

CCA FOR JOINT BLIND SOURCE SEPARATION OF MULTIPLE DATASETS WITH APPLICATION TO GROUP FMRI ANALYSIS *

Yi-Ou Li, Wei Wang, and Tülay Adalı

University of Maryland Baltimore County
Baltimore, MD

Vince D. Calhoun

The MIND Institute and the
University of New Mexico, Albuquerque, NM
Yale University, New Haven, CT

ABSTRACT

In this work, we propose a scheme for joint blind source separation (BSS) of multiple datasets using canonical correlation analysis (CCA). The proposed scheme jointly extracts sources from each dataset in the order of between-set source correlations. We show that, when sources are uncorrelated within each dataset and correlated across different datasets only on corresponding indices, (i) CCA on two datasets achieves BSS when the sources from the two datasets have distinct between-set correlation coefficients, and (ii) CCA on multiple datasets (M-CCA) achieves BSS with a more relaxed condition on the between-set source correlation coefficients compared to CCA on two datasets. We present simulation results to demonstrate the properties of CCA and M-CCA on joint BSS. We apply M-CCA to group functional magnetic resonance imaging (fMRI) data acquired from several subjects performing a visuomotor task and obtain interesting brain activations as well as their correlation profiles across different subjects in the group.

Index Terms— Canonical correlation analysis, blind source separation, magnetic resonance imaging, group analysis

1. INTRODUCTION

Analysis of data from multiple subjects is an important problem in biomedical applications such as the study of brain activations through functional magnetic resonance imaging (fMRI) data. Group analyses mainly focus on comparing experimental results from two contrast conditions [1],[2], or making inference on group behavior with datasets from multiple subjects/sessions [3],[4]. Blind source separation methods such as ICA have proved to be a useful tool in such analyses as they decompose the data into individual components for expository data representation. Group independent component analysis (ICA) [3] performs ICA on multiple datasets to obtain independent component estimates

associated with each dataset. Since the estimated components for each individual dataset are derived from a set of common components using concatenated data, the inherent correlation among the components may reduce the statistical power of the post analysis such as t-test on group activation level. Tensorial ICA [4] decomposes the group data into a set of components consisting of a spatial activation map, a time course, and a subject loading vector. Although tensorial ICA is based on a three-way data model, the subject-dependent variations are represented only by the loading vector of each component, which limits the subject difference to a scaling factor.

In this work, we propose to apply canonical correlation analysis (CCA) on multiple datasets in attempt to separate sources within each dataset as well as to provide the correlation profiles of the separated sources across different datasets. In CCA, because the demixing matrices are formed individually for each dataset, the independence of source estimates for each dataset is better preserved. Meanwhile, the correlation profiles of the estimated sources across different datasets can be appreciated through the estimated canonical correlation values.

CCA is a statistical method to summarize the correlation structure between two multivariate data by linear transformations [5]. In more recent efforts, CCA has been applied to solving the blind source separation (BSS) problem on a multi-dimensional dataset, see, *e.g.*, [6],[7]. In such methods, CCA is used to maximize the autocorrelation structure of the recovered sources. The CCA approach can be posed as a generalized eigenvalue decomposition based on cross-correlation matrices at different delay points [7]. As an eigenvalue based method, CCA has the limitation of achieving BSS only when the eigenvalues, *i.e.*, the autocorrelation coefficients of the latent sources, are *all distinct*. Methods such as utilizing the optimal weighted sum of the cross-correlation matrices at different delay points have been investigated [7], [8], so as to improve the solution of demixing matrix that recovers the latent sources.

To study the utility of CCA in achieving BSS on multiple datasets, we first show that CCA on two datasets can be

*THIS RESEARCH IS SUPPORTED IN PART BY THE NIH GRANT R01 EB 000840.

posed as a special case of the generalized eigenvalue decomposition and it achieves source separation when the correlation coefficients of the sources across the two datasets are all distinct. Secondly, we propose to apply M-CCA [9] to more than two datasets and show that M-CCA improves the chance to achieve source separation for multiple datasets compared with CCA on two datasets. We give the condition under which BSS can be achieved by M-CCA.

In the next section, we propose a generative source mixture model and show that CCA on two multidimensional datasets achieves BSS upon all distinct source correlation coefficients. We study conditions for M-CCA to achieve source separation in Section 3. In Section 4, we compare experimental results of CCA and M-CCA on simulated and true fMRI data, and we conclude our work with a discussion in the last section.

