



**HAL**  
open science

# Canonical correlation analysis based on sparse penalty and through rank-1 matrix approximation

Abdeldjalil Aissa El Bey, Abd-Krim Seghouane

► **To cite this version:**

Abdeldjalil Aissa El Bey, Abd-Krim Seghouane. Canonical correlation analysis based on sparse penalty and through rank-1 matrix approximation. [Research Report] RR-2015-01-SC, Dépt. Signal et Communications (Institut Mines-Télécom-Télécom Bretagne-UEB); Laboratoire en sciences et technologies de l'information, de la communication et de la connaissance (UMR CNRS 6285 - Télécom Bretagne - Université de Bretagne Occidentale - Université de Bretagne Sud); Department of Electrical and Electronic Engineering (Melbourne) (University of Melbourne). 2015, pp.20. hal-01202844

**HAL Id: hal-01202844**

**<https://hal.science/hal-01202844>**

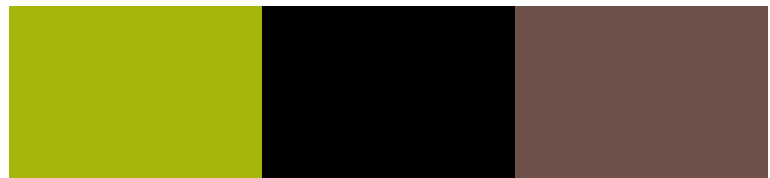
Submitted on 21 Sep 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Collection des rapports de recherche de  
Télécom Bretagne

**RR-2015-01-SC**



**Canonical correlation analysis based on  
sparse penalty and through rank-1 matrix  
approximation**

*Abdeljalil Aïssa-El-Bey (Télécom Bretagne)  
Abd-Krim Seghouane (University of Melbourne)*

---

# Canonical Correlation Analysis Based on Sparse Penalty and Through Rank-1 Matrix Approximation

Abdeldjalil Aïssa-El-Bey and Abd-Krim Seghouane

## 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 one matrix approximation and the orthogonal projectors onto the space spanned by the columns of 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 outperforms the state of the art sparse CCA algorithms.

## 1 Introduction

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

---

In the context of high-dimensional data, there are usually a large portion of features that are not informative in data analysis. When the canonical variables involve all features in the original space, the canonical projectors are in general not sparse. Therefore, it is not easy to interpret canonical variables in such high-dimensional data analysis. These problems may be tackled by selecting sparse subsets of variables, i.e. obtaining sparse canonical projectors in the linear combinations of variables of each data set [6, 9–11]. For example, in [10] the authors propose a new criterion for sparse CCA and applied a penalized matrix decomposition approach to solve the sparse CCA problem and in [9] the presented sparse CCA approach compute the canonical projectors from primal and dual representations.

In this paper, we adopt an alternative formulation of CCA problem which is based on rank-1 matrix approximation of the orthogonal projectors of data sets [12]. Based on this new formulation of CCA problem, we developed a new sparse CCA based on penalized rank-1 matrix approximation which aims to overcome the drawback of CCA in the context of high-dimensional data and improve interpretability. The proposed sparse CCA seeks to obtain iteratively a sparse pairwise of canonical projectors by solving a penalized rank-1 matrix approximation via sparse coding method. Also, we present in this paper a smoothed version of CCA problem based on rank-1 matrix approximation where we impose some smoothness on the projections of the variables in order to avoid abrupt or sudden variations. These proposed algorithms differ from the existing ones in their derivation which is based on penalized rank one matrix approximation and the orthogonal projectors onto the space spanned by the columns of the two sets of multidimensional variables instead of the simple cross-matrix product [6, 9–11].

The rest of the paper is organized as follow: 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  $i^{th}$  component of  $\mathbf{x}$ .

## 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 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  $\mathbf{w}_y^T \mathbf{y}$  are 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:

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

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

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

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

This 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 [5] 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 equation (6) and assuming  $C_{xx}$  is invertible gives

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

It has been shown in [5] that we can choose the associated eigenvectors corresponding to the top eigenvalues of the generalized eigenvalue problem in (9) and then use (8) to 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  $C_{xy}$  for the derivation of new CCA variant algorithms [6, 9–11]. For the derivation of the proposed CCA variants we adopt an alternative description of CCA which is based on the orthogonal projectors onto the space spanned by the columns of the two sets of multidimensional variables [12].

### 3 Canonical Correlation Analysis based on rank-1 matrix approximation

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

$$\begin{aligned} \arg \max_{\mathbf{w}_x, \mathbf{w}_y} \quad & \mathbf{w}_x^T \mathbf{X} \mathbf{Y}^T \mathbf{w}_y \\ \text{subject to} \quad & \mathbf{w}_x^T \mathbf{X} \mathbf{X}^T \mathbf{w}_x = 1, \quad \mathbf{w}_y^T \mathbf{Y} \mathbf{Y}^T \mathbf{w}_y = 1, \end{aligned} \quad (10)$$

and the generalized eigenvalue problem given by equations (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 the both side of equations (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 columns of  $\mathbf{X}$  and  $\mathbf{Y}$  respectively. So substituting  $\mathbf{X}^T\mathbf{w}_x$  in equation (14) and

