$ in the CCA objective can be obtained from $B_N$. The selection of basis functions of $B_N$ will be discussed in section IV. For simplicity of notation, $B_N$ will be denoted as $B$.

Expanding the objective function given in (5) we obtain
\[
\mathcal{L} = \text{tr}[T^T T] - 2s \psi^T B^T B \theta \psi + s^2 \theta^T B^T B \theta (\psi^T B^T B \psi)
\]  
(6)
For fixed $s$ and $\theta$, the minimization over $\psi$ is obtained by setting the gradient of (6) to zero (i.e., $\frac{\partial \mathcal{L}}{\partial \psi} = 0$) resulting in the estimate of $\psi$
\[
\psi = (s^2 \theta^T B^T B \theta \psi B^T B)^{-1} s B^T B \theta
\]  
(7)
under the constraint $\theta^T B^T B \theta = 1$, the estimate of $\psi$ becomes
\[
\psi = s^{-1} (B^T B)^{-1} B^T T B \theta
\]  
(8)
Substituting the estimate of $\psi$ in (6) gives
\[
\mathcal{L} = \text{tr}[T^T T] - 2s \psi^T B^T B \theta (\psi^T B^T B)^{-1} B^T T B \theta + s^2 \theta^T B^T B \theta (\psi^T B^T B)^{-1} (B^T B)^{-1} B^T T B \theta
\]  
(9)
This leads to optimization problem
\[
\max_{\theta}(\theta^T B^T T B (B^T B)^{-1} B^T T B \theta)
\]  
(10)
s.t. $\theta^T B^T B \theta = 1$

For the sake of computational efficiency, the matrix $B^T B$ is decomposed using Cholesky decomposition [37]. The Cholesky decomposed matrix is denoted by $R_B$, where $B^T B = R_B^T R_B$. Defining $\tilde{\theta} = R_B \theta$ implies $\theta = R_B^{-1} \tilde{\theta}$. Substituting $\theta$ in (10), leads to the maximization of the following optimization problem
\[
\max_{\tilde{\theta}}(\tilde{\theta}^T (R_B^{-1})^T B^T T B (B^T B)^{-1} B^T T B \theta)
\]  
(11)
s.t. $\tilde{\theta}^T \tilde{\theta} = 1$

where $\tilde{\theta}$ is the right singular vector of the matrix $(R_B^{-1})^T B^T T B R_B^{-1}$. Similarly, to estimate $\psi$ a recursion of procedure above is followed, where the objective function in (6) is minimized over $\psi$ (i.e., $\frac{\partial \mathcal{L}}{\partial \psi} = 0$) for fixed $s$, $\psi$ under the constraint $\psi^T B^T B \psi = 1$, then the estimate of $\theta$ is expressed as
\[
\hat{\theta} = s^{-1} \psi^T B^T T B \theta
\]  
(12)
Rearranging the terms in (12) by transposition to obtain the estimate of $\theta$:
\[
\hat{\theta} = s^{-1} (B^T B)^{-1} B^T T B \psi
\]  
(13)
Substituting the estimate of $\theta$ into (6) yields
\[
\mathcal{L} = \text{tr}[T^T T] - \psi^T B^T T B (B^T B)^{-1} B^T T B \psi
\]  
(14)
This leads to the maximization of following optimization problem
\[
\max_{\psi}(\psi^T B^T T B (B^T B)^{-1} B^T T B \psi)
\]  
(15)
s.t. $\psi^T B^T B \psi = 1$

Using the Cholesky decomposition of $B^T B$ and setting $\psi = R_B^{-1} \hat{\psi}$ leads to
\[
\max_{\psi}(\psi^T (R_B^{-1})^T B^T T B R_B^{-1} (R_B^{-1})^T B^T T B \psi)
\]  
(16)
s.t. $\psi^T \psi = 1$

where $\hat{\psi}$ is the right singular vector obtained from singular value decomposition (SVD) of the matrix $(R_B^{-1})^T B^T T B R_B^{-1}$.

