

RESEARCH

Open Access

# Sparse and smooth canonical correlation analysis through rank-1 matrix approximation

Abdeldjalil Aïssa-El-Bey<sup>1\*</sup>  and Abd-Krim Seghouane<sup>2</sup>

## Abstract

Canonical correlation analysis (CCA) is a well-known technique used to characterize the relationship between two sets of multidimensional variables by finding linear combinations of variables with maximal correlation. Sparse CCA and smooth or regularized CCA are two widely used variants of CCA because of the improved interpretability of the former and the better performance of the later. So far, the cross-matrix product of the two sets of multidimensional variables has been widely used for the derivation of these variants. In this paper, two new algorithms for sparse CCA and smooth CCA are proposed. These algorithms differ from the existing ones in their derivation which is based on penalized rank-1 matrix approximation and the orthogonal projectors onto the space spanned by the two sets of multidimensional variables instead of the simple cross-matrix product. The performance and effectiveness of the proposed algorithms are tested on simulated experiments. On these results, it can be observed that they outperform the state of the art sparse CCA algorithms.

**Keywords:** Canonical correlation analysis, Sparse representation, Rank-1 matrix approximation

## 1 Introduction

Canonical correlation analysis (CCA) [1] is a multivariate analysis method, the aim of which is to identify and quantify the association between two sets of variables. The two sets of variables can be associated with a pair of linear transforms (projectors) such that the correlation between the projections of the variables in lower dimensional space through these linear transforms are mutually maximized. The pair of canonical projectors are easily obtained by solving a simple generalized eigenvalue decomposition problem, which only involves the covariance and cross-covariance matrices of the considered random vectors. CCA has been widely applied in many important fields, for instance, facial expression recognition [2, 3], detection of neural activity in functional magnetic resonance imaging (fMRI) [4, 5], machine learning [6, 7] and blind source separation [8, 9].

In the context of high-dimensional data, there is usually a large portion of features that is not informative in data analysis. When the canonical variables involve all features in the original space, the canonical projectors are, in

general, not sparse. Therefore, it is not easy to interpret canonical variables in such high-dimensional data analysis. These problems may be tackled by selecting sparse subsets of variables, i.e. obtaining sparse canonical projectors in the linear combinations of variables of each data set [7, 10–12]. For example, in [11], the authors propose a new criterion for sparse CCA and applied a penalized matrix decomposition approach to solve the sparse CCA problem, and in [10], the presented sparse CCA approach computes the canonical projectors from primal and dual representations.

In this paper, we adopt an alternative formulation of the CCA problem which is based on rank-1 matrix approximation of the orthogonal projectors of data sets [13]. Based on this new formulation of the CCA problem, we developed a new sparse CCA based on penalized rank-1 matrix approximation which aims to overcome the drawback of CCA in the context of high-dimensional data and improved interpretability. The proposed sparse CCA seeks to obtain iteratively a sparse pair of canonical projectors by solving a penalized rank-1 matrix approximation via a sparse coding method. Also, we present in this paper a smoothed version of the CCA problem based on rank-1 matrix approximation where we impose some smoothness on the projections of the variables in order to

\*Correspondence: [abdeldjalil.aissaelbey@imt-atlantique.fr](mailto:abdeldjalil.aissaelbey@imt-atlantique.fr)

<sup>1</sup>IMT Atlantique, UMR CNRS 6285 Lab-STICC, Université Bretagne Loire, Brest 29238, France

Full list of author information is available at the end of the article

avoid abrupt or sudden variations. These proposed algorithms differ from the existing ones in their derivation which is based on penalized rank-1 matrix approximation and the orthogonal projectors onto the space spanned by the two sets of multidimensional variables instead of the simple cross-matrix product [7, 10–12].

The rest of the paper is organized as follows: In Section 2, we give a brief review of the CCA problem. In Section 3, we present a formulation of CCA using a rank-1 matrix approximation of the orthogonal projectors of data sets and derive the smoothed solution. In Section 4, we introduce our new sparse CCA algorithm. In Section 5, we present some simulation results to demonstrate the effectiveness of the proposed method compared to state of the art CCA algorithms. Finally, Section 6 concludes the paper.

Henceforth, bold lower cases denote real-valued vectors and bold upper cases denote real-valued matrices. The transpose of a given matrix  $\mathbf{A}$  is denoted by  $\mathbf{A}^T$ . All vectors will be column vectors unless transposed. Throughout the paper,  $\mathbf{I}_n$  stands for  $n \times n$  identity matrix,  $\mathbf{0}$  stands for the null vector and  $\mathbf{1}_n$  is the (column) vector of  $\mathbb{R}^n$  with one entries only. For a vector  $\mathbf{x}$ , the notation  $x_i$  will stand for the  $i^{\text{th}}$  component of  $\mathbf{x}$ . As usual, for any integer  $m$ ,  $\llbracket 1, m \rrbracket$  stands for  $\{1, 2, \dots, m\}$ .

## 2 Canonical correlation analysis

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

$$\rho(\mathbf{w}_x, \mathbf{w}_y) = \frac{\text{cov}(\mathbf{w}_x^T \mathbf{x}, \mathbf{w}_y^T \mathbf{y})}{\sqrt{\text{var}(\mathbf{w}_x^T \mathbf{x}) \text{var}(\mathbf{w}_y^T \mathbf{y})}}. \quad (1)$$

Then, the objective function  $\rho$  can be rewritten as:

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

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

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

Incorporating these two constraints, the Lagrangian is given by:

$$\begin{aligned} \mathcal{J}(\lambda_x, \lambda_y, \mathbf{w}_x, \mathbf{w}_y) = & \mathbf{w}_x^T \mathbf{C}_{xy} \mathbf{w}_y - \lambda_x (\mathbf{w}_x^T \mathbf{C}_{xx} \mathbf{w}_x - 1) \\ & - \lambda_y (\mathbf{w}_y^T \mathbf{C}_{yy} \mathbf{w}_y - 1). \end{aligned} \quad (3)$$

Taking derivatives with respect to  $\mathbf{w}_x$  and  $\mathbf{w}_y$ , we obtain

$$\frac{\partial \mathcal{J}}{\partial \mathbf{w}_x} = \mathbf{C}_{xy} \mathbf{w}_y - 2\lambda_x \mathbf{C}_{xx} \mathbf{w}_x = 0 \quad (4)$$

$$\frac{\partial \mathcal{J}}{\partial \mathbf{w}_y} = \mathbf{C}_{xy}^T \mathbf{w}_x - 2\lambda_y \mathbf{C}_{yy} \mathbf{w}_y = 0. \quad (5)$$

These equations lead to the following generalized eigenvalue problem

$$\mathbf{C}_{xy} \mathbf{w}_y = \lambda \mathbf{C}_{xx} \mathbf{w}_x \quad (6)$$

$$\mathbf{C}_{xy}^T \mathbf{w}_x = \lambda \mathbf{C}_{yy} \mathbf{w}_y, \quad (7)$$

where  $\lambda = 2\lambda_x = 2\lambda_y$ . One way to solve this problem is as proposed in [6] by assuming  $\mathbf{C}_{yy}$  is invertible; we can write

$$\mathbf{w}_y = \frac{1}{\lambda} \mathbf{C}_{yy}^{-1} \mathbf{C}_{xy}^T \mathbf{w}_x, \quad (8)$$

and so substituting in Eq. (6) and assuming  $\mathbf{C}_{xx}$  is invertible gives

$$\mathbf{C}_{xx}^{-1} \mathbf{C}_{xy} \mathbf{C}_{yy}^{-1} \mathbf{C}_{xy}^T \mathbf{w}_x = \lambda^2 \mathbf{w}_x. \quad (9)$$

It has been shown in [6] that we can choose the associated eigenvectors corresponding to the top eigenvalues of the generalized eigenvalue problem in (9) and then use (8) for find the corresponding  $\mathbf{w}_y$ . A number of existing methods for sparse and smooth CCA have used the description provided above of CCA and focused on the use of the cross matrix  $\mathbf{C}_{xy}$  for the derivation of new CCA variant algorithms [7, 10–12]. For the derivation of the proposed CCA variants, we adopt an alternative description of CCA which is based on the orthogonal projectors onto the space spanned by the two sets of multidimensional variables [13].

## 3 Canonical correlation analysis based on rank-1 matrix approximation

In practice, the covariance matrices  $\mathbf{C}_{xx}$ ,  $\mathbf{C}_{yy}$  and  $\mathbf{C}_{xy}$  are usually not available. Instead, the estimated covariance matrices are constructed based on given sample data set. Let  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{d_x \times N}$  and  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N] \in \mathbb{R}^{d_y \times N}$  be the two sets of instances of  $\mathbf{x}$  and  $\mathbf{y}$ , respectively. Without loss of generality, we can assume both

$\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  and  $\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$  have zero mean, i.e.,  $\sum_{i=1}^N \mathbf{x}_i = \mathbf{0}$  and  $\sum_{i=1}^N \mathbf{y}_i = \mathbf{0}$ . Or, we can center the data sets such that  $\mathbf{x}_i \leftarrow \mathbf{x}_i - \boldsymbol{\mu}_x$  and  $\mathbf{y}_i \leftarrow \mathbf{y}_i - \boldsymbol{\mu}_y$  for all  $i \in \llbracket 1, N \rrbracket$ , where

$\boldsymbol{\mu}_x = N^{-1} \sum_{i=1}^N \mathbf{x}_i$  and  $\boldsymbol{\mu}_y = N^{-1} \sum_{i=1}^N \mathbf{y}_i$ . Then, the optimization problem for CCA based on estimated covariance matrices is given by

$$\arg \max_{\mathbf{w}_x, \mathbf{w}_y} \quad \mathbf{w}_x^T \mathbf{X} \mathbf{Y}^T \mathbf{w}_y \quad (10)$$

$$\text{subject to} \quad \mathbf{w}_x^T \mathbf{X} \mathbf{X}^T \mathbf{w}_x = 1, \quad \mathbf{w}_y^T \mathbf{Y} \mathbf{Y}^T \mathbf{w}_y = 1,$$

and the generalized eigenvalue problem given by Eqs. (6) and (7) can be rewritten as

$$\mathbf{X} \mathbf{Y}^T \mathbf{w}_y = \lambda \mathbf{X} \mathbf{X}^T \mathbf{w}_x \quad (11)$$

$$\mathbf{Y} \mathbf{X}^T \mathbf{w}_x = \lambda \mathbf{Y} \mathbf{Y}^T \mathbf{w}_y. \quad (12)$$

