Projection-Pursuit-Based Method for Blind Separation of Nonnegative Sources

Zuyuan Yang, Yong Xiang, *Senior Member, IEEE*, Yue Rong, *Senior Member, IEEE*, and Shengli Xie, *Senior Member, IEEE*

***Abstract*— This paper presents a projection pursuit (PP) based method for blind separation of nonnegative sources. First, the available observation matrix is mapped to construct a new mixing model, in which the inaccessible source matrix is normalized to be column-sum-to-1. Then, the PP method is proposed to solve this new model, where the mixing matrix is estimated column by column through tracing the projections to the mapped observations in specified directions, which leads to the recovery of the sources. The proposed method is much faster than Chan's method, which has similar assumptions to ours, due to the usage of optimal projection. It is also more advantageous in separating cross-correlated sources than the independence- and uncorrelation-based methods, as it does not employ any statistical information of the sources. Furthermore, the new method does not require the mixing matrix to be nonnegative. Simulation results demonstrate the superior performance of our method.**

***Index Terms*— Blind source separation, linear programming (LP), nonnegative sources, projection pursuit (PP).**

NOTATIONS

$\mathbf{x}$, x_i	Column vector, the ith element of $\mathbf{x}$.
$\mathbf{X}$, $\mathbf{x}_j$, x_{ij}	Matrix, the jth column of $\mathbf{X}$, the (i, j)th entry of $\mathbf{X}$.
$\mathbf{X}_t$	Matrix with t columns.
$\mathbf{X}(i:j, b:t)$	A submatrix of $\mathbf{X}$ with rows from i to j and columns from b to t.
$\mathbf{X}_\Phi$	A submatrix of $\mathbf{X}$ with column index set Φ.
$\mathbf{X}^\perp$	The basis of a subspace orthogonal to $\mathbf{X}$.
$\Re$	Real number set.
$\mathbf{0}$	All-zero column vector.
$\mathbf{1}$	All-1 column vector.

Manuscript received December 29, 2011; revised October 1, 2012; accepted October 4, 2012. Date of publication November 29, 2012; date of current version December 18, 2012. This work was supported in part by the Australian Research Council under Grant DP 110102076, the National Natural Science Foundation of China under Grant 61104053, Grant 61273192, and Grant 61172180, the National Basic Research Program of China (937 Program) under Grant 2010CB731800, and the Natural Science Foundation of Guangdong Province under Grant S2011030002886.

Z. Yang is with the Faculty of Automation, Guangdong University of Technology, Guangzhou 510006, China, and also with the School of Information Technology, Deakin University, Melbourne 3125, Australia (e-mail: yangzuyuan@yahoo.com.cn).

Y. Xiang is with the School of Information Technology, Deakin University, Melbourne 3125, Australia (e-mail: yxiang@deakin.edu.au).

Y. Rong is with the Department of Electrical and Computer Engineering, Curtin University, Bentley 6102, Australia (e-mail: y.rong@curtin.edu.au).

S. Xie is with the Faculty of Automation, Guangdong University of Technology, Guangzhou 510006, China (e-mail: eeoshlxie@scut.edu.cn).

Digital Object Identifier 10.1109/TNNLS.2012.2224124

I. INTRODUCTION

BLIND source separation (BSS) has attracted much attention for over a decade due to its great potential in a wide range of applications such as digital communication, speech identification, biological data analysis, remote sensing, and human image processing [1]–[5]. Since BSS requires no or minimum prior knowledge of the sources and the mixing matrix, a statistical approach is usually required to perform BSS. Based on the concept of independent component analysis (ICA), various ICA methods have been proposed to separate independent sources from their measured mixtures, such as the quaternion ICA [6], fast ICA [7], ICA by kurtosis contrast maximization [8], ICA based on entropy-bound minimization (ICA-EBM) [9], and nonnegative ICA (NICA) [10], [11]. Since the ICA-based methods employ the higher order statistics of the measured data, a large number of data samples are needed to obtain satisfactory BSS performance [12]. In applications where fewer data samples are available, the second-order statistics (SOS) methods are preferable [13]. The SOS-based BSS methods often require the source signals to be mutually uncorrelated and have different frequency spectra [14], [15].

Although independent or mutually uncorrelated sources are often encountered in practice, spatially correlated sources also occur in many applications [3], [16]–[21]. For these applications, the BSS methods exploiting the statistical properties of the sources will fail. The existing works for cross-correlated sources are far from mature. Some methods tend to achieve BSS through enhancing the spatial diversity of the sources using certain preprocessing techniques. For example, a precoding scheme is used in [16] to reduce the cross-correlation of the sources, and a prefiltering approach is employed in [17] to enhance source independence. Other methods achieve BSS by analyzing the time-frequency (TF) points. They can be extended to the underdetermined cases but the computational cost is usually very high [18]–[21].

A different approach to BSS is to exploit the nonnegative signal characteristic. Nonnegative signals exist in various applications such as biomedical data analysis and image processing [2], [22], [23]. Since this approach does not utilize the statistical information of the sources, it can deal with both independent (or mutually uncorrelated) sources and spatially correlated sources. A typical approach to processing nonnegative signals is the nonnegative matrix factorization [24]–[29]. However, it assumes that all entries of the mixing matrix

are nonnegative. This assumption is also required by other nonnegativity-based methods [4], [5], [30].

In [31], Chan *et al*. present a BSS method known as convex analysis of mixtures of nonnegative sources by linear programming (CAMNS-LP), which relaxes the constraint on the mixing matrix to allow it to have negative entries. This method exploits the nonnegativity of sources and adopts a special deterministic assumption called local dominance. The local dominance assumption is a good assumption or approximation for source signals exhibiting sparsity or high contrast such as human portraits and some biomedical images [31], [32]. To estimate all source signals, this method needs to solve up to $2(n-1)$ LP problems, where n is the number of sources. It is known that solving an LP problem is time consuming if the data sample size is large [31], [33]. Therefore, the CAMNS-LP method becomes inefficient in computation with the increase of the source signals.

In this paper, we propose a more efficient method in the computation for blind separation of nonnegative sources. In our method, the observation matrix is first mapped into a superplane such that one of the rows of the mapped observation matrix has equal elements with value 1. The mapping process ensures that the unaccessible source matrix is normalized to be column-sum-to-1. Consequently, those columns of the normalized source matrix that satisfy the local dominance condition are unit vectors. For a unit vector, one element is 1 and all other elements are zero. Then, based on the property of the normalized source matrix, one column of the mixing matrix can be estimated by searching an optimal projection vector for the mapped observation matrix. After that, another column of the mixing matrix can be obtained by searching another optimal vector in the subspace orthogonal to the already estimated column. This process is repeated until all columns of the mixing matrix are obtained. After the estimation of the mixing matrix, the sources can be easily recovered. Since the new method estimates the mixing matrix by tracing the projection of the mapped observation matrix, it is called the projection pursuit (PP) method.

Unlike the traditional ICA-based BSS methods in [6]–[11], and [17], the proposed PP method does not require the sources to be independent or subband-independent. Moreover, different from the sparsity based methods in [34]–[36], it does not need the sources to be sparse in many time instants for efficient source recovery. The PP method is also advantageous over the existing BSS methods for nonnegative sources. Compared with the methods in [5] and [24]–[29], it does not require the mixing matrix to be nonnegative. In relation to the CAMNS-LP method in [31], our method has much less computational complexity, as it only needs to solve one LP problem. It is also more efficient than the TF-based methods which need to search the sparse TF points [18], [19]. In addition, if the mixing matrix is nonnegative, the normalized model in this paper is similar to those in the processing of spectral unmixing, where the sources satisfy nonnegativity and column-sum-to-1 physically [37]–[40]. By using the high contrast of the sources, the pure pixel index (PPI) method achieves a good result for end-member extraction in spectral unmixing but it needs many iterations [38]. To improve PPI, a fast iterative PPI (FIPPI) algorithm is developed in [39]. However, the FIPPI algorithm often requires an efficient initialization, such as the automatic target generation process (ATGP) approach in [40], to speed up its convergence.

The remainder of this paper is organized as follows. Section II formulates the problem through examining the computational cost of the CAMNS-LP method. The PP method is presented in Section III, together with the analysis of its computational complexity. Section IV provides simulation results to compare the performances of the PP method and the existing benchmark methods. Finally, Section V concludes this paper.

II. Problem Formulation

The instantaneous mixing model under consideration is as follows [9], [31], [41]:

$$\mathbf{X} = \mathbf{AS} \tag{1}$$

where $\mathbf{X} \in \Re^{m\times N}$ is the observation matrix, $\mathbf{A} \in \Re^{m\times n}$ is the mixing matrix, $\mathbf{S} \in \Re^{n\times N}$ is the source matrix, and m, n, and N denote the numbers of the observations (or outputs), the sources (or inputs), and the samples, respectively. The aim of BSS is to recover $\mathbf{S}$ from $\mathbf{X}$ without the information of $\mathbf{A}$. To achieve BSS, some assumptions are made in [31].

- A1) All sources are nonnegative, i.e., $s_{j,t} \geq 0$, where $j = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, N$.
- A2) Each source signal is local dominant, i.e., for each $j \in \{1, 2, \ldots, n\}$, there exists an unknown index l_j such that $s_{j,l_j} > 0$ and $s_{i,l_j} = 0\ \forall i \neq j$.
- A3) $m \geq n$ and $\mathbf{A}$ is of full column rank.

The rationality of the above assumptions is justified in [31], where the CAMNS-LP method has estimated successfully the source signals one by one.

The main ideas of the CAMNS-LP method are summarized as follows, first of all, the affine hull spanned by the sources and that by the mixtures are equivalent[1]:

$$\text{aff}\{\mathbf{S}^T\} = \text{aff}\{\mathbf{X}^T\} \tag{2}$$

where the superscript ^{T} denotes transpose operation. Then it is further shown that $\text{aff}\{\mathbf{S}^T\}$ can be written as [31]

$$\text{aff}\{\mathbf{S}^T\} = \left\{\mathbf{d} + \mathbf{C}\boldsymbol{\alpha} \mid \boldsymbol{\alpha} \in \Re^{(n-1)\times 1}\right\} \tag{3}$$

where $\boldsymbol{\alpha}$ denotes the representative coefficient vector, $\mathbf{d} \in \Re^{N\times 1}$ denotes the mean of the columns of $\mathbf{X}^T$, $\mathbf{C} \in \Re^{N\times(n-1)}$ is composed of the eigenvectors associated with the first $n-1$ principal eigenvalues of $\bar{\mathbf{X}}^T\bar{\mathbf{X}}$, and $\bar{\mathbf{X}} = \mathbf{X} - \mathbf{1}_{n\times 1}\mathbf{d}^T$. Considering the nonnegativity assumption of the sources, it is further shown that the intersection of $\text{aff}\{\mathbf{S}^T\}$ with positive quadrant is a convex hull. As a result, the sources can be found by searching the extreme points of this convex hull.