The matrix $(R_B^{-1})^T B^T T B R_B^{-1}$ is a transposed form of matrix obtained from optimization problem in (11). Using
a single SVD of \((R_{B}^{-1})^\top B^\top TBR_{B}^{-1}\) yields \(\theta\) as the right and \(\psi\) as the left singular vector. The spatial maps were obtained by projecting estimated canonical vectors onto original datasets (i.e., \(u_i^\top X, v_i^\top Y\), which are of dimension \(P\)). The resulting optimizations in (11) and (16) solve for \(K\) dimensional vectors \((\theta_i, \psi_i)\) instead of standard CCA, which solves for \(N\) and \(N - 1\) dimensional vectors \((u, v)\), respectively, where \(K \ll N\). The projection obtained using \(B\) incorporates a priori information in CCA objective function, which is the advantage of our proposed method. The next section illustrates a method through which the low frequency structure of fMRI data is employed in \(B\). The stepwise procedure of the proposed pCCA algorithm is presented in Algorithm 1.

**Algorithm 1:** Stepwise procedure of the proposed pCCA for estimating canonical vectors

<table>
<thead>
<tr>
<th>Data:</th>
<th>Training data (X \in \mathbb{R}^{N \times P}), (Y \in \mathbb{R}^{(N-1) \times P}).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Result:</td>
<td>(\Theta) and (\Psi) containing (r) pairs of right singular vectors (\theta_i \in \mathbb{R}^{K \times 1}) and (\psi_i \in \mathbb{R}^{K \times 1}).</td>
</tr>
<tr>
<td>1</td>
<td>(\mu_x = \frac{1}{P} \sum_{p=1}^{P} x_p) and (\mu_y = \frac{1}{P} \sum_{p=1}^{P} y_p)</td>
</tr>
<tr>
<td>2</td>
<td>for (p \in [1, P]) do</td>
</tr>
<tr>
<td>3</td>
<td>(x_p \leftarrow x_p - \mu_x)</td>
</tr>
<tr>
<td>4</td>
<td>(y_p \leftarrow y_p - \mu_y)</td>
</tr>
<tr>
<td>5</td>
<td>end</td>
</tr>
<tr>
<td>6</td>
<td>Compute (C_{xx}, C_{xy}, C_{yx}, C_{yy}) and (T = C_{xx}^{-1/2}C_{xy}C_{yx}^{-1/2}).</td>
</tr>
<tr>
<td>7</td>
<td>Perform Cholesky decomposition of (B^\top B = R_B^\top R_B);</td>
</tr>
<tr>
<td>8</td>
<td>Decompose ((R_{B}^{-1})^\top B^\top TBR_{B}^{-1}) using SVD (\Theta = V(:, 1 : r));</td>
</tr>
<tr>
<td>9</td>
<td>Estimate (\Theta = R_B^{-1} \Theta)</td>
</tr>
<tr>
<td>10</td>
<td>(\Psi = U(:, 1 : r))</td>
</tr>
<tr>
<td>11</td>
<td>Estimate (\Psi = R_B^{-1} \Psi)</td>
</tr>
</tbody>
</table>

IV. SELECTION OF BASES

There are many possible choices to construct the columns of the basis matrix (e.g., Fourier, wavelet, spline, etc.). In this study, we present a methodology to derive the columns of matrix \(B\) from DCT basis functions [38] as they can be used to enforce a predefined temporal structure for the subspace on which the data are projected. To model low frequency information in fMRI data, DCT provides a continuous periodic signal structure, which has the benefit of compactness [39]. DCT basis functions \(b_k(n)\) are defined for \(0 \leq n < (N - 1)\) time points as:

\[
b_k(n) = \lambda_k \sqrt{\frac{2}{N}} \left( \cos \left( \frac{k \pi}{2N} \right), \cos \left( \frac{3k \pi}{2N} \right), \cdots, \cos \left( \frac{k \pi (2n + 1)}{2N} \right) \right)^\top \quad (17)
\]

with \(\lambda_k = \begin{cases} 2^{-1/2} & \text{if } k = 0 \\ 1 & \text{otherwise} \end{cases}\)

where \(k = 1, \cdots, N - 1\) corresponds to frequency components. In what follows, we will describe a method through which \(K\)-DCT basis functions \((K \ll N)\) are selected to form the columns of \(B\). The \(p\)th voxel, \(x_p \in \mathbb{R}^N\), corresponding to a region of interest (ROI) in the brain, is projected onto the DCT subspace, resulting in coefficients:

\[
d_p^x(k) = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x_p(n) \lambda_k \cos \left( \frac{(2n+1)k}{2N} \right) \quad (18)
\]

where \(d_p^x(k)\) is the \(k\)th DCT coefficient. DCT coefficients are calculated for several voxels within a neighborhood of the ROI. The average spectrum obtained from the DCT coefficients in the ROI is used to determine high energy frequency components \(K\). The desired temporal structure is enforced onto the subspace by selecting those DCT basis functions, \(b_k\), from (17), which represent 90% of the fMRI time course energies. The matrix \(B\) consisting of \(K\)-DCT basis vectors selected using indices of coefficients obtained using (18) is embedded in the CCA objective function for estimating canonical vectors. The algorithmic procedure is presented in Algorithm 2 for computing priori information.

**Algorithm 2:** Stepwise procedure for computing priori information

<table>
<thead>
<tr>
<th>Data:</th>
<th>Training data (x_p \in \mathbb{R}^N), (T = \text{compactness threshold.})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Result:</td>
<td>Matrix (B_K) of size (N \times K).</td>
</tr>
<tr>
<td>1</td>
<td>Select ROI matrix (X) of size (N \times P) containing (P) averaged voxel time series ((\text{i.e., } X = (x_1, \cdots, x_P)^\top));</td>
</tr>
<tr>
<td>2</td>
<td>Begin with a DCT basis matrix (B) of size (N \times N);</td>
</tr>
<tr>
<td>3</td>
<td>for (p = 1) to (P) do</td>
</tr>
<tr>
<td>4</td>
<td>Calculate (d_p^x(k)) using (18);</td>
</tr>
<tr>
<td>5</td>
<td>Sort elements of (d_p^x(k)) in descending order;</td>
</tr>
<tr>
<td>6</td>
<td>Set (K = 1);</td>
</tr>
<tr>
<td>7</td>
<td>while (</td>
</tr>
<tr>
<td>8</td>
<td>(K = K + 1);</td>
</tr>
<tr>
<td>9</td>
<td>end</td>
</tr>
<tr>
<td>10</td>
<td>Set (d_p^x(K + 1 : end) = 0);</td>
</tr>
<tr>
<td>11</td>
<td>Find the indices of nonzero coefficients (d_p^x) and set it to (\text{ind});</td>
</tr>
<tr>
<td>12</td>
<td>for (k = 1) to (\text{length(ind)}) do</td>
</tr>
<tr>
<td>13</td>
<td>(B_K.\text{append}(B(:, \text{ind}(k))))</td>
</tr>
<tr>
<td>14</td>
<td>end</td>
</tr>
<tr>
<td>15</td>
<td>end</td>
</tr>
</tbody>
</table>