Then, by multiplying both sides of Eqs. (11) and (12) by  $\mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1}$  and  $\mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T)^{-1}$ , respectively, we obtain:

$$\mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{Y}^T \mathbf{w}_y = \mathbf{P}_x \mathbf{Y}^T \mathbf{w}_y = \lambda \mathbf{X}^T \mathbf{w}_x \quad (13)$$

$$\mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T)^{-1} \mathbf{Y} \mathbf{X}^T \mathbf{w}_x = \mathbf{P}_y \mathbf{X}^T \mathbf{w}_x = \lambda \mathbf{Y}^T \mathbf{w}_y, \quad (14)$$

where  $\mathbf{P}_x = \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X}$  and  $\mathbf{P}_y = \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T)^{-1} \mathbf{Y}$  are the orthogonal projectors onto the linear spans of the rows of  $\mathbf{X}$  and  $\mathbf{Y}$  respectively. So substituting  $\mathbf{X}^T \mathbf{w}_x$  in Eq. (14) and  $\mathbf{Y}^T \mathbf{w}_y$  in Eq. (13) gives

$$\mathbf{P}_x \mathbf{P}_y \mathbf{X}^T \mathbf{w}_x = \mathbf{K}_{xy} \mathbf{X}^T \mathbf{w}_x = \lambda^2 \mathbf{X}^T \mathbf{w}_x$$

$$\mathbf{P}_y \mathbf{P}_x \mathbf{Y}^T \mathbf{w}_y = \mathbf{K}_{yx} \mathbf{Y}^T \mathbf{w}_y = \lambda^2 \mathbf{Y}^T \mathbf{w}_y.$$

Therefore, we can observe that  $\mathbf{X}^T \mathbf{w}_x$  and  $\mathbf{Y}^T \mathbf{w}_y$  are the left singular vectors associated to the largest singular values of the matrices  $\mathbf{K}_{xy} = \mathbf{P}_x \mathbf{P}_y$  and  $\mathbf{K}_{yx} = \mathbf{P}_y \mathbf{P}_x$  respectively. By using the symmetric property of the matrices  $\mathbf{P}_x$  and  $\mathbf{P}_y$  we have:

$$\mathbf{K}_{yx} = \mathbf{P}_y \mathbf{P}_x = \mathbf{P}_y^T \mathbf{P}_x^T = (\mathbf{P}_x \mathbf{P}_y)^T = \mathbf{K}_{xy}^T. \quad (15)$$

The singular value decomposition of the matrices  $\mathbf{K}_{xy}$  and  $\mathbf{K}_{yx}$  is given by:

$$\mathbf{K}_{xy} = \mathbf{U} \mathbf{D} \mathbf{V}^T \quad (16)$$

$$\mathbf{K}_{yx} = \mathbf{V} \mathbf{D} \mathbf{U}^T, \quad (17)$$

where  $\mathbf{u}_i$  and  $\mathbf{v}_i$  are the  $i^{\text{th}}$  column vectors of the matrices  $\mathbf{U}$  and  $\mathbf{V}$ , respectively, and  $\mathbf{D} = \text{diag}(d_1, \dots, d_N)$  such that  $d_1 \geq d_2 \geq \dots \geq d_N$  represent the singular values of  $\mathbf{K}_{xy}$  and  $\mathbf{K}_{yx}$ . We can deduce from Eqs. (16), (17) and (15) that the left singular vectors of  $\mathbf{K}_{yx}$  correspond to the right singular vectors of  $\mathbf{K}_{xy}$ .

In order to estimate the canonical projectors, we define the nearest rank-1 matrix approximation of  $\mathbf{K}_{xy}$  by:

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

where the nearest means that the squared Frobenius norm between  $\mathbf{K}_{xy}$  and  $\mathbf{K}_1$ , defined by  $\|\mathbf{K}_{xy} - \mathbf{K}_1\|_F^2$ , is minimal.

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

$$\arg \min_{\mathbf{w}_x, \mathbf{w}_y} \|\mathbf{K}_{xy} - \mathbf{X}^T \mathbf{w}_x \mathbf{w}_y^T \mathbf{Y}\|_F^2. \quad (18)$$

Consequently, the projected data  $\mathbf{X}^T \mathbf{w}_x$  and  $\mathbf{Y}^T \mathbf{w}_y$  consist of the left and right singular vectors, respectively, associated to the largest singular value of the matrix  $\mathbf{K}_{xy}$ . Therefore, after estimating the left and right singular vectors  $\mathbf{u}_1$  and  $\mathbf{v}_1$ , respectively, and associated singular value  $d_1$  of the matrix  $\mathbf{K}_{xy}$ , we can obtain the projectors  $\mathbf{w}_x$  and  $\mathbf{w}_y$  by solving the following least square equations (see Step 5 in Algorithm 1):

$$\arg \min_{\mathbf{w}_x} \|\sqrt{d_1} \mathbf{u}_1 - \mathbf{X}^T \mathbf{w}_x\|_2^2$$

$$\arg \min_{\mathbf{w}_y} \|\sqrt{d_1} \mathbf{v}_1 - \mathbf{Y}^T \mathbf{w}_y\|_2^2.$$

Hence, for multiple projected data, the solution consist of the associated singular vectors corresponding to the top singular values of the matrix  $\mathbf{K}_{xy}$ .

---

**Algorithm 1** Rank-1 matrix approximation CCA algorithm

---

**Require:** Training data  $\mathbf{X} \in \mathbb{R}^{d_x \times N}$  and  $\mathbf{Y} \in \mathbb{R}^{d_y \times N}$ .

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

- 1: Compute the centred data  $\mathbf{x}_i \leftarrow \mathbf{x}_i - \boldsymbol{\mu}_x$  and  $\mathbf{y}_i \leftarrow \mathbf{y}_i - \boldsymbol{\mu}_y$  for all  $i \in \llbracket 1, N \rrbracket$ , where  $\boldsymbol{\mu}_x = N^{-1} \sum_{i=1}^N \mathbf{x}_i$  and  $\boldsymbol{\mu}_y = N^{-1} \sum_{i=1}^N \mathbf{y}_i$ ;
  - 2: Compute  $\mathbf{P}_x = \mathbf{X}^T (\mathbf{X} \mathbf{X}^T + \gamma_x \mathbf{I}_{d_x})^{-1} \mathbf{X}$ ,  $\mathbf{P}_y = \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + \gamma_y \mathbf{I}_{d_y})^{-1} \mathbf{Y}$  and  $\mathbf{K}_{xy} = \mathbf{P}_x \mathbf{P}_y$ ;
  - 3: Perform the SVD of  $\mathbf{K}_{xy}$ :  $\mathbf{K}_{xy} = \mathbf{U} \mathbf{D} \mathbf{V}^T$ ;
  - 4: Form  $\tilde{\mathbf{U}} = [\mathbf{u}_1, \dots, \mathbf{u}_r]$ ,  $\tilde{\mathbf{D}} = \text{diag}(\sqrt{d_1}, \dots, \sqrt{d_r})$  and  $\tilde{\mathbf{V}} = [\mathbf{v}_1, \dots, \mathbf{v}_r]$ ;
  - 5: Set  $\mathbf{W}_x = (\mathbf{X} \mathbf{X}^T + \gamma_x \mathbf{I}_{d_x})^{-1} \mathbf{X} \tilde{\mathbf{U}} \tilde{\mathbf{D}}$  and  $\mathbf{W}_y = (\mathbf{Y} \mathbf{Y}^T + \gamma_y \mathbf{I}_{d_y})^{-1} \mathbf{Y} \tilde{\mathbf{V}} \tilde{\mathbf{D}}$ .
- 

From (18), we can observe that the optimization problem (10) that involves the two constraints  $\|\mathbf{w}_x^T \mathbf{X}\|_2 = 1$  and  $\|\mathbf{w}_y^T \mathbf{Y}\|_2 = 1$  has now been transformed into a rank-1 matrix approximation problem free of constraints and which can be solved with an SVD. With this approach, the proposed algorithm avoids the need of using these constraints and hence also avoids their relaxations as it was proposed in [11].

One disadvantage of the above approach is the restriction that  $\mathbf{X} \mathbf{X}^T$  and  $\mathbf{Y} \mathbf{Y}^T$  must be non-singular. In order to prevent overfitting and avoid the singularity of  $\mathbf{X} \mathbf{X}^T$  and  $\mathbf{Y} \mathbf{Y}^T$  [6], two regularization terms,  $\gamma_x \mathbf{I}_{d_x}$  and  $\gamma_y \mathbf{I}_{d_y}$ , with

$\gamma_x > 0$ ,  $\gamma_y > 0$  are added in (10). Therefore, the regularized version solves the generalized eigenvalue problem with  $\mathbf{P}_x = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T + \gamma_x \mathbf{I}_{d_x})^{-1}\mathbf{X}$  and  $\mathbf{P}_y = \mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \gamma_y \mathbf{I}_{d_y})^{-1}\mathbf{Y}$ . We summarized the method of solving the entire rank-1 matrix approximation CCA in Algorithm 1.

