Two dimensional CCA via penalized matrix decomposition for structure preserved fMRI data analysis

Muhammad Ali Qadar[a,*], Abdeldjalil Aïssa-El-Bey[b], Abd-Krim Seghouane[a]

[a] *Department of Electrical and Electronic Engineering, The University of Melbourne, VIC 3010, Australia*
[b] *IMT Atlantique, Brest, France*

Abstract

Two dimensional canonical correlation analysis (2DCCA) is a data driven method that has been used in image analysis to preserve the spatial structure of images. 2DCCA finds pairs of left and right linear transforms by directly operating on two dimensional data (i.e., image data) such that the correlation between their projections is maximized without neglecting the local spatial structure of the data. However, in context to high dimensional data, the performance of 2DCCA suffers from interpretability of learned projection variables. In this study, to improve the interpretability of projection variables while preserving the local structure of two dimensional data, we propose two new 2DCCA approaches, sparse 2DCCA and regularized 2DCCA. The idea is to solve a penalized rank-1 matrix approximation problem obtained by incorporating the orthogonal projector variables derived from the data instead of employing their cross-matrix product. The validity of the proposed approaches in comparison with standard CCA, 2DCCA and existing sparse 2DCCA(S2DCCA) have been evaluated on three different datasets from: simulated, event-related right finger-tapping task functional magnetic resonance imaging (fMRI) and resting-state

*Corresponding author
Email addresses: mqadar@student.unimelb.edu.au (Muhammad Ali Qadar), abdeldjalil.aissaelbey@imt-atlantique.fr (Abdeldjalil Aïssa-El-Bey), abd-krim.seghouane@unimelb.edu.au (Abd-Krim Seghouane)

fMRI experiments. The experimental results reveal that the proposed methods can successfully extract latent components from both task-dependent as well as resting-state datasets with improved brain connectivity estimation performance. The proposed methods are capable of adapting to the sudden variations in brain activity patterns and exhibit are robust in modelling multivariate functional connections. The proposed methods are directly applicable to image data with improved computation time and can serve as efficient brain connectivity analysis algorithms.

Keywords: Canonical correlation analysis, functional magnetic resonance imaging, sparse decomposition, two dimensional, rank-1 approximation.

---

1. Introduction

A number of statistical methods (hypothesis-driven and data-driven) have been used to detect signal variation with respect to a given experimental paradigm in fMRI data. Signal variations in fMRI data arise due to blood-oxygenation-level dependencies (BOLD) that originate from the difference in blood oxygenation levels between activated and non-activated brain areas Friman et al. (2003). Data-driven methods for the fMRI data analysis extract subspace components related to structures or patterns, such as brain activations, low frequency drifts, or motion related artefacts. These methods decompose observed data based on specific objectives imposed on extracted components; for example, principal component analysis (PCA) results in maximum variance components Andersen et al. (1999), independent component analysis (ICA)Mckeown et al. (1998) yields independent components, and dictionary learning (DL) extracts sparse components Lee et al. (2011). Canonical correlation analysis (CCA) Hotelling (1936), which extracts maximally correlated components between two data matrices, has also been employed for fMRI data analysis Friman et al. (2002a,b); Khalid and Seghouane (2013); Nandy and Cordes (2004); Zhuang et al. (2017). In recent years, CCA has become a popular data analysis tool due to increased interest in identifying relatioships between

pairs of datasets Reshef et al. (2011).

CCA is a multivariate data analysis method used to analyse the correlation between two datasets by seeking optimal projections (called canonical projectors) in such a way that the correlation between projections is maximized Correa et al. (2010). One of the drawbacks of CCA is that it cannot be applied to two dimensional (2D) data such as images. CCA in its original form requires image data to be vectorized before analysis. However, vectorization of the image data may break its underlying structure Yang et al. (2004). In addition, vectorization yields large covariance matrices which may be ill-conditioned, resulting in an unstable solution with high computational complexity. To address the above mentioned issues, a two dimensional CCA (2DCCA) was proposed in Lee and Choi (2007), which directly used image data. 2DCCA is based on the 2D image representation of data and it defines two separate canonical projection vectors corresponding to the row and column directions of the image data, therefore, image-to-vector transformation is not required Desai et al. (2018). In the case of fMRI data, 2DCCA suffers from high dimensionality associated with the large number of volumized pixels (i.e., voxels). Furthermore, the canonical projectors are non-informative, since they are estimated as a dense linear combination of the samples in the original feature space Yan et al. (2012). The issues of high dimensionality and interpretability may be addressed by selecting a subset of sparse variables, thereby estimating sparse canonical projectors from the linear combinations of observed variables in each dataset Hardoon and Shawe-Taylor (2011); Sun et al. (2011); Witten et al. (2009). Variants of sparse CCA have been employed to analyse fMRI data Ahn et al. (2015); Aissa-El-Bey and Seghouane (2017); Avants et al. (2014); Fang et al. (2016); Le Floch et al. (2012); Lin et al. (2014); Mohammadi-Nejad et al. (2017); Seghouane et al. (2017). In Lin et al. (2014), a sparse CCA using group constraints was used to investigate the correspondence between DNA (deoxyribonucleic acid) sequences (namely single nucleotide polymorphism - SNP) and fMRI data to analyse the effects of genetic variations on brain activity. In Mohammadi-Nejad et al. (2017), a sparse CCA approach was applied to the fusion of fMRI and structural MRI

data to preserve spatial structure by imposing non-negativity on the canonical projectors. In Aissa-El-Bey and Seghouane (2017), a sparse CCA method, derived via penalized rank one matrix approximation was applied to resting-state fMRI data analysis in order to find maximally autocorrelated regions of the brain during rest.

Despite the popularity of 2DCCA in a number of applications, such as face recognition Kukharev and Kamenskaya (2010); Sun et al. (2010); Wang (2010); Yan et al. (2012), image inpainting Ogawa and Haseyama (2014), image fusion and feature extraction Gao et al. (2018); Tang et al. (2015), there are only a limited number of studies of 2DCCA in fMRI analysis Desai et al. (2018); Khalid et al. (2012). In this study, we analyze fMRI data by proposing two 2DCCA methods, sparse-2DCCA and regularized 2DCCA (reg-2DCCA). Sparse-2DCCA and reg-2DCCA aim to overcome the aforementioned drawbacks of CCA in high dimensional data and improve the interpretability of canonical projectors. In the formulation of sparse-2DCCA, we solve a penalized rank-1 matrix approximation problem obtained by incorporating orthogonal projectors in the 2DCCA objective function derived from the datasets. The proposed sparse-2DCCA approach iteratively estimates a pairwise sparse canonical projectors by minimizing a rank-1 approximation problem using sparse coding methods. Furthermore, to force learned canonical projectors to be smooth, we propose the reg-2DCCA approach based on rank-1 matrix approximation where smoothness constraints are imposed on the canonical projectors to model sudden variations. The proposed methods have been evaluated in competition with CCA, 2DCCA Lee and Choi (2007), and S2DCCA Yan et al. (2012) on three different fMRI datasets from: simulated, event-related task-based right fingertapping and resting-state experiments.

The remainder of this study is organized as follows. Section II reviews the 2DCCA approach and its solution. Section III presents the derivations and the algorithmic procedure of the proposed sparse-2DCCA approach. Section IV presents the formulation and the algorithmic procedure for the proposed reg-2DCCA approach. Section V evaluate the proposed methods on simulated and

real fMRI datasets. Finally, section VI concludes the study.

2. Methods

Notation:. In remainder of the paper, we denote vectors by boldface lower case letters, (e.g., $\mathbf{u}$), matrices by boldface upper case letters, (e.g., $\mathbf{X}$), and scalars by either italic or greek letters (e.g., d or γ_x). A single image is denoted by $\mathbf{X}_i$, where i represents index of image. Penalty functions are denoted by $\mathcal{F}_x(\ .\ )$ and $\mathcal{F}_y(\ .\ )$ with sparsity parameters λ_x and λ_y.

2.1. Background

2.1.1. Two dimensional canonical correlation analysis

Let $\{\mathbf{X}_i \in \mathbb{R}^{p_x \times q_x}, i = 1, 2, \cdots, N\}$ and $\{\mathbf{Y}_i \in \mathbb{R}^{p_y \times q_y}, i = 1, 2, \cdots, N\}$ be two sets of image data respectively corresponding to N observations from two random image variables $\mathbf{X}$ and $\mathbf{Y}$. Assume that the matrices $\mathbf{X}$ and $\mathbf{Y}$ are centered (i.e., $\mathbb{E}[\mathbf{X}] = 0$ and $\mathbb{E}[\mathbf{Y}] = 0$, where $\mathbb{E}[\ .\ ]$ is the expectation term). Then, the 2DCCA seeks a pairwise left projection directions $\boldsymbol{\alpha}_x \in \mathbb{R}^{p_x}$ and $\boldsymbol{\alpha}_y \in \mathbb{R}^{p_y}$ and pairwise right projection directions $\boldsymbol{\beta}_x \in \mathbb{R}^{q_x}$ and $\boldsymbol{\beta}_y \in \mathbb{R}^{q_y}$, such that the correlation (ρ) between two bilateral projections $\boldsymbol{\alpha}_x^\top \mathbf{X} \boldsymbol{\beta}_x$ and $\boldsymbol{\alpha}_y^\top \mathbf{Y} \boldsymbol{\beta}_y$ is maximized. The 2DCCA objective function is given by

$$\rho(\mathbf{X}, \mathbf{Y}; \boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y, \boldsymbol{\beta}_x, \boldsymbol{\beta}_y) = \frac{\mathrm{cov}(\boldsymbol{\alpha}_x^\top \mathbf{X} \boldsymbol{\beta}_x, \boldsymbol{\alpha}_y^\top \mathbf{Y} \boldsymbol{\beta}_y)}{\sqrt{\mathrm{Var}(\boldsymbol{\alpha}_x^\top \mathbf{X} \boldsymbol{\beta}_x) \mathrm{Var}(\boldsymbol{\alpha}_y^\top \mathbf{Y} \boldsymbol{\beta}_y)}}. \tag{1}$$

We define the right sample covariance matrices and cross-covariance matrices as follows

$$\mathbf{C}_{xy}^r = \frac{1}{N} \sum_{i=1}^{N} \mathbf{X}_i \boldsymbol{\beta}_x \boldsymbol{\beta}_y^\top \mathbf{Y}_i^\top$$

$$\mathbf{C}_{xx}^r = \frac{1}{N} \sum_{i=1}^{N} \mathbf{X}_i \boldsymbol{\beta}_x \boldsymbol{\beta}_x^\top \mathbf{X}_i^\top$$

$$\mathbf{C}_{yy}^r = \frac{1}{N} \sum_{i=1}^{N} \mathbf{Y}_i \boldsymbol{\beta}_y \boldsymbol{\beta}_y^\top \mathbf{Y}_i^\top.$$