To estimate one of the sources, the searching of extreme points is cast into the following LP problems (with respect to $\boldsymbol{\alpha}$):

$$\begin{aligned} \min &: \mathbf{r}^T(\mathbf{d} + \mathbf{C}\boldsymbol{\alpha}) \\ \text{s.t. } &\mathbf{d} + \mathbf{C}\boldsymbol{\alpha} \succeq \mathbf{0} \end{aligned} \tag{4}$$

[1]This result is based on an additional assumption that A has unit row sum, which can be ensured by transforming the available observations [31].

$$\begin{aligned} &\max: \mathbf{r}^T(\mathbf{d}+\mathbf{C}\boldsymbol{\alpha}) \\ &\text{s.t. } \mathbf{d}+\mathbf{C}\boldsymbol{\alpha} \succeq \mathbf{0} \end{aligned} \tag{5}$$

where $\mathbf{r} \in \Re^{N\times 1}$ denotes a vector which is chosen randomly before beginning the optimization process and $\succeq$ denotes componentwise inequality. Let $\boldsymbol{\alpha}_1^*$ and $\boldsymbol{\alpha}_2^*$ be the optimal solutions of (4) and (5), respectively. Then either $(\mathbf{d}+\mathbf{C}\boldsymbol{\alpha}_1^*)^T$ or $(\mathbf{d}+\mathbf{C}\boldsymbol{\alpha}_2^*)^T$ is the estimate of one of the sources, depending on the calculated values of the cost functions in (4) and (5). To estimate the other sources, new LP problems need to be constructed through updating $\mathbf{r}$ by using the well-known QR decomposition. In total, $2(n-1)$ LP problems need to be solved to recover all source signals [31]. The computational complexity of solving $2(n-1)$ LP problems by using the typical primal-dual interior-point method is up to $O\left(2\eta(n-1)^2\left(N^{1.5}+(n-1)^2N^{0.5}\right)\right)$, where η is the number of iterations. Clearly, with increase in the number of sources, the computational efficiency of the CAMNS-LP method will decrease dramatically.

In the next section, we shall propose the PP method, which is more efficient in computation. For the sake of simplicity, we only consider the case of $m = n$ in the sequel. If $m > n$, one can reduce the dimension m of the observations into n by using the PCA method [5], [44]. We also assume that there is no zero column in $\mathbf{X}$ (if any, they can be easily detected and removed in advance).

III. PP Method

While the CAMNS-LP method directly estimates the sources at the cost of high computational complexity, an intuition is whether a more efficient method can be developed by first estimating the mixing matrix and then recovering the sources from (1). The vector form of (1) can be written as

$$\mathbf{x}_t = s_{1t}\mathbf{a}_1 + s_{2t}\mathbf{a}_2 + \cdots + s_{nt}\mathbf{a}_n \tag{6}$$

where $t = 1, 2, \ldots, N$. If the ith source signal dominates at the time instant j, i.e., $s_{ij} \neq 0$ and $s_{1j} = \cdots = s_{i-1,j} = s_{i+1,j} = \cdots = s_{nj} = 0$, then, it follows from (6) that $\mathbf{x}_j = s_{ij}\mathbf{a}_i$, i.e., $\mathbf{x}_j = \mathbf{a}_i$ up to a scalar. Thus, based on the source local dominance assumption A2, any column of $\mathbf{A}$ is equal to at least one column of $\mathbf{X}$ up to a scalar. In other words, the columns of $\mathbf{A}$ are hidden in the columns of $\mathbf{X}$. So one can estimate $\mathbf{A}$ by finding the columns of $\mathbf{X}$ that are related to the columns of $\mathbf{A}$, up to scalar and permutation ambiguities. However, since the number of the columns of $\mathbf{X}$ is often very large, how to find those special columns of $\mathbf{X}$ is a big challenge. In the PP method, we first construct a new mixing model by mapping $\mathbf{X}$ such that the corresponding new source matrix is column-sum-to-1. Thus, any column of the new source matrix satisfying the local dominance condition will be a unit vector. The special property of the new source matrix is beneficial for tracing the indices of those special columns of $\mathbf{X}$, and thus helpful for the estimation of the mixing matrix $\mathbf{A}$ and the recovery of the sources.

A. Constructing a New Mixing Model

From Assumptions A1 and A3, there must exist a nonzero vector $\mathbf{u} \neq \mathbf{0}$ such that $\mathbf{u}^T\mathbf{X}$ is a positive vector, e.g., $\mathbf{u}^T$ can be the product of a random positive row vector and the inverse of $\mathbf{A}$. This $\mathbf{u}$ can be calculated by solving the following LP problem which is often used to get an initial value in the feasible region [42]:

$$\begin{aligned} &\text{Min}: \gamma \\ &\text{s.t.} \begin{cases} \gamma - \sum_{i=1}^{n} u_i x_{ij} < 0 & \forall j \\ \gamma > \delta \end{cases} \end{aligned} \tag{7}$$

where x_{ij} is the (i, j)th entry of $\mathbf{X}$ and δ is a small positive constant (typically, $\delta = 1e-6$). Let $\mathbf{D}$ be a diagonal matrix whose diagonal entries consist of the reciprocal of the elements in $\mathbf{u}^T\mathbf{X}$

$$\mathbf{D} = \text{diag}(\mathbf{1}^T \oslash (\mathbf{u}^T\mathbf{X})) \tag{8}$$

where $\oslash$ denotes the componentwise division. Then, all elements of $\mathbf{u}^T\mathbf{X}\mathbf{D}$ are equal to 1

$$\mathbf{u}^T\mathbf{X}\mathbf{D} = (\mathbf{1}_{N\times 1})^T. \tag{9}$$

Also, since $\mathbf{u}^T\mathbf{X} = \mathbf{u}^T\mathbf{A}\mathbf{S}$ is a positive vector and $\mathbf{S}$ is nonnegative satisfying local dominance condition, $\mathbf{u}^T\mathbf{A}$ is a positive vector. Thus, there exists a positive-definite diagonal matrix $\mathbf{L}$ such that all elements of $\mathbf{u}^T\mathbf{A}\mathbf{L}$ equal 1

$$\mathbf{u}^T\mathbf{A}\mathbf{L} = (\mathbf{1}_{n\times 1})^T. \tag{10}$$

Let

$$\tilde{\mathbf{S}} = \mathbf{L}^{-1}\mathbf{S}\mathbf{D}. \tag{11}$$

Clearly, $\tilde{\mathbf{S}}$ is nonnegative because $\mathbf{L}^{-1}$, $\mathbf{S}$, and $\mathbf{D}$ are nonnegative. From (1), and (9)–(11), it holds that for any $t \in \{1, 2, \ldots, N\}$

$$\begin{aligned} \sum\nolimits_{j=1}^{n} \tilde{s}_{jt} &= \mathbf{u}^T\mathbf{A}\mathbf{L}\tilde{\mathbf{s}}_t \\ &= \mathbf{u}^T\mathbf{A}\mathbf{L}\mathbf{L}^{-1}\mathbf{s}_t d_{tt} \\ &= \mathbf{u}^T\mathbf{A}\mathbf{s}_t d_{tt} \\ &= \mathbf{u}^T\mathbf{x}_t d_{tt} \\ &= 1. \end{aligned} \tag{12}$$

Equation (12) means that each column of $\tilde{\mathbf{S}}$ is sum-to-1, i.e., $\mathbf{S}$ can be normalized to be column-sum-to-1 by (12). Note that, if the mixing matrix is nonnegative, the simple normalization method in [43] can also be used. However, if there exist some negative entries in the mixing matrix, this simple method cannot ensure the nonnegativity of $\tilde{\mathbf{S}}$. Now, we are in the position to construct a new mixing model with respect to $\tilde{\mathbf{S}}$.

Since $\mathbf{u} \neq \mathbf{0}$, suppose, without loss of generality, that the qth element of $\mathbf{u}$ is nonzero. Let $\mathbf{U}$ be the $n \times n$ identity matrix with the qth row replaced by $\mathbf{u}^T$. Clearly, $\mathbf{U}$ is full rank. Let $\tilde{\mathbf{X}}$ be the map of $\mathbf{X}$ given by

$$\tilde{\mathbf{X}} = \mathbf{U}\mathbf{X}\mathbf{D}. \tag{13}$$

Then the new mixing model about $\tilde{\mathbf{S}}$ results from (11) and (13) as follows:

$$\tilde{\mathbf{X}} = \tilde{\mathbf{A}}\tilde{\mathbf{S}} \tag{14}$$

where

$$\tilde{\mathbf{A}} = \mathbf{U}\mathbf{A}\mathbf{L}. \tag{15}$$

Here, $\tilde{\mathbf{A}}$ is full rank because $\mathbf{U}$, $\mathbf{A}$, and $\mathbf{L}$ are full rank. Also, since the qth row of $\mathbf{U}$ is $\mathbf{u}^T$, it follows from (9), (10), (13), and (15) that all of the elements in the qth row of $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{X}}$ are equal to 1. Furthermore, as we previously mentioned, $\tilde{\mathbf{S}}$ is nonnegative and its columns are sum-to-1. This implies that any column of $\tilde{\mathbf{S}}$ satisfying the local dominance condition is a unit vector. As a result, those columns of $\tilde{\mathbf{X}}$ that correspond to these unit vectors of $\tilde{\mathbf{S}}$, or equivalently the columns of $\tilde{\mathbf{A}}$, will be highlighted and thus facilitate the estimation of the mixing matrix $\mathbf{A}$. A geometric illustration of the map to the observations in the case of $n = 3$ is given in Fig. 1, where the nonzero elements of the source columns at the local dominant instants are assumed to be 1 for better visualization. We can see from Fig. 1(a) that the columns of the mixing matrix $\mathbf{A}$ (denoted by circles) are included but hidden in the columns of the observation matrix $\mathbf{X}$ (denoted by dots). However, it can be seen from Fig. 1(b) that, after mapping, the columns of the new mixing matrix $\tilde{\mathbf{A}}$ are highlighted as they locate at the vertices of the triangle region formed by all columns of the mapped observation matrix $\tilde{\mathbf{X}}$.