V. EXPERIMENTAL RESULTS

In this section, we evaluate the proposed method using four different experiments on: simulated data, real fMRI data obtained from right finger tapping (RFT) using a block paradigm, event-related right finger tapping task, and resting state fMRI (the rsfMRI dataset was obtained from Human Connectome Project repository).
As a comparison benchmark, rCCA was also adopted as described in [34] and included in our results. To select an optimal value of regularization parameters, we carried out 10-folds cross-validation on the data by splitting the data spatially into training and test sets [40]. A value of $\gamma_x = \gamma_y = 13$ was selected to estimate the canonical vectors of rCCA over values of $\gamma_x$ and $\gamma_y$ ranging between 0 and 20 sampled at 0.5. We then carried out standard CCA, rCCA, and the proposed pCCA (Algorithm 1) on the dataset to recover the underlying sources. To demonstrate the efficiency of the proposed pCCA, we computed the average correlations of recovered sources against the original sources, presented in Table I, where $\{R_1, \cdots, R_k\}$ are recovered sources. Similarly, Table II contains mean correlations of recovered spatial maps with original ones, where $\{S_1, \cdots, S_k\}$ denotes recovered spatial maps. It is evident that pCCA exhibits higher correlations corresponding to recovered sources as compared to standard CCA and rCCA at low $SNR = 10\, dB$. This is mainly due to the fact that standard CCA and rCCA model ignore temporal smoothness, whereas, the proposed pCCA incorporates low frequency information through basis expansion, resulting in regularized canonical variates. Fig. 1(b), (c), and (d) show the recovered spatial maps and source signals from standard CCA, rCCA, and the proposed pCCA, respectively. The sources extracted using the proposed algorithm are smooth and signal peaks are localized, whereas, the recovered sources by standard CCA and rCCA are noisy with deformed peaks. The spatial maps recovered by the proposed pCCA accurately classify the activated voxels in each region and shows higher functional connectivity with other regions as compared to standard CCA and rCCA. In fact, the aim here is not merely blind source separation, but to estimate connected activation patterns between highly correlated distinct regions.
To test the significance of detected voxels, we created receiver operating characteristic curves (ROC) from simulated data (response data + noise versus noise only) using Bartlett’s test statistic defined in section VI as activation scores [19]. The scores \( L \) for each voxel were used to calculate the ROC curves shown in Fig. 2. It is observed that for low SNR pCCA significantly outperforms standard CCA and rCCA. The detection probability of pCCA is always higher, however, at higher values of SNR pCCA ROC curves starts to closely track the ROC curve corresponding to standard CCA and rCCA. The detection capability of pCCA is higher as compared to standard CCA and rCCA even at \( SNR = 15 \) dB as shown in Fig. 2.

Furthermore, we examined the estimated eigenvalues, which are not reported here. The trend of the eigenvalues corresponding to pCCA turns out to be consistent with larger values as compared to standard CCA and rCCA. This reveals that the proposed pCCA is efficient in extracting low frequency components and is able to provide a smoother representation of underlying source signals.

After validating the proposed method on simulated data, we establish 3 more examples on real fMRI datasets to synthesize the validity of proposed method for real world problems.

![Fig. 2. Four sets of receiver operating characteristic (ROC) curves obtained using both standard CCA, rCCA[34], and pCCA (proposed) methods at signal-to-noise ratio (SNR) values \( \in \{0, 5, 10, 15\} \).](image)

**B. Block-paradigm right finger tapping task fMRI data**

The fMRI dataset used for analysis consists of single subject data acquired while performing a finger-tapping task over auditory instructions. The block-paradigm consists of alternating 30s ON(fingertapping) and 30s OFF(rest) periods. The data were acquired in real and imaginary parts through a standard 3T Siemens TIM TRIO system with 12-channel radio frequency (RF) coils in a gradient-echo EPI (Echo Planar Imaging) sequence experiment. The data collection was initialized with a 12s rest period to stabilize T1 effects.