Therefore, the correlation between $\boldsymbol{\alpha}_x^\top \mathbf{X} \boldsymbol{\beta}_x$ and $\boldsymbol{\alpha}_y^\top \mathbf{Y} \boldsymbol{\beta}_y$ can be rewritten as

$$\mathrm{cov}(\boldsymbol{\alpha}_x^\top \mathbf{X} \boldsymbol{\beta}_x, \boldsymbol{\alpha}_y^\top \mathbf{Y} \boldsymbol{\beta}_y) = \boldsymbol{\alpha}_x^\top \mathbf{C}_{xy}^r \boldsymbol{\alpha}_y$$

where $\mathrm{Var}(\boldsymbol{\alpha}_x^\top \mathbf{X}\boldsymbol{\beta}_x) = \boldsymbol{\alpha}_x^\top \mathbf{C}_{xx}^r \boldsymbol{\alpha}_x$, $\mathrm{Var}(\boldsymbol{\alpha}_y^\top \mathbf{Y}\boldsymbol{\beta}_y) = \boldsymbol{\alpha}_y^\top \mathbf{C}_{yy}^r \boldsymbol{\alpha}_y$. Then, the function ρ in (1) is given by

$$\rho(\mathbf{X}, \mathbf{Y}; \boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y, \boldsymbol{\beta}_x, \boldsymbol{\beta}_y) = \frac{\boldsymbol{\alpha}_x^\top \mathbf{C}_{xy}^r \boldsymbol{\alpha}_y}{\sqrt{(\boldsymbol{\alpha}_x^\top \mathbf{C}_{xx}^r \boldsymbol{\alpha}_x)(\boldsymbol{\alpha}_y^\top \mathbf{C}_{yy}^r \boldsymbol{\alpha}_y)}}. \tag{2}$$

Since, the value of $\rho(\mathbf{X}, \mathbf{Y}; \boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y, \boldsymbol{\beta}_x, \boldsymbol{\beta}_y)$ is invariant to the scaling of the projection directions, we reformulate the objective function in (2) to maximize the optimization problem

$$\begin{aligned} &\underset{\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y}{\arg\max} && \boldsymbol{\alpha}_x^\top \mathbf{C}_{xy}^r \boldsymbol{\alpha}_y \\ &\text{subject to} && \boldsymbol{\alpha}_x^\top \mathbf{C}_{xx}^r \boldsymbol{\alpha}_x = 1, \quad \boldsymbol{\alpha}_y^\top \mathbf{C}_{yy}^r \boldsymbol{\alpha}_y = 1. \end{aligned} \tag{3}$$

On the other hand, the 2DCCA problem can also be rewritten in terms of left covariance and cross-covariance matrices, such as:

$$\begin{aligned} &\underset{\boldsymbol{\beta}_x, \boldsymbol{\beta}_y}{\arg\max} && \boldsymbol{\beta}_x^\top \mathbf{C}_{xy}^l \boldsymbol{\beta}_y \\ &\text{subject to} && \boldsymbol{\beta}_x^\top \mathbf{C}_{xx}^l \boldsymbol{\beta}_x = 1, \quad \boldsymbol{\beta}_y^\top \mathbf{C}_{yy}^l \boldsymbol{\beta}_y = 1. \end{aligned} \tag{4}$$

where $\mathbf{C}_{xy}^l = \frac{1}{N}\sum_{i=1}^N \mathbf{X}_i \boldsymbol{\alpha}_x \boldsymbol{\alpha}_y^\top \mathbf{Y}_i^\top$, $\mathbf{C}_{xx}^l = \frac{1}{N}\sum_{i=1}^N \mathbf{X}_i \boldsymbol{\alpha}_x \boldsymbol{\alpha}_x^\top \mathbf{X}_i^\top$, and $\mathbf{C}_{yy}^l = \frac{1}{N}\sum_{i=1}^N \mathbf{Y}_i \boldsymbol{\alpha}_y \boldsymbol{\alpha}_y^\top \mathbf{Y}_i^\top$.

2DCCA iteratively estimates the pair of left canonical projectors $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$ by maximizing (3) with $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_y$ fixed. Similarly, the pair of right canonical projectors $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_y$ are estimated by maximizing (4) with $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$ fixed. Then, $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_y$ are fixed, the maximization of (3) leads to the following generalized eigenvalue decomposition (GEVD) problem

$$\mathbf{C}_{xy}^r \boldsymbol{\alpha}_y = \lambda \mathbf{C}_{xx}^r \boldsymbol{\alpha}_x \tag{5}$$

$$\mathbf{C}_{yx}^r \boldsymbol{\alpha}_x = \lambda \mathbf{C}_{yy}^r \boldsymbol{\alpha}_y \tag{6}$$

In a similar manner, with fixed $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$, the maximization of (4) leads to the following GEVD

$$\mathbf{C}_{xy}^l \boldsymbol{\beta}_y = \lambda \mathbf{C}_{xx}^l \boldsymbol{\beta}_x \tag{7}$$

$$\mathbf{C}^{l}_{yx}\boldsymbol{\beta}_x = \lambda \mathbf{C}^{l}_{yy}\boldsymbol{\beta}_y \tag{8}$$

2DCCA Lee and Choi (2007) is a two stage CCA procedure, where the vectors $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$ are obtained by applying the CCA procedure on the two matrices $\mathbf{X}_{\boldsymbol{\beta}_x} = [\mathbf{X}_1\boldsymbol{\beta}_x, \cdots, \mathbf{X}_N\boldsymbol{\beta}_x]$ and $\mathbf{Y}_{\boldsymbol{\beta}_y} = [\mathbf{Y}_1\boldsymbol{\beta}_y, \cdots, \mathbf{Y}_N\boldsymbol{\beta}_y]$, whereas the vectors $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$ are obtained by applying the CCA procedure on the two matrices $\mathbf{X}_{\boldsymbol{\alpha}_x} = [\mathbf{X}_1\boldsymbol{\alpha}_x, \cdots, \mathbf{X}_N\boldsymbol{\alpha}_x]$ and $\mathbf{Y}_{\boldsymbol{\alpha}_y} = [\mathbf{Y}_1\boldsymbol{\alpha}_y, \cdots, \mathbf{Y}_N\boldsymbol{\alpha}_y]$.

2.2. Proposed sparse two dimensional CCA algorithm

In this section, we derive our proposed sparse two dimensional CCA (S2DCCA) method by formulating a rank-1 matrix approximation problem. We also provide a pseudo-iterative algorithm to solve the proposed optimization problem. Let us consider two datasets $\mathbf{X} = [\mathbf{X}_1, \cdots, \mathbf{X}_N] \in \mathbb{R}^{p_x \times Nq_x}$ and $\mathbf{Y} = [\mathbf{Y}_1, \cdots, \mathbf{Y}_N] \in \mathbb{R}^{p_y \times Nq_y}$. Both $\mathbf{X}$ and $\mathbf{Y}$ are zero mean such that $\mathbf{M}_x = 0$ and $\mathbf{M}_y = 0$, where $\mathbf{M}_x = N^{-1}\sum_{i=1}^{N}\mathbf{X}_i$ and $\mathbf{M}_y = N^{-1}\sum_{i=1}^{N}\mathbf{Y}_i$ for $i = 1, \cdots, N$. The following reformulates the 2DCCA problem to that of finding the best rank-1 matrix approximation under the Frobenius norm of the product of the two orthogonal projectors derived from the datasets.

First, the optimization problem of 2DCCA for estimating left projection directions $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$ using the estimated covariances is formulated by

$$\begin{aligned} \underset{\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y}{\arg\max} \quad & \boldsymbol{\alpha}_x^{\top}\mathbf{X}_r\mathbf{Y}_r^{\top}\boldsymbol{\alpha}_y \\ \text{subject to} \quad & \boldsymbol{\alpha}_x^{\top}\mathbf{X}_r\mathbf{X}_r^{\top}\boldsymbol{\alpha}_x = 1, \quad \boldsymbol{\alpha}_y^{\top}\mathbf{Y}_r\mathbf{Y}_r^{\top}\boldsymbol{\alpha}_y = 1. \end{aligned} \tag{9}$$

where

$$\mathbf{X}_r = \mathbf{X}\mathbf{R}_x, \mathbf{R}_x = \mathbf{I}_N \otimes \boldsymbol{\beta}_x \text{ and } \mathbf{Y}_r = \mathbf{Y}\mathbf{R}_y, \mathbf{R}_y = \mathbf{I}_N \otimes \boldsymbol{\beta}_y \tag{10}$$

such that $\mathbf{X}_r\mathbf{X}_r^{\top} = \sum_{i=1}^{N}\mathbf{X}_i\boldsymbol{\beta}_x\boldsymbol{\beta}_x^{\top}\mathbf{X}_i^{\top}$ and $\mathbf{Y}_r\mathbf{Y}_r^{\top} = \sum_{i=1}^{N}\mathbf{Y}_i\boldsymbol{\beta}_y\boldsymbol{\beta}_y^{\top}\mathbf{Y}_i^{\top}$. Then, the GEVD problems given in (5) and (6) are written as

$$\mathbf{X}_r\mathbf{Y}_r^{\top}\boldsymbol{\alpha}_y = \lambda\mathbf{X}_r\mathbf{X}_r^{\top}\boldsymbol{\alpha}_x \tag{11}$$

$$\mathbf{Y}_r\mathbf{X}_r^\top\boldsymbol{\alpha}_x = \lambda\mathbf{Y}_r\mathbf{Y}_r^\top\boldsymbol{\alpha}_y. \tag{12}$$

To proceed, we multiply each side of (11) and (12) by $\mathbf{X}_r^\top(\mathbf{X}_r\mathbf{X}_r^\top)^{-1}$ and $\mathbf{Y}_r^\top(\mathbf{Y}_r\mathbf{Y}_r^\top)^{-1}$, respectively, to obtain:

$$\mathbf{X}_r^\top(\mathbf{X}_r\mathbf{X}_r^\top)^{-1}\mathbf{X}_r\mathbf{Y}_r^\top\boldsymbol{\alpha}_y = \mathbf{P}_{X_r}\mathbf{Y}_r^\top\boldsymbol{\alpha}_y = \lambda\mathbf{X}_r^\top\boldsymbol{\alpha}_x \tag{13}$$

$$\mathbf{Y}_r^\top(\mathbf{Y}_r\mathbf{Y}_r^\top)^{-1}\mathbf{Y}_r\mathbf{X}_r^\top\boldsymbol{\alpha}_x = \mathbf{P}_{Y_r}\mathbf{X}_r^\top\boldsymbol{\alpha}_x = \lambda\mathbf{Y}_r^\top\boldsymbol{\alpha}_y, \tag{14}$$

where $\mathbf{P}_{X_r} = \mathbf{X}_r^\top(\mathbf{X}_r\mathbf{X}_r^\top)^{-1}\mathbf{X}_r$ and $\mathbf{P}_{Y_r} = \mathbf{Y}_r^\top(\mathbf{Y}_r\mathbf{Y}_r^\top)^{-1}\mathbf{Y}_r$ correspond to the orthogonal projector matrices onto the row-space of $\mathbf{X}$ and $\mathbf{Y}$, respectively. By substituting $\mathbf{X}_r^\top\boldsymbol{\alpha}_x = \lambda^{-1}\mathbf{P}_{X_r}\mathbf{Y}^\top\boldsymbol{\alpha}_y$ in (14) and $\mathbf{Y}_r^\top\boldsymbol{\alpha}_y = \lambda^{-1}\mathbf{P}_{Y_r}\mathbf{X}_r^\top\boldsymbol{\alpha}_x$ in (13), we obtain

$$\mathbf{P}_{X_r}\mathbf{P}_{Y_r}\mathbf{X}_r^\top\boldsymbol{\alpha}_x = \lambda^2\mathbf{X}_r^\top\boldsymbol{\alpha}_x$$

$$\mathbf{P}_{Y_r}\mathbf{P}_{X_r}\mathbf{Y}_r^\top\boldsymbol{\alpha}_y = \lambda^2\mathbf{Y}_r^\top\boldsymbol{\alpha}_y$$

