COMMON COMPONENTS ANALYSIS VIA LINKED BLIND SOURCE SEPARATION

Guoxu Zhou[⋆] *Andrzej Cichocki*[⋆], IEEE Fellow *Danilo P. Mandic*[†], IEEE Fellow

[⋆] Lab for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Wako-shi Saitama, 351-0105 Japan. Email: {zhouguoxu,cia}@brain.riken.jp.
[†] Department of Electrical and Electronic Engineering, Imperial College, London, United Kingdom. E-mail: d.mandic@imperial.ac.uk.

ABSTRACT

Very often data we encounter in practice is a collection of matrices rather than a single matrix. These multi-block data often share some common features, due to the background in which they are measured. In this study we propose a new concept of linked blind source separation (BSS) that aims at discovering and extracting unique and physically meaningful common components from multi-block data, which also contain strong individual components. The validity and potential of the proposed method is justified by simulations.

Index Terms— Linked blind source separation, group independent component analysis, nonnegative matrix factorization

1. INTRODUCTION

High-dimensional data are increasingly prevalent in many areas of science. To this end, a variety of data analysis tools have been proposed to perform data compression, interpretation, information retrieve, etc. Recently, multi-block data analysis has also attracted attention [1, 2, 3, 4]. Multi-block data is encountered when multiple measurements are taken in a set of experiments from the same subject (using various modalities or from multiple subjects under similar configurations. For example, in biomedical studies, human electrophysiological signals responding to some pre-designed stimuli are collected from different individuals and trials. A number of different technologies and devices may be used to collect diverse information reflecting different aspects of data. All these result in naturally linked multi-block data. These data typically share some common information due to the background in which they are collected, and at the same time they also possess their individual features. It is therefore very meaningful to analyze such data in a connected and linked way and not individually. This study is devoted to such an interesting and promising topic.

Methods specifically developed for multi-block data analysis already exist. By assuming that the multi-block data are spanned by a set of common components, group independent component analysis (group ICA) and independent vector analysis have achieved some success in multi-block fMRI data analysis [5, 6]. Besides sharing common components, however, the multi-block data also contain their individual components, a key issue for which existing methods are not suitable. To this end, a recent method named Joint and Individual Variation Explained (JIVE) was proposed for integrated analysis of multiple data types [2], together with a new algorithm which separates their joint and individual subspaces. To our best knowledge, however, the potential of unique common feature analysis is not yet fully exploited, in particular the intimate link with blind source separation has not yet been established.

2. COMMON ORTHOGONAL BASES EXTRACTION (COBE)

2.1. The Model

Given a set of matrices $\mathcal{Y} = \{\mathbf{Y}_n \in \mathbb{R}^{I\times J_n} : n \in \mathcal{N}\}$, $\mathcal{N} = \{1, 2, \ldots, N\}$, we are interested in the following linked matrix factorization problem

$$\mathbf{Y}_n \approx \bar{\mathbf{A}}\bar{\mathbf{B}}_n^T + \breve{\mathbf{A}}_n\breve{\mathbf{B}}_n^T \doteq \bar{\mathbf{Y}}_n + \breve{\mathbf{Y}}_n,\ n \in \mathcal{N}, \qquad (1)$$

where $\bar{\mathbf{A}} \in \mathbb{R}^{I\times C}$, $\breve{\mathbf{A}}_n \in \mathbb{R}^{I\times(R_n-C)}$, $R_n = \text{rank}(\mathbf{Y}_n)$, and $C \leq \min\{R_n : n \in \mathcal{N}\}$. As such, each matrix $\mathbf{Y}_n$ is represented by two parts: the common space $\bar{\mathbf{Y}}_n = \bar{\mathbf{A}}\bar{\mathbf{B}}_n^T$ and the individual space $\breve{\mathbf{Y}}_n = \breve{\mathbf{A}}_n\breve{\mathbf{B}}_n^T$, which are spanned by the common components (i.e., columns of $\bar{\mathbf{A}}$) existing in all $\mathbf{Y}_k$ ($k \in \mathcal{N}$) and the individual components $\breve{\mathbf{A}}_n$ only present in $\mathbf{Y}_n$, respectively. Our aim is to find the common components $\bar{\mathbf{A}}$ from a given set of matrices $\mathbf{Y}_n$, $n \in \mathcal{N}$.