Next, we shall employ the new mixing model (14) to estimate the mixing matrix $\mathbf{A}$.

B. Estimating One Column of A

First of all, let us analyze the relationship between the columns of $\mathbf{A}$ and the column index of $\tilde{\mathbf{S}}$. It is easy to see from (11) that for any index j, if $\tilde{\mathbf{s}}_j$ is a unit vector, then $\mathbf{s}_j$ is a vector with only one nonzero element. Consequently, $\mathbf{x}_j$ is the estimate of a column of $\mathbf{A}$, neglecting possible scaling and permutation. Therefore, one can estimate a column of $\mathbf{A}$ through finding the index j such that $\tilde{\mathbf{s}}_j$ is a unit vector. Furthermore, this index j corresponds to the index of the maximum of the scalar projections of all column vectors $\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \ldots, \tilde{\mathbf{s}}_N$ on $[0, 0, \ldots, 0, 1]^T$ (or the minimum on $[1, 1, \ldots, 1, 0]^T$), due to the nonnegativity and column-sum-to-1 features of $\tilde{\mathbf{S}}$. However, since $\tilde{\mathbf{S}}$ is unknown, it is impossible to calculate any scalar projections of its columns.

Let $\tilde{\mathbf{v}}$ be a nonzero vector (i.e., $\tilde{\mathbf{v}} \neq \mathbf{0}$) and $\mathbf{v} = (\tilde{\mathbf{v}}^T \tilde{\mathbf{A}}^{-1})^T$. For any i, we have

$$\begin{aligned} \tilde{\mathbf{v}}^T \tilde{\mathbf{s}}_i / \|\tilde{\mathbf{v}}\| &= \tilde{\mathbf{v}}^T \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{A}} \tilde{\mathbf{s}}_i / \|\tilde{\mathbf{v}}\| \\ &= \tilde{\mathbf{v}}^T \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{x}}_i / \|\tilde{\mathbf{v}}\| \\ &= (\mathbf{v}^T \tilde{\mathbf{x}}_i / \|\mathbf{v}\|)(\|\mathbf{v}\| / \|\tilde{\mathbf{v}}\|). \end{aligned} \tag{16}$$

Equation (16) shows that the scalar projections of $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_N$ on $\mathbf{v}$ are proportional to that of $\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \ldots, \tilde{\mathbf{s}}_N$ on $\tilde{\mathbf{v}}$. Motivated by (16), the task here is to find a nonzero vector $\mathbf{v}$ such that the index j of the extreme of the scalar projections of $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_N$ on $\mathbf{v}$ is the desired index, i.e., it satisfies the condition that $\tilde{\mathbf{s}}_j$ is a unit vector. To proceed, for $\mathbf{v} \neq \mathbf{0}$, we first define two index sets as follows:

$$\begin{cases} \Phi(\mathbf{v}, \tilde{\mathbf{X}}) = \{t \mid \mathbf{v}^T \tilde{\mathbf{x}}_t = \max(\mathbf{v}^T \tilde{\mathbf{X}})\} \\ \Psi(\mathbf{v}, \tilde{\mathbf{X}}) = \{t \mid \mathbf{v}^T \tilde{\mathbf{x}}_t = \min(\mathbf{v}^T \tilde{\mathbf{X}})\}. \end{cases} \tag{17}$$

Clearly, in the case of $\mathbf{v} \neq \mathbf{0}$, $\Phi(\mathbf{v}, \tilde{\mathbf{X}})$ and $\Psi(\mathbf{v}, \tilde{\mathbf{X}})$ are equivalent to $\{t \mid \mathbf{v}^T \tilde{\mathbf{x}}_t / \|\mathbf{v}\| = \max(\mathbf{v}^T \tilde{\mathbf{X}} / \|\mathbf{v}\|)\}$ and $\{t \mid \mathbf{v}^T \tilde{\mathbf{x}}_t / \|\mathbf{v}\| = \min(\mathbf{v}^T \tilde{\mathbf{X}} / \|\mathbf{v}\|)\}$, respectively. Hence, for the convenience of description, we call $\mathbf{v}$ the projection vector and $\mathbf{v}^T \tilde{\mathbf{x}}_t$ the projection of $\tilde{\mathbf{x}}_t$ on $\mathbf{v}$. Based on (17), we propose the following theorem.

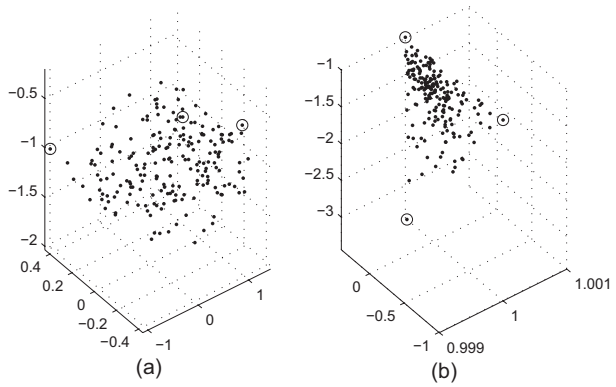


Fig. 1. Geometric illustration of the map to the observations in the case of $n = 3$, where dots and circles denote the columns of the observation matrix and the mixing matrix, respectively. (a) Before mapping. (b) After mapping.

Theorem 1: If the mixing matrix is full rank and the sources are nonnegative satisfying the local dominance condition, then for any $\mathbf{v} \neq \mathbf{0}$, there exists a $j \in \Phi(\mathbf{v}, \tilde{\mathbf{X}})$ or $\Psi(\mathbf{v}, \tilde{\mathbf{X}})$ such that $\tilde{\mathbf{s}}_j$ is a unit vector.

Proof: See Appendix A. ■

From Theorem 1 and its proof, we have the following corollary.

Corollary 1: For any $\mathbf{v} \neq \mathbf{0}$, if $\max(\mathbf{v}^T \tilde{\mathbf{X}}) > 0$, then $\exists j \in \Phi(\mathbf{v}, \tilde{\mathbf{X}})$ such that $\tilde{\mathbf{s}}_j$ is a unit vector. If $\min(\mathbf{v}^T \tilde{\mathbf{X}}) < 0$, then $\exists j \in \Psi(\mathbf{v}, \tilde{\mathbf{X}})$ such that $\tilde{\mathbf{s}}_j$ is a unit vector.

Theorem 1 and Corollary 1 show that the desired index that corresponds to a unit vector in $\tilde{\mathbf{S}}$ is indeed included in the index set corresponding to the maximum or minimum of the projections of $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_N$ on $\mathbf{v}$. Specifically, if $\max(\mathbf{v}^T \tilde{\mathbf{X}}) > 0$, the desired index must be in $\Phi(\mathbf{v}, \tilde{\mathbf{X}})$. If one can find a proper vector $\mathbf{v}$ such that

$$\tilde{\mathbf{s}}_j = \tilde{\mathbf{s}}_k \quad \text{for all } j, k \in \Phi(\mathbf{v}, \tilde{\mathbf{X}}) \tag{18}$$

then the desired index can be any element of $\Phi(\mathbf{v}, \tilde{\mathbf{X}})$. On the other hand, if $\min(\mathbf{v}^T \tilde{\mathbf{X}}) < 0$, then the desired index can be any element of $\Psi(\mathbf{v}, \tilde{\mathbf{X}})$ if

$$\tilde{\mathbf{s}}_j = \tilde{\mathbf{s}}_k \quad \text{for all } j, k \in \Psi(\mathbf{v}, \tilde{\mathbf{X}}). \tag{19}$$

Therefore, in order to obtain the desired index to estimate a column of $\mathbf{A}$, we need to find an optimal projection vector $\mathbf{v}$ for $\tilde{\mathbf{X}}$ such that (18) holds in the case of $\max(\mathbf{v}^T \tilde{\mathbf{X}}) > 0$ or (19) holds in the case of $\min(\mathbf{v}^T \tilde{\mathbf{X}}) < 0$. This falls into the PP problem which involves finding the most "interesting" possible projections in multidimensional data [45], [46]. PP-based methods have been widely used for data analysis such as PCA [7], [44], [47]. And the exact PP algorithms are often developed based on the chosen projections which are related to the practical applications.

In the present scenario, the most "interesting" possible projection vector $\mathbf{v}$ is the one satisfying (18) or (19). In order to obtain such projection vector $\mathbf{v}$, we first randomly generate a full-rank square matrix $\mathbf{W}$. Then, we construct $\mathbf{v}$ by using the columns of $\mathbf{W}$ iteratively according to the projections of the

columns of $\tilde{\mathbf{X}}$ on the columns of $\mathbf{W}$. Since $\mathbf{W}$ and $\tilde{\mathbf{A}}$ are full rank, (18) and (19) are equivalent to the equation $\mathbf{W}\tilde{\mathbf{x}}_j = \mathbf{W}\tilde{\mathbf{x}}_k$. Therefore, (18) and (19) can be replaced by

$$\mathbf{W}\tilde{\mathbf{x}}_j = \mathbf{W}\tilde{\mathbf{x}}_k \quad \text{for all } j,k \in \Phi(\mathbf{v},\tilde{\mathbf{X}}) \tag{20}$$

and

$$\mathbf{W}\tilde{\mathbf{x}}_j = \mathbf{W}\tilde{\mathbf{x}}_k \quad \text{for all } j,k \in \Psi(\mathbf{v},\tilde{\mathbf{X}}) \tag{21}$$

respectively. Now, the PP problem becomes how to find a projection vector $\mathbf{v}$ satisfying (20) or (21).