It can be seen that the above equations yield an eigenvalue decomposition (EVD) problem where $\mathbf{X}_r^\top\boldsymbol{\alpha}_x$ and $\mathbf{Y}_r^\top\boldsymbol{\alpha}_y$ are the eigenvectors corresponding to the largest eigenvalues of the projection matrices $\mathbf{K}_{XY} = \mathbf{P}_{X_r}\mathbf{P}_{Y_r}$ and $\mathbf{K}_{YX} = \mathbf{P}_{Y_r}\mathbf{P}_{X_r}$, respectively. Let us consider the EVD of $\mathbf{K}_{XY}$ and $\mathbf{K}_{YX}$, such that:

$$\mathbf{K}_{XY} = \mathbf{U}\mathbf{D}_{XY}\mathbf{U}^{-1} \quad \text{and} \quad \mathbf{K}_{YX} = \mathbf{V}\mathbf{D}_{YX}\mathbf{V}^{-1} \tag{15}$$

where $\mathbf{U}$ and $\mathbf{V}$ are the matrices of eigenvectors, $\mathbf{D}_{XY}$ and $\mathbf{D}_{YX}$ are diagonal matrices of size $N \times N$, containing the eigenvalues of $\mathbf{K}_{XY}$ and $\mathbf{K}_{YX}$ in descending order, respectively. By virtue of the symmetry property of the projection matrices (i.e., $\mathbf{P}_{XY}$ and $\mathbf{P}_{YX}$), we can write:

$$\mathbf{K}_{YX} = \mathbf{P}_{Y_r}\mathbf{P}_{X_r} = \mathbf{P}_{Y_r}^\top\mathbf{P}_{X_r}^\top = (\mathbf{P}_{X_r}\mathbf{P}_{Y_r})^\top = \mathbf{K}_{XY}^\top \tag{16}$$

Thus, from the equations (15) and (16) we deduce that $\mathbf{D}_{XY} = \mathbf{D}_{YX}$, $\mathbf{U}^{-1} = \mathbf{V}^\top$, therefore, the decomposition of the matrix $\mathbf{K}_{XY}$ can be rewritten as a

$$\mathbf{K}_{XY} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$$

which corresponds to the singular value decomposition (SVD) of $\mathbf{K}_{XY}$. To proceed with the estimation of the canonical projectors, we formulate a rank-1 approximation of the matrix $\mathbf{K}_{XY}$, given by:

$$\mathbf{K}_1 = d_1 \mathbf{u}_1 \mathbf{v}_1^\top$$

where the aim is to minimize the squared Frobenius norm between $\|\mathbf{K}_{XY} - \mathbf{K}_1\|_F^2$. Therefore, we can define the rank-1 approximation of $\mathbf{K}_{XY}$ as the following optimization problem:

$$\underset{\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y}{\arg\min} \left\| \mathbf{K}_{XY} - d_1 \mathbf{X}_r^\top \boldsymbol{\alpha}_x \boldsymbol{\alpha}_y^\top \mathbf{Y}_r \right\|_F^2 \tag{17}$$

The projection of linear directions onto the data (i.e., $\mathbf{X}_r^\top \boldsymbol{\alpha}_x$ and $\mathbf{Y}_r^\top \boldsymbol{\alpha}_y$) corresponds to the left and right singular vectors relating to the largest singular value of $\mathbf{K}_{XY}$, where d_1 is defined as $\boldsymbol{\alpha}_x^\top \mathbf{X}_r \mathbf{K}_{XY} \boldsymbol{\alpha}_y^\top \mathbf{Y}_r$. The remaining singular values/vectors can be estimated via an iterative deflation procedure.

In a similar way, the optimization problem related to the estimation of the pair of right linear projection directions $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_y$ based on the estimated covariances is

$$\begin{aligned} &\underset{\boldsymbol{\beta}_x, \boldsymbol{\beta}_y}{\arg\max} \quad \boldsymbol{\beta}_x^\top \mathbf{X}_l^\top \mathbf{Y}_l \boldsymbol{\beta}_y \\ &\text{subject to} \quad \boldsymbol{\beta}_x^\top \mathbf{X}_l^\top \mathbf{X}_l \boldsymbol{\beta}_x = 1,\ \boldsymbol{\beta}_y^\top \mathbf{Y}_l^\top \mathbf{Y}_l \boldsymbol{\beta}_y = 1 \end{aligned} \tag{18}$$

where

$$\mathbf{X}_l = \mathbf{X}^\top \mathbf{L}_x, \mathbf{L}_x = \mathbf{I}_N \otimes \boldsymbol{\alpha}_x \ \mathbf{Y}_l = \mathbf{Y}^\top \mathbf{L}_y, \mathbf{L}_y = \mathbf{I}_N \otimes \boldsymbol{\alpha}_y \tag{19}$$

such that $\mathbf{X}_l^\top \mathbf{X}_l = \sum_{i=1}^N \mathbf{X}_i^\top \boldsymbol{\alpha}_x \boldsymbol{\alpha}_x^\top \mathbf{X}_i$ and $\mathbf{Y}_l^\top \mathbf{Y}_l = \sum_{i=1}^N \mathbf{Y}_i^\top \boldsymbol{\alpha}_y \boldsymbol{\alpha}_y^\top \mathbf{Y}_i$. The objective function (18) can be rewritten as a rank-1 matrix approximation problem:

$$\underset{\boldsymbol{\beta}_x, \boldsymbol{\beta}_y}{\arg\min} \left\| \mathbf{Q}_{XY} - s_1 \mathbf{X}_l^\top \boldsymbol{\beta}_x \boldsymbol{\beta}_y^\top \mathbf{Y}_l \right\|_F^2 \tag{20}$$

where $\mathbf{Q}_{XY} = \mathbf{P}_{X_l} \mathbf{P}_{Y_l}$ with $\mathbf{P}_{X_l} = \mathbf{X}_l^\top (\mathbf{X}_l \mathbf{X}_l^\top)^{-1} \mathbf{X}_l$ and $\mathbf{P}_{Y_l} = \mathbf{Y}_l^\top (\mathbf{Y}_l \mathbf{Y}_l^\top)^{-1} \mathbf{Y}_l$, $\mathbf{X}_l^\top \boldsymbol{\beta}_x$ and $\mathbf{Y}_l^\top \boldsymbol{\beta}_y$ are the singular vectors corresponding to the singular value s_1 of $\mathbf{Q}_{XY}$ defined as $\boldsymbol{\beta}_x^\top \mathbf{X}_l \mathbf{Q}_{XY} \boldsymbol{\beta}_y^\top \mathbf{Y}_l$.

After obtaining the rank-1 matrix approximation problems for left and right

linear transforms in (17) and (20), we now impose the penalty functions onto the optimization problems in (17) and (20).

In general, the canonical projectors $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$ obtained by solving (17) are not sparse, in order to estimate sparse $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$, we adopt a techniques similar to those of Chu et al. (2013); Sun et al. (2011) by introducing the penalties on the problem (17) and (20). The objective function (17) with sparsity constraints is given by:

$$\begin{aligned} &\underset{\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y}{\arg\min} \left\| \mathbf{K}_{XY} - \mathbf{X}_r^\top \boldsymbol{\alpha}_x \boldsymbol{\alpha}_y^\top \mathbf{Y}_r \right\|_F^2 \\ &\text{subject to} \quad \mathcal{F}_x(\boldsymbol{\alpha}_x) \leq \lambda_x \quad \text{and} \quad \mathcal{F}_y(\boldsymbol{\alpha}_y) \leq \lambda_y \end{aligned} \tag{21}$$

with λ_x and λ_y are trade-off parameters, $\mathcal{F}_x(\ .\ )$ and $\mathcal{F}_y(\ .\ )$ are the penalty functions, these function may take various forms, such as the l_0-quasi-norm (defined as $\mathcal{F}(\mathbf{z}) = \|\mathbf{z}\|_0$) or the Lasso penalty (defined as l_1-norm $\mathcal{F}(\mathbf{z}) = \|\mathbf{z}\|_1$). To solve problem (21), we first fix $\boldsymbol{\alpha}_y$ and solve (21) for $\boldsymbol{\alpha}_x$. Similarly, we can then solve (21) for $\boldsymbol{\alpha}_y$ with fixed $\boldsymbol{\alpha}_x$. The above two step procedure is then repeated until convergence.