$\mathbf{Y}^T \mathbf{w}_y$  in equation (13) gives

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

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

$$\arg \min_{\mathbf{w}_x, \mathbf{w}_y} \left\| \mathbf{K}_{xy} - \mathbf{X}^T \mathbf{w}_x \mathbf{w}_y^T \mathbf{Y} \right\|_F^2 \quad (15)$$

where  $\|\cdot\|_F^2$  is the squared Frobenius norm. Consequently, the projected data  $\mathbf{w}_x^T \mathbf{X}$  and  $\mathbf{w}_y^T \mathbf{Y}$  consist on the eigenvectors associated to the largest eigenvalue of the matrix  $\mathbf{K}_{xy}$ . Hence, for multiple projected data the solution consist on the associated eigenvectors corresponding to the top eigenvalues of the matrix  $\mathbf{K}_{xy}$ .

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  $\mathbf{Y} \mathbf{Y}^T$  [5], 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

---

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

---

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

**Output:** 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  $\mathbf{P}_x = \mathbf{X}^T (\mathbf{X} \mathbf{X}^T + \gamma_x \mathbf{I}_{d_x})^{-1} \mathbf{X}$ ,  $\mathbf{P}_y = \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + \gamma_y \mathbf{I}_{d_y})^{-1} \mathbf{Y}$  and  $\mathbf{K}_{xy} = \mathbf{P}_x \mathbf{P}_y$ ;
  - 2: Perform the SVD of  $\mathbf{K}_{xy}$ :  $\mathbf{K}_{xy} = \mathbf{U} \mathbf{D} \mathbf{V}^T$ ;
  - 3: Form  $\tilde{\mathbf{U}} = [\mathbf{u}_1, \dots, \mathbf{u}_r]$ ,  $\tilde{\mathbf{D}} = \text{diag}(\sqrt{d_1}, \dots, \sqrt{d_r})$  and  $\tilde{\mathbf{V}} = [\mathbf{v}_1, \dots, \mathbf{v}_r]$ ;
  - 4: 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  $\mathbf{W}_y = (\mathbf{Y} \mathbf{Y}^T + \gamma_y \mathbf{I}_{d_y})^{-1} \mathbf{Y} \tilde{\mathbf{V}} \tilde{\mathbf{D}}$ .
- 

### 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 equation (15) such that:

$$\arg \min_{\mathbf{w}_x, \mathbf{w}_y} \left\| \mathbf{K}_{xy} - \mathbf{X}^T \mathbf{w}_x \mathbf{w}_y^T \mathbf{Y} \right\|_F^2 + \alpha_x \mathbf{w}_x^T \mathbf{X} \Omega_x \mathbf{X}^T \mathbf{w}_x + \alpha_y \mathbf{w}_y^T \mathbf{Y} \Omega_y \mathbf{Y}^T \mathbf{w}_y \quad (16)$$

In many cases, the matrices  $\Omega_x$  and  $\Omega_y$  are chosen as a multiple of the identity matrix, giving preference to solutions with smaller norms. In our case, the matrices  $\Omega_x$  and  $\Omega_y$  are a non-negative

definite roughness penalty matrices used to penalize the second differences [13] [14] and  $\alpha_x > 0$  and  $\alpha_y > 0$  are trade-off parameters such as:

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

The choice of such matrices may be used to enforce smoothness if the underlying vector is believed to be mostly continuous. Therefore, the criterion of equation (16) can be rewritten as

$$\arg \min_{\mathbf{w}_x, \mathbf{w}_y} \|\mathbf{X}^T \mathbf{w}_x\|_2^2 \|\mathbf{Y}^T \mathbf{w}_y\|_2^2 - 2 \mathbf{w}_x^T \mathbf{X} \mathbf{K}_{xy} \mathbf{Y}^T \mathbf{w}_y + \alpha_x \mathbf{w}_x^T \mathbf{X} \boldsymbol{\Omega}_x \mathbf{X}^T \mathbf{w}_x + \alpha_y \mathbf{w}_y^T \mathbf{Y} \boldsymbol{\Omega}_y \mathbf{Y}^T \mathbf{w}_y \quad (17)$$

The optimization problem (17) 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 (17). Then, we fix  $\mathbf{w}_x$  and minimize (17) to obtain  $\mathbf{w}_y$ . The above two procedures are repeated until convergence. Taking derivatives in respect to  $\mathbf{w}_x$  and  $\mathbf{w}_y$ , we obtain

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

Therefore, we obtain  $\mathbf{w}_x$  and  $\mathbf{w}_y$  by solving the above equations in least square sense. For multiple canonical projectors, we propose to use a deflation procedure where the second pairwise of canonical projectors are defined by using the corresponding residual matrices  $\mathbf{K}_{xy} - \mathbf{X}^T \mathbf{w}_x \mathbf{K}_{xy} \mathbf{w}_y^T \mathbf{Y} \mathbf{w}_x^T \mathbf{X} \mathbf{Y}^T \mathbf{w}_y$ . Then, we can define the other pairwise of sparse projectors. The method for solving the smoothed rank-1 matrix approximation CCA is summarized by Algorithm 2.