2. CCA ON TWO DATASETS

2.1. Generative model

We assume that:

(i) Dataset \mathbf{x}_k , $k = 1, 2$, is a linear mixture of p sources, \mathbf{s}_k , mixed by a nonsingular matrix, \mathbf{A}_k , *i.e.*,

$$\mathbf{x}_k = \mathbf{A}_k \mathbf{s}_k; \quad (1)$$

(ii) Sources are uncorrelated within each dataset and have zero mean and unit variance, *i.e.*,

$$E\{\mathbf{s}_k \mathbf{s}_k^T\} = \mathbf{I}, \quad k = 1, 2, \quad (2)$$

where \mathbf{I} stands for the identity matrix;

(iii) Sources from two datasets have nonzero correlation only on their *corresponding* indices, and have correlation coefficients, $r_{1,2}(1) \leq r_{1,2}(2), \dots, \leq r_{1,2}(p)$, where $r_{1,2}(i) = E\{s_1(i)s_2(i)\}$, *i.e.*,

$$E\{\mathbf{s}_1 \mathbf{s}_2^T\} = \mathbf{R}_{1,2} \quad (3)$$

where $\mathbf{R}_{1,2} = \text{diag}([r_{1,2}(1), \dots, r_{1,2}(p)])$.

The condition for CCA to recover the sources in datasets \mathbf{x}_1 and \mathbf{x}_2 is that strict inequality holds for the correlation coefficients defined in assumption (iii).

2.2. CCA for BSS of two datasets

We show that, when data follow the model assumptions and the separability condition given in Section 2.1 is satisfied, CCA on \mathbf{x}_1 and \mathbf{x}_2 produces two sets of sources \mathbf{s}_1 and \mathbf{s}_2 as the estimated canonical variates and the between-set source correlation coefficients $r_{1,2}(i)$, $i = 1, 2, \dots, p$ are estimated as the canonical correlations, *i.e.*, CCA achieves BSS for the two datasets.

For convenience, we whiten each dataset by principal component analysis (PCA) and normalize the variance so that the demixing matrix to be estimated for each whitened dataset

is an orthonormal matrix with the assumed data model. The whitening step can always be achieved given that the true mixing matrices are nonsingular.

Matrix \mathbf{B}_k is the whitening matrix for \mathbf{x}_k and the whitened data are given by:

$$\mathbf{y}_k = \mathbf{B}_k \mathbf{x}_k, \quad (4)$$

such that

$$E\{\mathbf{y}_k \mathbf{y}_k^T\} = \mathbf{I}, \quad k = 1, 2. \quad (5)$$

Eqs.(1), (2), and (5) together imply:

$$\mathbf{B}_k \mathbf{A}_k \mathbf{A}_k^T \mathbf{B}_k^T = \mathbf{I}, \quad k = 1, 2. \quad (6)$$

When CCA is applied to \mathbf{y}_1 and \mathbf{y}_2 , the two canonical transformation matrices, \mathbf{E}_1 and \mathbf{E}_2 , are calculated by solving the following eigenvalue decomposition problem:

$$\mathbf{E}_1 E\{\mathbf{y}_1 \mathbf{y}_2^T\} E\{\mathbf{y}_2 \mathbf{y}_1^T\} \mathbf{E}_1^T = \mathbf{D}, \quad (7)$$

where \mathbf{D} is a diagonal matrix with the estimated canonical correlation values d_1, d_2, \dots, d_p as the diagonal entries, and

$$\mathbf{E}_2 = \mathbf{D}^{-1/2} \mathbf{E}_1 E\{\mathbf{y}_1 \mathbf{y}_2^T\}. \quad (8)$$

Defining the global transformation matrix for dataset \mathbf{x}_k as

$$\mathbf{G}_k = \mathbf{E}_k \mathbf{B}_k \mathbf{A}_k, \quad k = 1, 2, \quad (9)$$

and substituting (3), (6), and (9) into (7), we have:

$$\mathbf{G}_1 \mathbf{R}_{1,2} \mathbf{G}_1^T = \mathbf{D}. \quad (10)$$

Since \mathbf{G}_1 is an orthonormal matrix, we have

$$\mathbf{R}_{1,2} \mathbf{G}_1^T = \mathbf{G}_1^T \mathbf{D}, \quad (11)$$

which implies that the i th column of \mathbf{G}_1^T , *i.e.*, $\mathbf{g}_1(i)$ satisfies the following equation,

$$\mathbf{R}_{1,2} \mathbf{g}_1(i) = d_i \mathbf{g}_1(i). \quad (12)$$

When $\mathbf{R}_{1,2}$ has all distinct diagonal entries, the only nontrivial solution of $\mathbf{g}_1(i)$ is that all but the i th element is nonzero, and hence, $\mathbf{g}_1(i)$ extracts the i th source in \mathbf{x}_1 up to a scaling factor.