Let us start with defining the following index set for a given $\mathbf{W}$:

$$\Gamma_i = \begin{cases} \{t \mid \mathbf{w}_i^T\tilde{\mathbf{x}}_t = \max(\mathbf{w}_i^T\tilde{\mathbf{X}}_{\Gamma_{i-1}}), \quad t \in \Gamma_{i-1}\}, \\ \text{if } \max(\mathbf{w}_1^T\tilde{\mathbf{X}}) > 0 \\ \{t \mid \mathbf{w}_i^T\tilde{\mathbf{x}}_t = \min(\mathbf{w}_i^T\tilde{\mathbf{X}}_{\Gamma_{i-1}}), \quad t \in \Gamma_{i-1}\}, \\ \text{if } \max(\mathbf{w}_1^T\tilde{\mathbf{X}}) \le 0 \end{cases} \tag{22}$$

where $\Gamma_0 = \{1,2,\ldots,N\}$ and $1 \le i \le n$. Then we propose a lemma as follows.

Lemma 1: If $\mathbf{W}$ is a full-rank square matrix, there exists $l \le n$ such that $\mathbf{W}\tilde{\mathbf{x}}_j = \mathbf{W}\tilde{\mathbf{x}}_k$ for all $j,k \in \Gamma_l$.

Proof: See Appendix B. ∎

From Lemma 1 and (22), it follows that the projection vector $\mathbf{v}$ satisfying (20) or (21) can be obtained in $l(l \le n)$ iterations if $\mathbf{v}$ is updated in such a way that

$$\begin{cases} \Phi(\mathbf{v},\tilde{\mathbf{X}}) = \Gamma_i & \forall i \le l, \quad \text{if } \max(\mathbf{w}_1^T\tilde{\mathbf{X}}) > 0 \\ \Psi(\mathbf{v},\tilde{\mathbf{X}}) = \Gamma_i & \forall i \le l, \quad \text{if } \max(\mathbf{w}_1^T\tilde{\mathbf{X}}) \le 0. \end{cases} \tag{23}$$

We propose the following scheme to calculate $\mathbf{v}$:

1) initialize $\mathbf{v}$ to be a zero vector;
2) update $\mathbf{v}$ by

$$\mathbf{v} := \mathbf{v} + \lambda_i\mathbf{w}_i, \qquad \text{for any } \ i \le l \tag{24}$$

where λ_i is a positive constant that ensures (23).

Clearly, selecting λ_i such that (23) holds is critical in this scheme. We first consider the case of $\max(\mathbf{w}_1^T\tilde{\mathbf{X}}) > 0$, which can be further divided into two subcases: $i = 1$ and $i > 1$.

1) *Subcase 1* ($i = 1$): Since the initial value of $\mathbf{v}$ is $\mathbf{0}$, it results from (24) that the updated $\mathbf{v} = \lambda_1\mathbf{w}_1$. Based on (17) and (22), it holds that

$$\begin{aligned} \Gamma_1 &= \{t \mid \mathbf{w}_1^T\tilde{\mathbf{x}}_t = \max(\mathbf{w}_1^T\tilde{\mathbf{X}}_{\Gamma_0})\} \\ &= \{t \mid \mathbf{w}_1^T\tilde{\mathbf{x}}_t = \max(\mathbf{w}_1^T\tilde{\mathbf{X}})\} \\ &= \Phi(\mathbf{w}_1,\tilde{\mathbf{X}}). \end{aligned} \tag{25}$$

Besides, if λ_1 is a positive constant, we have

$$\Phi(\mathbf{w}_1,\tilde{\mathbf{X}}) = \Phi(\lambda_1\mathbf{w}_1,\tilde{\mathbf{X}}) = \Phi(\mathbf{v},\tilde{\mathbf{X}}). \tag{26}$$

It follows from (25) and (26) that the first equation in (23) holds for $i = 1$. Therefore, any positive constant can be chosen as λ_1.

2) *Subcase 2* ($i > 1$): We use a recurrence approach to analyze the choice of λ_i. As shown in Subcase 1, $\Phi(\mathbf{v},\tilde{\mathbf{X}}) = \Gamma_1$ is guaranteed after the first iteration. Suppose that $\Phi(\mathbf{v},\tilde{\mathbf{X}}) = \Gamma_{i-1}$ after the $(i-1)$th iteration. Then, based on (17), $\mathbf{v}^T\tilde{\mathbf{x}}_j = \mathbf{v}^T\tilde{\mathbf{x}}_k$ for any $j,k \in \Gamma_{i-1}$. In the ith iteration, since $\Gamma_i \subseteq \Gamma_{i-1}$, then $\mathbf{v}^T\tilde{\mathbf{x}}_j$ is equal to $\mathbf{v}^T\tilde{\mathbf{x}}_k$ for any $j,k \in \Gamma_i$. Also, it follows from (22) that for any $j,k \in \Gamma_i$, $\mathbf{w}_i^T\tilde{\mathbf{x}}_j = \mathbf{w}_i^T\tilde{\mathbf{x}}_k$. Therefore, for any $\lambda_i > 0$, it holds that

$$(\mathbf{v} + \lambda_i\mathbf{w}_i)^T\tilde{\mathbf{x}}_j = (\mathbf{v} + \lambda_i\mathbf{w}_i)^T\tilde{\mathbf{x}}_k \qquad \forall j,k \in \Gamma_i. \tag{27}$$

Denote $\Gamma_j - \Gamma_k = \{t \mid t \in \Gamma_j, t \notin \Gamma_k\}$ and choose the positive λ_i such that

$$\max\left((\mathbf{v} + \lambda_i\mathbf{w}_i)^T\tilde{\mathbf{X}}_{\Gamma_{i-1}}\right) > \max\left((\mathbf{v} + \lambda_i\mathbf{w}_i)^T\tilde{\mathbf{X}}_{\Gamma_0-\Gamma_{i-1}}\right) \tag{28}$$

that is

$$\begin{cases} \lambda_i > 0, \quad \text{and} \\ \lambda_i\left(\max(\mathbf{w}_i^T\tilde{\mathbf{X}}_{\Gamma_{i-1}}) - \max(\mathbf{w}_i^T\tilde{\mathbf{X}}_{\Gamma_0-\Gamma_{i-1}})\right) \\ > \max(\mathbf{v}^T\tilde{\mathbf{X}}_{\Gamma_0-\Gamma_{i-1}}) - \max(\mathbf{v}^T\tilde{\mathbf{X}}_{\Gamma_{i-1}}). \end{cases} \tag{29}$$

If $\Gamma_i = \Gamma_{i-1}$, it yields from (28) that

$$\max\left((\mathbf{v} + \lambda_i\mathbf{w}_i)^T\tilde{\mathbf{X}}_{\Gamma_i}\right) > \max\left((\mathbf{v} + \lambda_i\mathbf{w}_i)^T\tilde{\mathbf{X}}_{\Gamma_0-\Gamma_i}\right). \tag{30}$$

From (27) and (30), one can see that for the updated $\mathbf{v}$ using $\mathbf{v} + \lambda_i\mathbf{w}_i$ where λ_i satisfies (29), $\Phi(\mathbf{v},\tilde{\mathbf{X}})$ is equal to Γ_i, i.e., the first equation in (23) holds.

If $\Gamma_i \ne \Gamma_{i-1}$, we have $\Gamma_i \subset \Gamma_{i-1}$ as $\Gamma_i \subseteq \Gamma_{i-1}$. According to (22), $\max(\mathbf{w}_i^T\tilde{\mathbf{X}}_{\Gamma_i}) > \max(\mathbf{w}_i^T\tilde{\mathbf{X}}_{\Gamma_{i-1}-\Gamma_i})$, which leads to

$$\max(\lambda_i\mathbf{w}_i^T\tilde{\mathbf{X}}_{\Gamma_i}) > \max(\lambda_i\mathbf{w}_i^T\tilde{\mathbf{X}}_{\Gamma_{i-1}-\Gamma_i}) \tag{31}$$

for any $\lambda_i > 0$. It can be deduced from (31) that

$$\max(\lambda_i\mathbf{w}_i^T\tilde{\mathbf{X}}_{\Gamma_i}) = \max(\lambda_i\mathbf{w}_i^T\tilde{\mathbf{X}}_{\Gamma_{i-1}}). \tag{32}$$

Since $\Gamma_i \subset \Gamma_{i-1}$ and $\mathbf{v}^T\tilde{\mathbf{x}}_j = \mathbf{v}^T\tilde{\mathbf{x}}_k$ for any $j,k \in \Gamma_{i-1}$, it follows from (31) and (32), respectively, that

$$\max\left((\mathbf{v} + \lambda_i\mathbf{w}_i)^T\tilde{\mathbf{X}}_{\Gamma_i}\right) > \max\left((\mathbf{v} + \lambda_i\mathbf{w}_i)^T\tilde{\mathbf{X}}_{\Gamma_{i-1}-\Gamma_i}\right) \tag{33}$$

and

$$\max\left((\mathbf{v} + \lambda_i\mathbf{w}_i)^T\tilde{\mathbf{X}}_{\Gamma_i}\right) = \max\left((\mathbf{v} + \lambda_i\mathbf{w}_i)^T\tilde{\mathbf{X}}_{\Gamma_{i-1}}\right). \tag{34}$$

Furthermore, since $\Gamma_0 - \Gamma_i = (\Gamma_0 - \Gamma_{i-1}) + (\Gamma_{i-1} - \Gamma_i)$, one can see from (28), (33), and (34) that (30) also holds when $\Gamma_i \ne \Gamma_{i-1}$. This implies that the first equation in (23) holds for $\Gamma_i \ne \Gamma_{i-1}$.

Based on the above analysis, in the case of $\max(\mathbf{w}_1^T\tilde{\mathbf{X}}) > 0$, the optimal projection vector $\mathbf{v}$ can be found by using (24) where λ_1 can be any positive constant and $\lambda_i (i > 1)$ is chosen according to (29). Finally, one column of $\mathbf{A}$ can be estimated as follows:

$$\hat{\mathbf{a}}_1 = \mathbf{x}_j, \qquad j = \arg\max\left(\mathbf{v}^T\tilde{\mathbf{X}}\right) \tag{35}$$

where $\arg\max(\mathbf{v}^T\tilde{\mathbf{X}})$ denotes the index corresponding to the maximum of $\mathbf{v}^T\tilde{\mathbf{X}}$.