## 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 (15). Then, we propose an efficient iterative algorithm to solve the sparse solution of the proposed criterion.

In general cases, the canonical projectors  $\mathbf{w}_x$  and  $\mathbf{w}_y$  solutions of equation (15) are not sparse, i.e., the entries of both  $\mathbf{w}_x$  and  $\mathbf{w}_y$  are nonzeros. To obtain the sparse solution, we adopt the similar trick used in [6, 10, 11] by imposing penalty functions on the optimization problem (15). Therefore, we can write the new optimization problem as:

$$\arg \min_{\mathbf{w}_x, \mathbf{w}_y} \|\mathbf{K}_{xy} - \mathbf{X}^T \mathbf{w}_x \mathbf{w}_y^T \mathbf{Y}\|_F^2 \quad \text{subject to} \quad \mathcal{F}_x(\mathbf{w}_x) \leq \beta_x \quad \text{and} \quad \mathcal{F}_y(\mathbf{w}_y) \leq \beta_y \quad (18)$$



---

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

---

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

**Output:** 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  $\mathbf{P}_x = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T + \gamma_x \mathbf{I}_{d_x})^{-1}\mathbf{X}$ ,  $\mathbf{P}_y = \mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \gamma_y \mathbf{I}_{d_y})^{-1}\mathbf{Y}$  and  $\mathbf{K}_{xy} = \mathbf{P}_x \mathbf{P}_y$ ;
  - 2: **for**  $i = 1, 2, \dots, r$  **do**
  - 3:     Perform the SVD of  $\mathbf{K}_{xy}$  :  $\mathbf{K}_{xy} = \mathbf{U} \mathbf{D} \mathbf{V}^T$ ;
  - 4:     Initialize  $\tilde{\mathbf{u}} = \mathbf{u}_1$  and  $\tilde{\mathbf{v}} = \mathbf{v}_1$ ;
  - 5:     **repeat**
  - 6:         Update the  $i$ -th column of  $\mathbf{W}_x$  :  $\mathbf{W}_x(:, i) = \left( \mathbf{X}\mathbf{X}^T + \alpha_x \mathbf{X}\Omega_x \mathbf{X}^T + \gamma_x \mathbf{I}_{d_x} \right)^{-1} \mathbf{X} \mathbf{K}_{xy} \tilde{\mathbf{v}}$ ;
  - 7:         Update  $\tilde{\mathbf{u}} = \frac{\mathbf{X}^T \mathbf{W}_x(:, i)}{\|\mathbf{X}^T \mathbf{W}_x(:, i)\|_2}$ ;
  - 8:         Update the  $i$ -th column of  $\mathbf{W}_y$  :  $\mathbf{W}_y(:, i) = \left( \mathbf{Y}\mathbf{Y}^T + \alpha_y \mathbf{Y}\Omega_y \mathbf{Y}^T + \gamma_y \mathbf{I}_{d_y} \right)^{-1} \mathbf{Y} \mathbf{K}_{xy}^T \tilde{\mathbf{u}}$ ;
  - 9:         Update  $\tilde{\mathbf{v}} = \frac{\mathbf{Y}^T \mathbf{W}_y(:, i)}{\|\mathbf{Y}^T \mathbf{W}_y(:, i)\|_2}$ ;
  - 10:        **until** convergence
  - 11:     Update  $\mathbf{K}_{xy}$  :  $\mathbf{K}_{xy} \leftarrow \mathbf{K}_{xy} - \tilde{\mathbf{u}}^T \mathbf{K}_{xy} \tilde{\mathbf{v}} \tilde{\mathbf{v}}^T$ ;
  - 12: **end for**
- 

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 nonzero entries of a vector; Lasso penalty with  $\ell_1$ -norm  $\mathcal{F}(z) = \|z\|_1$  and so on.

The optimization problem (18) 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 (18). Then, we fix  $\mathbf{w}_x$  and minimize (18) to obtain  $\mathbf{w}_y$ . 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  $\mathbf{w}_y$  the problem (18) is equivalent to much simpler sparse coding problem

$$\arg \min_{\mathbf{w}_x} \left\| \mathbf{K}_{xy} \mathbf{Y}^T \mathbf{w}_y - \mathbf{X}^T \mathbf{w}_x \right\|_2^2 \quad \text{subject to} \quad \mathcal{F}_x(\mathbf{w}_x) \leq \beta_x$$

which can be solved by using any sparse approximation method. In the same way, we can solve the problem (18) regarding  $\mathbf{w}_y$  for a fix  $\mathbf{w}_x$  by minimizing the following criterion:

$$\arg \min_{\mathbf{w}_y} \left\| \mathbf{K}_{xy}^T \mathbf{X}^T \mathbf{w}_x - \mathbf{Y}^T \mathbf{w}_y \right\|_2^2 \quad \text{subject to} \quad \mathcal{F}_y(\mathbf{w}_y) \leq \beta_y$$

Based on the above description, we can obtain the first pairwise of sparse projectors  $\mathbf{w}_x$  and  $\mathbf{w}_y$ . For multiple projection vectors, we propose to use a deflation procedure as presented in Section 3.1 where the second pairwise of sparse projectors are defined by using the corresponding resid-