To solve (1), the JIVE method [2] considers the following optimization problem (although not explicitly stated in [2]):

$$\begin{aligned} \min \quad & \sum_{n\in\mathcal{N}} \left\| \mathbf{Y}_n - \bar{\mathbf{A}}\bar{\mathbf{B}}_n^T - \breve{\mathbf{A}}_n\breve{\mathbf{B}}_n^T \right\|_F^2 \\ s.t. \quad & \bar{\mathbf{A}}^T\bar{\mathbf{A}} = \mathbf{I}_C,\ \breve{\mathbf{A}}_n^T\breve{\mathbf{A}}_n = \mathbf{I}_{R_n-C}, \\ & \bar{\mathbf{A}}^T\breve{\mathbf{A}}_n = \mathbf{0}, \quad n \in \mathcal{N}. \end{aligned} \qquad (2)$$

The physical meaning of the orthogonality constraints $\bar{\mathbf{A}}^T\breve{\mathbf{A}}_n = \mathbf{0}$ is that there is no any interaction between common and indi-

vidual subspaces. In each iteration JIVE effectively performs joint principal component analysis (PCA) and individual PCA in an alternating manner: (i) in the joint PCA step, the individual subspaces are removed prior to applying PCA to all data; (ii) the individual PCA step is performed on each single individual data after the common subspace has been removed. The existence of both common and individual components distinguishes (2) from existing group ICA methods [5, 6, 7].

The JIVE method suffers from two major drawbacks. First, notice that the matrix $\mathbf{A}_n = \begin{bmatrix}\bar{\mathbf{A}} & \breve{\mathbf{A}}_n\end{bmatrix}$ essentially gives the optimal rank-R_n approximation of $\mathbf{Y}_n$ but with separated common and individual subspaces. In other words, (2) consists of two major functions: *dimensionality reduction* of each single data matrix $\mathbf{Y}_n$ and *separation of common and individual subspaces* of all data matrices. Apparently, the dimensionality reduction of $\mathbf{Y}_n$ depends on each $\mathbf{Y}_n$ only, while the latter requires an integrated analysis of all data matrices. For this reason, it is more natural to realize these two stpes separately rather than simultaneously. The JIVE method performs these two functions simultaneously, as a result, the ambiguity associated with JIVE is in giving principal components of all data rather than common components, especially when the common components in a data ensemble are relatively weak but are consistently present in all data sets. Second, the factors given by JIVE are not unique and lack specific physical meaning, this can be observed from $\bar{\mathbf{Y}}_n = (\bar{\mathbf{A}}\mathbf{U})(\mathbf{U}^{-1}\bar{\mathbf{B}}_n^T)$, where $\mathbf{U}$ is any invertible matrix of appropriate size. This paper introduce a method to overcome these two drawbacks.

2.2. The COBE Algorithm

As discussed above, it is more reasonable to perform *dimensionality reduction* and *common and individual feature separation* separately rather than simultaneously. Hence, we consider a two-step approach to solve (2):

Step 1: *Dimensionality reduction:* update the matrices $\mathbf{Y}_n$ in (2) by their optimal rank-R_n approximation $\mathbf{A}_n\mathbf{B}_n^T$. We call the original $\mathbf{Y}_n$ *raw* data and the reduced version $\mathbf{Y}_n \leftarrow \mathbf{A}_n\mathbf{B}_n^T$ *cleaned* data;

Step 2: *Common and Individual Subspaces Separation:* solve (2) using the cleaned data.