Alternatively, (21) can be formulated as an ordinary sparse coding task by splitting it into two optimization problems. Then, for fixed $\boldsymbol{\alpha}_y$, (21) is given by

$$\underset{\boldsymbol{\alpha}_x}{\arg\min} \left\| \mathbf{K}_{XY} \mathbf{Y}_r^\top \boldsymbol{\alpha}_y - \mathbf{X}_r^\top \boldsymbol{\alpha}_x \right\|_2^2 \ \text{ subject to } \ \mathcal{F}_x(\boldsymbol{\alpha}_x) \leq \lambda_x \tag{22}$$

the problem (22) may be minimized using any sparse approximation approach. Similarly, (21) with fixed $\boldsymbol{\alpha}_x$ is:

$$\underset{\boldsymbol{\alpha}_y}{\arg\min} \left\| \mathbf{K}_{XY}^\top \mathbf{X}_r^\top \boldsymbol{\alpha}_x - \mathbf{Y}_r^\top \boldsymbol{\alpha}_y \right\|_2^2 \ \text{ subject to } \ \mathcal{F}_y(\boldsymbol{\alpha}_y) \leq \lambda_y \tag{23}$$

The first pairwise sparse projectors (i.e., $\boldsymbol{\alpha}_x$, $\boldsymbol{\alpha}_y$) can be estimated by solving the optimization problems (22) and (23) using sparse coding techniques used in Chu et al. (2013); Sun et al. (2011). To obtain several canonical projectors, consider the SVD of $\mathbf{K}_{XY} = \mathbf{U}\mathbf{D}\mathbf{V}^\top = \sum_{i=1}^{N} d_i \mathbf{u}_i \mathbf{v}_i^\top$, where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the column vectors of the matrices of singular vectors $\mathbf{U}$ and $\mathbf{V}$, respectively, and $\mathbf{D} = \text{diag}(d_1, \cdots, d_N)$, with (i.e., $d_1 \geq d_2 \geq \cdots \geq d_N$). The second pair of the canonical projectors are estimated by removing the contribution of the first pair

by:

$$\mathbf{K}_{XY} - d_1 \mathbf{u}_1 \mathbf{v}_1^\top = \sum_{i=2}^{N} d_i \mathbf{u}_i \mathbf{v}_i^\top \tag{24}$$

The singular vectors $\mathbf{u}_1$ and $\mathbf{v}_1$ represent the projection data $\mathbf{X}_r^\top \boldsymbol{\alpha}_x$ and $\mathbf{Y}_r^\top \boldsymbol{\alpha}_y$, respectively. By virtue of the unitary property of matrices $\mathbf{U}$ and $\mathbf{V}$, the singular values corresponding to the vectors $\mathbf{u}_1$ and $\mathbf{v}_1$ can be computed by $d_1 = \mathbf{u}_1^\top \mathbf{K}_{XY} \mathbf{v}_1$. By using d_1, $\mathbf{u}_1$ and $\mathbf{v}_1$, the second pairwise canonical projectors can be obtained by the residual matrix $\mathbf{K}_{XY} - \boldsymbol{\alpha}_x^\top \mathbf{X}_r \mathbf{K}_{Xy} \mathbf{Y}_r^\top \boldsymbol{\alpha}_y \mathbf{X}_r^\top \boldsymbol{\alpha}_x \boldsymbol{\alpha}_y^\top \mathbf{Y}_r$ in a deflationary procedure. We can then find the remaining pairwise canonical projectors.

In the same way, the estimated $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$ are used to find the right pair of projection directions $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_y$ by solving the following optimization problem:

$$\begin{aligned} &\underset{\boldsymbol{\beta}_x, \boldsymbol{\beta}_y}{\arg\min} \left\| \mathbf{Q}_{XY} - \mathbf{X}_l^\top \boldsymbol{\beta}_x \boldsymbol{\beta}_y^\top \mathbf{Y}_l \right\|_F^2 \\ &\text{subject to} \quad \mathcal{F}_x(\boldsymbol{\beta}_x) \leq \delta_x \quad \text{and} \quad \mathcal{F}_x(\boldsymbol{\beta}_y) \leq \delta_y. \end{aligned} \tag{25}$$

The proposed algorithm for sparse 2DCCA based on rank-1 matrix approximation is described in Algorithm 1 and 2. The computational cost of each iteration of our algorithm is $O(p_x p_y)$ for computing $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$ and $O(q_x q_y)$ for computation of $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_y$. The computational cost of computing both $\mathbf{K}_{XY}$ and $\mathbf{Q}_{XY}$ is dominated by the computational cost of the matrix inversions $(\mathbf{X}_r \mathbf{X}_r^\top)^{-1}$ and $(\mathbf{Y}_r \mathbf{Y}_r^\top)^{-1}$ which is $O(p_x^3)$ and $O(p_y^3)$ for $\mathbf{K}_{XY}$ and $(\mathbf{X}_l \mathbf{X}_l^\top)^{-1}$ and $(\mathbf{Y}_l \mathbf{Y}_l^\top)^{-1}$ which is $O(q_x^3)$ and $O(q_y^3)$ for $\mathbf{Q}_{XY}$. The proposed approach also allows different levels of sparsity for different $\boldsymbol{\alpha}_x$, $\boldsymbol{\alpha}_y$, $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_y$ using different λ_x, λ_y, δ_x, and δ_y.

2.3. Regularized two dimensional CCA algorithm

In this section, we propose regularized 2DCCA (R2DCCA) approach as another form of (21) that enforces smoothness constraints on the canonical projection directions. Additionally, we present an efficient iterative algorithm to estimate the left and right projection directions.

Algorithm 1: Overview of proposed sparse 2DCCA algorithm

Input: Training data $\mathbf{X} \in \mathbb{R}^{p_x \times Nq_x}$, $\mathbf{Y} \in \mathbb{R}^{p_y \times Nq_y}$, desired number of components k.

Output:

- The k pairs of canonical projectors $\boldsymbol{\alpha}_x \in \mathbb{R}^{p_x \times k}$ and $\boldsymbol{\alpha}_y \in \mathbb{R}^{p_y \times k}$.
- The k pairs of canonical projectors $\boldsymbol{\beta}_x \in \mathbb{R}^{q_x \times k}$ and $\boldsymbol{\beta}_y \in \mathbb{R}^{q_y \times k}$.

repeat
Compute $\mathbf{X}_r = \mathbf{X}\mathbf{R}_x$, $\mathbf{Y}_r = \mathbf{Y}\mathbf{R}_y$, and $\mathbf{K}_{XY}$.
Use **Algorithm 2** to obtain $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$,
Compute $\mathbf{X}_l = \mathbf{X}\mathbf{L}_x$, $\mathbf{Y}_l = \mathbf{Y}\mathbf{L}_y$, and $\mathbf{Q}_{XY}$.
Use **Algorithm 2** to obtain $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_y$.
until convergence

In an approach similar to that of Section III, we introduce a Tikhonove regularization term in the rank-1 matrix approximation problem (17), such that:

$$\begin{aligned}\underset{\boldsymbol{\alpha}_x,\boldsymbol{\alpha}_y}{\arg\min} \left\|\mathbf{K}_{XY} - \mathbf{X}_r^\top \boldsymbol{\alpha}_x \boldsymbol{\alpha}_y^\top \mathbf{Y}_r\right\|_F^2 \\ + \gamma_x \boldsymbol{\alpha}_x^\top \mathbf{X}_r \boldsymbol{\Omega}_x \mathbf{X}_r^\top \boldsymbol{\alpha}_x + \gamma_y \boldsymbol{\alpha}_y^\top \mathbf{Y}_r \boldsymbol{\Omega}_y \mathbf{Y}_r^\top \boldsymbol{\alpha}_y\end{aligned} \tag{26}$$

where $\gamma_x > 0$ and $\gamma_y > 0$ are smoothness controlling parameters. $\boldsymbol{\Omega}_x$ and $\boldsymbol{\Omega}_y$ are the non-negative roughness penalty matrices Ciuciu et al. (2003); Ramsay and Silverman (2005) such that:

$$\forall \mathbf{z} \in \mathbb{R}^N, \qquad \mathbf{z}^\top \boldsymbol{\Omega} \mathbf{z} = z_1^2 + z_N^2 + \sum_{i=2}^{N-1} (z_{i+1} - 2z_i + z_{i-1})^2 \tag{27}$$

To proceed with the minimization, we can rewrite (26) as:

$$\begin{aligned}\underset{\boldsymbol{\alpha}_x,\boldsymbol{\alpha}_y}{\arg\min} \left\|\mathbf{X}_r^\top \boldsymbol{\alpha}_x\right\|_2^2 \left\|\mathbf{Y}_r^\top \boldsymbol{\alpha}_y\right\|_2^2 - 2\boldsymbol{\alpha}_x^\top \mathbf{X}_r \mathbf{K}_{XY} \mathbf{Y}_r^\top \boldsymbol{\alpha}_y \\ + \gamma_x \boldsymbol{\alpha}_x^\top \mathbf{X}_r \boldsymbol{\Omega}_x \mathbf{X}_r^\top \boldsymbol{\alpha}_x + \gamma_y \boldsymbol{\alpha}_y^\top \mathbf{Y}_r \boldsymbol{\Omega}_y \mathbf{Y}_r^\top \boldsymbol{\alpha}_y\end{aligned} \tag{28}$$

Problem (28) is then minimized for $\boldsymbol{\alpha}_x$ by fixing $\boldsymbol{\alpha}_y$, and vice versa. This two step minimization procedure is then repeated until convergence. By taking the

Algorithm 2: Iterative sparse 2DCCA procedure

Input: Training data $\mathbf{X}_r$, $\mathbf{Y}_r$, and $\mathbf{K}_{XY}$ (or $\mathbf{X}_l$, $\mathbf{Y}_l$, and $\mathbf{Q}_{XY}$).

Output: The k pairs of canonical projectors $\boldsymbol{\alpha}_x \in \mathbb{R}^{p_x \times k}$ and $\boldsymbol{\alpha}_y \in \mathbb{R}^{p_y \times k}$.

for $i = 1, \cdots, k$ **do**

  $SVD(\mathbf{K}_{XY}) = \mathbf{U}\mathbf{D}\mathbf{V}^\top$,

  Initialize $\tilde{\mathbf{u}} = \mathbf{u}_1$ and $\tilde{\mathbf{v}} = \mathbf{v}_1$;

  repeat

  Update the i-th column of $\boldsymbol{\alpha}_x$ according to (22):

  $\boldsymbol{\alpha}_x(:,i) = \arg\min_{\boldsymbol{\alpha}_x(:,i)} \left\|\mathbf{K}_{XY}\tilde{\mathbf{v}} - \mathbf{X}_r^\top \boldsymbol{\alpha}_x(:,i)\right\|_2^2$

    subject to $\mathcal{F}_x(\boldsymbol{\alpha}_x(:,i)) \leq \beta_x$

  Update $\tilde{\mathbf{u}} = \frac{\mathbf{X}_r^\top \boldsymbol{\alpha}_x(:,i)}{\|\mathbf{X}_r^\top \boldsymbol{\alpha}_x(:,i)\|_2}$;

  Update the i-th column of $\boldsymbol{\alpha}_y$ according to (23):

  $\boldsymbol{\alpha}_y(:,i) = \arg\min_{\boldsymbol{\alpha}_y(:,i)} \left\|\mathbf{K}_{XY}^\top\tilde{\mathbf{u}} - \mathbf{Y}_r^\top \boldsymbol{\alpha}_y(:,i)\right\|_2^2$

    subject to $\mathcal{F}_y(\boldsymbol{\alpha}_y(:,i)) \leq \beta_y$

  Update $\tilde{\mathbf{v}} = \frac{\mathbf{Y}_r^\top \boldsymbol{\alpha}_y(:,i)}{\|\mathbf{Y}_r^\top \boldsymbol{\alpha}_y(:,i)\|_2}$;

  until convergence

  Update $\mathbf{K}_{XY} : \mathbf{K}_{XY} \leftarrow \mathbf{K}_{XY} - \tilde{\mathbf{u}}^\top \mathbf{K}_{XY}\tilde{\mathbf{v}}\tilde{\mathbf{u}}\tilde{\mathbf{v}}^\top$

end

derivative of (28) with respect to $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$, we obtain:

$$\left(\left\|\mathbf{Y}_r^\top \boldsymbol{\alpha}_y\right\|_2^2 \mathbf{X}_r\mathbf{X}_r^\top + \gamma_x \mathbf{X}_r \boldsymbol{\Omega}_x \mathbf{X}_r^\top\right)\boldsymbol{\alpha}_x = \mathbf{X}_r\mathbf{K}_{XY}\mathbf{Y}_r^\top\boldsymbol{\alpha}_y$$

$$\left(\left\|\mathbf{X}_r^\top \boldsymbol{\alpha}_x\right\|_2^2 \mathbf{Y}_r\mathbf{Y}_r^\top + \gamma_y \mathbf{Y}_r \boldsymbol{\Omega}_y \mathbf{Y}_r^\top\right)\boldsymbol{\alpha}_y = \mathbf{Y}_r\mathbf{K}_{XY}^\top\mathbf{X}_r^\top\boldsymbol{\alpha}_x.$$

The first left projectors $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$ can be obtained by solving the above system of equations as a least squares problem (see algorithm 3 and 4). The remaining canonical projectors can be estimated by an iterative deflation approach where the second pair of the canonical projectors are found from the residual matrix $\mathbf{K}_{XY} - \boldsymbol{\alpha}_x^\top \mathbf{X}_r \mathbf{K}_{XY} \mathbf{Y}_r^\top \boldsymbol{\alpha}_y \mathbf{X}_r^\top \boldsymbol{\alpha}_x \boldsymbol{\alpha}_y^\top \mathbf{Y}_r$. Using the residual matrix, we can find the other pair of canonical projectors. In a similar manner to that presented in

the Section III, given $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$, we can estimate the right pair of canonical projectors $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_y$. The algorithmic procedure to obtain regularized rank-1 matrix approximation of 2DCCA is presented in Algorithm 3 and 4.