### 3.1 Smoothed rank-1 matrix approximation CCA algorithm

In order to give preference to a particular solution with desirable properties for the proposed CCA problem, a regularization term (Tikhonov regularization) can be included in Eq. (18) such that:

$$\begin{aligned} \arg \min_{\mathbf{w}_x, \mathbf{w}_y} & \|\mathbf{K}_{xy} - \mathbf{X}^T \mathbf{w}_x \mathbf{w}_y^T \mathbf{Y}\|_F^2 + \alpha_x \mathbf{w}_x^T \mathbf{X} \mathbf{\Omega}_x \mathbf{X}^T \mathbf{w}_x \\ & + \alpha_y \mathbf{w}_y^T \mathbf{Y} \mathbf{\Omega}_y \mathbf{Y}^T \mathbf{w}_y. \end{aligned} \quad (19)$$

In many cases, the matrices  $\mathbf{\Omega}_x$  and  $\mathbf{\Omega}_y$  are chosen as a multiple of the identity matrix, giving preference to solutions with smaller norms. In our case, the matrices  $\mathbf{\Omega}_x$  and  $\mathbf{\Omega}_y$  are non-negative definite roughness penalty matrices used to penalize the second differences [14, 15], and  $\alpha_x > 0$  and  $\alpha_y > 0$  are trade-off parameters such as:

$$\forall \mathbf{z} \in \mathbb{R}^N, \quad \mathbf{z}^T \mathbf{\Omega} \mathbf{z} = z_1^2 + z_N^2 + \sum_{i=2}^{N-1} (z_{i+1} - 2z_i + z_{i-1})^2. \quad (20)$$

The choice of such matrices may be used to enforce smoothness if the underlying vector is believed to be mostly continuous. The criterion of Eq. (19) can be rewritten as

$$\begin{aligned} \arg \min_{\mathbf{w}_x, \mathbf{w}_y} & \|\mathbf{X}^T \mathbf{w}_x\|_2^2 \|\mathbf{Y}^T \mathbf{w}_y\|_2^2 - 2 \mathbf{w}_x^T \mathbf{X} \mathbf{K}_{xy} \mathbf{Y}^T \mathbf{w}_y \\ & + \alpha_x \mathbf{w}_x^T \mathbf{X} \mathbf{\Omega}_x \mathbf{X}^T \mathbf{w}_x + \alpha_y \mathbf{w}_y^T \mathbf{Y} \mathbf{\Omega}_y \mathbf{Y}^T \mathbf{w}_y. \end{aligned} \quad (21)$$

The optimization problem (21) can be alternatively solved by optimizing  $\mathbf{w}_x$  and  $\mathbf{w}_y$ . Specifically, we first fix  $\mathbf{w}_y$  and solve  $\mathbf{w}_x$  by minimizing (21). Then, we fix  $\mathbf{w}_x$  and minimize (21) to obtain  $\mathbf{w}_y$ . The above two procedures are repeated until convergence. Taking derivatives with respect to  $\mathbf{w}_x$  and  $\mathbf{w}_y$ , we obtain

$$\begin{aligned} \left( \|\mathbf{Y}^T \mathbf{w}_y\|_2^2 \mathbf{X} \mathbf{X}^T + \alpha_x \mathbf{X} \mathbf{\Omega}_x \mathbf{X}^T \right) \mathbf{w}_x &= \mathbf{X} \mathbf{K}_{xy} \mathbf{Y}^T \mathbf{w}_y \\ \left( \|\mathbf{X}^T \mathbf{w}_x\|_2^2 \mathbf{Y} \mathbf{Y}^T + \alpha_y \mathbf{Y} \mathbf{\Omega}_y \mathbf{Y}^T \right) \mathbf{w}_y &= \mathbf{Y} \mathbf{K}_{xy}^T \mathbf{X}^T \mathbf{w}_x. \end{aligned}$$

Therefore, we obtain  $\mathbf{w}_x$  and  $\mathbf{w}_y$  by solving the above equations in the least square sense (see Steps 7 and 9 in Algorithm 2). For multiple canonical projectors, let us consider the singular value decomposition of  $\mathbf{K}_{xy} = \mathbf{U} \mathbf{D} \mathbf{V}^T = \sum_{i=1}^N d_i \mathbf{u}_i \mathbf{v}_i^T$ , where  $\mathbf{u}_i$  and  $\mathbf{v}_i$  are the  $i^{\text{th}}$  column vectors of the matrices  $\mathbf{U}$  and  $\mathbf{V}$ , respectively, and

$\mathbf{D} = \text{diag}(d_1, \dots, d_N)$  such that  $d_1 \geq d_2 \geq \dots \geq d_N$ . In order to estimate the second pair of canonical projectors, we must remove the contribution of the first pair of canonical projectors from the matrix  $\mathbf{K}_{xy}$ . To this end, we must remove the contribution of the singular vectors associated to the largest singular value  $d_1$  using:

$$\mathbf{K}_{xy} - d_1 \mathbf{u}_1 \mathbf{v}_1^T = \sum_{i=2}^N d_i \mathbf{u}_i \mathbf{v}_i^T.$$

As presented in Section 3, the singular vectors  $\mathbf{u}_1$  and  $\mathbf{v}_1$  represent the projected data  $\mathbf{X}^T \mathbf{w}_x$  and  $\mathbf{Y}^T \mathbf{w}_y$ , respectively. Then, by using the unitary property of matrices  $\mathbf{U}$  and  $\mathbf{V}$ , we can compute the singular value associated to the singular vectors  $\mathbf{u}_1$  and  $\mathbf{v}_1$  by  $d_1 = \mathbf{u}_1^T \mathbf{K}_{xy} \mathbf{v}_1$ . Therefore, we propose to use a deflation procedure where the second pair of canonical projectors are defined by using the corresponding residual matrix  $\mathbf{K}_{xy} - \mathbf{w}_x^T \mathbf{X} \mathbf{K}_{xy} \mathbf{Y}^T \mathbf{w}_y \mathbf{X}^T \mathbf{w}_x \mathbf{w}_y^T \mathbf{Y}$ . Then, we can define the other pair of projectors. The method for solving the smoothed rank-1 matrix approximation CCA is summarized by Algorithm 2.

---

#### Algorithm 2 Smoothed rank-1 matrix approximation CCA algorithm

---

**Require:** Training data  $\mathbf{X} \in \mathbb{R}^{d_x \times N}$  and  $\mathbf{Y} \in \mathbb{R}^{d_y \times N}$ .

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

- 1: Compute the centred data  $\mathbf{x}_i \leftarrow \mathbf{x}_i - \boldsymbol{\mu}_x$  and  $\mathbf{y}_i \leftarrow \mathbf{y}_i - \boldsymbol{\mu}_y$  for all  $i \in \llbracket 1, N \rrbracket$ , where  $\boldsymbol{\mu}_x = N^{-1} \sum_{i=1}^N \mathbf{x}_i$  and  $\boldsymbol{\mu}_y = N^{-1} \sum_{i=1}^N \mathbf{y}_i$
  - 2: Compute  $\mathbf{P}_x = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T + \gamma_x \mathbf{I}_{d_x})^{-1}\mathbf{X}$ ,  $\mathbf{P}_y = \mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \gamma_y \mathbf{I}_{d_y})^{-1}\mathbf{Y}$  and  $\mathbf{K}_{xy} = \mathbf{P}_x \mathbf{P}_y$
  - 3: **for**  $i = 1, 2, \dots, r$  **do**
  - 4:     Perform the SVD of  $\mathbf{K}_{xy}$ :  $\mathbf{K}_{xy} = \mathbf{U} \mathbf{D} \mathbf{V}^T$
  - 5:     Initialize  $\tilde{\mathbf{u}} = \mathbf{u}_1$  and  $\tilde{\mathbf{v}} = \mathbf{v}_1$
  - 6:     **repeat**
  - 7:         Update the  $i$ -th column of  $\mathbf{W}_x$ :  $\mathbf{W}_x(:, i) = (\mathbf{X}\mathbf{X}^T + \alpha_x \mathbf{X} \mathbf{\Omega}_x \mathbf{X}^T + \gamma_x \mathbf{I}_{d_x})^{-1} \mathbf{X} \mathbf{K}_{xy} \tilde{\mathbf{v}}$
  - 8:         Update  $\tilde{\mathbf{u}} = \frac{\mathbf{X}^T \mathbf{W}_x(:, i)}{\|\mathbf{X}^T \mathbf{W}_x(:, i)\|_2}$
  - 9:         Update the  $i$ -th column of  $\mathbf{W}_y$ :  $\mathbf{W}_y(:, i) = (\mathbf{Y}\mathbf{Y}^T + \alpha_y \mathbf{Y} \mathbf{\Omega}_y \mathbf{Y}^T + \gamma_y \mathbf{I}_{d_y})^{-1} \mathbf{Y} \mathbf{K}_{xy}^T \tilde{\mathbf{u}}$
  - 10:         Update  $\tilde{\mathbf{v}} = \frac{\mathbf{Y}^T \mathbf{W}_y(:, i)}{\|\mathbf{Y}^T \mathbf{W}_y(:, i)\|_2}$
  - 11:     **until** convergence
  - 12:     Update  $\mathbf{K}_{xy}$ :  $\mathbf{K}_{xy} \leftarrow \mathbf{K}_{xy} - \tilde{\mathbf{u}}^T \mathbf{K}_{xy} \tilde{\mathbf{v}} \tilde{\mathbf{v}} \tilde{\mathbf{u}}^T$
  - 13: **end for**
-

For illustrating the advantage of the proposed smoothed CCA approach over standard CCA, we generated for three distinct simulated activation cases; spatially independent case  $S_1$ , partial spatial overlap  $S_2$ , and complete spatial overlap case  $S_3$  as done in [16]. Three temporal sources, with 120 s duration, were constructed to represent the brain hemodynamics, i.e. block design activation ( $T_1$ ), and two sinusoids ( $T_2$  and  $T_3$ ) with frequencies  $\in \{1.5, 9.5\}$  Hz, respectively, and box signals were used as brain activation patterns [16]. Three distinct visual patterns of size  $10 \times 10$  voxels were created with amplitudes of 1 at voxel indexes  $\{2, \dots, 6\} \times \{2, \dots, 6\}$  for pattern  $A$ ,  $\{8, 9\} \times \{8, 9\}$  for pattern  $B$ , and  $\{5, \dots, 9\} \times \{5, \dots, 9\}$  for pattern  $C$ , and 0 elsewhere. The three simulated cases are shown in Figs. 1, 2 and 3: spatially independent events in Fig. 1, partial spatial overlapping events in Fig. 2 and complete spatial overlapping events in Fig. 3.

We can observe from Figs. 1, 2 and 3 that the proposed smoothed CCA algorithm have recovered both the temporal signal and spatial maps with better accuracy than CCA for the three presented cases  $S_1$ ,  $S_2$  and  $S_3$ . This demonstrates the effectiveness of the proposed smoothed CCA approach in regularization when the estimated signals are believed to be continuous and smooth.

#### 4 Sparse CCA algorithm based on rank-1 matrix approximation

In this section, we will propose the sparse CCA method based on rank-1 matrix approximation by penalizing the optimization problem (18). Then, we propose an efficient iterative algorithm to solve the sparse solution of the proposed criterion.