The key point is that, as $\mathbf{Y}_n$ are cleaned and of exact rank-R_n, in theory, we ensure the desire separation into the common and individual subspaces, that is,

$$\mathbf{Y}_n = \mathbf{A}_n\mathbf{B}_n^T = \bar{\mathbf{A}}\bar{\mathbf{B}}_n^T + \breve{\mathbf{A}}_n\breve{\mathbf{B}}_n^T,\ \forall n \in \mathcal{N}, \tag{3}$$

where $\mathbf{A}_n = \begin{bmatrix}\bar{\mathbf{A}} & \breve{\mathbf{A}}_n\end{bmatrix}$, $\mathbf{B}_n = \begin{bmatrix}\bar{\mathbf{B}}_n & \breve{\mathbf{B}}_n\end{bmatrix}$, and (3) holds if and only if $\bar{\mathbf{A}}$ contain only common components. To illustrate this, suppose that $\bar{\mathbf{A}}$ contains a component $\bar{\mathbf{a}}_k$ which is not contained in a data block, say $\mathbf{Y}_{n_0}$. Obviously it is impossible to represent $\mathbf{Y}_{n_0}$ exactly using $\bar{\mathbf{a}}_k$ and another $(R_{n_0}-1)$ vectors. With (3), the optimal $\bar{\mathbf{A}}$ and $\breve{\mathbf{A}}_n$ can be expressed as

$$\begin{bmatrix}\bar{\mathbf{A}} & \breve{\mathbf{A}}_n\end{bmatrix} = \mathbf{Y}_n{\mathbf{B}_n^T}^{\dagger}, \quad \mathbf{A}_n^T\mathbf{A}_n = \mathbf{I}_{R_n}, \quad n \in \mathcal{N}, \tag{4}$$

where $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudo-inverse of a matrix. Let $\mathbf{Y}_n = \mathbf{Q}_n\mathbf{R}_n$ such that $\mathbf{Q}_n^T\mathbf{Q}_n = \mathbf{I}$ (this can be achieved by using, e.g., QR decomposition), and define $\mathbf{Z}_n \doteq \mathbf{R}_n{\mathbf{B}_n^T}^{\dagger}$. Then (4) is equivalent to

$$\begin{bmatrix}\bar{\mathbf{A}} & \breve{\mathbf{A}}_n\end{bmatrix} = \mathbf{Q}_n\mathbf{Z}_n, \quad n \in \mathcal{N}, \tag{5}$$

and hence for any $n_1, n_2 \in \mathcal{N}$, $n_1 \neq n_2$, the following holds

$$\begin{cases}\mathbf{Q}_{n_1}\mathbf{z}_{n_1,k} = \mathbf{Q}_{n_2}\mathbf{z}_{n_2,k} = \bar{\mathbf{a}}_k & \text{if } k \leq C;\\ \mathbf{Q}_{n_1}\mathbf{z}_{n_1,k} \neq \mathbf{Q}_{n_2}\mathbf{z}_{n_2,k} & \text{if } k > C,\end{cases} \tag{6}$$

where $\mathbf{z}_{n,k}$ and $\bar{\mathbf{a}}_k$ are the kth column of $\mathbf{Z}_n$ and $\bar{\mathbf{A}}$, respectively. This motivates us to solve the following model:

$$\begin{aligned}\min_{\mathbf{Z}_n,\,\bar{\mathbf{A}}} \quad & f(\mathbf{Z}_n,\ \bar{\mathbf{A}}) = \sum_{n=1}^{N}\left\|\mathbf{Q}_n\mathbf{Z}_n - \bar{\mathbf{A}}\right\|_F^2\\ s.t. \quad & \bar{\mathbf{A}}^T\bar{\mathbf{A}} = \mathbf{I},\end{aligned} \tag{7}$$

where f should be sufficiently small in order to guarantee the extracted $\bar{\mathbf{A}}$ are true common components. We then optimize with respect to $\mathbf{Z}_n$ and $\bar{\mathbf{A}}$ alternatively. In this way, when $\bar{\mathbf{A}}$ is fixed, the optimal $\mathbf{Z}_n$ is computed using least squares

$$\mathbf{Z}_n \leftarrow \mathbf{Q}_n^T\bar{\mathbf{A}},\ n \in \mathcal{N}. \tag{8}$$