---

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

---

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

**Output:** 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  $\mathbf{P}_x = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T + \gamma_x \mathbf{I}_{d_x})^{-1}\mathbf{X}$ ,  $\mathbf{P}_y = \mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \gamma_y \mathbf{I}_{d_y})^{-1}\mathbf{Y}$  and  $\mathbf{K}_{xy} = \mathbf{P}_x \mathbf{P}_y$ ;
  - 2: **for**  $i = 1, 2, \dots, r$  **do**
  - 3:     Perform the SVD of  $\mathbf{K}_{xy}$  :  $\mathbf{K}_{xy} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ ;
  - 4:     Initialize  $\tilde{\mathbf{u}} = \mathbf{u}_1$  and  $\tilde{\mathbf{v}} = \mathbf{v}_1$ ;
  - 5:     **repeat**
  - 6:         Update the  $i$ -th column of  $\mathbf{W}_x$  :  $\mathbf{W}_x(:, i) = \arg \min_{\mathbf{W}_x(:, i)} \|\mathbf{K}_{xy}\tilde{\mathbf{v}} - \mathbf{X}^T \mathbf{W}_x(:, i)\|_2^2$  subject to  $\mathcal{F}_x(\mathbf{W}_x(:, i)) \leq \beta_x$ ;
  - 7:         Update  $\tilde{\mathbf{u}} = \frac{\mathbf{X}^T \mathbf{W}_x(:, i)}{\|\mathbf{X}^T \mathbf{W}_x(:, i)\|_2}$ ;
  - 8:         Update the  $i$ -th column of  $\mathbf{W}_y$  :  $\mathbf{W}_y(:, i) = \arg \min_{\mathbf{W}_y(:, i)} \|\mathbf{K}_{xy}^T \tilde{\mathbf{u}} - \mathbf{Y}^T \mathbf{W}_y(:, i)\|_2^2$  subject to  $\mathcal{F}_y(\mathbf{W}_y(:, i)) \leq \beta_y$ ;
  - 9:         Update  $\tilde{\mathbf{v}} = \frac{\mathbf{Y}^T \mathbf{W}_y(:, i)}{\|\mathbf{Y}^T \mathbf{W}_y(:, i)\|_2}$ ;
  - 10:        **until** convergence
  - 11:     Update  $\mathbf{K}_{xy}$  :  $\mathbf{K}_{xy} \leftarrow \mathbf{K}_{xy} - \tilde{\mathbf{u}}^T \mathbf{K}_{xy} \tilde{\mathbf{v}} \tilde{\mathbf{u}} \tilde{\mathbf{v}}^T$ ;
  - 12: **end for**
- 

ual matrices  $\mathbf{K}_{xy} - \mathbf{X}^T \mathbf{w}_x \mathbf{K}_{xy} \mathbf{w}_y^T \mathbf{Y} \mathbf{w}_x^T \mathbf{X} \mathbf{Y}^T \mathbf{w}_y$ . Using the same way, we can define the other pairwise of sparse projectors.

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

## 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 [10], relying on a penalized matrix decomposition denoted PMD. An R package implementing this algorithm, called PMA, is available at <http://cran.r-project.org/web/packages/PMA/index.html>. Sparsity parameters are selected using the permutation approach presented in [15] of which the code is pro-

---

vided in PMA package.

- The sparse CCA presented in [6] 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 [11] where the sparse canonical projectors are computed by solving two  $\ell_1$ -minimization problems by using Linearized Bregman iterative method [16]. This algorithm is denoted CCA LB (Linearized Bregman). We re-implemented the sparse CCA algorithm proposed [11] in Matlab.

For proposed sparse CCA algorithm, we have used  $\mathcal{F}_x(\mathbf{z}) = \mathcal{F}_y(\mathbf{z}) = \|\mathbf{z}\|_0$  as penalty functions.

## 5.1 Synthetic data

This simulation setup is inspired from [17]. 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  $N_r = 1000$ . We compare the performance of our algorithm to methods of the state of the art by estimating the precision accuracy of the space spanned by estimated canonical projectors. Then, we compute for each simulation run  $r$  the angle  $\theta^r(\hat{\mathbf{W}}_x^r, \mathbf{W}_x)$  between the subspace spanned by the estimated canonical projectors contained in the columns of  $\hat{\mathbf{W}}^r$  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  $N_r$  Monte-Carlo run such that:

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

Table 1: Simulation settings

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

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 Table 2.

Table 2: Simulation results

		$\theta_x$	$\theta_y$	$\theta_x$	$\theta_y$	$\theta_x$	$\theta_y$
Method		$N = 50$		$N = 100$		$N = 200$	
<b>Scenario 1:</b>	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
<b>Scenario 2:</b>	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
<b>Scenario 3:</b>	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
<b>Scenario 4:</b>	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

We can observe that the simulation accuracy of the proposed sparse CCA method is significantly better compared to other CCA methods, especially for a large number of observations  $N$ . In the case of low number of observations the proposed sparse CCA method is still doing well and where

---

the performance gain increases with increasing number of observations. This demonstrates the robustness of our sparse CCA method with respect to the number of available observations and the benefit of using our sparse CCA method in the context of a relatively low number of observations

## 5.2 Blind 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), underwater acoustic communications and wireless communications. Then, 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 [18–23].