In general, the canonical projectors  $w_x$  and  $w_y$  found as solutions in Eq. (18) are not sparse, i.e., the entries of both

$w_x$  and  $w_y$  are non-zeros. To obtain the sparse solution, we adopt the similar trick used in [7, 11, 12, 17] by imposing penalty functions on the optimization problem (18). Therefore, we can write the new optimization problem as:

$$\arg \min_{w_x, w_y} \|K_{xy} - X^T w_x w_y^T Y\|_F^2 \quad \text{subject to} \quad (22)$$

$$\mathcal{F}_x(w_x) \leq \beta_x \quad \text{and} \quad \mathcal{F}_y(w_y) \leq \beta_y,$$

where  $\mathcal{F}_x(\cdot)$  and  $\mathcal{F}_y(\cdot)$  are penalty functions, which can take on a variety of forms. Useful examples are  $\ell_0$ -quasi-norm  $\mathcal{F}(z) = \|z\|_0$  which count the non-zero entries of a vector; Lasso penalty with  $\ell_1$ -norm  $\mathcal{F}(z) = \|z\|_1$  and so on.

The optimization problem (22) can be alternatively solved by optimizing  $w_x$  and  $w_y$ . Specifically, we first fix  $w_y$  and solve for  $w_x$  by minimizing (22). Then, we fix  $w_x$  and minimize (22) to obtain  $w_y$ . The above two procedures are repeated until convergence.

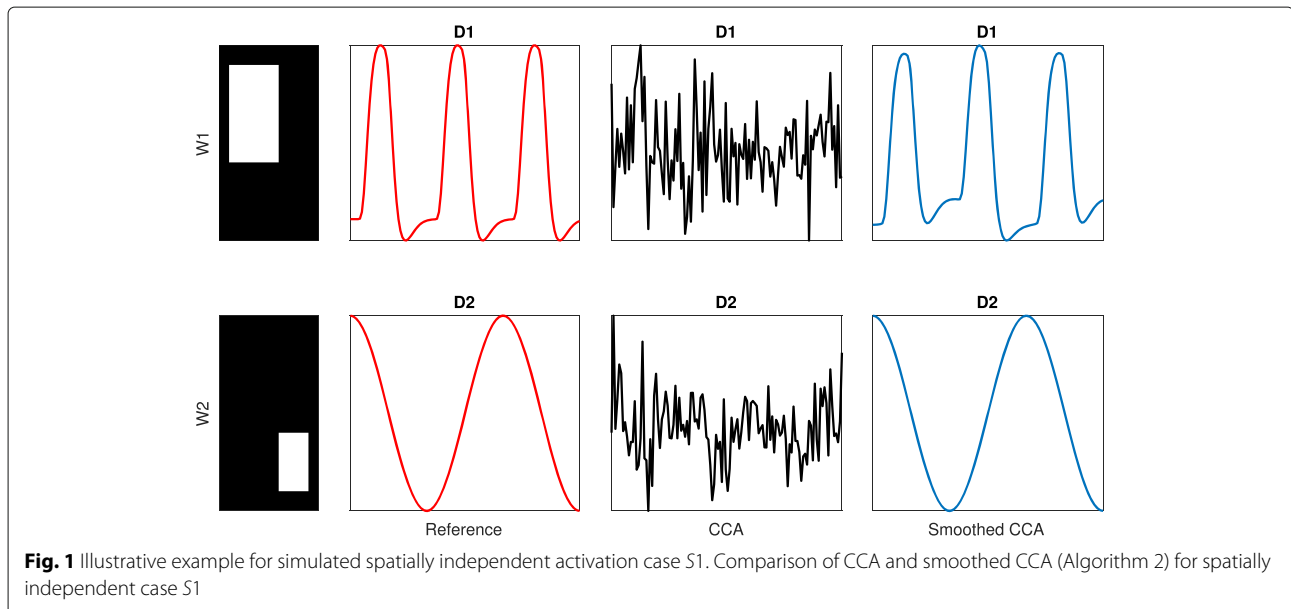
The straightforward approach to solve this problem is to formulate it as an ordinary sparse coding task. Then, for a fix  $w_y$  the problem (22) is equivalent to much simpler sparse coding problem

$$\arg \min_{w_x} \|K_{xy} Y^T w_y - X^T w_x\|_2^2 \quad \text{subject to} \quad \mathcal{F}_x(w_x) \leq \beta_x,$$

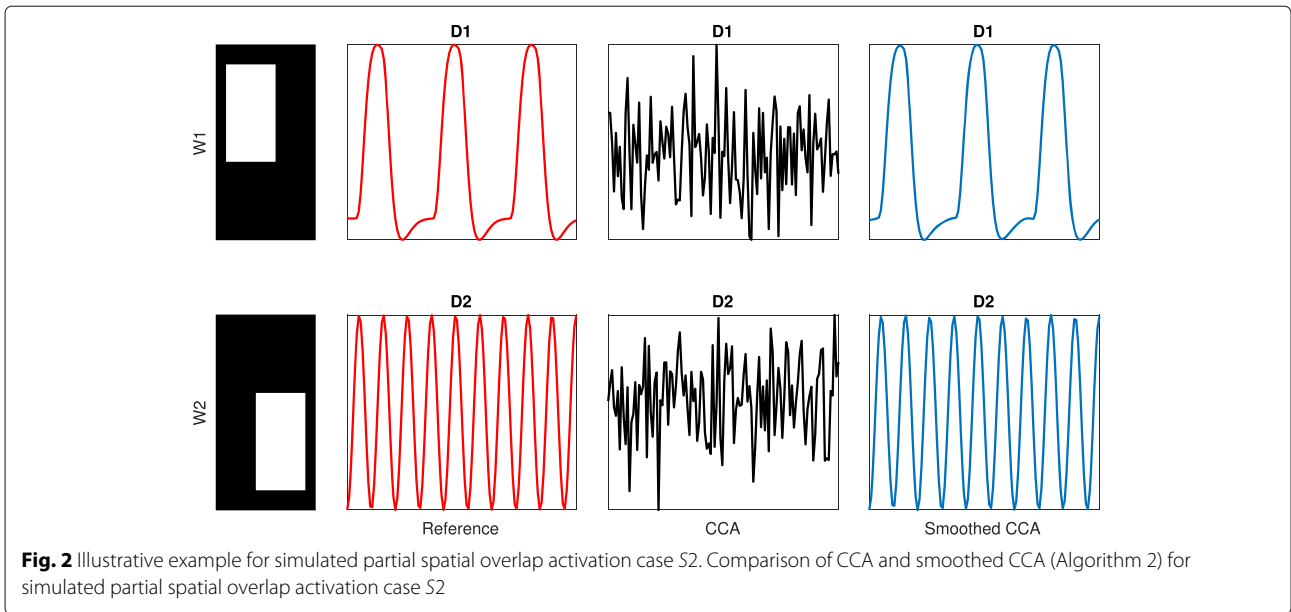
which can be solved by using any sparse approximation method. In the same way, we can solve the problem (22) regarding  $w_y$  for a fix  $w_x$  by minimizing the following criterion:

$$\arg \min_{w_y} \|K_{xy}^T X^T w_x - Y^T w_y\|_2^2 \quad \text{subject to} \quad \mathcal{F}_y(w_y) \leq \beta_y.$$

Based on the above description, we can obtain the first pair of sparse projectors  $w_x$  and  $w_y$ . For multiple projection vectors, we propose to use a deflation procedure as



**Fig. 1** Illustrative example for simulated spatially independent activation case  $S_1$ . Comparison of CCA and smoothed CCA (Algorithm 2) for spatially independent case  $S_1$

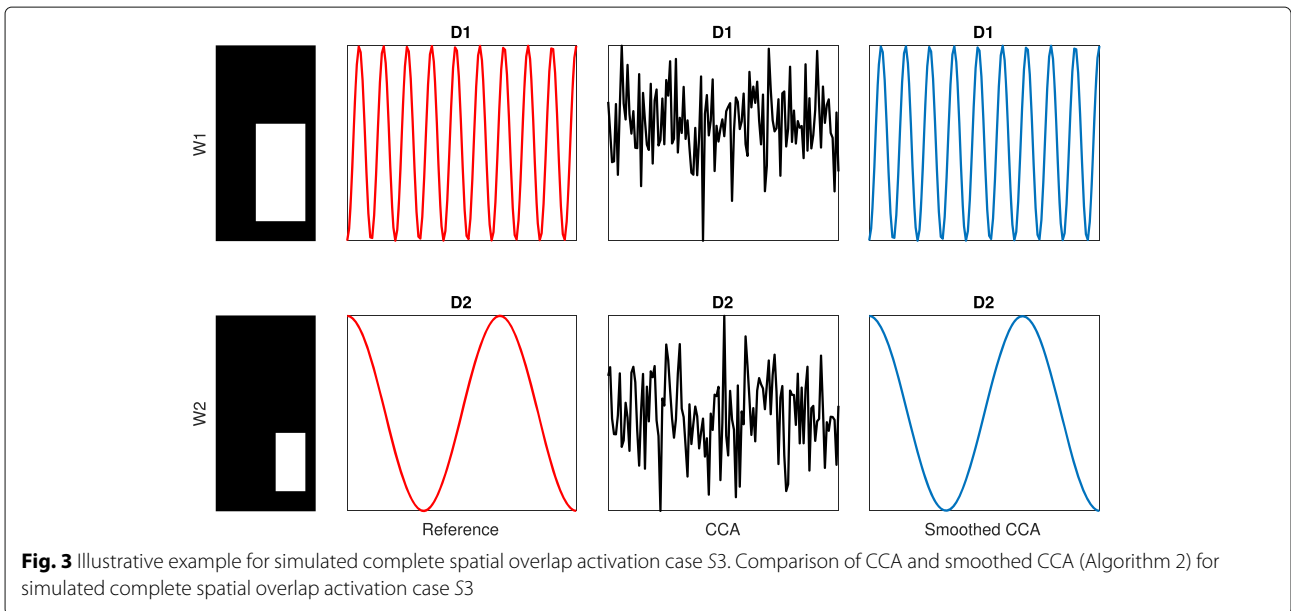


presented in Section 3.1 where the second pair of sparse projectors are defined by using the corresponding residual matrices  $K_{xy} - X^T w_x K_{xy} w_y^T Y w_x^T X Y^T w_y$ . Using the same way, we can define the other pair of sparse projectors.

The uncorrelated entries of the projected vector is obtained due to the orthogonality of the canonical components. The orthogonality among these components is lost due to the constraints added to the cost (18), a nice property enjoyed by standard CCA. Several other CCA procedures lose this property as well; this is just the price to pay for using the other constraints (sparsity or smoothness).