Data acquisition parameters include: repetition time (TR = 2s), echo time (TE = 2ms), flip angle (70 degree), slice thickness = 3.5mm, matrix size (64 \times 64), number of slices = 32, field-of-view (FOV = 24cm) and slice gap = 1mm. Magnitude and phase images were written out as four dimensional (4-D) nifti files using statistical parameter mapping (SPM12) [41]. The data was then processed through various preprocessing stages as described in [42], [43]. The movements in fMRI time series data were compensated by co-registering the magnitude data. Co-registration was followed by spatial normalization where images were normalized into standard MNI (Montreal Neurological Institute) space. The data was sub-sampled to 53 \times 63 \times 46 voxels. We applied spatial smoothing to both real and imaginary parts of the data with a 10 \times 10 \times 10 \text{mm}^3 full-width at half-maximum (FWHM) Gaussian kernel. In order to get rid of low frequency drifts, we used DCT basis with a cut-off frequency of \( (1/128) \text{Hz} \). Note that, CCA estimation was performed on magnitude-only data. The magnitude data was then masked to remove extraneous brain regions outside of the scalp resulting in the reduced dataset by a factor of 3. This operation is useful to reduce the computational time.

To prepare for CCA, the magnitude-only data was parcellated using a spatially constrained functional clustering technique [44]. The clustering performed here is a two stage procedure, in first stage, a gray matter mask is used to construct connectivity graphs by treating each voxel as a node while edges correspond to temporal correlation between voxels. In the second stage, normalized cut spectral clustering is used to segment the whole brain into 129 regions. The purpose of parcellation is to obtain averaged voxel time series data corresponding to each brain region. Voxels from each region were then averaged to obtain a matrix \( \mathbf{X} \) of size \( N \times Q \) where \( N = 165 \), corresponds to the number of timepoints and \( Q = 129 \) corresponds to seed regions.

Using the seed regions of interest, the prior information matrix \( \mathbf{B} \) (as defined in section IV) is calculated by averaging the DCT coefficients in the seed regions of interest (algorithm 2). We present the extracted temporal components and spatial maps exhibiting motor component in the right motor activation region in Fig. 3 using standard CCA, rCCA, and the proposed pCCA (algorithm 1), respectively. For rCCA, the regularization parameters were set to a value of 2 by a 10-folds cross-validation over values of \( \gamma_x \) and \( \gamma_y \) ranging between 0 and 5 sampled at 0.5. The estimated time courses and spatial maps were thresholded at a p-value of less than \( 1 \times 10^{-3} \) to show significantly activated voxels in the motor area, consistent with the locations of human cortical regions responsible for task-related processing [45]. It can be seen that the temporal correlations between recovered time courses and the task-paradigm is higher in case of the proposed pCCA with a value of 0.8564, whereas correlation value for standard CCA and rCCA are 0.8017 and 0.8452, respectively, with all p-values less than \( 1 \times 10^{-3} \). The recovered time courses using the proposed pCCA closely resembles the task-paradigm, however, standard CCA and rCCA exhibits higher fluctuations in peaks of the recovered time courses. The spatial maps of brain activations in motor area estima-
C. Event-related right finger tapping task fMRI data

We evaluate the proposed CCA method using event-related right finger-tapping task data. A 3.0 T fMRI system was used to acquire the BOLD/EPI images. The system acquired 30 contiguous slices of matrix size 64 × 64 with voxel size of 3.4mm × 3.44mm × 4mm, where slice thickness was 4mm and flip angle set to 80°. The data were recorded for total 650s with TR/TE = 2000/35ms. After discarding 30s of dummy scans, the task and the rest periods consisted of 14s window that was repeated 40 times, followed by an additional 30s rest period. For each period of resting-state after task, the interstimulus interval (ISI) ranged between 40s and 30s. The system acquired 143 00307 fMRI images, the data were then processed through various preprocessing stages, 1) spatial alignment, 2) normalization, 3) spatial smoothing, 4) masking, and 5) temporal smoothing. The fMRI images were realigned to correct for head motion, which affects the signal intensity during the course of an fMRI experiment. In the subsequent steps, all fMRI images were spatially normalized to a standard Talairach template, the voxels resampled to the size 2mm × 2mm × 2mm followed by spatial smoothing with a 5mm × 5mm × 5mm full-width at half-maximum (FWHM) Gaussian kernel. The BOLD time series dataset was collected in 4-dimensional images, the data were then masked to remove any data outside the scalp, where voxels with more than masking threshold were retained. Masking improves the computation complexity by reducing the data with a factor of 8 along the spatial direction. Then, the data were organized in a 2-D matrix X. We then detrended the dataset by removing the low frequency drifts using DCT basis set at a cut-off frequency of (1/128)Hz. The detrended data were smoothed with a 1.5s FWHM kernel to eliminate high frequency noise.