Compared to Algorithm 1, the computation cost of each iteration of this algorithm is dominated by the cost of computing matrix inversions. The computational cost for $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_x$ is $O(p_x^3)$ and $O(p_y^3)$ and for $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_x$ is $O(q_x^3)$ and $O(q_y^3)$.

Algorithm 3: Overview of proposed regularized 2DCCA algorithm

Input: Training data $\mathbf{X} \in \mathbb{R}^{p_x \times Nq_x}$, $\mathbf{Y} \in \mathbb{R}^{p_y \times Nq_y}$, desired number of components k.

Output:

- The k pairs of canonical projectors $\boldsymbol{\alpha}_x \in \mathbb{R}^{p_x \times k}$ and $\boldsymbol{\alpha}_y \in \mathbb{R}^{p_y \times k}$.
- The k pairs of canonical projectors $\boldsymbol{\beta}_x \in \mathbb{R}^{q_x \times k}$ and $\boldsymbol{\beta}_y \in \mathbb{R}^{q_y \times k}$.

repeat
Compute $\mathbf{X}_r = \mathbf{X}\mathbf{R}_x$, $\mathbf{Y}_r = \mathbf{Y}\mathbf{R}_y$, and $\mathbf{K}_{XY}$.
Use **Algorithm 4** to obtain $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$,
Compute $\mathbf{X}_l = \mathbf{X}\mathbf{L}_x$, $\mathbf{Y}_l = \mathbf{Y}\mathbf{L}_y$, and $\mathbf{Q}_{XY}$.
Use **Algorithm 4** to obtain $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_y$.
until convergence

3. Experimental results

This section assesses the performance of the proposed methods against existing CCA algorithms, CCA, 2DCCA Lee and Choi (2007), S2DCCA Yan et al. (2012). We present a simulation experiment and two real-world experiments on fMRI data (event-related task and resting-state fMRI). In the simulated experiments, we compare the CCA algorithms in a blind source separation setting. In real fMRI analysis, we use CCA algorithms to extract the most dominantly activated brain areas and use the extracted brain areas to generate spatial maps.

Algorithm 4: Iterative regularized 2DCCA procedure

Input: $\mathbf{X}_r, \mathbf{Y}_r$, and $\mathbf{K}_{XY}$ (or $\mathbf{X}_l$, $\mathbf{Y}_l$, and $\mathbf{Q}_{XY}$).

Output: The k pairs of canonical projectors $\boldsymbol{\alpha}_x \in \mathbb{R}^{p_x \times k}$ and $\boldsymbol{\alpha}_y \in \mathbb{R}^{p_y \times k}$.

for $i = 1, \cdots, r$ **do**

  $SVD(\mathbf{K}_{XY}) = \mathbf{U}\mathbf{D}\mathbf{V}^\top$,

  Initialize $\tilde{\mathbf{u}} = \mathbf{u}_1$ and $\tilde{\mathbf{v}} = \mathbf{v}_1$;

  repeat

  Update the i-th column of $\boldsymbol{\alpha}_x$:

  $\boldsymbol{\alpha}_x(:,i) = \left(\mathbf{X}_r\mathbf{X}_r^\top + \alpha_x\mathbf{X}_r\boldsymbol{\Omega}_x\mathbf{X}_r^\top + \gamma_x\mathbf{I}_{p_x}\right)^{-1}\mathbf{X}_r\mathbf{K}_{XY}\tilde{\mathbf{v}}$

  Update $\tilde{\mathbf{u}} = \frac{\mathbf{X}_r^\top\boldsymbol{\alpha}_x(:,i)}{\|\mathbf{X}_r^\top\boldsymbol{\alpha}_x(:,i)\|_2}$;

  Update the i-th column of $\boldsymbol{\alpha}_y$:

  $\boldsymbol{\alpha}_y(:,i) = \left(\mathbf{Y}_r\mathbf{Y}_r^\top + \alpha_y\mathbf{Y}_r\boldsymbol{\Omega}_y\mathbf{Y}_r^\top + \gamma_y\mathbf{I}_{q_x}\right)^{-1}\mathbf{Y}_r\mathbf{K}_{XY}^\top\tilde{\mathbf{u}}$

  Update $\tilde{\mathbf{v}} = \frac{\mathbf{Y}_r^\top\boldsymbol{\alpha}_y(:,i)}{\|\mathbf{Y}_r^\top\boldsymbol{\alpha}_y(:,i)\|_2}$;

  until convergence

  Update $\mathbf{K}_{XY} : \mathbf{K}_{XY} \leftarrow \mathbf{K}_{XY} - \tilde{\mathbf{u}}^\top\mathbf{K}_{XY}\tilde{\mathbf{v}}\tilde{\mathbf{u}}\tilde{\mathbf{v}}^\top$

end

3.1. Blind source separation

In this section, we create simulated data to compare the performance of our two proposed methods sparse-2DCCA and reg-2DCCA with standard CCA, 2DDCA, S2DCCA in terms of their ability to recover the underlying sources from a pair of linear mixtures of correlated sources. We generated three distinct boxed activation patterns S_1, S_2 and S_3 in a grid of size (21×21) pixels along with three distinct time courses R_1, R_2 and R_3 with $N = 120$ time points. The data matrices $\mathbf{X}$ and $\mathbf{Y}$ were generated from combinations of these distinct spatial and temporal signals. Each activation pattern exhibits an amplitude of 1 at locations: $\{5, \cdots, 14\} \times \{5, \cdots, 14\}$ for S_1, $\{15, \cdots, 19\} \times \{15, \cdots, 19\}$ for S_2, and $\{8, \cdots, 17\} \times \{8, \cdots, 17\}$ for S_3 and zeros elsewhere. The ground truth sources and activation patterns are shown in the Fig. 1. We also gene-

rated three cases of distinct activations patterns as done in Lee et al. (2011); Seghouane and Iqbal (2018), with a) independent spatial activation patterns, b) partial spatial overlap of activation patterns, and c) complete spatial overlap of activation patterns. To create image datasets for 2DCCA, the activation patterns $\mathbf{S} \in \mathbb{R}^{21\times21\times2}$ and sources $\mathbf{R} \in \mathbb{R}^{2\times120}$ were mixed to generate a dataset $\mathbf{X} = \mathbf{SR}$ while $\mathbf{X} \in \mathbb{R}^{21\times21\times120}$. We then corrupted $\mathbf{X}$ with additive white Gaussian noise (AWGN) at different noise levels $(\sigma^2) \in \{0.1, 0.2, 0.3, 0.4\}$. Since CCA requires two datasets, we generated a second dataset $\mathbf{Y}$ of the same size as $\mathbf{X}$ by sliding a 3×3 spatial filter across $\mathbf{X}$ Friman et al. (2002b). When applying standard 1-dimensional CCA, the datasets $\mathbf{X}$ and $\mathbf{Y}$ were reshaped into matrices of size (441×120).

The resulting corrupted datasets were then decomposed using standard CCA, 2DCCA, S2DCCA, and our proposed methods sparse-2DCCA and reg-2DCCA to recover the underlying sources (i.e., $\mathbf{R}$s and $\mathbf{S}$). All 2DCCA algorithms were iterated 20 times to estimate the left and right canonical projectors. The number of canonical projectors to be learned were set to $k = 9$ for each 2D-CCA algorithms. The convergence for each 2D-CCA algorithm was calculated using $\left\|\boldsymbol{\beta}_{x_i} - \boldsymbol{\beta}_{x_{(i-1)}}\right\|_2 < \epsilon_x$ and $\left\|\boldsymbol{\beta}_{y_i} - \boldsymbol{\beta}_{y_{(i-1)}}\right\|_2 < \epsilon_y$ for i-th iteration, where ϵ_x and ϵ_y are the convergence tolerance. The regularization parameters were set according to the highest correlation between ground truth and recovered sources/activation patterns. These parameters were $\lambda_u = \lambda_v = 5$ for S2DCCA, $\lambda_x = \lambda_y = 4$ for the proposed sparse-2DCCA and $\gamma_x = 6, \gamma_y = 2$ for the proposed reg-2DCCA.

The recovered activation patterns and sources were then correlated with the ground truth to compute average correlations over 25 trials. We calculated the average of these correlations over all noise level σ^2. For each case, the average correlations of the recovered sources and the activation patterns are presented in Tables 1 and 2, respectively. In these tables, the best results are made bold. It can be observed that the data decomposition performed by the proposed algorithms were closer to the underlying generating sources. In most noise levels and cases, the proposed algorithms exhibit higher correlation with the

ground truths. For all cases the reg-2DCCA algorithm outperforms all other algorithms. To illustrate each case visually, we have presented the most correlated sources/activation patterns in Fig. 2. It can be seen that, for case (c), CCA and 2DCCA exhibit poor performance, where the recovered sources contains a linear combination of both R_2 and R_3, while S2DCCA was able separate both activation patterns, however, the recovered sources suffer from noise and exhibit lower correlations. For case (c), the proposed algorithms show improvement in separating both dynamics (R_2, R_3) and their corresponding overlapping activation patterns, which can also be observed from the results presented in Table 1 and 2. We also calculated the running time of each algorithm over a single trial, presented in Table 3.

In this section, we have demonstrated the efficiency of the proposed algorithms in competition with CCA, 2DCCA, S2DCCA Yan et al. (2012) algorithms. In the next sections, we evaluate the proposed algorithms on real-world data, where we establish two examples on experimental fMRI data from an event-related right finger-tapping task and resting-state experiments.