Then, we summarized the method of solving the entire sparse rank-1 matrix approximation CCA in Algorithm 3

In terms of difference between the proposed approach to achieve sparse CCA and the method proposed in [11]; the method proposed in [11] uses a penalized matrix decomposition on the cross-product matrix  $XY^T$ , whereas our proposed approach is based on a rank-1 matrix approximation of  $K_{xy}$  as defined in (18). Furthermore, the method proposed in [11] makes the assumption that  $XX^T$  and  $YY^T$  are identities to replace the constraints  $w_x^T XX^T w_x \leq 1$  and  $w_y^T YY^T w_y \leq 1$  by  $\|w_x\|_2 \leq 1$  and  $\|w_y\|_2 \leq 1$  (Eqs. (4.2) and (4.3) of [11]). This assumption is



relaxed in the proposed sparse CCA algorithm presented in Section 4. This is obtained by directly including these constraints  $\mathbf{w}_x^T \mathbf{X} \mathbf{X}^T \mathbf{w}_x = 1$  and  $\mathbf{w}_y^T \mathbf{Y} \mathbf{Y}^T \mathbf{w}_y = 1$  in the derivation of the matrix  $\mathbf{K}_{xy}$  used in the penalized rank-1 matrix approximation via Eq. (3).

The same argument is valid for [7] and [12] as both these papers are based on the cross-product matrix  $\mathbf{X} \mathbf{Y}^T$ ; furthermore, their approaches used for regularization is similar to the one described in Algorithm 1 and therefore different from the regularization adopted in this paper given in Algorithm 2.

---

**Algorithm 3** Sparse rank-1 matrix approximation CCA algorithm

---

**Require:** Training data  $\mathbf{X} \in \mathbb{R}^{d_x \times N}$  and  $\mathbf{Y} \in \mathbb{R}^{d_y \times N}$ .

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

1: Compute the centred data  $\mathbf{x}_i \leftarrow \mathbf{x}_i - \boldsymbol{\mu}_x$  and  $\mathbf{y}_i \leftarrow \mathbf{y}_i - \boldsymbol{\mu}_y$  for all  $i \in \llbracket 1, N \rrbracket$ , where  $\boldsymbol{\mu}_x = N^{-1} \sum_{i=1}^N \mathbf{x}_i$  and

$$\boldsymbol{\mu}_y = N^{-1} \sum_{i=1}^N \mathbf{y}_i;$$

2: Compute  $\mathbf{P}_x = \mathbf{X}^T (\mathbf{X} \mathbf{X}^T + \gamma_x \mathbf{I}_{d_x})^{-1} \mathbf{X}$ ,  $\mathbf{P}_y = \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + \gamma_y \mathbf{I}_{d_y})^{-1} \mathbf{Y}$  and  $\mathbf{K}_{xy} = \mathbf{P}_x \mathbf{P}_y$ ;

3: **for**  $i = 1, 2, \dots, r$  **do**

4:   Perform the SVD of  $\mathbf{K}_{xy}$ :  $\mathbf{K}_{xy} = \mathbf{U} \mathbf{D} \mathbf{V}^T$ ;

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

6:   **repeat**

7:     Update the  $i$ -th column of  $\mathbf{W}_x$ :

$$\mathbf{W}_x(:, i) = \arg \min_{\mathbf{W}_x(:, i)} \|\mathbf{K}_{xy} \tilde{\mathbf{v}} - \mathbf{X}^T \mathbf{W}_x(:, i)\|_2^2 \quad \text{subject to}$$

$$\mathcal{F}_x(\mathbf{W}_x(:, i)) \leq \beta_x;$$

8:     Update  $\tilde{\mathbf{u}} = \frac{\mathbf{X}^T \mathbf{W}_x(:, i)}{\|\mathbf{X}^T \mathbf{W}_x(:, i)\|_2}$ ;

9:     Update the  $i$ -th column of  $\mathbf{W}_y$ :

$$\mathbf{W}_y(:, i) = \arg \min_{\mathbf{W}_y(:, i)} \|\mathbf{K}_{xy}^T \tilde{\mathbf{u}} - \mathbf{Y}^T \mathbf{W}_y(:, i)\|_2^2 \quad \text{subject to}$$

$$\mathcal{F}_y(\mathbf{W}_y(:, i)) \leq \beta_y;$$

10:     Update  $\tilde{\mathbf{v}} = \frac{\mathbf{Y}^T \mathbf{W}_y(:, i)}{\|\mathbf{Y}^T \mathbf{W}_y(:, i)\|_2}$ ;

11:     **until** convergence

12:     Update  $\mathbf{K}_{xy}$ :  $\mathbf{K}_{xy} \leftarrow \mathbf{K}_{xy} - \tilde{\mathbf{u}}^T \mathbf{K}_{xy} \tilde{\mathbf{v}} \tilde{\mathbf{v}} \tilde{\mathbf{u}}^T$ ;

13: **end for**

---

## 5 Experiments

In this section, we present several computer simulations in the context of blind channel estimation in single-input multiple-output (SIMO) systems and blind source separation to demonstrate the effectiveness of the proposed algorithm. We compare the performance of the proposed

algorithm with existing state of the art sparse CCA methods:

- The sparse CCA presented in [11], relying on a penalized matrix decomposition denoted PMD. An R package implementing this algorithm, called PMA, is available at <http://cran.r-project.org/web/packages/PMA/index.html>. Sparsity parameters are selected using the permutation approach presented in [18] of which the code is provided in PMA package.
- The sparse CCA presented in [7] where the CCA is reformulated as a least-squares problem denoted LS CCA. A Matlab package implementing this algorithm is available at <http://www.public.asu.edu/~jye02/Software/CCA/>.
- The sparse CCA presented in [12] where the sparse canonical projectors are computed by solving two  $\ell_1$ -minimization problems by using the Linearized Bregman iterative method [19]. This algorithm is denoted CCA LB (Linearized Bregman). We re-implemented the sparse CCA algorithm proposed in [12] using Matlab.

For the proposed sparse CCA algorithm, we have used  $\mathcal{F}_x(\mathbf{z}) = \mathcal{F}_y(\mathbf{z}) = \|\mathbf{z}\|_0$  as penalty functions. We solve the sparse coding problem by using orthogonal matching pursuit (OMP) algorithm [20, 21]. For proposed smoothed CCA algorithm, we chose  $\boldsymbol{\Omega}_x = \boldsymbol{\Omega}_y$  and given by Eq. (20).

### 5.1 Synthetic data

This simulation setup is inspired from [22]. The synthetic data  $\mathbf{X}$  and  $\mathbf{Y}$  were generating according to multivariate normal distribution, with covariance matrices described in Table 1. The number of simulations with each configuration was  $N_k = 1000$ . We compare the performance of our algorithm to the state of the art methods by estimating the precision accuracy of the space spanned by  $r$  estimated canonical projectors. We compute for each simulation run  $k$  the angle  $\theta^k(\hat{\mathbf{W}}_x^k, \mathbf{W}_x)$  between the subspace<sup>1</sup> spanned by the estimated canonical projectors contained in the columns of  $\hat{\mathbf{W}}^k$  and the subspace spanned by the true canonical projectors contained in the columns of  $\mathbf{W}_x$  solution of the eigenproblem (9). The same criterion is used for the canonical projectors  $\mathbf{W}_y$ . The average angles are estimated over  $N_k$  Monte-Carlo run such that:

$$\theta_x = \frac{1}{N_k} \sum_{k=1}^{N_k} \theta^k(\hat{\mathbf{W}}_x^k, \mathbf{W}_x) \quad \text{and}$$

$$\theta_y = \frac{1}{N_k} \sum_{k=1}^{N_k} \theta^k(\hat{\mathbf{W}}_y^k, \mathbf{W}_y).$$

**Table 1** Simulation settings

Parameters	$d_x$	$d_y$	$r$	$N$	$C_{xx}$	$C_{yy}$	$C_{xy}$
Scenario 1	4	4	3	{50, 100, 200}	$I_4$	$I_4$	$\begin{bmatrix} \frac{9}{10} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Scenario 2	4	6	2	{50, 100, 200}	$I_4$	$I_6$	$\begin{bmatrix} \frac{3}{5} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
Scenario 3	4	6	2	{50, 100, 200}	$I_4$	$I_6$	$\begin{bmatrix} \frac{2}{5} & \frac{4}{25} & 0 & 0 & 0 & 0 \\ \frac{4}{25} & \frac{2}{5} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
Scenario 4	6	10	2	{50, 100, 200}	$I_6$	$\begin{bmatrix} M & \mathbf{0} \\ \mathbf{0} & I_7 \end{bmatrix}$ with $M(i,j) = 0,3^{ i-j }$	$\frac{1}{2} \begin{bmatrix} I_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$
Scenario 5	20	20	10	{50, 100, 200}	$I_{20}$	$I_{20}$	$\frac{7}{10} \begin{bmatrix} I_{10} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$
Scenario 6	20	20	10	{50, 100, 200}	$I_{20}$	$I_{20}$	$\begin{bmatrix} S_{10} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ with $S_{10}(i,j) = 0,4^{ i-j+1 }$

For each algorithm, we used the following parameters: LS CCA algorithm with  $\lambda_x = \lambda_y = 0.5$ , CCA LB algorithm with  $\mu_x = \mu_y = 2$ ; Algorithm 2 with  $\alpha_x = \alpha_y = 10^{-2}$ ; and Algorithm 3 with  $\beta_x = \beta_y = 3$ . The simulation performance on the estimated angle between the subspace spanned by the true canonical projectors and the estimated one by the different methods are reported in Tables 2 and 3, and plotted in Fig. 4. Note that the true canonical projectors  $W_x$  and  $W_y$  are sparse due to the structure of the matrices  $C_{xy}$  (see Eqs. (8) and (9)).

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

### 5.2 Blind channel identification for SIMO systems

Blind channel identification is a fundamental signal processing technology aimed at retrieving a system's unknown information from its outputs only. Estimation of sparse long channels (i.e. channels with small number of nonzero coefficients but a large span of delays) is

considered in this simulation. Such sparse channels are encountered in many communication applications: high-definition television (HDTV) [23], underwater acoustic communications [24] and wireless communications [25, 26]. The problem addressed in this section is to determine the sparse impulse response of a SIMO system in a blind way, i.e. only the observed system outputs are available and used without assuming knowledge of the specific input signal.