Note that since the model and the method are symmetric for \mathbf{x}_1 and \mathbf{x}_2 , the same procedure can be used to demonstrate the source separation for \mathbf{x}_2 .

It is evident from the above discussion that when the between-set source correlation coefficients are not all distinct, eigenvalue-based CCA method fails to separate sources whose correlation coefficients are equal. However, this problem can be significantly mitigated by incorporating more datasets into the model, which is studied in the following section.

3. M-CCA ON MULTIPLE DATASETS

3.1. Generative model

We extend the generative model given in Section 2.1 by including M datasets with each dataset, \mathbf{x}_k , $k = 1, 2, \dots, M$ following assumptions (i), (ii) stated in Section 2.1. We extend assumption (iii) in Section 2.1 such that $E\{\mathbf{s}_k \mathbf{s}_l^T\} = \mathbf{R}_{k,l} = \text{diag}([r_{k,l}(1), \dots, r_{k,l}(p)])$, $\forall k \neq l; k, l \in \{1, 2, \dots, M\}$, and the correlation coefficients $r_{k,l}$ are in decreasing order.

The condition for M-CCA to recover the sources in the datasets \mathbf{x}_k , $k = 1, 2, \dots, M$ is that: $\forall 1 \leq i < j \leq p, \exists l \neq k$ for each $k \in \{1, \dots, M\}$, such that $r_{k,l}(i) > r_{k,l}(j)$.

3.2. M-CCA for BSS of multiple datasets

Lemma 3.1 *Given a group of datasets \mathbf{x}_k , $k = 1, 2, \dots, M$ following the generative model and the separability condition stated in Section 3.1, the i th source in each dataset $s_k(i)$ is recovered by a demixing vector $\mathbf{e}_k(i)$, when $\mathbf{e}_k(i)$ is chosen such that the sum of the correlation coefficients $r_{k,l}(i) \triangleq \text{corr}(\mathbf{e}_k^T(i)\mathbf{x}_k, \mathbf{e}_l^T(i)\mathbf{x}_l)$ for all $k, l \in \{1, 2, \dots, M\}$ is maximized, i.e., $s_k(i) = \mathbf{e}_k^T(i)\mathbf{x}_k$ where*

$$\{\mathbf{e}_1(i), \mathbf{e}_2(i), \dots, \mathbf{e}_M(i)\} = \arg \max \left\{ \sum_{k,l=1}^M r_{k,l}(i) \right\}.$$

Proof The proof is carried out in deflationary mode, i.e., the first source recovered from each dataset achieves the largest sum of between-set correlations, then these sources are subtracted from each dataset, or being avoided by putting constraints on the demixing vectors recovering subsequent sources.

In the i th stage of the deflationary procedure, we first show that, when $r_{k,l}(i) > r_{k,l}(j)$, any linear mixture of the remaining sources in \mathbf{x}_k and \mathbf{x}_l achieves a correlation value lower than $r_{k,l}(i)$.

Without loss of generality, we whiten each dataset by PCA and normalize the variance so that the demixing matrix to be estimated for each whitened dataset is an orthonormal matrix. For linear mixtures $\bar{s}_k = \alpha_k s_k(i) + \beta_k s_k(j)$ and $\bar{s}_l = \alpha_l s_l(i) + \beta_l s_l(j)$, $i < j$, $\{k, l\} \in 1, 2, \dots, M$, their correlation coefficient can be written as

$$\bar{r}_{k,l} \triangleq E\{\bar{s}_k \bar{s}_l\} = \alpha_k \alpha_l r_{k,l}(i) + \beta_k \beta_l r_{k,l}(j). \quad (13)$$

Given that the datasets are whitened, all demixing vectors have unit norm, i.e., $\alpha_i^2 + \beta_i^2 = 1, i = k, l$. Hence, Eq. (13) is the inner product of two vectors having the same elliptical trace on semimajor axis $\sqrt{r_{k,l}(i)}$ and semiminor axis $\sqrt{r_{k,l}(j)}$. Therefore, we have

$$\left[\sqrt{r_{k,l}(i)} \alpha_k \quad \sqrt{r_{k,l}(j)} \beta_k \right] \left[\sqrt{r_{k,l}(i)} \alpha_l \quad \sqrt{r_{k,l}(j)} \beta_l \right]^T \leq r_{k,l}(i),$$

where the equality holds only if $\alpha_k, \alpha_l = 1$ and $\beta_k, \beta_l = 0$. The same conclusion is obtained in case of more than two

source mixtures with the elliptical analogy in higher dimension.