Let 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 2 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  $L$  is the maximal channel length (which is assumed to be known). Such a system model can be described as follows :

$$\begin{cases} x_1(t) = s(t) * h_1(t) + \eta_1(t) \\ x_2(t) = s(t) * h_2(t) + \eta_2(t) \end{cases} \quad (19)$$

where  $*$  denotes linear convolution,  $\boldsymbol{\eta}(t) = [\eta_1(t), \eta_2(t)]^T$  is an additive spatial white Gaussian noise, i.e.  $\mathbb{E}[\boldsymbol{\eta}(t)\boldsymbol{\eta}(t)^T] = \sigma^2 \mathbf{I}_2$  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 is to estimate the channel coefficients vector  $\mathbf{h}$ . The identification method presented by Xu et al. in [24] which is closely related to linear prediction, exploits the commutativity of the convolution. Based on this approach and inspired from [25], we present in the following an experience to asses the performance of blind channel identification methods based on CCA.

Then, from equation (19), the noise-free outputs  $x_i(n)$ ,  $i = 1, 2$  and using the commutativity of convolution, it follows :

$$h_2(t) * x_1(t) = h_1(t) * x_2(t) , \quad (20)$$

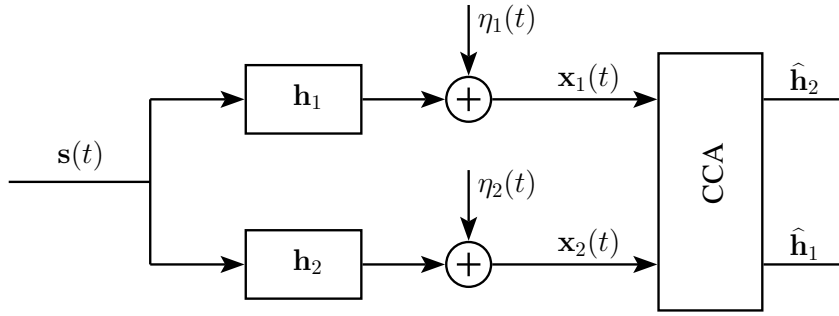


Figure 1: The block diagram of a SIMO system A linear SIMO system and the corresponding blind identification diagram.

In case the outputs  $x_i(t)$  are corrupted by additive noise, this property inspired the design of the identification diagram shown in Figure 1, 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

$$\begin{aligned} & \arg \min_{h_1, h_2} \quad \|\mathbf{X}_1 \mathbf{h}_2 - \mathbf{X}_2 \mathbf{h}_1\|_2 \\ & \text{subject to} \quad \|\mathbf{X}_1 \mathbf{h}_1\|_2 = \|\mathbf{X}_2 \mathbf{h}_2\|_2 = 1. \end{aligned}$$

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 2 outputs represented by polynomial transfer function of degree  $L = 66$ . The channel impulse response is generated following 3GPP ETU (Extended Typical Urban) channel model [26] 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 the Table 3:

Table 3: 3GPP Extended Typical Urban channel model [26]

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

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

$$\text{NMSE} = \frac{1}{N_r} \sum_{r=1}^{N_r} 1 - \left( \frac{\widehat{\mathbf{h}}_r^T \mathbf{h}}{\|\widehat{\mathbf{h}}_r\| \|\mathbf{h}\|} \right)^2,$$

where  $\widehat{\mathbf{h}}_r$  denotes the estimated channel coefficient vector at the  $r^{\text{th}}$  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  $\beta_x = \beta_y = 10$ .