Let us consider a mathematical model where the input and the output are discrete, the system is driven by a single-input sequence  $s(t)$  and yields two output sequences  $x_1(t)$  and  $x_2(t)$  and the system has finite impulse responses (FIR's)  $h_i(t)$ , for  $t = 0, \dots, L$  and  $i = 1, 2$  with  $L$  as the maximal channel length (which is assumed to be known). Such a system model can be described as follows :

$$\begin{cases} x_1(t) = s(t) * h_1(t) + \eta_1(t) \\ x_2(t) = s(t) * h_2(t) + \eta_2(t) \end{cases} \quad (23)$$

where  $*$  denotes linear convolution,  $\eta(t) = [\eta_1(t), \eta_2(t)]^T$  is an additive spatial white Gaussian noise, i.e.  $\mathbb{E}[\eta(t)\eta(t)^T] = \sigma^2 I_2$ , and  $\mathbf{h} = [\mathbf{h}_1^T \mathbf{h}_2^T]^T$  with  $\mathbf{h}_i = [h_i(0), \dots, h_i(L)]^T$  ( $i = 1, 2$ ) denotes the impulse response vector of the  $i^{th}$  channel. Given a finite set of observation of length  $T$ , the objective in this experience

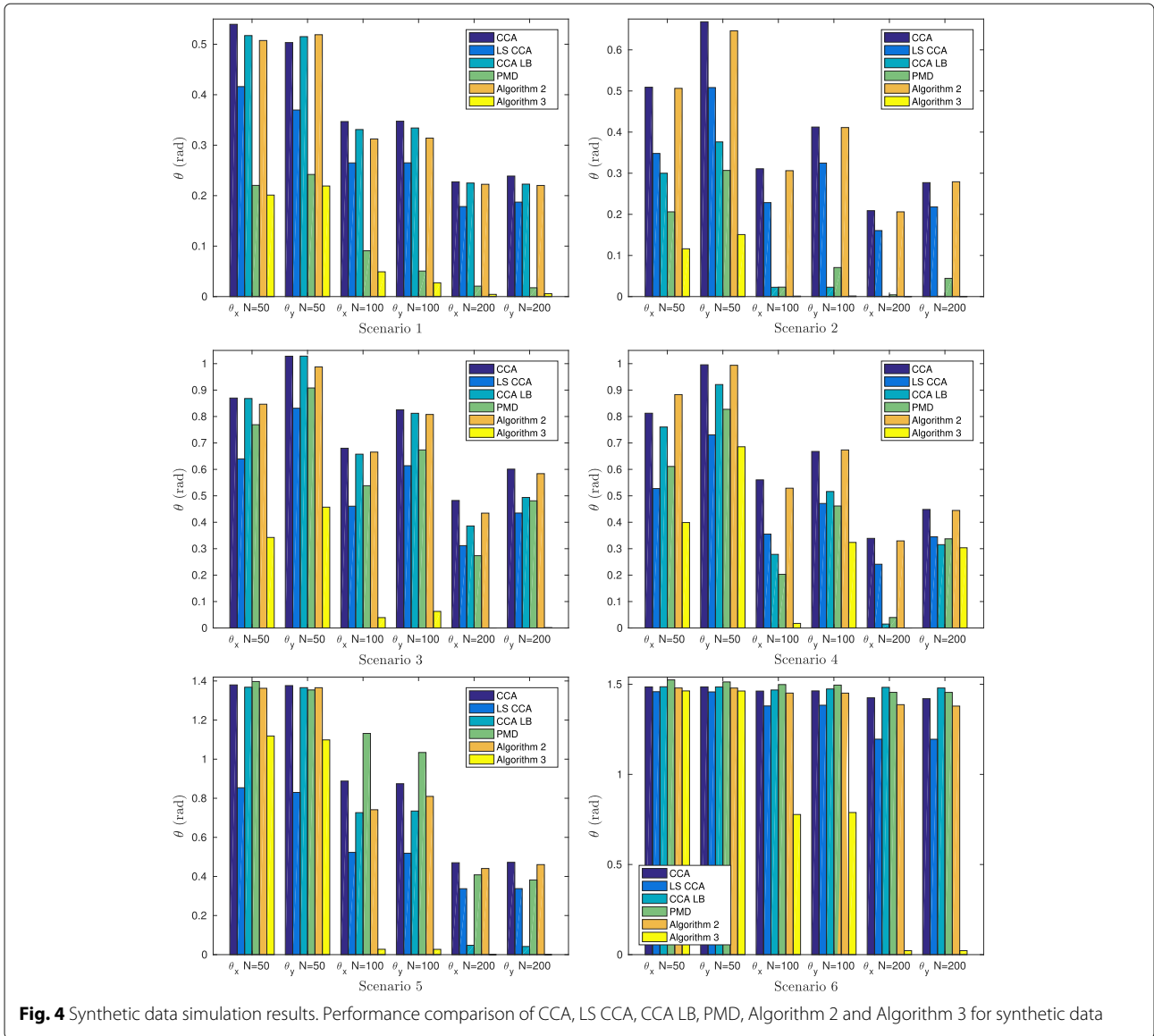


**Table 2** Simulation results part 1

	Method	$\theta_x$ (rad)	$\theta_y$ (rad)	$\theta_x$ (rad)	$\theta_y$ (rad)	$\theta_x$ (rad)	$\theta_y$ (rad)
		$N = 50$		$N = 100$		$N = 200$	
Scenario 1:	CCA	0.5395	0.5033	0.3468	0.3475	0.2273	0.2388
	LS CCA	0.4161	0.3697	0.2649	0.2650	0.1784	0.1872
	CCA LB	0.5172	0.5151	0.3310	0.3341	0.2250	0.2228
	PMD	0.2203	0.2420	0.0908	0.0506	0.0207	0.0175
	Algorithm 2	0.5074	0.5189	0.3123	0.3140	0.2225	0.2202
	Algorithm 3	0.2011	0.2191	0.0491	0.0273	0.0044	0.0057
Scenario 2:	CCA	0.5091	0.6682	0.3108	0.4123	0.2089	0.2771
	LS CCA	0.3481	0.5083	0.2285	0.3247	0.1605	0.2182
	CCA LB	0.3000	0.3761	0.0227	0.0228	0.0008	0.0009
	PMD	0.2061	0.3068	0.0230	0.0706	0.0043	0.0443
	Algorithm 2	0.5064	0.6462	0.3062	0.4111	0.2061	0.2792
	Algorithm 3	0.1162	0.1508	0.0012	0.0015	0.0001	0.0001
Scenario 3:	CCA	0.8699	1.0281	0.6800	0.8254	0.4823	0.6009
	LS CCA	0.6398	0.8314	0.4608	0.6139	0.3116	0.4348
	CCA LB	0.8681	1.0285	0.6575	0.8122	0.3859	0.4938
	PMD	0.7690	0.9080	0.5382	0.6736	0.2736	0.4811
	Algorithm 2	0.8465	0.9876	0.6654	0.8078	0.4345	0.5839
	Algorithm 3	0.3424	0.4571	0.0393	0.0628	0.0001	0.0016

**Table 3** Simulation results part 2

	Method	$\theta_x$ (rad)	$\theta_y$ (rad)	$\theta_x$ (rad)	$\theta_y$ (rad)	$\theta_x$ (rad)	$\theta_y$ (rad)
		$N = 50$		$N = 100$		$N = 200$	
Scenario 4:	CCA	0.8125	0.9956	0.5603	0.6678	0.3390	0.4484
	LS CCA	0.5275	0.7305	0.3553	0.4711	0.2412	0.3449
	CCA LB	0.7603	0.9209	0.2785	0.5163	0.0149	0.3152
	PMD	0.6111	0.8273	0.2031	0.4616	0.0397	0.3373
	Algorithm 2	0.8829	0.9938	0.5288	0.6735	0.3295	0.4447
	Algorithm 3	0.3990	0.6856	0.0173	0.3237	0.0001	0.3035
Scenario 5:	CCA	1.3798	1.3764	0.8879	0.8744	0.4700	0.4722
	LS CCA	0.8538	0.8298	0.5231	0.5187	0.3373	0.3378
	CCA LB	1.3681	1.3659	0.7264	0.7347	0.0478	0.0417
	PMD	1.3972	1.3542	1.1316	1.0342	0.4082	0.3820
	Algorithm 2	1.3627	1.3655	0.7413	0.8096	0.4407	0.4605
	Algorithm 3	1.1185	1.0986	0.0275	0.0271	0.0001	0.0001
Scenario 6:	CCA	1.4853	1.4854	1.4624	1.4633	1.4249	1.4199
	LS CCA	1.4589	1.4578	1.3797	1.3838	1.1954	1.1951
	CCA LB	1.4862	1.4851	1.4684	1.4740	1.4830	1.4793
	PMD	1.5244	1.5130	1.4985	1.4954	1.4553	1.4551
	Algorithm 2	1.4794	1.4791	1.4512	1.4509	1.3869	1.3790
	Algorithm 3	1.4633	1.4628	0.7775	0.7885	0.0220	0.0221



is to estimate the channel coefficients vector  $\mathbf{h}$ . The identification method presented by Xu et al. in [27] which is closely related to linear prediction exploits the commutativity of the convolution. Based on this approach and inspired from [28], we present in the following an experience to assess the performance of blind channel identification methods based on CCA.

From Eq. (23), the noise-free outputs  $x_i(n)$ ,  $i = 1, 2$  and using the commutativity of convolution, it follows :

$$h_2(t) * x_1(t) = h_1(t) * x_2(t). \tag{24}$$

In case the outputs  $x_i(t)$  are corrupted by additive noise, this property inspired the design of the identification diagram shown in Fig. 5, which allows to find estimates of the channels impulse response,  $\hat{h}_1$  and  $\hat{h}_2$ , by collecting

$T$  observations sample and minimizing the following cost function

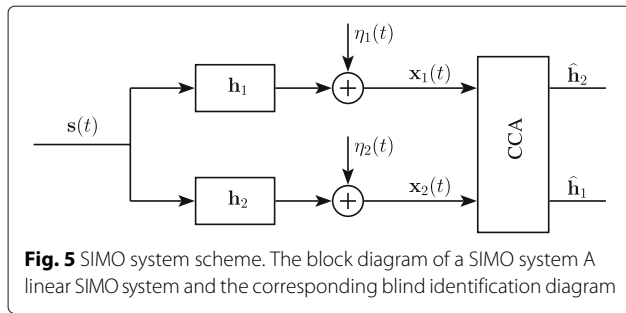
$$\begin{aligned} & \arg \min_{\mathbf{h}_1, \mathbf{h}_2} \|\mathbf{X}_1 \mathbf{h}_2 - \mathbf{X}_2 \mathbf{h}_1\|_2 \\ & \text{subject to} \quad \|\mathbf{X}_1 \mathbf{h}_1\|_2 = \|\mathbf{X}_2 \mathbf{h}_2\|_2 = 1, \end{aligned}$$

where

$$\mathbf{X}_i = \begin{bmatrix} x_i(L) & \dots & x_i(0) \\ \vdots & \ddots & \vdots \\ x_i(T-1) & \dots & x_i(T-L-1) \end{bmatrix} \quad i = 1, 2.$$

This problem is a canonical correlation analysis (CCA) problem.