And when $\mathbf{Z}_n$, $n \in \mathcal{N}$, are fixed, (7) is equivalent to

$$\max_{\bar{\mathbf{A}}} \quad \text{trace}(\mathbf{P}^T\bar{\mathbf{A}}) \quad s.t. \quad \bar{\mathbf{A}}^T\bar{\mathbf{A}} = \mathbf{I} \tag{9}$$

where trace$(\cdot)$ denotes the trace of a matrix and

$$\mathbf{P} = \sum\nolimits_{n=1}^{N}\mathbf{Q}_n\mathbf{Z}_n. \tag{10}$$

Let $\mathbf{P} = \mathbf{E}\mathbf{\Lambda}\mathbf{V}^T \in \mathbb{R}^{I\times c}$ be the truncated SVD (tSVD) of $\mathbf{P}$, where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_c) \in \mathbb{R}^{c\times c}$ is a diagonal matrix with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_c > 0$. Motivated by the work in [8] (page 601)[1], we show that the optimal solution of (9) is

$$\bar{\mathbf{A}} = \mathbf{E}\mathbf{V}^T. \tag{11}$$

In fact,

$$\text{trace}(\mathbf{P}^T\bar{\mathbf{A}}) = \text{trace}(\mathbf{V}\mathbf{\Lambda}\mathbf{E}^T\bar{\mathbf{A}}) = \text{trace}(\mathbf{\Lambda}(\mathbf{E}^T\bar{\mathbf{A}}\mathbf{V})).$$

As $\mathbf{E}^T\mathbf{E} = \mathbf{I}$ and $(\bar{\mathbf{A}}\mathbf{V})^T(\bar{\mathbf{A}}\mathbf{V}) = \mathbf{I}$, we have $[\mathbf{E}^T\bar{\mathbf{A}}\mathbf{V}]_{ii} \leq 1$, which means that trace$(\mathbf{\Lambda}(\mathbf{E}^T\bar{\mathbf{A}}\mathbf{V})) \leq \sum_{i=1}^{c}\lambda_i$. Clearly, given $\bar{\mathbf{A}} = \mathbf{E}\mathbf{V}^T$ and $\bar{\mathbf{A}}^T\bar{\mathbf{A}} = \mathbf{I}$, we have $\mathbf{E}^T\bar{\mathbf{A}}\mathbf{V} = \mathbf{I}$ and hence trace$(\mathbf{P}^T\bar{\mathbf{A}})$ reaches its upper bound $\sum_{i=1}^{c}\lambda_i$.

The above procedure is named Common Orthogonal Bases Extraction (COBE) and the pseudo-code is presented in Algorithm 1. Once $\bar{\mathbf{A}}$ has been estimated, from (3), the coefficient matrices $\bar{\mathbf{B}}_n$ can be computed from

$$\bar{\mathbf{B}}_n = \left(\mathbf{Y}_n^T - \breve{\mathbf{B}}_n\breve{\mathbf{A}}_n^T\right)\bar{\mathbf{A}}(\bar{\mathbf{A}}^{\mathbf{T}}\bar{\mathbf{A}})^{-1} = \mathbf{Y}_n^T\bar{\mathbf{A}}, \tag{12}$$

this then allows us to compute the common subspace $\bar{\mathbf{A}}\bar{\mathbf{B}}_n^T$ and the individual subspace $\breve{\mathbf{Y}}_n = \mathbf{Y}_n - \bar{\mathbf{A}}\bar{\mathbf{B}}_n^T, \forall n \in \mathcal{N}$.