Secondly, under the separation condition, for each dataset \mathbf{x}_k , there exists a dataset \mathbf{x}_l such that $r_{k,l}(i) > r_{k,l}(j)$. Therefore, $\sum_{k,l=1}^M r_{k,l}(i)$ achieves its global maximum *only if* the i th source is recovered for each dataset. ■

The demixing matrices that maximize correlation among the sources can be obtained by the M-CCA methods proposed in [9]. M-CCA is developed as an extension of CCA to maximize the *overall* correlation among sources recovered from multiple datasets. Five criteria are defined to evaluate the overall correlation among estimated sources. Three of them are eigenvalue-based criteria, which maximize similarity of rows/columns of the matrix composed by the correlation coefficients of sources from datasets of all pairwise combinations. The other two criteria are based on, respectively, sum and sum of square, of the correlation coefficients on sources from datasets of all pairwise combinations.

4. SIMULATIONS

4.1. Experiment on simulated datasets

Three sets of Laplacian distributed random sources are generated to assume the specified correlation profile show in Fig. 1. Sources in each dataset are mixed by a random square mixing matrix.

Three experiments are performed: (i) CCA on \mathbf{x}_1 and \mathbf{x}_2 , (ii) CCA on \mathbf{x}_2 and \mathbf{x}_3 , and (iii) M-CCA on \mathbf{x}_1 , \mathbf{x}_2 , and \mathbf{x}_3 . To implement M-CCA, one of the eigenvalue-based criterion is used, which maximizes the largest eigenvalue of the matrix with the between-set correlations [9]. Twenty Monte Carlo trials are performed with different realizations of the sources and mixing matrices in each trial.

The inter-symbol interference (ISI) is calculated based on the product of the demixing matrix and the true mixing matrix, and shown in Table 1. Mean and standard deviation of ISI are calculated based on results from twenty trials. ISI values range from 0 (perfect separation) to 1 (uniform mixture).

It is observed that in experiment (i), sources are not well separated due to the existence of non-distinct correlation coefficients between sources in \mathbf{x}_1 and \mathbf{x}_2 , while source separation is achieved in experiment (ii) since source correlation coefficients between \mathbf{x}_2 and \mathbf{x}_3 are all distinct. In (iii), M-CCA on three datasets successfully separates all the sources due to the fact that, for each dataset, there exists pairwise combination(s) such that the source correlation values are all distinct, i.e., the separation condition in Section 3.2 is satisfied.

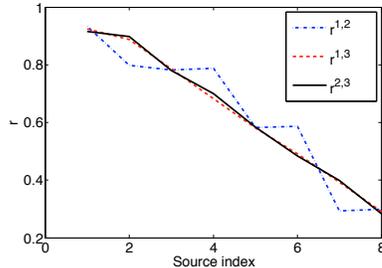


Fig. 1. Source correlation profiles of the three simulated datasets

Table 1. Comparison of separation performance by ISI

	CCA		M-CCA
	(i) $\mathbf{x}_1 : \mathbf{x}_2$	(ii) $\mathbf{x}_2 : \mathbf{x}_3$	(iii) $\mathbf{x}_1 : \mathbf{x}_2 : \mathbf{x}_3$
\mathbf{G}_1	0.10 ± 0.04		0.04 ± 0.04
\mathbf{G}_2	0.09 ± 0.03	0.04 ± 0.04	0.04 ± 0.03
\mathbf{G}_3		0.04 ± 0.05	0.04 ± 0.04

4.2. Experiment on group fMRI data from a visuomotor task

The fMRI data is acquired from eleven subjects performing a visuomotor task and preprocessed according to the typical fMRI analysis procedures [10]. Each dataset is then whitened by PCA and dimension reduced to 25 normalized principal components.