Then, we present here some numerical simulations to assess the performance of the proposed algorithm. We consider a SIMO system with two outputs represented by



polynomial transfer function of degree  $L = 66$ . The channel impulse response is generated following 3GPP ETU (Extended Typical Urban) channel model [29] with frequency sampling 15.36 MHz which is used to model a channel impulse response for urban area in the context of wireless communications. The multipath delay profile for this channel is shown in Table 4.

The input signal is a BPSK i.i.d. sequence of length  $T = \{256, 1024\}$ . The observation is corrupted by the additive white Gaussian noise with a variance  $\sigma^2$  chosen such that the signal to noise ratio  $SNR = \frac{\|h\|^2}{\sigma^2}$  varies in the range  $[0, 40]$  in dB. Statistics are evaluated over  $N_k = 100$  Monte Carlo runs, and estimation performance are given by the normalized mean square error criterion :

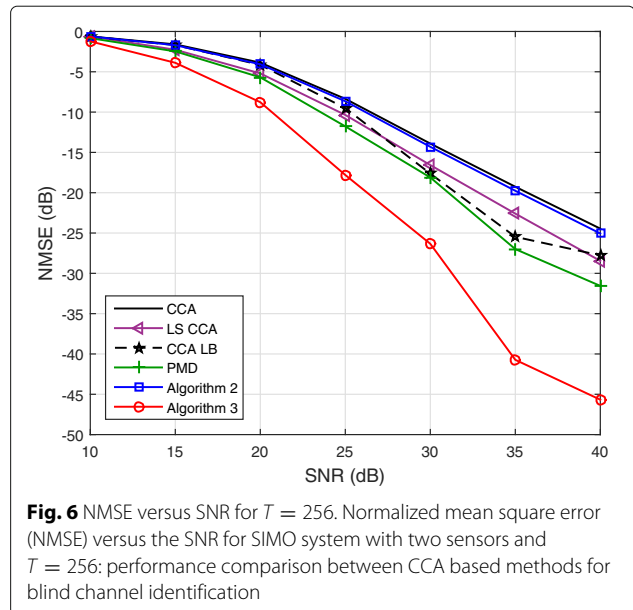
$$NMSE = \frac{1}{N_k} \sum_{k=1}^{N_k} 1 - \left( \frac{\hat{h}_k^T h}{\|\hat{h}_k\| \|h\|} \right)^2,$$

where  $\hat{h}_k$  denotes the estimated channel coefficient vector at the  $k^{th}$  Monte Carlo run. For each algorithm, we used the following parameters: LS CCA algorithm with  $\lambda_x = \lambda_y = 10^{-2}$ , CCA LB algorithm with  $\mu_x = \mu_y = 10^{-1}$ ; Algorithm 2 with  $\alpha_x = \alpha_y = 10^{-3}$ ; and Algorithm 3 with  $\beta_x = \beta_y = 10$ .

In Figs. 6 and 7, the normalized mean square error is plotted versus the SNR for the proposed approaches and state of the art algorithm. It is clearly shown that our sparse CCA based on rank-1 matrix approximation provide the best results for all SNR range and all observation length. Especially, we can observe that the proposed method outperforms the PMD algorithm [11] by 9 dB for moderate and high SNR. This results show the robustness of the proposed method against the additive noise and its fast convergence. Indeed, from Fig. 6, we can observe that the proposed sparse CCA method provide for moderate and high SNR a near-optimal performance even in the case of low observation size.

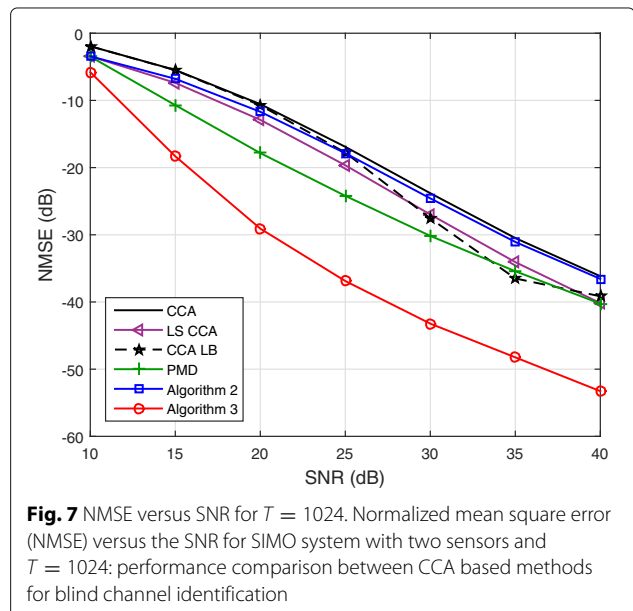
**Table 4** 3GPP extended typical urban channel model [29]

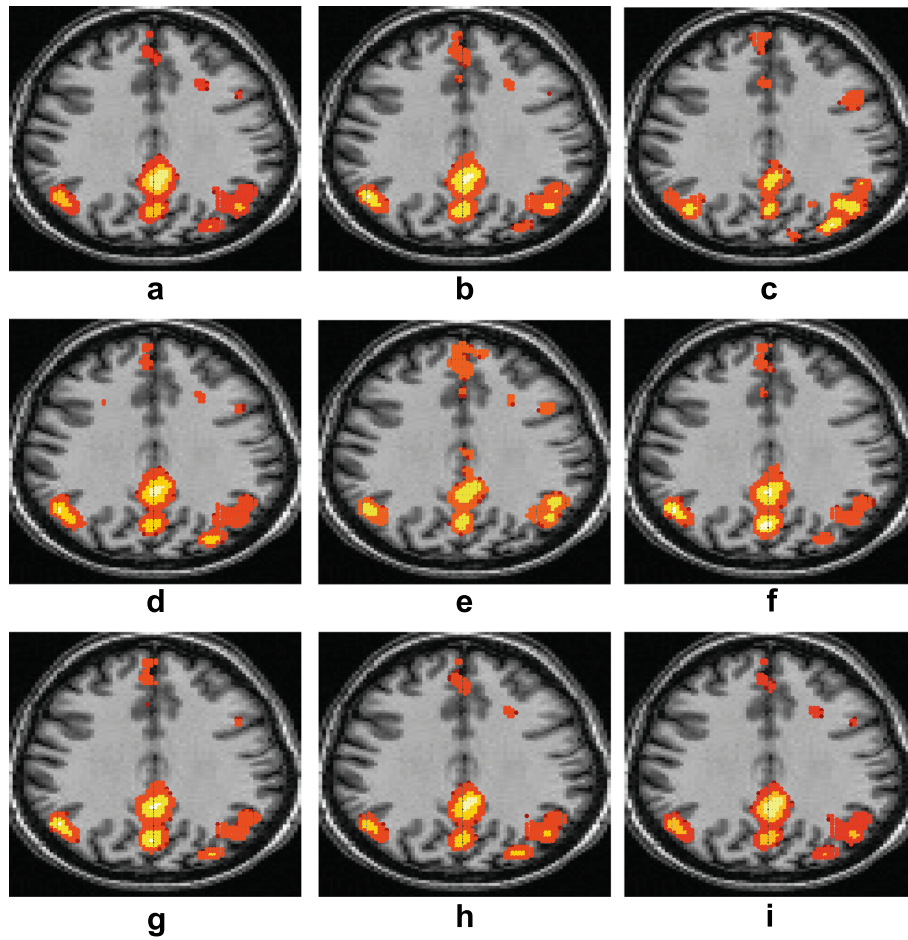
Excess tap delay (ns)	0	50	120	200	230	500	1600	2300	5000
Relative power (dB)	-1.0	-1.0	-1.0	0.0	0.0	0.0	-3.0	-5.0	-7.0



### 5.3 Blind source separation for fMRI signals

In this section, we evaluate the performance of the proposed CCA variant algorithms on a problem of functional magnetic resonance imaging (fMRI) resting state experiment (see Fig. 8 and Table 5). In this case, we are interested in functional connectivity and recovering a resting state network, i.e. the default mode network from a data matrix  $Y$  formed by vectorizing each time series observed in every voxel creating a matrix  $n \times N$  where  $n$  is the number of time points and  $N$  the number of voxels ( $\approx 10,000 - 100,000$ ) [30].





**Fig. 8** FMRI simulation results. The functional connectivity results of a single subject for default mode network (DMN) using eight different CCA variant algorithms. **a** Reference. **b** CCA. **c** LS CCA,  $\lambda_x = \lambda_y = 0.5$ . **d** CCA LB,  $\mu_x = \mu_y = 10$ . **e** PMD, **f** Algorithm 2,  $\alpha_x = \alpha_y = 10^{-4}$ . **g** Algorithm 2,  $\alpha_x = \alpha_y = 10^{-3}$ . **h** Algorithm 3,  $\beta_x = \beta_y = 3$ . **i** Algorithm 3,  $\beta_x = \beta_y = 4$

To use CCA, either a second data set obtained from a different subject is used or the second data set is obtained from the original data  $Y$  by time delay [31]. This last option is used in this application example. Instead of taking  $N$  as the total number of voxels, only the cortical, subcortical and cerebellum regions in the brain obtained by parcellating the whole brain into 116 ROIs using automated anatomical labelling [32] were considered. For each considered region, the average time series was generated and used.

The single subject (id 100307) rsfMRI dataset used in this section was obtained from the Human Connectome Project Q1 release [33]. The acquisition parameters of rsfMRI data are  $90 \times 104$  matrix, 220 mm FOV, 72

slices, TR = 0.72 s, TE = 33.1 ms, flip angle =  $52^\circ$ , BW = 2290Hz/Px, in-plane FOV =  $208 \times 180$  mm with 2.0 mm isotropic voxels. The obtained data was already preprocessed with the preprocessing pipeline consisting of motion correction, temporal pre-whitening, slice time correction and global drift removal, and the scans were spatially normalized to a standard MNI152 template and were resampled to  $2 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$  voxels. The reader is referred to [33, 34] for more details regarding data acquisition and preprocessing.