In the case of $\max(\mathbf{w}_1^T\tilde{\mathbf{X}}) \le 0$, it must hold that $\min(\mathbf{w}_1^T\tilde{\mathbf{X}}) < 0$ as $\mathbf{w}_1 \ne \mathbf{0}$ and $\tilde{\mathbf{X}}$ is of full row rank. Then, similarly, one column of $\mathbf{A}$ can be estimated by

$$\hat{\mathbf{a}}_1 = \mathbf{x}_j, \qquad j = \arg\min\left(\mathbf{v}^T\tilde{\mathbf{X}}\right) \tag{36}$$

where $\arg\min(\mathbf{v}^T\tilde{\mathbf{X}})$ denotes the index corresponding to the minimum of $\mathbf{v}^T\tilde{\mathbf{X}}$, and $\mathbf{v}$ is also calculated using (24). However, in this case, λ_1 is any positive constant but $\lambda_i (i > 1)$ should be chosen according to the following criterion:

$$\begin{cases} \lambda_i > 0, \quad \text{and} \\ \lambda_i \left(\min(\mathbf{w}_i^T \tilde{\mathbf{X}}_{\Gamma_{i-1}}) - \min(\mathbf{w}_i^T \tilde{\mathbf{X}}_{\Gamma_0 - \Gamma_{i-1}}) \right) \\ < \min(\mathbf{v}^T \tilde{\mathbf{X}}_{\Gamma_0 - \Gamma_{i-1}}) - \min(\mathbf{v}^T \tilde{\mathbf{X}}_{\Gamma_{i-1}}). \end{cases} \tag{37}$$

As for the existence of λ_i in (29) and (37), we have the following proposition.

Proposition 1: For a full-rank square matrix $\mathbf{W}$, 1) if $\max(\mathbf{w}_1^T\tilde{\mathbf{X}}) > 0$, then for any $i > 1$, there exists λ_i such that (29) holds; 2) if $\max(\mathbf{w}_1^T\tilde{\mathbf{X}}) \leq 0$, then for any $i > 1$, there exists λ_i such that (37) holds.

Proof: See Appendix C. ■

C. Estimating the Other Columns of A

According to the PP scheme, after a projection vector is obtained, the data are reduced by removing the component along that vector direction such that a new vector can be found [45], [46]. This process will be repeated until all desired projection vectors are acquired. Thus, once $r (1 \leq r \leq n-1)$ columns of $\mathbf{A}$ are estimated, the next column of $\mathbf{A}$ can be estimated by finding the most "interesting" projection vector in the subspace orthogonal to $\hat{\mathbf{A}}_r$ defined as

$$\hat{\mathbf{A}}_r = [\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_r] \tag{38}$$

where $\hat{\mathbf{a}}_i$ denotes the ith estimated column. Based on the generalized inverse of matrix, the basis of the subspace orthogonal to $\hat{\mathbf{A}}_r$ can be computed by [48]

$$\hat{\mathbf{A}}_r^{\perp} = \left(\mathbf{I} - \hat{\mathbf{A}}_r (\hat{\mathbf{A}}_r^T \hat{\mathbf{A}}_r)^{-1} \hat{\mathbf{A}}_r^T \right) \mathbf{H} \tag{39}$$

where $\mathbf{H} \in \Re^{n \times (n-r)}$ is a matrix of full column rank. Therefore, one can use the scheme shown in Section III-B to estimate the next column of $\mathbf{A}$ but the auxiliary matrix $\mathbf{W}$ should be constructed as follows:

$$\begin{cases} \mathbf{W}(1:n, 1:n-r) = \hat{\mathbf{A}}_r^{\perp} \\ \mathbf{W}(1:n, n-r+1:n) = \hat{\mathbf{A}}_r. \end{cases} \tag{40}$$

From (40), we can see that for any $r (1 \leq r \leq n-1)$, the updated $\mathbf{W}$ is always square and full rank. As a result, the conclusions in Lemma 1 and Proposition 1 hold for all r. This guarantees that all other columns of $\mathbf{A}$ can be obtained by using the scheme in Section III-B together with constructing $\mathbf{W}$ by (40).

D. Summary and Analysis

Based on the discussions in the Sections III-B and III-C, the proposed PP method is formulated as follows.

1) *Step 1:* Calculate $\mathbf{u}$ by (7) and suppose $u_q \neq 0$. Let $\mathbf{U}$ be the $n \times n$ identity matrix with the qth row replaced by $\mathbf{u}^T$ and compute $\mathbf{D}$ by (8).
2) *Step 2:* Map $\mathbf{X}$ into $\tilde{\mathbf{X}}$ by (13) and randomly generate a full-rank square matrix $\mathbf{W}$.
3) *Step 3:* Set $\mathbf{v} = \mathbf{0}$. If $\max(\mathbf{w}_1^T\tilde{\mathbf{X}}) > 0$, select λ_i by (29), update $\mathbf{v}$ by (24) until (20) holds and then estimate $\hat{\mathbf{a}}_1$ by (35). Otherwise, select λ_i by (37), update $\mathbf{v}$ by (24) until (21) holds, and then estimate $\hat{\mathbf{a}}_1$ by (36).
4) *Step 4:* Estimate the other columns of $\mathbf{A}$ in the following way:

```
for r = 1, 2, ..., n − 1;
    update Â_r by (38);
    update Â_r^⊥ by (39);
    update W by (40);
    estimate â_{r+1} by (35) [or (36)] using the
    method in Step 3;
end for
```

5) *Step 5:* Let $\hat{\mathbf{A}}_n = [\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \ldots, \hat{\mathbf{a}}_n]$. Estimate the source matrix by $\hat{\mathbf{S}} = \hat{\mathbf{A}}_n^{-1}\mathbf{X}$.

In the PP method, since the vector $\mathbf{u}$ can be accurately calculated by solving an LP problem in Step 1, the matrix $\tilde{\mathbf{X}}$ can be precisely obtained in Step 2. Moreover, the vector $\hat{\mathbf{a}}_1$ estimated in Step 3 is a global optimum. Similarly, the vectors $\hat{\mathbf{a}}_2, \hat{\mathbf{a}}_3, \ldots, \hat{\mathbf{a}}_n$ are perfectly estimated in Step 4. Thus, the estimated mixing matrix $\hat{\mathbf{A}}_n$ is a global optimum, which ensures the perfect recovery of the sources if the model assumptions are perfectly fulfilled.

As far as the computational complexity is concerned, the computational cost in Step 1 is dominated by the calculation of $\mathbf{u}$ through solving the LP problem in (7), which has a complexity up to $O(\eta(n-1)(N^{1.5} + (n-1)^2 N^{0.5}))$ by using the primal-dual interior-point method, where η is the number of iterations [31], [33]. In Step 2, the computation of $\tilde{\mathbf{X}}$ using (13) requires $O(n^2 N)$ multiplications. In the Step 3, the complexity of updating the projection vector $\mathbf{v}$ in (24) is $O(lnN)$, where $l \leq n$ denotes the number of iterations involved. The computational costs in Steps 4 and 5 are $O(l(n-1)nN)$ and $O(n^2 N + n^3)$, respectively. Hence, the total computational complexity of our method is no more than $O(\eta(n-1)(N^{1.5} + (n-1)^2 N^{0.5}) + ln^2 N + 2n^2 N + n^3)$. As we mentioned in Section II, the complexity of the CAMNS-LP method is up to $O(2\eta(n-1)^2(N^{1.5} + (n-1)^2 N^{0.5}))$. In practice, it often holds that $N > n^2 > n > 1$ and $\eta > n \geq l$. Thus, the approximate computational costs of the PP and CAMNS-LP methods can be written as $O(\eta(n-1)N^{1.5})$ and $O(2\eta(n-1)^2 N^{1.5})$, respectively. Clearly, our method is much more efficient in computation than the CAMNS-LP method.

IV. Simulation Results

In this section, we provide three simulation examples to illustrate the computational complexity and source separation performance of the PP method, in comparison with the WPS-DICA algorithm based on subband independence [17], the TF method using sparsity [18], the ATGP-FIPPI approach using high contrast [39], and the CAMNS-LP method [31] which requires similar assumptions as those in our method. Each method is implemented using MATLAB R2009 a installed in a personal computer with Intel(R) Celeron(R) 2.4 GHz CPU, 2 GB memory and Microsoft Windows 7 operational system. The elapsed CPU time is used to measure the computing speed. The source separation performance is measured by the mean of the sum square error (M-SSE) index defined as

TABLE I
e_{ave} VERSUS NUMBER OF SOURCES GENERATED BY COMPUTER SOFTWARE

n	e_{ave} (dB)				
	PP	WPSDICA	TF	ATGP-FIPPI	CAMNS-LP
4	−293.3578	6.4710	7.3711	−152.1610	−289.0798
8	−292.9375	10.1570	11.1690	140.0098	−280.2750
12	−290.9316	12.2901	13.3695	−127.1011	−269.0701
16	−284.5937	13.6598	14.2018	−120.0984	−256.3028
20	−283.9879	14.6601	15.3491	−107.2290	−247.3905

follows [31]:

$$e(\mathbf{S}, \hat{\mathbf{S}}) = \frac{1}{n} \min_{\boldsymbol{\pi} \in \Pi_n} \sum_{i=1}^{n} \|\mathbf{s}^i - \hat{\mathbf{s}}^{\pi_i}\|^2 \tag{41}$$

where $\mathbf{s}^i$ is the ith row of the source matrix $\mathbf{S}$, $\hat{\mathbf{s}}^i$ is the ith row of the estimated source matrix $\hat{\mathbf{S}}$, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)^T$, and $\Pi_n = \{\boldsymbol{\pi} \in \Re^{n\times 1} \mid \pi_i \in \{1, 2, \ldots, n\}, \pi_i \neq \pi_j, \forall i \neq j\}$ is the set of all permutations of $\{1, 2, \ldots, n\}$. Here, the L2-norms of $\mathbf{s}^i$ and $\hat{\mathbf{s}}^i$, $\forall i$ are normalized to be 1. The optimization in (41) aims to find the best match between the original sources and the estimated sources, which can be solved by the Hungarian algorithm in [49].

A. Separation of Computer-Generated Signals

In the first simulation, computer-generated signals are used as source signals, which are mutually correlated but satisfy the local dominance condition. Specifically, they are generated by using the rand() function of MATLAB, followed by zeroing some nonzero samples such that the local dominant condition is satisfied. Each signal is with a uniform distribution on [0 1] and contains 30 000 samples. Five scenarios corresponding to 4, 8, 12, 16, and 20 sources are considered. For each scenario, 1000 independent runs are carried out to compute the average M-SSE index e_{ave} (in dB) and the average CPU-time T_{ave} (in seconds). In each run, a different mixing matrix is used, which is randomly generated. Table I shows the e_{ave} values obtained by PP, WPSDICA, TF, ATGP-FIPPI, and CAMNS-LP under different numbers of sources. One can see that the PP, ATGP-FIPPI, and CAMNS-LP methods achieve excellent source separation performance whilst the PP method performs the best. This result is not surprising, as the sources satisfy the local dominance condition required by these three methods. Besides, our method is almost immune to the number of sources, making it suitable for dealing with a large number of sources. In contrast, WPSDICA and TF fail to separate the sources.