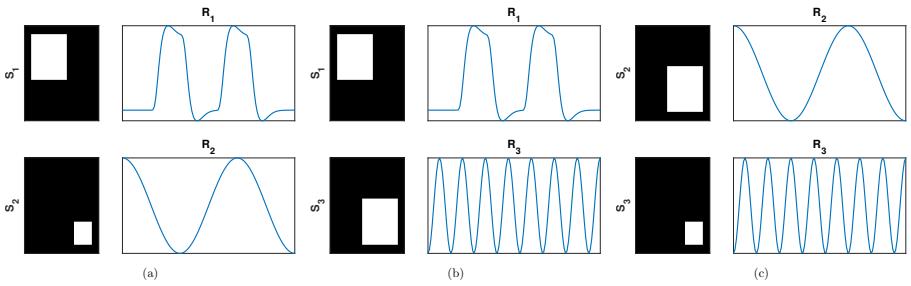

Figure 1: Ground truth time courses and activation maps corresponding to a) independent spatial activation patterns, b) partial spatial overlap, and complete spatial overlap of activation patterns.

3.2. Event-related task fMRI analysis

We evaluate our proposed 2DCCA methods using fMRI data from an event-related right finger tapping task (the dataset has been used in Seghouane and

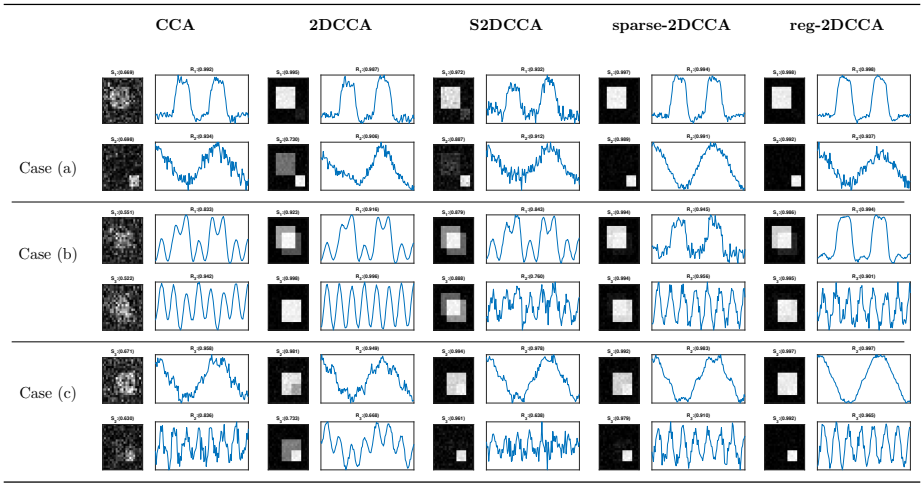

Figure 2: This illustrates the simulation results of three different cases shown in Fig. 1. The recovered time courses and activation patterns are shown for (a) independent spatial activation patterns, (b) partial spatial overlap of activation patterns, and (c) complete spatial overlap of activation patterns. Each column exhibit the results corresponding to different algorithms, whereas the correlation coefficients with respect to original activation pattern/sources pair are printed in parenthesis.

Iqbal (2017) and was presented first in Lee et al. (2011)). The BOLD/echo-planar imaging (EPI) sequence was acquired using a 3.0T fMRI scanner system at ISOL, Republic of Korea. This scanner system acquired 30 contiguous slices having matrix size of (64×64) with voxel size of 3.4 mm $\times$ 3.4 mm $\times$ 4 mm, where slice thickness was 4 mm and flip angle set to 180^o. The total recording time of the dataset was 650 s with TR/TE = 2000/35 ms. After 30 s of resting time at the beginning of the recording, alternating task and rest periods with a 14 s window were repeated 40 times with an additional 30 s period of rest at the end. The interstimulus interval (ISI) ranged between 4 and 20 s with an average ISI period of 12 s.

Prior to analysis, the dataset was processed through a preprocessing pipeline performed in Matlab using the SPM12 package Friston et al. (2009).This preprocessing pipeline involves, 1) spatial alignment, 2) normalization, 3) spatial smoothing, 4) masking and 5) temporal smoothing. In spatial alignment, all

Table 1: Average correlation coefficients of **recovered sources** in competition with the originals for case (a), (b), and (c) over 25 trials using CCA, 2DCCA, S2DCCA, the proposed sparse-2DCCA, and the proposed reg-2DCCA.

		CCA				**2DCCA**				**S2DCCA**				**sparse-2DCCA**				**reg-2DCCA**			
Var (σ^2)		0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
Case (a)	R1	0.97	0.98	0.94	0.88	0.99	0.98	0.97	0.82	0.97	0.93	0.83	0.76	0.82	0.99	0.98	0.79	0.99	0.99	0.98	0.98
	R2	0.96	0.91	0.71	0.67	0.86	0.83	0.75	0.68	0.99	0.91	0.76	0.48	0.95	0.99	0.98	0.74	0.95	0.85	0.84	0.65
	avg	**0.97**	0.94	0.82	0.77	0.92	0.91	0.86	0.75	0.98	0.92	0.79	0.62	0.88	**0.99**	**0.98**	0.77	**0.97**	0.92	0.91	**0.81**
Case (b)	R1	0.91	0.82	0.86	0.82	0.86	0.75	0.70	0.73	0.80	0.76	0.72	0.77	0.84	0.86	0.72	0.69	1.00	0.96	0.97	0.78
	R3	0.83	0.80	0.88	0.81	0.77	0.89	0.77	0.80	0.65	0.70	0.70	0.62	0.84	0.83	0.72	0.71	0.88	0.85	0.81	0.82
	avg	0.87	0.81	0.87	0.82	0.82	0.82	0.73	0.77	0.73	0.73	0.71	0.69	0.84	0.84	0.72	0.70	**0.94**	**0.90**	**0.89**	**0.80**
Case (c)	R2	0.99	0.97	0.92	0.86	0.93	0.95	0.92	0.91	0.98	0.97	0.93	0.90	0.98	0.98	0.98	0.96	0.89	0.97	0.95	0.87
	R3	0.94	0.83	0.64	0.60	0.85	0.63	0.59	0.33	0.62	0.43	0.23	0.18	0.97	0.89	0.80	0.53	0.90	0.95	0.57	0.57
	avg	0.97	0.90	0.78	0.73	0.89	0.79	0.76	0.62	0.80	0.70	0.58	0.54	**0.98**	**0.93**	**0.89**	**0.74**	0.90	**0.96**	0.76	0.72

Table 2: Average correlation coefficients of **recovered activation patterns** in competition with the originals for case (a), (b), and (c) over 25 trials using CCA, 2DCCA, S2DCCA, the proposed sparse-2DCCA, and the proposed reg-2DCCA.

		CCA				**2DCCA**				**S2DCCA**				**sparse-2DCCA**				**reg-2DCCA**			
Var (σ^2)		0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
Case (a)	S1	0.67	0.68	0.68	0.68	1.00	0.99	0.99	0.89	0.99	0.96	0.89	0.79	0.96	1.00	0.99	0.96	1.00	0.99	0.99	0.97
	S2	0.70	0.70	0.69	0.67	0.69	0.71	0.69	0.43	0.97	0.86	0.66	0.39	0.88	0.98	0.97	0.87	1.00	0.95	0.92	0.83
	avg	0.68	0.69	0.68	0.67	0.84	0.85	0.84	0.66	0.98	0.91	0.77	0.59	0.92	**0.99**	**0.98**	**0.91**	**1.00**	**0.97**	**0.95**	**0.90**
Case (b)	S1	0.53	0.55	0.53	0.52	0.93	0.90	0.83	0.92	0.91	0.85	0.82	0.83	0.92	0.92	0.81	0.82	1.00	0.96	0.96	0.91
	S3	0.56	0.56	0.54	0.56	0.84	0.91	0.85	0.85	0.89	0.83	0.80	0.75	0.94	0.90	0.83	0.81	1.00	0.99	0.95	0.94
	avg	0.54	0.55	0.54	0.54	0.89	0.91	0.84	0.88	0.90	0.84	0.81	0.79	**0.93**	0.91	0.82	0.82	**1.00**	**0.97**	**0.95**	**0.93**
Case (c)	S2	0.67	0.67	0.67	0.65	0.99	0.99	0.97	0.96	1.00	0.99	0.97	0.94	1.00	0.99	0.99	0.97	0.99	0.99	0.98	0.96
	S3	0.66	0.65	0.62	0.58	0.94	0.72	0.86	0.62	0.95	0.79	0.54	0.50	0.97	0.96	0.78	0.71	0.91	0.99	0.82	0.69
	avg	0.66	0.66	0.64	0.62	0.97	0.85	**0.92**	0.79	0.97	0.89	0.76	0.72	**0.98**	**0.98**	0.88	**0.84**	0.95	**0.99**	0.90	**0.82**

fMRI images were realigned to correct for the head motion that effects the underlying signal intensity during the course of an experiment. After realignment, all fMRI images spatially normalized to the standard Talairach template, where each voxel was resampled to (2mm $\times$ 2mm $\times$ 2mm). These images were then spatially smoothed with an 8mm $\times$ 8mm $\times$ 8mm full-width at half-maximum (FWHM) Gaussian kernel. In this dataset, the first 15 scans were acquired in dummy cycles and therefore, were discarded. After discarding the first 15 scans, we used remaining the 310 scans for analysis.

All images were then collected in a 4D dataset and were used for further processing and analysis. We applied masking on the dataset to get rid of the data outside the brain scalp, where voxels exceeding a masking threshold were retained. After masking number of voxels in the dataset was reduced by a factor of 8. The dataset was also detrended by removing low frequency drifts using a

Table 3: Running time(RT) (in seconds) of each algorithm over a single trial in simulated data.

CCA	**2DCCA**	**S2DCCA**	**sparse-2DCCA**	**reg-2DCCA**
0.963	0.0156	0.0297	0.0382	0.0389

discrete consine transform (DCT) basis set at a cut-off frequency of (1/128)Hz. The detrended data were then smoothed temporally with 1.5s FWHM Gaussian kernel to get rid of high frequency noise.

To apply 2D-CCA based algorithms, fMRI dataset $\mathbf{X}$ was reshaped in a 3-D matrix of size $(p_x \times q_x \times N)$, where p_x corresponds to the number of voxels in a slice, q_x represents the number of time points corresponding to a voxel and N is the number of 2D image slices in the dataset, respectively. A second dataset $\mathbf{Y}$ was created in a manner similar to that described in Section V-A. The datasets $\mathbf{X}$ and $\mathbf{Y}$ were centered before applying all algorithms.

To conduct fMRI analysis using standard 1-dimensional CCA, the fMRI volume was reshaped in a 2-D matrix of size $p \times N$, where p corresponds to the number of voxels, and N corresponds to the number of time points (i.e., $N = 310$). The projection directions computed using CCA were used to construct activation maps.