The second data set obtained by a single sample delay was used for CCA. The different CCA algorithms were applied on  $Y$  and  $Y_{t-1}$  of dimension  $n \times N$  to allow us to generate canonical correlation components representing

**Table 5** Performance comparison in terms of correlation with the reference Fig. 8 (a)

Algorithms	CCA	LS CCA	CCA LB	PMD	Algo 2 (f)	Algo 2 (g)	Algo 3 (h)	Algo 3 (i)
Correlation	0.9438	0.9054	0.9764	0.9235	0.9822	0.9852	0.9953	0.9995

maximally correlated temporal profile. The neural dynamics of interest can be obtained by correlating the modulation profile of the canonical correlation components with the time series representing average neural dynamics for regions of interest (ROIs). For functional connectivity analysis of the default mode network (DMN), the modulation profile that was most correlated with posterior cingulate cortex (PCC) representative time series is used. Using the neural dynamics of interest, sparsely distributed and clustered origin of the dynamics are obtained by converting the associated coefficient rows to  $z$ -scores.

Using the different CCA variant algorithms, the connected regions obtained for DMN are mostly PCC, medial pre-frontal cortex (MFC) and right inferior parietal lobe (IPL). As there is no gold standard reference for DMN connectivity available, therefore, we relied on the similarity of temporal dynamics of DMN-based modulation profile with PCC representative time series. The similarity measure used was correlation and estimated as  $> 0.9$  for all the algorithms.

## 6 Conclusions

In this paper, we have developed two new variants of CCA; more specifically, we have introduced new algorithms for sparse and smooth CCA. The proposed algorithms are based on penalized rank-1 matrix approximation and differ from the existing ones in the matrices they use for their derivation. Indeed, instead of focusing on the cross-matrix product of the two sets of multidimensional variables, we have used the product of the orthogonal projectors onto the space spanned by the columns of the two sets of multidimensional variables. Using this approach, the sparse and smooth CCA algorithms proposed differ only in the penalty used in the penalized rank-1 matrix approximation. Simulation results illustrating the effectiveness of the proposed CCA variant algorithms are provided where we can observe that proposed sparse CCA outperforms state of the art methods. As a continuation of the presented work and in order to fix the tuning parameters of the proposed approaches, the main idea of the permutation method presented in [18] will be studied and adapted.

## Endnotes

<sup>1</sup>Let  $A$  and  $B$  be two matrices. In order to compute the angle  $\theta$  between the subspaces spanned by the columns of  $A$  and  $B$ ; first, we compute an orthonormal basis  $A_{\perp}$  and  $B_{\perp}$  for the range of  $A$  and  $B$  respectively.  $\theta$  is computed by  $\theta = \arccos(\min(A_{\perp}^T B_{\perp}))$ .

## Funding

No funding was received or used to prepare this manuscript.

## Authors' contributions

All authors contributed equally to this work. All authors discussed the results and implications and commented on the manuscript at all stages. Both authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>IMT Atlantique, UMR CNRS 6285 Lab-STICC, Université Bretagne Loire, Brest 29238, France. <sup>2</sup>Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Melbourne 3010, Australia.

Received: 8 March 2016 Accepted: 24 February 2017

Published online: 09 March 2017

## References

- H Hotelling, Relations between two sets of variables. *Biometrika*. **28**(3–4), 321–377 (1936)
- W Zheng, X Zhou, C Zou, L Zhao, Facial expression recognition using kernel canonical correlation analysis (KCCA). *IEEE Trans. Neural Netw.* **17**(1), 233–238 (2006)
- XY Jing, S Li, C Lan, D Zhang, J Yang, Q Liu, Color image canonical correlation analysis for face feature extraction and recognition. *Signal Process.* **91**(8), 2132–2140 (2011)
- O Friman, J Carlsson, P Lundberg, M Borga, H Knutsson, Detection of neural activity in functional MRI using canonical correlation analysis. *Magn. Reson. Med.* **45**(2), 323–330 (2001)
- DR Hardoon, J Mourao-Miranda, M Brammer, J Shawe-Taylor, Unsupervised analysis of fMRI data using kernel canonical correlation. *NeuroImage*. **37**(4), 1250–1259 (2007)
- DR Hardoon, S Szedmak, J Shawe-Taylor, Canonical correlation analysis: an overview with application to learning methods. *Neural Comput.* **16**(12), 2639–2664 (2004)
- L Sun, S Ji, J Ye, Canonical correlation analysis for multilabel classification: a least-squares formulation, extensions, and analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(1), 194–200 (2011)
- W Liu, DP Mandic, A Cichocki, Analysis and online realization of the CCA approach for blind source separation. *IEEE Trans. Neural Netw.* **18**(5), 1505–1510 (2007)
- Y-O Li, T Adali, W Wang, VD Calhoun, Joint blind source separation by multiset canonical correlation analysis. *IEEE Trans. Signal Process.* **57**(10), 3918–3929 (2009)
- DR Hardoon, J Shawe-Taylor, Sparse canonical correlation analysis. *Mach. Learn.* **83**(3), 331–353 (2011). doi:10.1007/s10994-010-5222-7
- DM Witten, R Tibshirani, T Hastie, A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*. **10**(3), 515–534 (2009). doi:10.1093/biostatistics/kxp008
- D Chu, LZ Liao, MK Ng, X Zhang, Sparse canonical correlation analysis: new formulation and algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(12), 3050–3065 (2013)
- KV Mardia, JT Kent, JM Bibby, *Multivariate Analysis. Probability and mathematical statistics*, 1st edn. (Academic Press, University of Leeds, Leeds, 1979)
- AN Tikhonov, On the stability of inverse problems. *Doklady Akademii nauk SSSR*. **39**(5), 195–198 (1943)
- JO Ramsay, BW Silverman, *Functional Data Analysis*, 2nd edn. (Springer-Verlag, New York, 2005)
- K Lee, SK Tak, JC Yee, A data driven sparse GLM for fMRI analysis using sparse dictionary learning and MDL criterion. *IEEE Trans. Med. Imaging*. **30**, 1176–1089 (2011)
- A Aïssa-El-Bey, A-K Seghouane, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Sparse canonical correlation analysis based on rank-1 matrix approximation and its application for FMRI signals, (2016), pp. 4678–4682. doi:10.1109/ICASSP.2016.7472564
- S Gross, B Narasimhan, R Tibshirani, D Witten, Correlate: sparse canonical correlation analysis for the integrative analysis of genomic data. Technical Report User guide and technical document, Stanford University (2011)

19. JF Cai, S Osher, Z Shen, Convergence of the linearized bregman iteration for  $\ell_1$ -norm minimization. Technical Report CAM Report 08–52, University of California Los Angeles (2008)
20. YC Pati, R Rezaifar, PS Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, Proceedings of 27th Asilomar Conference on Signals, Systems and Computers, vol. 1, (Pacific Grove, 1993), pp. 40–44. doi:10.1109/ACSSC.1993.342465
21. G Davis, S Mallat, M Avellaneda, Adaptive greedy approximations. *Constr. Approximation*. **13**(1), 57–98 (1997). doi:10.1007/BF02678430
22. JA Branco, C Croux, P Filzmoser, MR Oliveira, Robust canonical correlations: a comparative study. *Comput. Stat.* **20**(2), 203–229 (2005). doi:10.1007/BF02789700
23. W Schreiber, Advanced television systems for terrestrial broadcasting: some problems and some proposed solutions. *Proc. IEEE*. **83**(6), 958–981 (1995)
24. M Kocic, D Brady, M Stojanovic, in *Proc. OCEANS*. Sparse equalization for real-time digital underwater acoustic communications, vol. 3, (San Diego, 1995), pp. 1417–1422
25. L Perros-Meilhac, E Moulines, K Abed-Meraim, P Chevalier, P Duhamel, Blind identification of multipath channels: a parametric subspace approach. *IEEE Trans. Signal Process.* **49**(7), 1468–1480 (2001)
26. S Ariyavisitakul, N Sollenberger, L Greenstein, Tap selectable decision-feedback equalization. *IEEE Trans. Commun.* **45**(12), 1497–1500 (1997)
27. G Xu, H Liu, L Tong, T Kailath, A least-squares approach to blind channel identification. *IEEE Trans. Signal Process.* **43**(12), 2982–2993 (1995)
28. S Van Vaerenbergh, J Via, I Santamaria, Blind identification of SIMO Wiener systems based on kernel canonical correlation analysis. *IEEE Trans. Signal Process.* **61**(9), 2219–2230 (2013)
29. 3GPP TS 36.104, Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) Radio Transmission and Reception (2015). 3GPP TS 36.104. [www.3gpp.org/dynareport/36104.htm](http://www.3gpp.org/dynareport/36104.htm)
30. NA Lazar, *Statistics for Biology and Health. The Statistical Analysis of Functional MRI Data*, 1st edn. (Springer, New York, 2008)
31. MU Khaled, AK Seghouane. Improving functional connectivity detection in fMRI by combining sparse dictionary learning and canonical correlation analysis, 10th IEEE International Symposium on Biomedical Imaging, (San Francisco, 2013), pp. 286–289. doi:10.1109/ISBI.2013.6556468
32. N Tzourio-Mazoyer, B Landeau, D Papathanassiou, F Crivello, O Etard, N Delcroix, B Mazoyer, M Joliot, Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the mni mri single-subject brain. *NeuroImage*. **15**, 273–289 (2002)
33. DM Barch, GC Burgess, MP Harms, SE Petersen, BL Schlaggar, M Corbetta, MF Glasser, S Curtiss, S Dixit, C Feldt, D Nolan, E Bryant, T Hartley, O Footer, JM Bjork, R Poldrack, S Smith, H Johansen-Berg, AZ Snyder, DCV Essen, Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage*. **80**, 169–189 (2013)
34. MF Glasser, SN Sotiropoulos, JA Wilson, TS Coalson, B Fischl, JL Andersson, J Xu, S Jbabdi, M Webster, JR Polimeni, DCV Essen, M Jenkinson, The minimal preprocessing pipelines for the human connectome project. *NeuroImage*. **80**, 105–124 (2013)

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---