The computational efficiency of the compared five methods is shown in Fig. 2. Since the T_{ave} values of these methods have large variance, they are scaled using $\log_{10}(1 + T_{\text{ave}})$ for better visual comparison. Clearly, the PP method is computationally much more efficient than the CAMNS-LP method. This is because only one LP problem is encountered in the PP method, while the latter needs to solve up to $2(n - 1)$ LP problems, which is very time consuming. Our method is also more efficient than the other three schemes, and the reasons are that WPSDICA needs a complex operation to extract the independent subband, TF spends much time on sparse time-frequency points searching, and ATGP-FIPPI uses additional initialization.

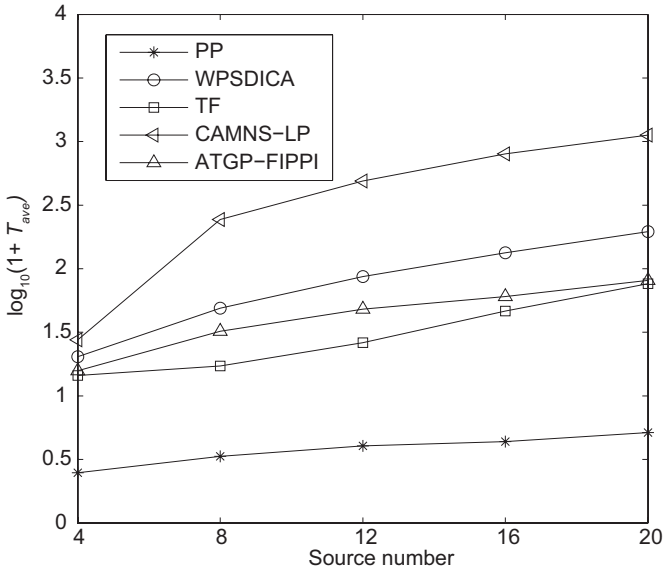


Fig. 2. Scaled average CPU time versus number of sources.

TABLE II
e_{ave} AND T_{ave} WHERE FINGERPRINT IMAGES ARE USED AS SOURCES

	PP	WPSDICA	TF	ATGP-FIPPI	CAMNS-LP
e_{ave} (dB)	−295.8921	4.1789	3.0012	−93.0231	−289.8763
T_{ave} (s)	6.3305	201.7410	698.2931	63.7009	105.1912

B. Separation of Human Fingerprint Images

In this simulation, three fingerprint images (640×480) are utilized as sources, which are from the DB1 dataset of the Fingerprint Verification Competition 2004.[2] Their correlation matrix is computed as follows:

$$\mathbf{R} = \begin{bmatrix} 1.0000 & 0.7910 & 0.5949 \\ 0.7910 & 1.0000 & 0.6549 \\ 0.5949 & 0.6549 & 1.0000 \end{bmatrix}.$$

Obviously, these fingerprint images are mutually correlated. It is also found that the local dominance condition holds for these images. Similar to the first simulation, 1000 independent runs are carried out to compute the indices e_{ave} and T_{ave}, where a different random mixing matrix is used in each run.

Table II shows the e_{ave} and T_{ave} values of the compared five methods. Since the fingerprint images satisfy the local dominance condition, the PP, ATGP-FIPPI, and CAMNS-LP methods perform very well in separating these images, much better than the TF and WPSDICA methods. Furthermore, the PP method uses much less time to separate the images than all other methods do.

[2]Available at http://biometrics.cse.msu.edu/fvc04db/index.html.

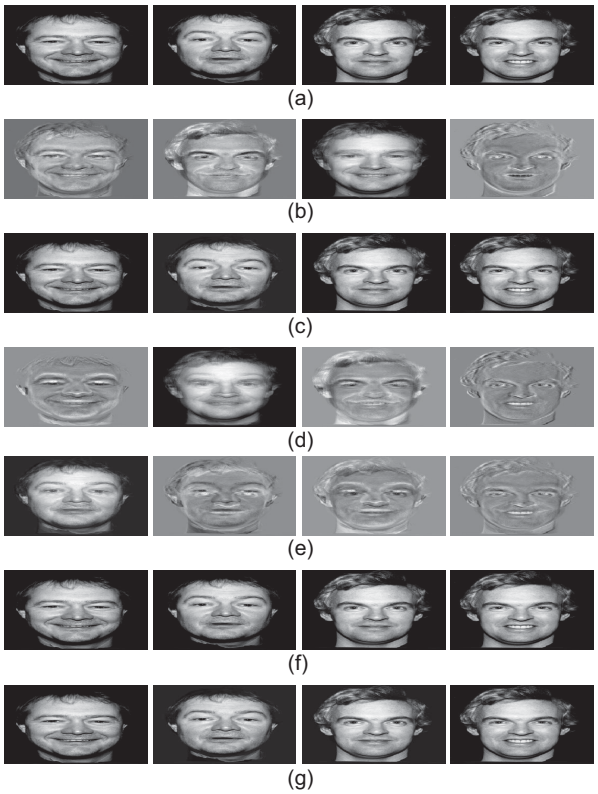


Fig. 3. (a) Original face images. (b) Mixed images. (c) Images recovered by the PP method. (d) Images recovered by the WPSDICA method. (e) Images recovered by the TF method. (f) Images recovered by the ATGP-FIPPI method. (g) Images recovered by the CAMNS-LP method.

C. Separation of Face Images

In the third simulation, we use four face images (275×350) as sources,[3] which are shown in Fig. 3(a). These images do not satisfy the local dominance condition and are also cross-correlated, having the following correlation matrix:

$$\mathbf{R} \approx \begin{bmatrix} 1.00 & 0.88 & 0.84 & 0.84 \\ 0.88 & 1.00 & 0.81 & 0.80 \\ 0.84 & 0.81 & 1.00 & 0.93 \\ 0.84 & 0.80 & 0.93 & 1.00 \end{bmatrix}.$$

For visual comparison, we first mix the face images by using the following randomly generated mixing matrix:

$$\mathbf{A} = \begin{bmatrix} 1.85 & -0.48 & -1.09 & 0.81 \\ -1.48 & -0.14 & 2.28 & 0.27 \\ 2.06 & 0.31 & 0.23 & 1.67 \\ -0.28 & 0.41 & 1.11 & -1.47 \end{bmatrix}$$

and the mixed images are shown in Fig. 3(b). The PP, WPSDICA, TF, ATGP-FIPPI, and CAMNS-LP methods are employed to recover the source images. Figs. 3(c)–(g) shows the recovered images by these methods, respectively. We can see that, although the local dominance assumption is violated in this simulation, the PP, ATGP-FIPPI, and CAMNS-LP methods achieve satisfactory separation performance. However, while the ATGP-FIPPI and CAMNS-LP methods take 50.0451 and 84.0469 s to perform source separation, respectively, the CPU time spent by the PP method is only 4.0616 s.

[3]Available at http://www.bsp.brain.riken.jp/ICALAB.

TABLE III
e_{ave} AND T_{ave} WHERE FACE IMAGES ARE USED AS SOURCES

	PP	WPSDICA	TF	ATGP-FIPPI	CAMNS-LP
e_{ave} (dB)	−30.0010	4.8050	4.4556	−29.9909	−29.9089
T_{ave} (s)	4.0598	98.6250	122.3579	50.1297	84.1026

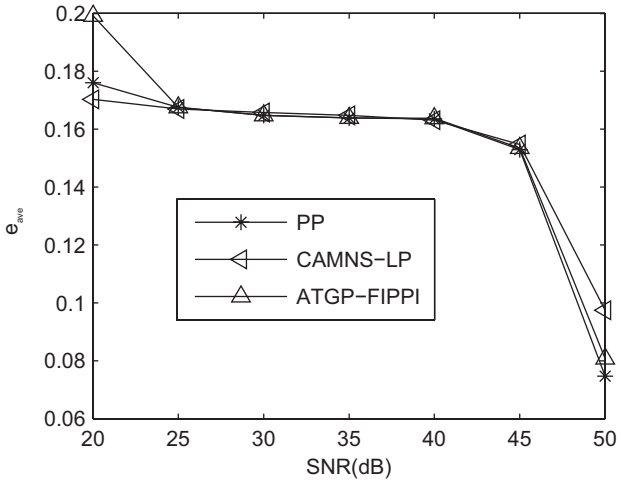


Fig. 4. e_{ave} indices of the different methods versus noise levels.

Then, in order to give a numerical comparison of the concerned methods, we carry out 1000 independent runs to compute the e_{ave} and T_{ave} indices. Here, different random mixing matrices are used in the simulation runs. From Table III, one can see that the numerical result is consistent with that of the visual comparison.

In addition, we test the sensitivity of the proposed PP method to noise using the above four face images, where the noise is with Gaussian distribution. The signal-to-noise ratio (SNR) is defined as $10 \log_{10}\|\mathbf{x}\|/\|\mathbf{x}-\mathbf{y}\|$, where $\mathbf{x}$ and $\mathbf{y}$ denote the original signal and the signal polluted by noise, respectively. Since the TF and ATGP-FIPPI methods cannot recover all the source images even in the noiseless case (Fig. 3 and Table III for reference), we mainly compare the recovery precision of the PP, ATGP-FIPPI, and CAMNS-LP methods in the noisy case. Fig. 4 gives the e_{ave} indices of these three methods against noise with SNR levels of 20, 25, 30, 35, 40, 45, and 50 dB, respectively. It shows that these three methods have similar performance against noise, due to their common assumption on the sources.