All 2D-CCA algorithms were then used to learn $k = 15$ left and right projection directions. The number of iterations for each 2D-CCA algorithm was set to 20. We tried different values of sparsity parameters, the optimal sparsity level was chosen based on the best results in terms correlation of the modelled hemodynamic response function (MHRF) with event-related task paradigm. For S2DCCA Yan et al. (2012), both λ_u and λ_v were set a to value of 0.01, for the proposed sparse-2DCCA, $\lambda_x = \lambda_y = 4$, and for the proposed reg-2DCCA, $\gamma_x = \gamma_y = 4$ in a range between 0 and 20. We then selected the most correlated projected data vector with respect to the MHRF as the recovered voxel time series. The recovered voxel series using, CCA, 2DCCA, S2DCCA, the proposed sparse-2DCCA and the proposed sparse-2DCCA are presented in the Fig. 3

(temporal correlations with MHRF are given in parenthesis). As shown in the Fig. 3, the highest temporal correlation of (0.629) is achieved by the proposed sparse-2DCCA algorithm. In order to inspect activations, we used these voxel time series to generate activations maps, Z-normalized and thresholded at $Z > 3$ corresponding to p-value less than $1e^{-3}$. The most significant activation maps for event-related data are presented in the Fig. 4. This figure illustrates that all CCA algorithms have characterized the neural activity in the task-related area (i.e., motor cortex). However, it can be observed that the proposed algorithms (sparse-2DCCA and reg-2DCCA) exhibit less spurious activations with increased specificity of activated voxels compared to CCA, 2DCCA and S2DCCA.

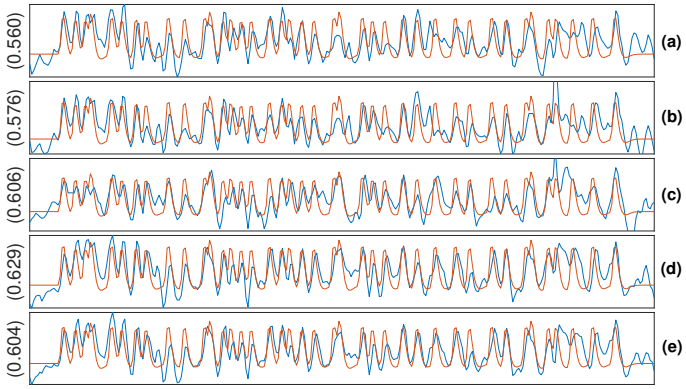

Figure 3: It illustrates the most correlated projected data vectors with respect to canonical HRF (in red) estimated by (a) CCA, (b) 2DCCA, (c) S2DCCA Yan et al. (2012), (d) the proposed sparse-2DCCA, and (e) the proposed reg-2DCCA. The values in the parenthesis corresponds to the correlation coefficients.

3.3. Resting-state fMRI analysis

Single subject (id 100307) resting-state fMRI (rsfMRI) data was obtained from the Human Connectome Project (HCP) Q1 release Barch et al. (2013). The dataset was acquired with the following parameters: TR/TE = 0.72 s/33.1 ms, slices = 72, field-of-view (FOV) = 220 mm, matrix size = 90 × 104, BW = 2290 Hz/Px, flip angle = 52^o, and in-place FOV = 208 × 180 mm isotropic voxels. The data was preprocessed according to preprocessing steps detailed in

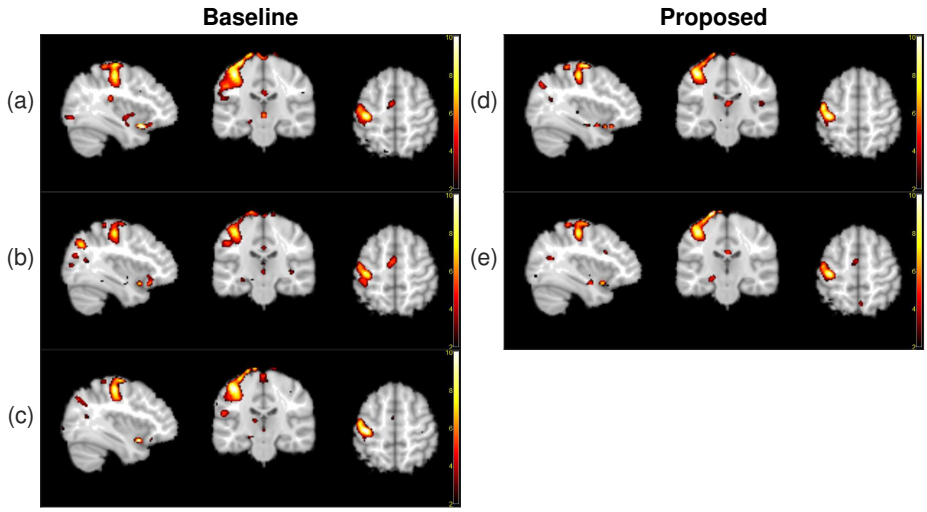

Figure 4: Most significant activation maps obtained using (a) CCA, (b) 2DCCA, (c) S2DCCA Yan et al. (2012), (d) the proposed sparse-2DCCA, and the proposed reg-2DCCA, Z-normalized and thresholded at a p-value less $1e^{-3}$.

Glasser et al. (2013). The dataset contained 1200 scans acquired in a duration of $14:33$ (min:sec). From 1200 scans, we selected the first 430 scans. The first 15 scans were discarded and the remaining 415 scans were employed for analysis. All scans were smoothed spatially using a $6 \times 6 \times 6$ mm^3 FWHM Gaussian kernel. The dataset was masked to remove any data outside the brain scalp, reducing the data by a factor of 4. We used a DCT basis set with cut off frequency of $1/170$ Hz to eliminate low frequency trends. The data was then temporally smoothed to remove high frequency noise using a 1.7 s Gaussian FWHM kernel. The format of the datasets (i.e., $\mathbf{X}$ and $\mathbf{Y}$) to be used in CCA as well as 2DCCA algorithms are as described in Section V-B.

For the analysis of rsfMRI data, we use the seed voxel-base correlation analysis Beckmann et al. (2005); Hale et al. (2010), which is a commonly used method restingCole et al. (2010); Lee et al. (2013-10-01); Van Den Heuvel and Hulshoff Pol (2010). In resting-state fMRI, it is assumed that functionally connected brain regions exhibit similar temporal fluctuations. If these brain regions correspond to a functional connectivity network (FCN) Beckmann et al. (2005),

then a voxel time series associated with the particular region can be obtained from that FCN and used as a reference time series to observe its connectivity with other FCNs Beckmann et al. (2005). Some examples of FCN are, the dorsal attention network (DAN), salience network, default mode network (DMN), etc. We extracted the seed-voxels from the DMN, for which the MNI coordinates Calhoun and Adali (2012); Leech et al. (2011) are presented in Table 4.

The 2DCCA based algorithms were used to learn $k = 20$ projection directions (i.e., $\boldsymbol{\alpha}_x$, $\boldsymbol{\alpha}_y$, $\boldsymbol{\beta}_x$, and $\boldsymbol{\beta}_y$), where the number of iterations for all algorithms were set to 20. The parameters for S2DCCA, the proposed sparse-2DCCA, and the proposed reg-2DCCA were set to $\lambda_u = \lambda_v = 0.1$, $\lambda_x = \lambda_y = 1$, and $\gamma_x = \gamma_y = 2$, respectively.

In order to analyze the datasets, the mean voxel time-series were extracted using a $6 \times 6 \times 6$ mm^3 cube centered at seed locations presented in Table 4. The mean voxel time series corresponding to different seed locations were then correlated with the estimated temporal dynamics (i.e., data projection onto CCA directions) to compare the performance of CCA algorithms. The recovered correlation coefficients are presented in Table 4. The proposed algorithms outperform CCA, 2DCCA, and S2DCCA. For each algorithm, the recovered time series with respect to mean voxel time series corresponding to the precuneus cortex are shown in Fig. 5. The activation maps corresponding to the DMN regions are shown in the Fig. 6, it can be seen that the activations recovered by the proposed algorithms show increased specificity of activated voxels, specifically in the precuneus cortex region where activations are tightly localized and contain distinct peaks.

4. Conclusion

We have proposed two new variants of 2DCCA applied on fMRI data. The proposed algorithms formulate a penalized rank-1 matrix approximation problem by incorporating orthogonal projectors in the 2DCCA objective function. Specifically, instead of using the cross-product of two multidimensional data ma-

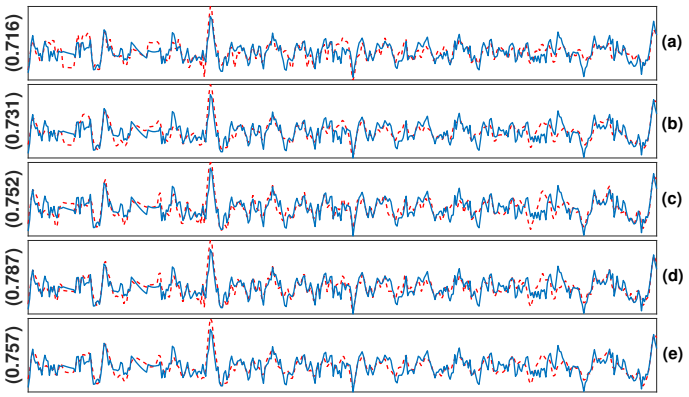

Figure 5: Illustrates the most correlated projected data vectors (in red) with respect to the mean voxel time series (in blue) extracted with a $6\times6\times6$ mm^3 cube centered at the seed-voxel corresponding to precuneus cortex obtained using (a) CCA, (b) 2DCCA, (c) S2DCCA Yan et al. (2012), (d) the proposed sparse-2DCCA, and (e) the proposed reg-2DCCA. The values in the parenthesis (on the left side) corresponds to the correlation coefficients.

Table 4: Correlation coefficients of the most correlated projected data vectors with respect to the seed voxel time series at the selected MNI coordinates Calhoun and Adali (2012) extracted by using a $6 \times 6 \times 6$ mm^3 cube.

	MNI coordinates			**CCA**	**2DCCA**	**S2DCCA**	**sparse-2DCCA**	**reg-2DCCA**
Ventral medial prefrontal cortex	6	70	14	0.500	0.590	0.639	**0.677**	**0.665**
Precuneus cortex	-8	-60	14	0.716	0.731	0.752	**0.787**	**0.757**
Superior frontal gyrus	8	50	38	0.587	**0.716**	**0.714**	0.691	0.695
Cingulate Gyrus	5	45	10	0.570	0.542	0.585	**0.625**	0.578
Mean				0.593	0.645	0.672	**0.695**	**0.674**

trices, we have proposed to use the product of orthogonal projectors onto the space spanned by these data matrices. The limitations of 2DCCA was tackled by learning the sparse canonical projectors to improve their interpretability and computational time. The performance of the proposed methods was evaluated on both simulated and real fMRI datasets. Experimental results show the improved performance of the proposed algorithms in comparison to standard CCA, 2DCCA, and S2DDCA algorithms.