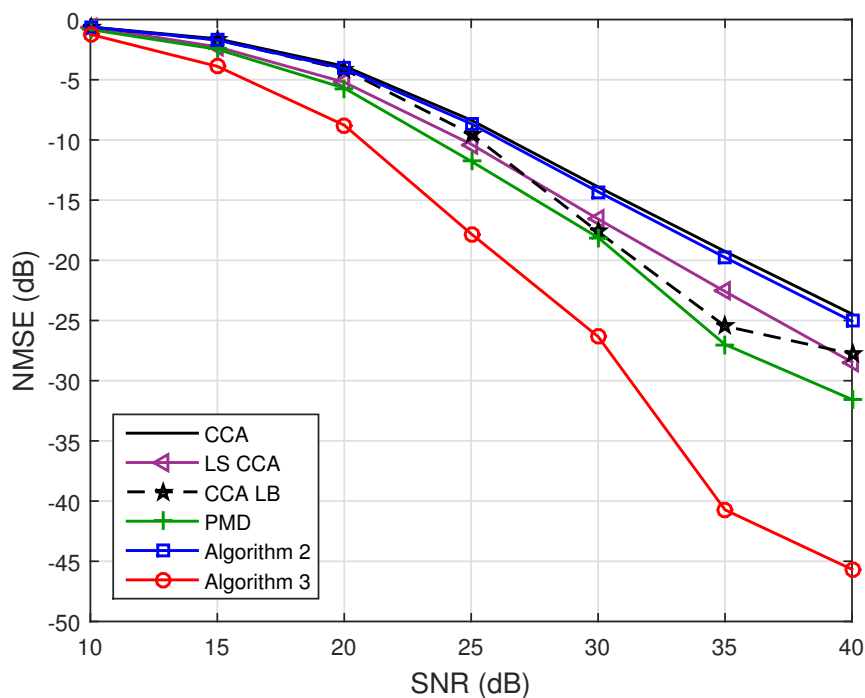


Figure 2: Normalized mean-square error (NMSE) versus the SNR for SIMO system with 2 sensors and  $T = 256$ : performance comparison between CCA based methods for blind channel identification.



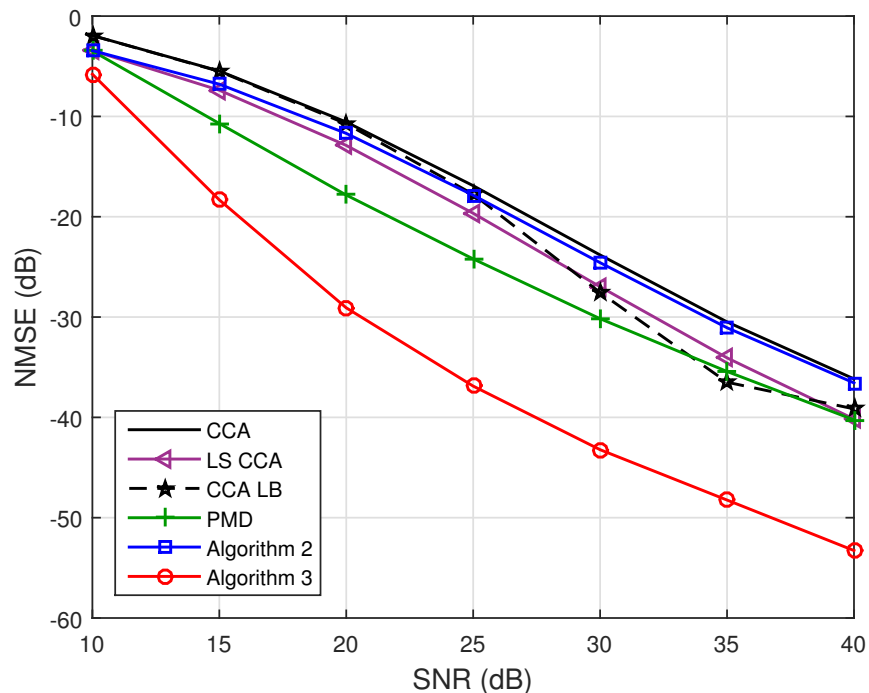


Figure 3: Normalized mean-square error (NMSE) versus the SNR for SIMO system with 2 sensors and  $T = 1024$ : performance comparison between CCA based methods for blind channel identification.

In Figures 2 and 3, 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 [10] 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 Figure 2 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.

### 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. 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  $\mathbf{Y}$  formed by vectorizing each 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

---

( $\approx 10,000 - 100,000$ ) [27].

To estimate functionally connected brain voxels, response signal strength known as coefficient matrix estimated as  $\mathbf{X}$  is considered [28]. According to the neural dynamics of interest, coefficient rows can be converted to z-scores to obtain sparsely distributed and clustered origin of the dynamics. The neural dynamics of interest can be obtained by correlating the modulation profile with the time-series representing average neural dynamics for regions of interest (ROIs). The representative time-series for cortical, subcortical, and cerebellum regions in the brain were obtained by parcelating the whole brain into 116 ROIs using automated anatomical labelling [29].

Only the first functional run from the first subject was used for functional connectivity analysis of a default mode network (DMN). The functional connectivity results of a single subject for DMN using eight different CCA variant algorithms are shown in figure 4. To obtain these results the modulation profile that was most correlated with posterior cingulate cortex (PCC) representative time-series is used. Using the different CCA variant algorithms, the connected regions obtained for DMN are mostly PCC, medial pre-frontal cortex (MFC), and right inferior parietal lobe (IPL), where as for rest of the proposed algorithms. As there is no gold standard reference for DMN connectivity available, therefore, we relied on the similarity of temporal dynamics of DMN based modulation profile with PCC representative time-series. The similarity measure used was correlation and estimated as  $> 0.9$  for all the algorithms.