V. CONCLUSION

In this paper, a PP-based BSS method was proposed to separate nonnegative sources. By exploiting the nonnegativity and local dominance of the sources, the BSS problem was cast into finding the most interesting projections of the mapped observations. As a result, the proposed PP method gives a globally optimal solution, and has a lower computational complexity than the well-known CAMNS-LP method does. Also, our method does not require the mixing matrix to be nonnegative. Furthermore, since the PP method does not rely on any statistical information of the sources, it separates the cross-correlated sources with higher precision than the subband independence based WPSDICA method. In addition,

unlike traditional sparse component analysis method, the proposed PP does not require the sources to be sparse in many time instants any more. However, similar to other local dominance based methods, it may degenerate if the noise is strong, especially if the local dominant samples happen to be polluted seriously. Our future work will aim to improve these kinds of methods in noisy scenarios.

APPENDIX A

PROOF OF THEOREM 1

Denote $\mathbf{g} = (\mathbf{v}^T\tilde{\mathbf{A}})^T$, and let g_b and g_t be the maximum and minimum of $\mathbf{g}$ with indices b and t, respectively. For $\mathbf{v} \neq \mathbf{0}$, if it is a scaled unit vector with only $v_q \neq 0$, then $\Phi = \Psi = \{1, 2, \ldots, N\}$. Thus all elements of the qth row of $\tilde{\mathbf{X}}$ equal to 1. In this case, Theorem 1 holds because of the local dominance assumption of the sources.

Next, we shall complete the proof by considering the case that $\mathbf{v}$ is neither a zero vector nor a scaled unit vector. Since $\tilde{\mathbf{A}}$ is full rank and its qth row consists of elements equal to 1, the elements of $\mathbf{g}$ are not the same, i.e., there must exist i, j such that $g_i \neq g_j$, which leads to $g_b \neq g_t$. Let $\mathbf{f} = (\mathbf{v}^T\tilde{\mathbf{X}})^T$ be the projection of $\tilde{\mathbf{X}}$. Then the jth element of $\mathbf{f}$ is $f_j = \mathbf{v}^T\tilde{\mathbf{x}}_j = \mathbf{v}^T\tilde{\mathbf{A}}\tilde{\mathbf{s}}_j = \mathbf{g}^T\tilde{\mathbf{s}}_j = \sum_{k=1}^n g_k\tilde{s}_{kj}$. The remainder of the proof is divided into the following two parts.

1) $g_b > 0$: Since $\mathbf{S}$ satisfies the local dominance condition, there exists an identity submatrix in $\tilde{\mathbf{S}}$. Thus, for the set $\{k \mid g_k = g_b\}$, there always exists a unit vector in $\tilde{\mathbf{S}}$ whose τth element is equal to 1, where $\tau \in \{k \mid g_k = g_b\}$. Denoting this vector by $\tilde{\mathbf{s}}_p$ with index p, it follows $\tilde{s}_{\tau p} = 1$ and $\tilde{s}_{kp} = 0$ for all $k \neq \tau$. Therefore, we have

$$\forall i \notin \{k \mid g_k = g_b\}, \qquad \tilde{s}_{ip} = 0. \tag{42}$$

Furthermore, since $\sum_{k=1}^n \tilde{s}_{kj} = 1$ and $\tilde{s}_{kj} \geq 0$ for any j, it holds that

$$\begin{aligned}
f_j &= \sum_{k=1}^{n} g_k\tilde{s}_{kj} \\
&= \sum_{k=1}^{t-1} g_k\tilde{s}_{kj} + \sum_{k=t+1}^{n} g_k\tilde{s}_{kj} \\
&\quad + g_t\left(1 - \left(\sum_{k=1}^{t-1}\tilde{s}_{kj} + \sum_{k=t+1}^{n}\tilde{s}_{kj}\right)\right) \\
&= \sum_{k=1}^{t-1}(g_k - g_t)\tilde{s}_{kj} + \sum_{k=t+1}^{n}(g_k - g_t)\tilde{s}_{kj} + g_t \\
&\leq \sum_{k=1}^{t-1}(g_b - g_t)\tilde{s}_{kj} + \sum_{k=t+1}^{n}(g_b - g_t)\tilde{s}_{kj} + g_t \\
&= (g_b - g_t)\left(\sum_{k=1}^{t-1}\tilde{s}_{kj} + \sum_{k=t+1}^{n}\tilde{s}_{kj}\right) + g_t \\
&= (g_b - g_t)(1 - \tilde{s}_{tj}) + g_t \\
&\leq g_b - g_t + g_t \\
&= g_b.
\end{aligned} \tag{43}$$

If $\tilde{s}_{kj} = 0$ for all $k \notin \{k \mid g_k = g_b\}$, then $\tilde{s}_{tj} = 0$ due to $g_t \neq g_b$ and thus the two inequalities in (43) become equations. Combined with (42), one can see that $f_p = g_b$. Note that for any j, if $f_j = g_b$, then f_j is the maximum of the projection $\mathbf{f}$, i.e., j belongs to Φ. Therefore, in the case of $g_b > 0$, there exists $j \in \Phi$ such that $\tilde{\mathbf{s}}_j$ is a unit vector, e.g., when $j = p$.

2) $g_b \leq 0$: From $g_b \leq 0$, it follows $g_t < 0$. Similar to (42), there exists an index q such that

$$\forall i \notin \{k \mid g_k = g_t\}, \quad \tilde{s}_{iq} = 0. \tag{44}$$

Also, similar to (43), we have

$$\begin{aligned}
f_j &= \sum_{k=1}^{n} g_k\tilde{s}_{kj} \\
&= \sum_{k=1}^{b-1}(g_k - g_b)\tilde{s}_{kj} + \sum_{k=b+1}^{n}(g_k - g_b)\tilde{s}_{kj} + g_b \\
&\geq \sum_{k=1}^{b-1}(g_t - g_b)\tilde{s}_{kj} + \sum_{k=b+1}^{n}(g_t - g_b)\tilde{s}_{kj} + g_b \\
&= (g_t - g_b)\left(\sum_{k=1}^{b-1}\tilde{s}_{kj} + \sum_{k=b+1}^{n}\tilde{s}_{kj}\right) + g_b \\
&\geq g_t - g_b + g_b \\
&= g_t.
\end{aligned} \tag{45}$$

If $\tilde{s}_{kj} = 0$ for all $k \notin \{k \mid g_k = g_t\}$, then $\tilde{s}_{bj} = 0$ as $g_b \neq g_t$ and the inequalities in (45) turn into equations. Combined with (44), it follows that $f_q = g_t$. For any j, if $f_j = g_t$, then f_j is the minimum of the projection $\mathbf{f}$, i.e., j is in Ψ. Therefore, in the case of $g_b \leq 0$, there exists $j \in \Psi$ such that $\tilde{\mathbf{s}}_j$ is a unit vector, e.g., when $j = q$.

The proof is completed by combining 1) and 2).

APPENDIX B

PROOF OF LEMMA 1

We prove Lemma 1 by using apagoge. Let us first consider the case of $\max(\mathbf{w}_1^T\tilde{\mathbf{X}}) > 0$. Denote $\mathbf{G} = \mathbf{W}\tilde{\mathbf{A}}$. According to the analysis shown in the proof of Theorem 1, for any i, if

$$\mathbf{W}\tilde{\mathbf{x}}_j \neq \mathbf{W}\tilde{\mathbf{x}}_k, \quad \exists j, k \in \Gamma_i, \quad j \neq k \tag{46}$$

then the ith row of $\mathbf{G}$ must have two elements whose values are equal to the maximum of that row. From the definition of Γ_i in (22), it is clear that $\Gamma_i \subseteq \Gamma_{i-1}$. Thus, if (46) holds for all $i \leq n$, there must exist two identical columns in $\mathbf{G}$, which means that $\mathbf{G}$ is rank-deficient. However, on the other hand, $\mathbf{G}$ should be full rank because both $\mathbf{W}$ and $\tilde{\mathbf{A}}$ are full rank. This results in contradiction. Therefore, there exists $l \leq n$ such that $\mathbf{W}\tilde{\mathbf{x}}_j = \mathbf{W}\tilde{\mathbf{x}}_k \forall j, k \in \Gamma_l$.

The proof for the case of $\max(\mathbf{w}_1^T\tilde{\mathbf{X}}) \leq 0$ can be conducted similarly. This completes the proof.

APPENDIX C
PROOF OF PROPOSITION 1

We first prove the statement 1). Since $\mathbf{W}$ is full rank, $\mathbf{w}_i \neq \mathbf{0}, \forall i$. Then, based on the definition of Γ_i in (22), one can conclude that $\Gamma_i \subseteq \Gamma_{i-1}$, and

$$\forall i > 1, \ \begin{cases} \Gamma_{i-1} \neq \emptyset \\ \Gamma_0 - \Gamma_{i-1} \neq \emptyset \end{cases} \tag{47}$$

where $\emptyset$ denotes the empty set. Also, since $\Phi(\mathbf{v}, \tilde{\mathbf{X}}) = \Gamma_{i-1}$, it holds that

$$\max(\mathbf{v}^T \tilde{\mathbf{X}}_{\Gamma_0 - \Gamma_{i-1}}) - \max(\mathbf{v}^T \tilde{\mathbf{X}}_{\Gamma_{i-1}}) < 0. \tag{48}$$

Thus, if $\max(\mathbf{w}_i^T \tilde{\mathbf{X}}_{\Gamma_{i-1}}) - \max(\mathbf{w}_i^T \tilde{\mathbf{X}}_{\Gamma_0 - \Gamma_{i-1}}) \geq 0$, λ_i can be any positive number; otherwise, λ_i can be a positive number smaller than

$$\frac{\max(\mathbf{v}^T \tilde{\mathbf{X}}_{\Gamma_0 - \Gamma_{i-1}}) - \max(\mathbf{v}^T \tilde{\mathbf{X}}_{\Gamma_{i-1}})}{\max(\mathbf{w}_i^T \tilde{\mathbf{X}}_{\Gamma_{i-1}}) - \max(\mathbf{w}_i^T \tilde{\mathbf{X}}_{\Gamma_0 - \Gamma_{i-1}})}.$$

The proof of the statement 2) can be arrived at in a similar manner and is omitted here. This completes the proof.

REFERENCES

[1] M. I. Stanciu, S. Azou, and A. Serbănescu, "On the blind estimation of chip time of time-hopping signals through minimization of a multimodal cost function," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 842–847, Feb. 2011.

[2] C. Gouy-Pailler, M. Congedo, C. Brunner, C. Jutten, and G. Pfurtscheller, "Nonstationary brain source separation for multiclass motor imagery," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 2, pp. 469–478, Feb. 2010.