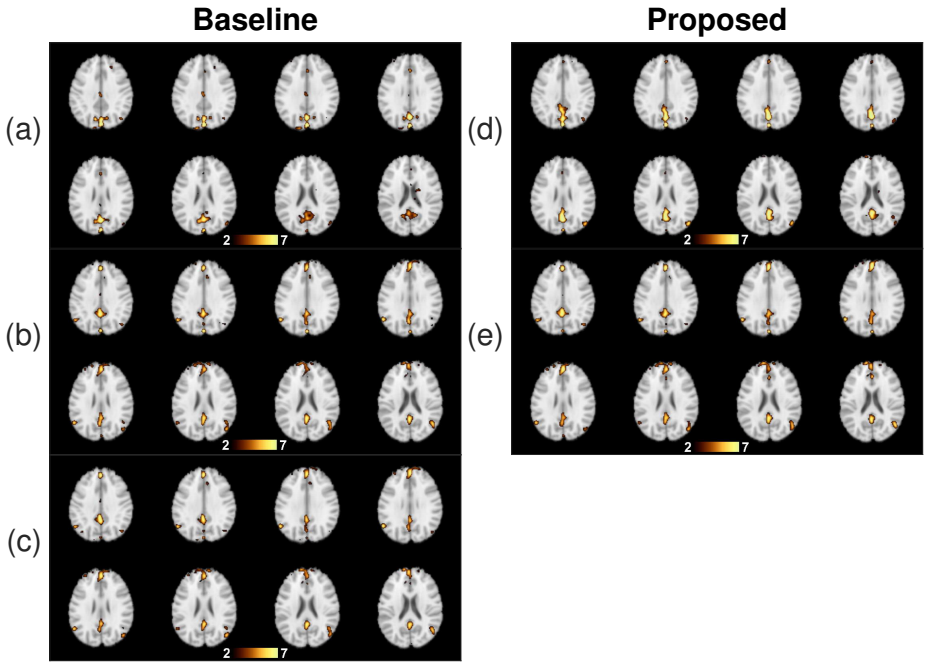

Figure 6: Activation maps exhibiting DMN regions extracted using (a) CCA, (b) 2DCCA, (c) 2DSCCA Yan et al. (2012), (d) the proposed sparse-2DCCA, and (e) the proposed reg-2DCCA, Z-normalized and thresholded at a p-value of less than $1e^{-3}$.

Acknowledgment

This work was supported in part by the Australian Research Council through Grant FT. 130101394.

Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

References

Ahn M, Shen H, Lin W, Zhu H. A sparse reduced rank framework for group analysis of functional neuroimaging data. Statistica Sinica 2015;25:295–312.

Aissa-El-Bey A, Seghouane A. Sparse and smooth canonical correlation analysis through rank-1 matrix approximation. EURASIP Journal on Advances in Signal Processing 2017;2017:25.

Andersen A, Gash D, Avion M. Principal component analysis of the dynamic response measured by fMRI: a generalized linear system framework. Magnetic Resonance Imaging 1999;17:795–815.

Avants BB, Libon DJ, Rascovsky et al. K. Sparse canonical correlation analysis relates network-level atrophy to multivariate cognitive measures in a neurodegenerative population. NeuroImage 2014;84:698–711.

Barch DM, et al. Function in the human connectome: Task-fmri and individual differences in behavior. NeuroImage 2013;80:169 –89. doi:`https://doi.org/10.1016/j.neuroimage.2013.05.033`.

Beckmann CF, Deluca M, Devlin et al. JT. Investigations into resting-state connectivity using independent component analysis. Philosophical Transactions of the Royal Society B: Biological Sciences 2005;360(1457):1001–13. doi:`10.1098/rstb.2005.1634`.

Calhoun VD, Adali T. Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery. IEEE Reviews in Biomedical Engineering 2012;5:60–73.

Chu D, Liao LZ, Ng MK, Zhang X. Sparse canonical correlation analysis: New formulation and algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence 2013;35:3050–65.

Ciuciu P, Poline J, Marrelec G, Idier J, Pallier C, Benali H. Unsupervised robust nonparametric estimation of the hemodynamic response function for

any fMRI experiment. IEEE Transactions on Medical Imaging 2003;22:1235–51.

Cole DM, Smith SM, Beckmann CF. Advances and Pitfalls in the Analysis and Interpretation of Resting-State FMRI Data. Frontiers in Systems Neuroscience 2010;4.

Correa NM, Adali T, Li et al. YO. Canonical correlation analysis for data fusion and group inferences. IEEE Signal Processing Magazine 2010;27:39–50.

Desai N, Seghouane A, Palaniswami M. Algorithms for two dimensional multi set canonical correlation analysis. Pattern Recognition Letters 2018;111:101–8.

Fang J, et al. Joint sparse canonical correlation analysis for detecting differential imaging genetics modules. Bioinformatics 2016;32:3480–8.

Friman O, Borga M, Lundberg et al. P. Detection of neural activity in fMRI using maximum correlation modeling. NeuroImage 2002a;15:386–95.

Friman O, Borga M, Lundberg et al. P. Exploratory fMRI analysis by autocorrelation maximization. NeuroImage 2002b;16:454–64.

Friman O, Borga M, Lundberg et al. P. Adaptive analysis of fMRI data. NeuroImage 2003;19:837–45.

Friston KJ, Holmes AP, Ashburner et al. J. SPM12. 2009. URL: `http://www.fil.ion.ucl.ac.uk/spm`.

Gao X, Sun Q, Yang J. MRCCA: A novel CCA based method and its application in feature extraction and fusion for matrix data. Applied Soft Computing 2018;62:45–56.

Glasser M, et al. The minimal preprocessing pipelines for the Human Connectome Project. NeuroImage 2013;80:105–24.

Hale JR, Brookes MJ, Hall EL, Zumer JM, Stevenson CM, Francis ST, Morris PG. Comparison of functional connectivity in default mode and sensorimotor networks at 3 and 7t. Magnetic Resonance Materials in Physics, Biology and Medicine 2010;23:339–49.

Hardoon DR, Shawe-Taylor J. Sparse canonical correlation analysis. Machine Learning 2011;83:331–53.

Hotelling H. Relations between two sets of variates. Biometrika 1936;28:321–77.

Khalid MU, Seghouane AK. Improving functional connectivity detection in fMRI by combining sparse dictionary learning and canonical correlation analysis. In: 2013 IEEE 10th International Symposium on Biomedical Imaging. 2013. p. 286–9.

Khalid MU, Shah A, Seghouane A. Adaptive 2DCCA Based Approach for Improving Spatial Specificity of Activation Detection in Functional MRI. In: 2012 International Conference on Digital Image Computing Techniques and Applications (DICTA). 2012. p. 1–6.

Kukharev G, Kamenskaya E. Application of two-dimensional canonical correlation analysis for face image processing and recognition. Pattern Recognition and Image Analysis 2010;20:210–9.

Le Floch E, et al. Significant correlation between a set of genetic polymorphisms and a functional brain network revealed by feature selection and sparse Partial Least Squares. NeuroImage 2012;63:11–24.

Lee K, Tak SK, Yee JC. A data driven sparse GLM for fMRI analysis using sparse dictionary learning and MDL criterion. IEEE Transactions on Medical Imaging 2011;30:1176–089.

Lee MH, Smyser CD, Shimony JS. Resting-state fMRI: A review of methods and clinical applications. American Journal of Neuroradiology 2013-10-01;34:1866–72.

Lee SH, Choi S. Two-dimensional canonical correlation analysis. IEEE Signal Processing Letters 2007;14:735–8.

Leech R, Kamourieh S, Beckmann et al. CF. Fractionating the default mode network: Distinct contributions of the ventral and dorsal posterior cingulate cortex to cognitive control. Journal of Neuroscience 2011;31(9):3217–24.

Lin D, Calhoun VD, Wang YP. Correspondence between fMRI and SNP data by group sparse canonical correlation analysis. Medical Image Analysis 2014;18:891–902.

Mckeown MJ, Makeig S, Brown et al. GG. Analysis of fMRI data by blind separation into independent spatial components. Human Brain Mapping 1998;6:160–88.

Mohammadi-Nejad AR, Hossein-Zadeh GA, Soltanian-Zadeh H. Structured and sparse canonical correlation analysis as a brain-wide multi-modal data fusion approach. IEEE Transactions on Medical Imaging 2017;36:1438–48.

Nandy R, Cordes D. Improving the spatial specificity of canonical correlation analysis in fMRI. Magnetic Resonance in Medicine 2004;52:947–52.

Ogawa T, Haseyama M. 2D semi-supervised CCA-based inpainting including new priority estimation. In: 2014 IEEE International Conference on Image Processing (ICIP). 2014. p. 1837–41.

Ramsay J, Silverman BW. Functional Data Analysis. 2nd ed. Springer Series in Statistics. New York: Springer-Verlag, 2005.

Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, Lander ES, Mitzenmacher M, Sabeti PC. Detecting novel associations in large data sets. Science 2011;334:1518–24.

Seghouane A, Iqbal A. Sequential Dictionary Learning From Correlated Data: Application to fMRI Data Analysis. IEEE Transactions on Image Processing 2017;26(6):3002–15. doi:10.1109/TIP.2017.2686014.

Seghouane A, Iqbal A. Consistent adaptive sequential dictionary learning. Signal Processing 2018;153:300–10.

Seghouane AK, Iqbal A, Desai N. BSmCCA: A block sparse multiple-set canonical correlation analysis algorithm for multi-subject fMRI data sets. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2017. p. 6324–8.

Sun L, Ji S, Ye J. Canonical correlation analysis for multilabel classification: A least-squares formulation, extensions, and analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 2011;33:194–200.

Sun N, Ji ZH, Zou C, Zhao L. Two-dimensional canonical correlation analysis and its application in small sample size face recognition. Neural Computing and Applications 2010;19:377–82.

Tang S, Xiao L, Huang W, Liu P, Wu H. Pan-sharpening using 2D CCA. Remote sensing letters 2015;.

Van Den Heuvel MP, Hulshoff Pol HE. Exploring the brain network: A review on resting-state fMRI functional connectivity. European Neuropsychopharmacology 2010;20:519–34.

Wang H. Local two-dimensional canonical correlation analysis. IEEE Signal Processing Letters 2010;17:921–4.

Witten DM, Tibshirani R, Hastie T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. Biostatistics 2009;10:515–34.

Yan J, Zheng W, Zhou X, Zhao Z. Sparse 2-D Canonical Correlation Analysis via Low Rank Matrix Approximation for Feature Extraction. IEEE Signal Processing Letters 2012;19:51–4.

Yang J, Zhang D, Frangi AF, Yang J. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 2004;26(1):131–7.

Zhuang X, Yang Z, Curran et al. T. A family of locally constrained CCA models for detecting activation patterns in fMRI. NeuroImage 2017;149:63–84.