After preprocessing, the dataset were parcellated into 192 voxel regions using similar technique as described in section V-B. The parcellation yielded data matrix X of size $N \times Q$ where $N = 310$, each voxel region averaged to obtain single time series of Q regions. The prior information matrix

B was computed using Algorithm 2. We then carried out standard CCA, rCCA, and the proposed pCCA (Algorithm 1) on the data. For rCCA, the regularization parameters were set to a value of 2 by performing a 10-folds cross validation procedure for values of $\gamma_x$ and $\gamma_y$ ranging between 0 and 5 sampled at 0.5. The estimated spatial maps were thresholded at a p-value of less than $1e^{-3}$.

The spatial maps and temporal components extracted using standard CCA, rCCA, and the proposed pCCA are shown in Fig. 4. The spatial maps obtained using the proposed pCCA are evidently better in terms of the specificity and show activations only in the motor area. The spatial maps estimated using pCCA exhibits higher voxel connectivity and activations are tightly localized in the motor area. Furthermore, the proposed pCCA successfully avoids incorporating extraneous voxels that do not contribute to the brain motor region, whereas standard CCA and rCCA has introduced spurious activations that affect the spatial specificity of activation detection. The temporal correlations between estimated time courses and task-paradigm were 0.5328 for standard CCA, 0.6000 for rCCA and 0.7058 for the proposed pCCA, respectively. Higher temporal correlation reveals the ability of the proposed pCCA in modeling signals exhibiting high fluctuations by regularizing the canonical components via basis expansion.

D. Resting-state fMRI data

Single subject (id 100307) rsfMRI dataset obtained from the Human Connectome Project Q1 release [46]. The acquisition parameters used were TR= 0.72s, TE = 33.1ms, BW = 2290 Hz/Px, matrix size = 90 × 104, slices = 72, field of view (FOV) = 220 mm , flip angle = 52°, in-plane FOV = 208 × 180mm isotropic voxels. The data were preprocessed in a stepwise preprocessing pipeline by performing motion correction, temporal whitening, slice-timing correction, removal of drift followed by spatial normalization to a standard MNI152 template and resampling to $2 \times 2 \times 2$ mm$^3$ voxel [46], [47]. A total 1200 of scans were acquired in the duration of 14 : 33 (min:sec). First 200 scans (143.3s) were selected for the analysis. The fMRI images were smoothed spatially...
Fig. 5. Spatial maps corresponding to DMN regions extracted using a) standard CCA, b) rCCA, and c) pCCA (proposed), with all three methods the p-value was less than $1e^{-3}$ ($P < 1e^{-3}$).

with $6 \times 6 \times 6$ mm$^3$ FWHM Gaussian kernel. The data were then vectorized and each image placed as rows of matrix $X$ of size $N \times Q$, where $N = 200$, which corresponds to the number of time points. Then, the data were detrended by removing the low frequency drifts using DCT basis set at a cut-off frequency of $1/128$ Hz, followed by temporal smoothing to remove high frequency noise using a 1.5s FWHM Gaussian kernel. To reduce computation time, the data were downsampled spatially by a factor of 4.