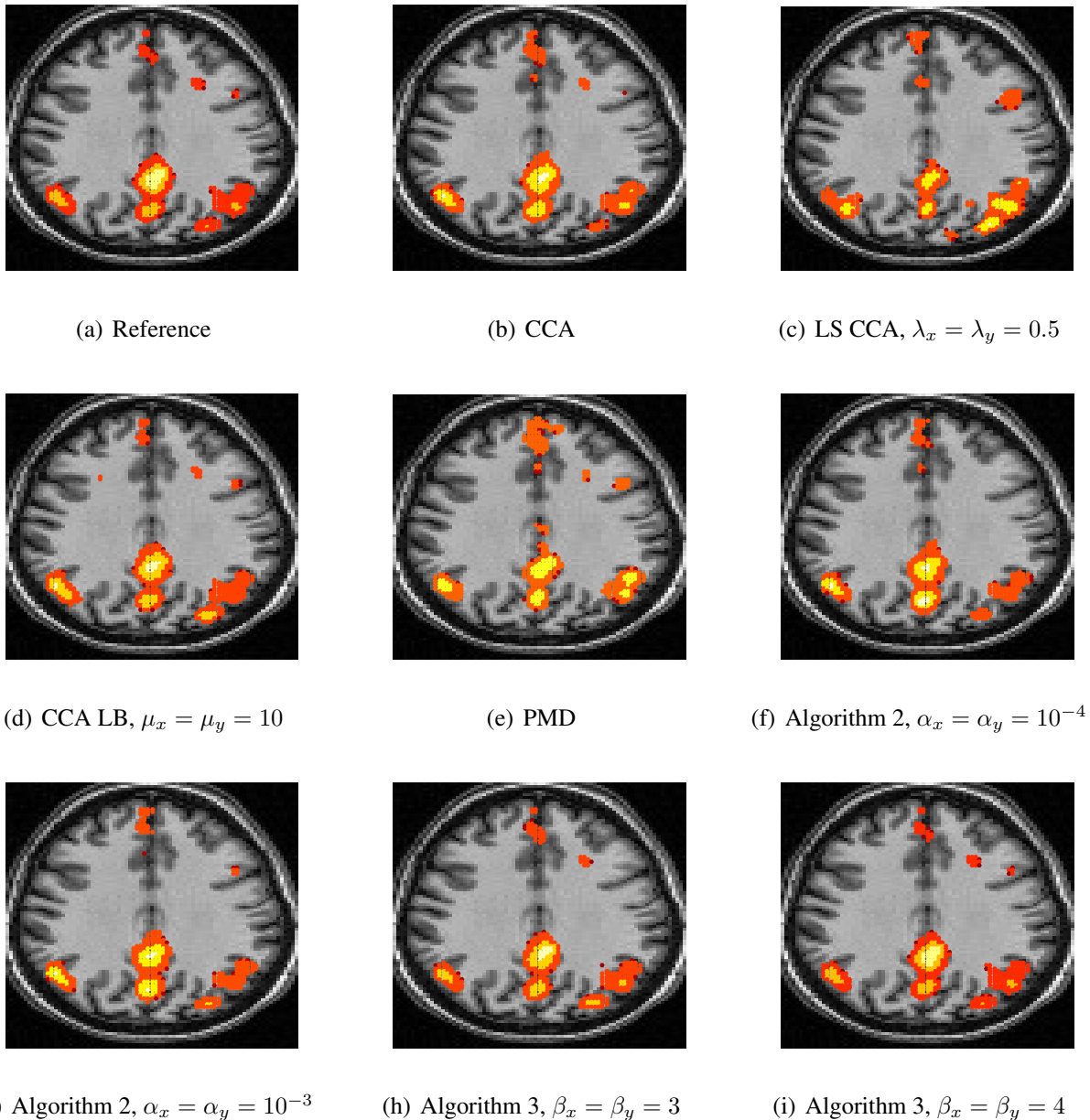


Figure 4: The functional connectivity results of a single subject for default mode network (DMN) using eight different CCA variant algorithms.

## 6 Conclusion

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 one approximation and differ from existing ones in the matrices they use for their derivation. Indeed instead of focusing on the cross-matrix product of the two sets of multidimensional variables we

---

have used the product of the orthogonal projectors onto the space spanned by the columns of the two sets of multidimensional variables. Using this approach the sparse and smooth CCA algorithms proposed differ only the penalty used in the penalized rank one 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.

## References

- [1] H. Hotelling, “Relations between two sets of variables,” *Biometrika*, vol. 28, no. 3-4, pp. 321–377, 1936.
- [2] W. Zheng, X. Zhou, C. Zou, and L. Zhao, “Facial expression recognition using kernel canonical correlation analysis (KCCA),” *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 233–238, January 2006.
- [3] O. Friman, J. Carlsson, P. Lundberg, M. Borga, and H. Knutsson, “Detection of neural activity in functional MRI using canonical correlation analysis,” *Magnetic Resonance in Medicine*, vol. 45, no. 2, pp. 323–330, February 2001.
- [4] D. R. Hardoon, J. Mourao-Miranda, M. Brammer, and J. Shawe-Taylor, “Unsupervised analysis of fmri data using kernel canonical correlation,” *NeuroImage*, vol. 37, no. 4, pp. 1250 – 1259, 2007.
- [5] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, “Canonical correlation analysis: An overview with application to learning methods,” *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, December 2004.
- [6] L. Sun, S. Ji, and J. Ye, “Canonical correlation analysis for multilabel classification: A least-squares formulation, extensions, and analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 194–200, January 2011.
- [7] Wei Liu, D.P. Mandic, and A. Cichocki, “Analysis and online realization of the cca approach for blind source separation,” *IEEE Transactions on Neural Networks*, vol. 18, no. 5, pp. 1505–1510, September 2007.
- [8] Yi-Ou Li, T. Adali, Wei Wang, and V.D. Calhoun, “Joint blind source separation by multiset canonical correlation analysis,” *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, October 2009.