[1]The main difference is that here $\bar{\mathbf{A}}$ is not necessarily square.

Algorithm 1 The COBE Algorithm

Input: c and $\mathbf{Y}_n, n \in \mathcal{N}$.

1: Let $\mathbf{Y}_n$=$\mathbf{Q}_n\mathbf{R}_n$ such that $\mathbf{Q}_n^T\mathbf{Q}_n = \mathbf{I}$ for all n.
2: Initialize $\mathbf{Z}_n$ randomly.
3: **while** not converged **do**
4: $\quad \mathbf{P} = \sum_{n\in\mathcal{N}} \mathbf{Q}_n\mathbf{Z}_n$.
5: $\quad \bar{\mathbf{A}}$=$\mathbf{E}\mathbf{V}^T$, where $[\mathbf{E},\ \mathbf{\Lambda},\ \mathbf{V}] = \text{tSVD}(\mathbf{P}, c)$.
6: $\quad \mathbf{Z}_n \leftarrow \mathbf{Q}_n^T\bar{\mathbf{A}}$.
7: **end while**
8: **return** $\bar{\mathbf{A}}$.

3. LINKED BSS IN COMMON SUBSPACE

3.1. Linked BSS with Pre-whitening

So far we have only imposed the orthogonality constraints on the components $\bar{\mathbf{A}}$. In this case, the common components are not unique as the columns of $\bar{\mathbf{A}}\mathbf{U}$ also form a common orthogonal basis for any orthogonal matrix $\mathbf{U}$ with appropriate size. Generally, we want to project the common components onto a feature space with some desired property or uniqueness. This can be done typically by, for example, blind source separation (BSS) [9]. BSS is a problem of finding latent variables $\mathbf{S}$ from their linear mixtures $\mathbf{Y} = \mathbf{S}\mathbf{M}^T$ such that

$$\hat{\mathbf{S}} = \Psi(\mathbf{Y}) = \Psi(\mathbf{S}\mathbf{M}^T) = \mathbf{S}\mathbf{P}\mathbf{D} \tag{13}$$

where Ψ denotes a BSS algorithm, $\mathbf{M}$ is the mixing matrix. $\mathbf{P}$ and $\mathbf{D}$ are a permutation matrix and a diagonal matrix, respectively, which model unavoidable ambiguities of BSS. In other words, by using BSS methods the sources can be exactly recovered from their mixtures, subject to the scale and permutation ambiguities without any knowledge of the mixing matrix $\mathbf{M}$. Hence, BSS is quite attractive and has severed as feature extraction tools in a wide range of applications, such as pattern recognition, classification, etc. Suppose the common subspace $\bar{\mathbf{Y}}_n$ is spanned by a set of latent common components (sources) $\bar{\mathbf{S}}$, that is,

$$\bar{\mathbf{Y}}_n = \bar{\mathbf{S}}\bar{\mathbf{M}}_n^T. \tag{14}$$

From $\bar{\mathbf{Y}}_n = \bar{\mathbf{A}}\bar{\mathbf{B}}_n^T$, we have

$$\bar{\mathbf{A}} = \bar{\mathbf{S}}(\bar{\mathbf{B}}_n^\dagger\bar{\mathbf{M}}_n)^T, \tag{15}$$

that is, the columns of $\bar{\mathbf{A}}$ are just the linear mixtures of $\bar{\mathbf{S}}$ with the mixing matrix $\bar{\mathbf{B}}_n^\dagger\bar{\mathbf{M}}_n$, which allows $\bar{\mathbf{S}}$ to be estimated via

$$\hat{\bar{\mathbf{S}}} = \Psi(\bar{\mathbf{A}}) = \bar{\mathbf{S}}\mathbf{P}\mathbf{D}, \tag{16}$$