We used seed-voxel based correlation analysis for rsfMRI data to extract most correlated time courses corresponding to specific brain regions. The dataset was analyzed based on the assumption that at the resting-state, low frequency signals are correlated in functionally connected areas of the brain. In seed-voxel based analysis, the seed-voxel, which is selected from functionally connected networks (FCN) [48] has to belong to set of correlated voxels from any of FCN (e.g. default mode network (DMN), salience network (SN), dorsal network (DN) etc.). In this study, we focus on using seed-voxel analysis for DMN. To extract voxel regions corresponding to DMN, we used a toolbox available for rsfMRI data processing [49]. The standard MNI coordinates [50] of seed-voxels are provided in Table III. We computed the average voxel time series corresponding to DMN (i.e., precuneus, medial frontal, inferior parietal cortical regions, medial temporal lobe etc.) to obtain a single voxel time series for each region. We then carried out standard CCA, rCCA and the proposed pCCA algorithm on the dataset and estimated the canonical components to reconstruct activations maps and time series. In case of rCCA, the regularization parameters were set to a value of 4 by a 10-folds cross-validation over values of $\gamma_x$ and $\gamma_y$ ranging between 0 and 5 sampled at 0.5. Fig. 5 shows the estimated spatial maps. Each spatial map obtained using standard CCA, rCCA, and the proposed pCCA was z-scored and thresholded at a p-value of less than $1e^{-3}$. The resulting spatial maps obtained from the proposed pCCA show high activations in the regions classified as DMN in [51] and [50]. It is evident from Fig. 5 that activations corresponding to prefrontal cortex diminished with an incremental progression of brain slices in case of standard CCA and rCCA, whereas the proposed pCCA successfully retains all functionally connected regions of DMN with stronger activations. Also, the activations are tightly localized around DMN and have distinct peaks in case of the proposed pCCA.

To test the recovery of temporal components corresponding to seed-voxels, we selected a $6 \times 6 \times 6$ mm$^3$ cube centered at the given MNI coordinates and correlated it with time series obtained from CCA algorithms. The correlation coefficients of mean time series versus estimated voxel time series presented in Table III. The results have shown that the proposed pCCA carries the highest temporal correlation in selected seed-voxel regions as compared to standard CCA and rCCA. The most correlated estimated voxel time series vs mean seed-voxel time series belonging to precuneus cortex is shown in Fig. 6, the proposed pCCA has outperformed in competition with standard CCA and rCCA, exhibiting a maximum overlap of two time series.

In order to further demonstrate the effectiveness of the proposed method, we extracted other resting-state networks (e.g. frontal network (FN)). Frontal networks are located at the front of the brain, and are responsible for the dopamine sensitive tasks (e.g., short term memory, motivation, planning etc). We ran CCA algorithms on the data and recovered activation maps. The estimated activation maps and most correlated estimated time courses with mean seed-voxel time course corresponding to left frontal network are shown in Fig. 7 and Fig. 8, respectively. The spatial maps extracted by standard CCA and rCCA look quite similar with the proposed pCCA, however, closer examination reveals that functional connectivity in the left frontal network is higher in case of the proposed pCCA. Furthermore, from the correlative perspective, it is evident that the proposed pCCA exhibits the highest temporal correlation with seed-voxel time course corresponding to left frontal network.

### VI. STATISTICAL SIGNIFICANCE ANALYSIS

After obtaining the proposed pCCA estimates, one can test for the level of significance of the estimated dependencies using either the maximum canonical correlation or all of the canonical correlations. However, one major issue in estimating the significance of voxels in fMRI data analysis is the intractability of the theoretical distribution corresponding to the largest canonical correlation value. Moreover, we assume observations are independent, which is not the case in fMRI data. Therefore, a statistical test (e.g., Bartlett test [52], [53]) must be used to test the level of significance instead of only using the largest canonical correlation [19]. Since the aim of CCA is to study the dependence of the estimated canonical variates, the statistics will test independence.