[3] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[4] Z. Yang, Y. Xiang, S. Xie, S. Ding, and Y. Rong, "Nonnegative blind source separation by sparse component analysis based on determinant measure," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 10, pp. 1601–1610, Oct. 2012.

[5] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 875–888, May 2010.

[6] J. Vía, D. P. Palomar, L. Vielva, and I. Santamaría, "Quaternion ICA from second-order statistics," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1586–1600, Apr. 2011.

[7] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.

[8] V. Zarzoso and P. Comon, "Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size," *IEEE Trans. Neural Netw.*, vol. 21, no. 2, pp. 248–261, Feb. 2010.

[9] X.-L. Li and T. Adali, "Independent component analysis by entropy bound minimization," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5151–5164, Oct. 2010.

[10] M. D. Plumbley, "Conditions for nonnegative independent component analysis," *IEEE Signal Process. Lett.*, vol. 9, no. 6, pp. 177–180, Jun. 2002.

[11] M. D. Plumbley and E. Oja, "A 'nonnegative PCA' algorithm for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 15, no. 1, pp. 66–76, Jan. 2004.

[12] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, 1st ed. Oxford, U.K.: Academic, 2010.

[13] S. An, Y. Hua, J. H. Manton, and Z. Fang, "Group decorrelation enhanced subspace method for identifying FIR MIMO channels driven by unknown uncorrelated colored sources," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4429–4441, Dec. 2005.

[14] K. Abed-Meraim, Y. Xiang, J. H. Manton, and Y. Hua, "Blind source separation using second-order cyclostationary statistics," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 694–701, Apr. 2001.

[15] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "Blind source separation using second order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[16] Y. Xiang, S. K. Ng, and V. K. Nguyen, "Blind separation of mutually correlated sources using precoders," *IEEE Trans. Neural Netw.*, vol. 21, no. 1, pp. 82–90, Jan. 2010.

[17] I. Kopriva and D. Seršić, "Wavelet packets approach to blind separation of statistically dependent sources," *Neurocomputing*, vol. 71, nos. 7–9, pp. 1642–1655, Mar. 2008.

[18] V. G. Reju, S. N. Koh, and I. Y. Soon, "An algorithm for mixing matrix estimation in instantaneous blind source separation," *Signal Process.*, vol. 89, no. 3, pp. 1762–1773, Mar. 2009.

[19] S. Kim and C. D. Yoo, "Underdetermined blind source separation based on subspace representation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2604–2614, Jul. 2009.

[20] S. Xie, L. Yang, J.-M. Yang, G. Zhou, and Y. Xiang, "Time-frequency approach to underdetermined blind source separation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 2, pp. 306–316, Feb. 2012.

[21] D. Peng and Y. Xiang, "Underdetermined blind separation of non-sparse sources using spatial time-frequency distributions," *Digital Signal Process.*, vol. 20, no. 2, pp. 581–596, Mar. 2010.

[22] W. Bian and X. Chen, "Smoothing neural network for constrained non-Lipschitz optimization with applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 3, pp. 399–411, Mar. 2012.

[23] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statisticals, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[24] A. Cichocki, R. Zdunek, and S. Amari, "Nonnegative matrix and tensor factorization," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 142–145, Jan. 2008.

[25] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Online nonnegative matrix factorization with robust stochastic approximation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1087–1099, Jul. 2012.

[26] G. Zhou, S. Xie, Z. Yang, J. Yang, and Z. He, "Minimum-volume-constrained nonnegative matrix factorization: Enhanced ability of learning parts," *IEEE Trans. Neural Netw.*, vol. 22, no. 10, pp. 1626–1637, Oct. 2011.

[27] Z. S. He, S. L. Xie, R. Zdunek, G. X. Zhou, and A. Cichocki, "Symmetric nonnegative matrix factorization: Algorithms and applications to probabilistic clustering," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2117–2131, Dec. 2011.

[28] D. D. Lee and H. S. Seung, "Learning the parts of objects by nonnegative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[29] G. Zhou, Z. Yang, S. Xie, and J. M. Yang, "Online blind source separation using incremental nonnegative matrix factorization with volume constraint," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 550–560, Apr. 2011.

[30] Z. Yang, S. Ding, and S. Xie, "Blind source separation by fully nonnegative constrained iterative volume maximization," *IEEE Signal Process. Lett.*, vol. 17, no. 9, pp. 799–802, Sep. 2010.

[31] T. H. Chan, W. K. Ma, C. Y. Chi, and Y. Wang, "A convex analysis framework for blind separation of non-negative sources," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5120–5134, Oct. 2008.

[32] F. Y. Wang, C. Y. Chi, T. H. Chan, and Y. Wang, "Blind separation of positive dependent sources by non-negative least-correlated component analysis," in *Proc. 16th IEEE Signal Process. Soc. Workshop Mach. Learn. Signal Process.*, Sep. 2006, pp. 73–78.

[33] I. J. Lustig, R. E. Marsten, and D. F. Shanno, "Interior point methods for linear programming: Computational state of the art," *ORSA J. Comput.*, vol. 6, no. 1, pp. 1–14, 1994.

[34] Y. Li, S. Amari, A. Cichocki, D. W. C. Ho, and S. Xie, "Underdetermined blind source separation based on sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 423–437, Feb. 2006.

[35] P. Georgiev, F. Theis, and A. Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 992–996, Jul. 2005.

[36] Y. Li, A. Cichocki, and S. Amari, "Analysis of sparse representation and blind source separation," *Neural Comput.*, vol. 16, no. 6, pp. 1193–1234, Jun. 2004.

[37] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1112–1125, Apr. 2011.
[38] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Proc. Summer JPL Airborne Earth Sci. Workshop*, 1995, pp. 23–26.
[39] C.-I. Chang and A. Plaza, "A fast iterative algorithm for implementation of pixel purity index," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 63–67, Jan. 2006.
[40] H. Ren and C.-I. Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1232–1249, Oct. 2003.
[41] A. T. Erdogan, "A simple geometric blind source separation method for bounded magnitude sources," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 438–449, Feb. 2006.
[42] S. Mehrotra, "On the implementation of a primal-dual interior point method," *SIAM J. Optim.*, vol. 2, no. 4, pp. 575–601, 1992.
[43] T.-H. Chan, L. Chen, P. L. Choyke, C.-Y. Chi, G. Wang, and Y. Wang, "Convex analysis for separation of functional patterns in DCE-MRI: A longitudinal study to antiangiogenic therapy," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Oct. 2008, pp. 261–266.
[44] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Berlin, Germany: Springer-Verlag, 2009.
[45] P. J. Huber, "Projection pursuit," *Annal. Stat.*, vol. 13, no. 2, pp. 435–475, Jun. 1985.
[46] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Trans. Comput.*, vol. 23, no. 9, pp. 881–890, Sep. 1974.
[47] V. Laparra, G. Camps-Valls, and J. Malo, "Iterative Gaussianization: From ICA to random rotations," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 537–549, Apr. 2011.
[48] X. Zhang, *Matrix Analysis and Applications*. Beijing, China: Springer-Verlag, 2004.
[49] P. Tichavský and Z. Koldovský, "Optimal pairing of signal components separated by blind techniques," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 119–122, Feb. 2004.

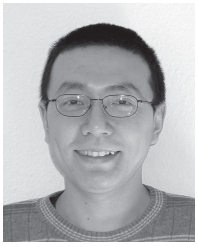
Zuyuan Yang received the B.E. degree from the Hunan University of Science and Technology, Xiangtan, China, and the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 2003 and 2010, respectively.

He is currently a Researcher with the Faculty of Automation, Guangdong University of Technology, Guangzhou. His current research interests include blind source separation, compressed sensing, nonnegative matrix factorization, and image processing.

Dr. Yang was a recipient of the Excellent Ph.D. Thesis Award of Guangdong Province in 2010.

Yong Xiang (SM'12) received the B.E. and M.E. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 1983 and 1989, respectively, and the Ph.D. degree from the University of Melbourne, Melbourne, Australia, in 2003.

He was with the Southwest Institute of Electronic Equipment of China, Chengdu, from 1983 to 1986. In 1989, he joined the University of Electronic Science and Technology of China, where he was a Lecturer from 1989 to 1992 and an Associate Professor from 1992 to 1997. He was a Senior Communications Engineer with Bandspeed Inc., Melbourne, from 2000 to 2002. He is currently an Associate Professor with the School of Information Technology, Deakin University, Melbourne. His current research interests include blind signal and system estimation, information and network security, wireless sensor networks, and multimedia, such as speech, image, and video, processing, compressed sensing, and biomedical signal processing.



Yue Rong (S'03–M'06–SM'11) received the Ph.D. degree (*summa cum laude*) in electrical engineering from the Darmstadt University of Technology, Darmstadt, Germany, in 2005.

He was a Post-Doctoral Researcher with the Department of Electrical Engineering, University of California, Riverside, from 2006 to 2007. Since 2007, he has been with the Department of Electrical and Computer Engineering, Curtin University, Bentley, Australia, where he is currently a Senior Lecturer. His current research interests include signal processing for communications, wireless communications, applications of linear algebra and optimization methods, and statistical and array signal processing.

Dr. Rong was a recipient of the Best Paper Award at the 2011 International Conference on Wireless Communications and Signal Processing, the Best Paper Award at the 2010 Asia-Pacific Conference on Communications, and the Young Researcher of the Year Award of the Faculty of Science and Engineering at Curtin University in 2010. He is an Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS, a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS special issue on theories and methods for advanced wireless relays, a Guest Editor of the *EURASIP Journal on Applied Signal Processing* special issue on signal processing methods for diversity and its applications, and was a TPC Member for the IEEE ICC, WCSP, IWCMC, and ChinaCom.

Shengli Xie (M'01–SM'02) was born in Hubei, China, in 1958. He received the M.S. degree in mathematics from Central China Normal University, Wuhan, China, and the Ph.D. degree in control theory and applications from the South China University of Technology, Guangzhou, China, in 1992 and 1997, respectively.

He is currently a Full Professor and the Director of the Laboratory for Intelligent Information Processing, Faculty of Automation, Guangdong University of Technology, Guangzhou. He has authored or coauthored more than 80 papers in journals and conference proceedings and two monographs. He holds 12 patents. His current research interests include automatic controls, signal processing, blind signal processing, and image processing.