by using a appropriate BSS algorithm Ψ. In this case, $\bar{\mathbf{A}}$ is the pre-whitened version of (14), from (15) and the fact that $\bar{\mathbf{A}}^T\bar{\mathbf{A}} = \mathbf{I}$. By using BSS, we may obtain unique common components with desired properties such as sparsity, independence, temporal correlations, etc, by imposing suitable penalties on $\bar{\mathbf{S}}$, or even nonlinear common features by using kernel tricks [10]. We call the above BSS procedure *linked BSS* because BSS is performed on multi-block linked data $\mathbf{Y}_n$. Note that the joint BSS (JBSS) method in [1] also performs BSS involving multi-block data. It extracts a group of signals with the highest correlation each time and it requires that the extracted groups have distinct correlations. In other words, the JBSS method is actually a way to realize BSS by applying multiple-set canonical correlation analysis. In contrast, the linked BSS method extracts common bases first and then applies any suitable BSS to discover common components with desired properties and diversities.

3.2. Common Nonnegative Features Extraction (CNFE)

In the case where $\bar{\mathbf{S}}$ is required to be nonnegative, we cannot apply NMF methods on $\bar{\mathbf{A}}$ directly. In this case, we may use two steps to extract nonnegative common components. First, the common space is estimated as $\bar{\mathbf{Y}}_n = \bar{\mathbf{A}}(\bar{\mathbf{A}}^T\mathbf{Y}_n)$. Then we consider the following low-rank approximation based (semi) nonnegative matrix factorization (NMF) [11, 12]:

$$\min \sum\nolimits_n \left\|\bar{\mathbf{S}}\bar{\mathbf{M}}_n^T - \bar{\mathbf{A}}\bar{\mathbf{B}}_n^T\right\|_F^2, \quad s.t.\ \bar{\mathbf{S}} \succeq \mathbf{0}. \tag{17}$$

By using low-rank NMF (if $\bar{\mathbf{M}}_n$ is also nonnegative) or low-rank semiNMF (where $\bar{\mathbf{M}}_n$ is arbitrary) we can then extract the common nonnegative components $\bar{\mathbf{S}}$. For example, by using the following multiplicative update rules iteratively, both $\bar{\mathbf{S}}$ and $\bar{\mathbf{M}}$ are nonnegative:

$$\begin{aligned} \bar{\mathbf{S}} &\leftarrow \bar{\mathbf{S}} \circledast [\bar{\mathbf{A}} \sum_n (\bar{\mathbf{B}}_n^T\bar{\mathbf{M}}_n)]_+ \oslash \bar{\mathbf{S}}(\sum_n \bar{\mathbf{M}}_n^T\bar{\mathbf{M}}_n), \\ \bar{\mathbf{M}}_n &\leftarrow \bar{\mathbf{M}}_n \circledast [\bar{\mathbf{B}}_n(\bar{\mathbf{A}}^T\bar{\mathbf{S}})]_+ \oslash \bar{\mathbf{M}}_n(\bar{\mathbf{S}}^T\bar{\mathbf{S}}),\ n \in \mathcal{N}, \end{aligned} \tag{18}$$

where $\circledast$ and $\oslash$ are element-wise product and division of matrices, and $[x]_+ = \max(x, 0)$ is element-wise defined on matrices. See [11] for detailed convergence analysis.

4. SIMULATIONS AND EXPERIMENTS

Linked BSS. In this simulation we generated a total of ten matrices $\mathbf{A}_n \in \mathbb{R}^{5000\times 10}$, $n = 1, 2, \ldots, 10$, whose first four columns were the speech signals included in the benchmark of ICALAB (named Speech4.mat) [13], and the other six components were drawn from independent standard normal distributions. The entries of the mixing matrices $\mathbf{B}_n \in \mathbb{R}^{50\times 10}$ were also drawn from independent standard normal distributions. Finally let $\mathbf{Y}_n = \mathbf{A}_n\mathbf{B}_n^T + \mathbf{E}_n$, where $\mathbf{E}_n$ models white Gaussian noise (SNR=20dB). We first used the COBE, JIVE [2], JBSS [1], and PCA methods to extract the common components. Then we ran the SOBI method [14] to recover the latent speech signals (as the JBSS performed not so good in this simulation we also used SOBI to improve its results). TABLE 1 shows the simulation results averaged

Table 1: Performance comparison in linked BSS. The latent signals were estimated by applying the SOBI method to the common subspaces extracted by the algorithms.