### TABLE III

| MNI coordinates of selected seed voxels of DMN and their correlation coefficients with estimated time series. |
|---------------------------------|----------------|----------------|----------------|
|                                | Standard CCA | Regularized CCA | Projection CCA |
| Precuneus cortex               | 0.66          | 0.72            | 0.74           |
| Ventral posterior cingulate    | 0.62          | 0.73            | 0.74           |
| Superior frontal gyrus         | 0.88          | 0.84            | 0.86           |
| Cingulate Gyrus                | 0.74          | 0.89            | 0.85           |
| Mean                            | 0.74          | 0.72            | 0.82           |

Ventral medial prefrontal cortex
Ventral posterior cingulate
Dorsal posterior cingulate
Frontal pole
Precuneus cortex
Cingulate Gyrus
Mean
in this paper. The obtained canonical variates were derived matrices structurally smooth in one domain was proposed approach via basis expansion a CCA method adapted for data mance. Taking a regularized rank-one matrix approximation domain and classical canonical correlation methods ignoring reshaped as a spatio-temporal matrix is structured in one freedom.

Consider testing the null hypothesis $H_0$ for independence of components between $X$ and $Y$ or equivalently the null hypothesis $C_{xy} = 0$. Also, the hypothesis is equivalent to the condition $s_1 \leq \cdots \leq s_r = 0$. Under $H_0$, Bartlett’s modification of $-2$ times the logarithm of the likelihood ratio criterion is defined as

$$L = -\left[P - \frac{1}{2}(N + N - 1 + 3)\right] \sum_{j=1}^{r} \log(1 - s_j^2)$$

$$= -[P - (N + 1)] \sum_{j=1}^{r} \log(1 - s_j^2) \quad (19)$$

where $s_j$ corresponds to $j^{th}$ nonzero canonical correlation coefficient and there are $r$ nonzero canonical correlations, where $s_1 \geq s_2, \cdots \geq s_r$, the test statistics $L$ follows a chi-square distribution ($\chi^2$) with $N \times (N - 1)$ degrees of freedom. In this case, the number of voxels ($P$) play the role of sample size ($N$). We take advantage of large number of voxels, since the role of samples and variables are reversed. Consider the case where $C_{xy} \neq 0$, the rank of $C_{xy}$ is $r < N$ and the number of correlated components to be computed is $l < r$. The likelihood ratio criterion for testing the null hypothesis $H_1: s_{l+1} = \cdots = s_r = 0$, that is, rank($C_{xy}$) $\leq l$ is given by

$$L = -[P - (N + 1)] \sum_{j=l+1}^{r} \log(1 - s_j^2) \quad (20)$$

has $\chi^2$ distribution with $(N - l)(N - l - 1)$ degrees of freedom.

VII. CONCLUSION

Datasets arising from spatio-temporal measurement fMRI studies can be structurally smooth. In this case, the dataset reshaped as a spatio-temporal matrix is structured in one domain and classical canonical correlation methods ignoring this structure in the data matrix will result in lower performance. Taking a regularized rank-one matrix approximation approach via basis expansion a CCA method adapted for data matrices structurally smooth in one domain was proposed in this paper. The obtained canonical variates were derived by introducing regularization in the left singular and right vectors when solving the rank one matrix approximation problem through alternating least square. The performance of the proposed CCA method was illustrated on both simulated and real fMRI datasets from resting-state and task-related experiments. The experimental results obtained showed that the proposed CCA method provided performance improvement compared to the classical and ridge regularized CCA approaches.

REFERENCES


Author/s: Qadar, MA; Seghouane, A-K

Title: A Projection CCA Method for Effective fMRI Data Analysis

Date: 2019-11-01

Citation: Qadar, MA; Seghouane, A-K, A Projection CCA Method for Effective fMRI Data Analysis, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, 2019, 66 (11), pp. 3247 - 3256

Persistent Link: http://hdl.handle.net/11343/233430

File Description: Accepted version