- 
- [9] D. R. Hardoon and J. Shawe-Taylor, “Sparse canonical correlation analysis,” *Machine Learning*, vol. 83, no. 3, pp. 331–353, June 2011.
- [10] D. M. Witten, R. Tibshirani, and T. Hastie, “A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis,” *Biostatistics*, vol. 10, no. 3, pp. 515–534, July 2009.
- [11] D. Chu, L.Z. Liao, M.K. Ng, and X. Zhang, “Sparse canonical correlation analysis: New formulation and algorithm,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 3050–3065, December 2013.
- [12] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, Academic Press, 1979.
- [13] A.N. Tikhonov, “On the stability of inverse problems,” *Doklady Akademii nauk SSSR*, vol. 39, no. 5, pp. 195–198, 1943.
- [14] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer-Verlag, 2005.
- [15] S. Gross, B. Narasimhan, R. Tibshirani, and D. Witten, “Correlate: Sparse canonical correlation analysis for the integrative analysis of genomic data,” Tech. Rep. User guide and technical document, Stanford University, 2011.
- [16] J.F. Cai, S. Osher, and Z. Shen, “Convergence of the linearized bregman iteration for  $\ell_1$ -norm minimization,” Tech. Rep. CAM Report 08-52, University of California Los Angeles, 2008.
- [17] J.A. Branco, C. Croux, P. Filzmoser, and M.R. Oliveira, “Robust canonical correlations: A comparative study,” *Computational Statistics*, vol. 20, no. 2, pp. 203–229, 2005.
- [18] A. Aïssa-El-Bey, M. Grebici, K. Abed-Meraim, and A. Belouchrani, “Blind system identification using cross-relation methods: further results and developments,” in *Proc. 7<sup>th</sup> International Symposium on Signal Processing and Its Applications ISSPA*, Paris, France, July 2003, vol. 1, pp. 649–652.
- [19] A. Aïssa-El-Bey and K. Abed-Meraim, “Blind SIMO channel identification using a sparsity criterion,” in *9th IEEE international workshop on signal processing advances in wireless communications, SPAWC*, Recife, Brazil, July 2008, pp. 271 – 275.
- [20] A. Aïssa-El-Bey and K. Abed-Meraim, “Blind identification of sparse SIMO channels using maximum a posteriori approach,” in *16th European Signal Processing Conference, EU-SIPCO*, Lausanne, Switzerland, August 2008.

- 
- [21] A. Kammoun, A. Aïssa-El-Bey, K. Abed-Meraim, and S. Affes, “Robustness of blind subspace based techniques using lp quasi-norms,” in *11th IEEE International Workshop on Signal Processing Advances in Wireless Communications SPAWC*, Marrakech, Morocco, June 2010.
- [22] A. Aïssa-El-Bey, K. Abed-Meraim, and C. Laot, “Adaptive blind estimation of sparse SIMO channels,” in *7th International Workshop on Systems, Signal Processing and their Applications WOSSPA*, Tipaza, Algeria, May 2011.
- [23] F-X. Socheleau, D. Pastor, and A. Aïssa-El-Bey, “Robust statistics based noise variance estimation : application to wideband interception of non-cooperative communications,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 1, pp. 746–755, January 2011.
- [24] G. Xu, H. Liu, L. Tong, and T. Kailath, “A least-squares approach to blind channel identification,” *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982–2993, December 1995.
- [25] S. Van Vaerenbergh, J. Via, and I. Santamaria, “Blind identification of SIMO Wiener systems based on kernel canonical correlation analysis,” *IEEE Transactions on Signal Processing*, vol. 61, no. 9, pp. 2219–2230, May 2013.
- [26] 3GPP TS 36.104, *Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception*, April 2015.
- [27] N. A. Lazar, *The Statistical Analysis of Functional MRI Data*, Springer, 2008.
- [28] N. M. Correa, T. Adali and Y. Li, and V. D. Calhoun, “Canonical correlation analysis for data fusion and group inference,” *Signal Processing Magazine*, vol. 20, pp. 39–50, 2010.
- [29] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, “Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the mni mri single-subject brain,” *NeuroImage*, vol. 15, pp. 273–289, 2002.

**Campus de Brest**

Technopôle Brest-Iroise  
CS 83818  
29238 Brest Cedex 3  
France  
+33 (0)2 29 00 11 11

**Campus de Rennes**

2, rue de la Chataigneraie  
CS 17607  
35576 Cesson Sévigné Cédex  
France  
+33 (0)2 99 12 70 00

**Campus de Toulouse**

10, avenue Edouard Belin  
BP 44004  
31028 Toulouse Cedex 04  
France  
+33 (0)5 61 33 83 65

[www.telecom-bretagne.eu](http://www.telecom-bretagne.eu)

© Télécom Bretagne, 2015  
Imprimé à Télécom Bretagne  
Dépôt légal : Septembre 2015  
ISSN : 1255-2275