Algorithm	SIR_1	SIR_2	SIR_3	SIR_4	Runtime (s)
COBE	22.4	26.0	31.5	32.0	0.1
JIVE	22.4	26.0	31.2	31.8	7.3
JBSS	22.6	23.1	23.5	25.0	1.6
PCA	15.5	17.2	18.0	19.4	0.5

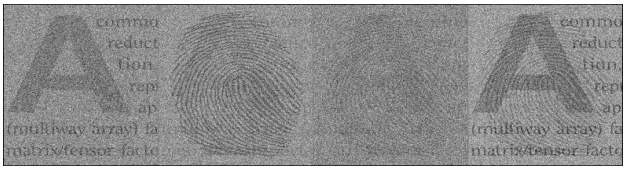
over 100 Monte-Carlo runs, with the signal-to-interference ratio of the cth estimated signal, i.e. SIR_c, defined as:

$$\text{SIR}_c = 10 \log_{10} \frac{\sum_t s_{t,c}^2}{\sum_t (s_{t,c} - \widehat{s}_{t,c})^2}, \tag{19}$$

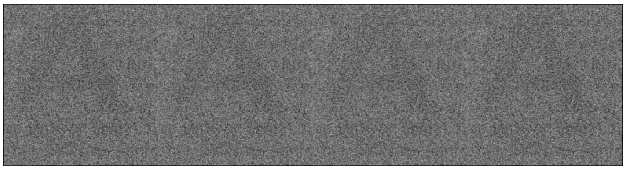
where $\mathbf{s}_c$, $\hat{\mathbf{s}}_c$ are normalized random variables with zero mean and unit variance, and $\hat{\mathbf{s}}_c$ is an estimate of $\mathbf{s}_c$. It can be seen that JIVE and COBE achieved higher SIRs than JBSS and PCA, although the performance of JBSS has been improved after incorporating the SOBI method compared with its original version. Moreover, although PCA has a close relation with COBE, it can be seen from the table that the common features extracted by PCA are often contaminated by individual features. COBE and JIVE almost achieved the same separation accuracy, but COBE was much faster. Next, all the common components were rescaled such that $\|\bar{\mathbf{a}}_c\|_2 / \|\breve{\mathbf{a}}_k\|_2 \approx 10\%$. This time only COBE was able to extract the true common components, illustrating enhanced robustness of COBE over JIVE, when the common components were relatively weak.

Common Nonnegative Components Extraction. In this experiment we generated four sets of nonnegative sources whose first two common components were respectively an image of fingerprint[2] and a patch of a document (see Figure1(c)), while to model interference the other eight components were drawn from independent uniform distributions between 0 and 1. They were mixed via different mixing matrices whose elements were drawn from independent uniform distributions between 0 and 1. It is known that the common sources in this example are dependent and consequently cannot be separated by using ICA methods; and because they are mixed with random dense components, they also cannot be separated by using ordinary NMF algorithms on each single set of mixtures. As the fingerprint and the patch existed in all images, we ran COBE to extract the bases of common sources and then used CNFE to extract the fingerprint and the patch. One typical realization is shown in Figure1(d). Figure1(b) displays four example images of nonnegative components extracted by using the nLCA-IVM method [15]. Due to the simultaneous presence of dense components (thus the identifiability conditions of nLCA-IVM were not satisfied here), it is not surprising that nLCA-IVM failed to extract the desired source images in this example. This experiment shows how the proposed method can be used to extract common nonnegative features, or equivalently, as nonnegative high correlation analysis method.