M-CCA is applied to the eleven sets of principal components. Two types of brain activations are estimated as the canonical variates from each dataset: (i) the activation on the posterior cingulate (related to the default mode), and (ii) the occipital lobe (primary visual area), cerebellar and motor cortex (motor task related) at the right and left hemisphere of the brain. The estimated mean activation maps, mean time courses, and estimated source correlation between subjects are displayed in Fig. 2. By comparing the time courses of component (i) and (ii), it is observed that the default mode is typically showing signal decreases with respect to the task. Based on the source correlation plots, the inter-subject variance of the visuomotor activations is estimated to be greater than that of the default mode component.

5. DISCUSSION

In this work we study the utility of canonical correlation analysis for joint blind source separation on multiple datasets. We pose CCA on two datasets as a generalized eigenvalue decomposition problem and show that it achieves joint BSS when the source correlation coefficients are all distinct. We study the conditions for maximization of source correlation among multiple datasets by M-CCA to achieve joint BSS. As an advantage, CCA and M-CCA also provide correlation profiles among the estimated sources from different datasets. Further-

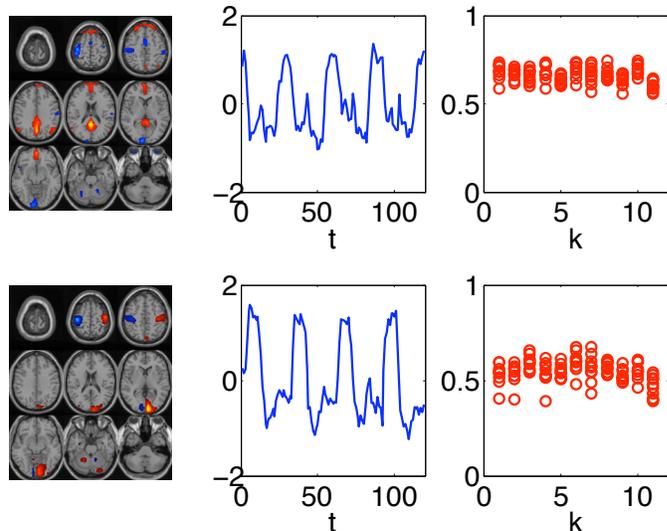


Fig. 2. Estimated mean activation maps (left), mean time courses (center), and the source correlation between subjects (right), of the default mode (top) and right visuomotor task-related (bottom) component by M-CCA on eleven fMRI datasets

more, because each dataset is processed individually, the extracted sources are more amenable to the post statistical analysis for making group inferences.

6. REFERENCES

- [1] A. S. Lukic, M. N. Wernick, L. K. Hansen, J. Anderson, and S. Strother, "A spatially robust ICA algorithm for multiple fMRI data sets," in *Proc. IEEE International Symposium on Biomedical Imaging (ISBI) 2002*, Arlington, VA.
- [2] Y.-O. Li, Tulay Adali, and Vince D. Calhoun, "A multivariate model for comparison of two datasets and its application to fMRI analysis," in *Proc. IEEE Workshop on Machine Learning for Signal Processing (MLSP) 2007*, Thessaloniki, Greece.
- [3] V. D. Calhoun, T. Adali, J. J. Pekar, and G. D. Pearlson, "A method for making group inferences from functional MRI data using independent component analysis," *Human Brain Mapping*, vol. 14, pp. 140–151, 2001.
- [4] C. F. Beckmann and S. M. Smith, "Tensorial extensions of independent component analysis for group fMRI data analysis," *NeuroImage*, vol. 25, 2005.
- [5] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, pp. 321–77, 1936.
- [6] O. Friman, M. Borga, P. Lundberg, and H. Knutsson, "Exploratory fMRI analysis by autocorrelation maximization," *NeuroImage*, vol. 16, pp. 454–64, 2002.
- [7] W. Liu, D. P. Mandic, and A. Cichocki, "Analysis and online realization of the CCA approach for blind source separation," *IEEE Trans. Neural Nets.*, vol. 18, pp. 1505–10, 2007.
- [8] M. Borga and H. Knutsson, *A canonical correlation approach to blind source separation*, Ph.D. thesis, Technical Report LiU-IMT-EX-0062, Linkoping University, 2001.
- [9] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, pp. 433–51, 1971.
- [10] Y.-O. Li, T. Adali, and V. D. Calhoun, "Estimating the number of independent components in fMRI data," *Human Brain Mapping*, vol. 28, pp. 1251–66, 2007.