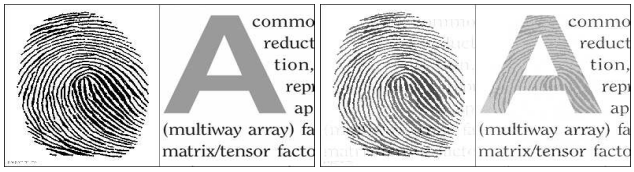
[2]Created by Phillip Martin, available at: `http://science.phillipmartin.info/science_fingerprint.htm`.



(a) Examples of the observations/mixtures



(b) Typical images extracted by nLCA-IVM



(c) The sources (d) The images extracted by CNFE

Fig. 1: Illustration of unique common nonnegative feature extraction. Common and individual subspaces separation priori to the unique common nonnegative components played a key role.

5. CONCLUSION

A novel linked blind source separation (BSS) scheme was proposed in this paper with the aim of discovering and extracting unique and physically meaningful common components from multi-block data, each of which contains both common components shared by all data and individual components possessed by very data item only. In the proposed method, the common subspace is extracted first and then any suitable BSS method can be applied to extract unique and physically meaningful common components. Simulations have shown the validity and efficiency of the proposed method compared to the state-of-the-art.

Acknowledgments

The work was supported by the JSPS KAKENHI (Grant No. 26730125) and the National Natural Science Foundation of China (Grant Nos. 61333013, 61273192, 61305028).

6. REFERENCES

[1] Y. Li, T. Adali, W. 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, oct. 2009.

[2] E. F. Lock, K. A. Hoadley, J. S. Marron, and A. B. Nobel, "Joint and individual variation explained (JIVE) for integrated analysis of multiple data types," *The Annals of Applied Statistics*, vol. 7, no. 1, pp. 523–542, 2013.

[3] C. M. Crainiceanu, B. S. Caffo, S. Luo, V. M. Zipunnikov, and N. M. Punjabi, "Population value decomposition, a framework for the analysis of image populations," *Journal of the American Statistical Association*, vol. 106, no. 495, pp. 775–790, 2011.

[4] M. Kim, J. Yoo, K. Kang, and S. Choi, "Nonnegative matrix partial co-factorization for spectral and temporal drum source separation," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 6, pp. 1192 –1204, oct. 2011.

[5] V.D. Calhoun, J Liu, and T. Adali, "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data," *NeuroImage*, vol. 45, no. 1, pp. 163 – 172, 2009.

[6] S. Ma, R. Phlypo, V. D. Calhoun, and T. Adali, "Capturing group variability using IVA: A simulation study and graph-theoretical analysis," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, BC, May 2013.

[7] M. Anderson, G. Fu, R. Phlypo, and T. Adali, "Independent vector analysis: Identification conditions and performance bounds," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4399–4410, Sept 2014.

[8] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 2715 North Charles Street, Baltimore, Maryland, 3 edition, 1996.

[9] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, John Wiley & Sons, Ltd, 2002.

[10] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *Journal of Machine Learning Research*, vol. 3, pp. 1–48, Mar. 2003.

[11] G. Zhou, A. Cichocki, and S. Xie, "Fast nonnegative matrix/tensor factorization based on low-rank approximation," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2928–2940, june 2012.

[12] Guoxu Zhou, Andrzej Cichocki, Qibin Zhao, and Shengli Xie, "Nonnegative matrix and tensor factorizations: An algorithmic perspective," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 54–65, May 2014.

[13] A. Cichocki, S. Amari, K. Siwek, and T. Tanaka, "ICALAB toolbox," 2007.

[14] A. Belouchrani, K. AbedMeraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, Feb 1997.

[15] F. Y. Wang, C. Y. Chi, T. H. Chan, and Y. Wang, "Nonnegative least-correlated component analysis for separation of dependent sources by volume maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 875–888, May 2010.