

A Convex Geometry Based Blind Source Separation Method for Separating Nonnegative Sources

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

Abstract—This paper presents a convex geometry (CG) based method for blind separation of nonnegative sources. First, the inaccessible source matrix is normalized to be column-sum-to-one by mapping the available observation matrix. Then, its zero-samples are found by searching the facets of the convex hull spanned by the mapped observations. By taken these zero-samples into account, a quadratic cost function with respect to each row of the unmixing matrix, together with a linear constraint in relation to the involved variables, is proposed. Upon which, an algorithm is presented to estimate the unmixing matrix by solving a classical convex optimization problem. Unlike the traditional blind source separation (BSS) methods, the CG-based method does not require the independence assumption, nor the un-correlation assumption. Compared with the BSS methods that are specifically designed to distinguish between nonnegative sources, the proposed method requires a weaker sparsity condition. Provided simulation results illustrate the performance of our method.

Index Terms—Blind source separation, convex geometry, correlated sources, nonnegative sources.

I. INTRODUCTION

BLIND source separation (BSS) aims to recover unknown sources only from their measurable mixtures. It is a fundamental signal processing problem that arises from various practical applications (e.g., digital communications, speech enhancement, medical data analysis, and remote sensing) [1]-[6]. Since BSS requires little-to-no a prior information of the sources and the mixing matrix, it is typically handled by exploiting the statistical properties of the sources. Under the assumption that the source signals are statistically independent [7], various higher-order statistics (HOS) methods have been proposed for BSS. Among these methods, the independent component analysis (ICA) based methods play an important role, which includes Fast ICA [8], quaternion ICA [9], and

entropy bound minimization ICA [10]. By employing the HOS of the measured data, the HOS-based methods demand lots of data samples in order to reach competitive BSS performance. Different from the HOS-based methods, BSS techniques based on second-order statistics (SOS) can achieve satisfying performance with the size of the sample set greatly reduced, making it the preferred approach when data is limited [11]. SOS-based BSS methods often require the source signals to be mutually uncorrelated, and to have different frequency spectra [12]-[14]. In addition to the statistics based methods, some other methods were developed to separate independent sources [15] and uncorrelated sources [16].

Although independent (or mutually uncorrelated) sources are often encountered in practice, this is also the case with spatially correlated. For example, in remote sensing image processing, the source matrix corresponds to the abundances, which should be column-sum-to-one. Thus, the sources are dependent [17]. Besides, in densely deployed wireless sensor networks, the density of sensors may be very high, and thus, signals from adjacent sensors are unavoidably cross correlated [18]-[20]. In order to separate mutually correlated sources, the sources must have certain special properties.

One of the special properties which can be exploited for BSS is sparsity [21]-[23]. It has been shown that BSS can be achieved by utilizing the sparsity of the sources in the time-domain [24], the frequency-domain [25], or the time-frequency (TF) domain [2], [26]. [24] requires that at least one source is silent at each time step. The sparsity condition required in [25] is that only one source is active at each frequency point. As for [2] and [26], it is required that for each source there exist a set of TF points, where only that source is active. Then clustering is applied to the eigenvectors corresponding to the TF points to estimate the mixing matrix. It is clear that all of these methods impose strong sparsity conditions on the sources, i.e., the sources must be very sparse in time, frequency, or TF domain.

Another one of the special properties that can be employed to perform BSS is the nonnegativity of sources. Nonnegative signals exist in various applications, such as biomedical data analysis and image processing [27]-[29]. A typical approach to processing nonnegative signals is nonnegative matrix factorization (NMF) [30]-[34], which aims to decompose a known nonnegative matrix into the product of two matrices, being both nonnegative. Since the NMF-based methods only exploit the nonnegativity of the sources and the mixing matrix, they can deal with sources that are either independent (or mutually uncorrelated) or spatially correlated. However, these methods

This work was supported in part by the Australian Research Council under grant DP 110102076, the NCET, and the National Natural Science Foundation of China under grants 61104053, 61333013, 61273192, U1201253.

Z. Yang is with the Faculty of Automation, Guangdong University of Technology, Guangzhou, 510006, China, he is also with the School of Information Technology, Deakin University, Burwood Campus, Melbourne, VIC 3125, Australia (tel: +61 3 52272905, fax: +61 3 52272167, e-mail: yangzuyuan@aliyun.com).

Y. Xiang is with the School of Information Technology, Deakin University, Burwood Campus, Melbourne, VIC 3125, Australia (Tel: +61 3 92517740, fax: +61 3 92446831, e-mail: yxiang@deakin.edu.au).

Y. Rong is with the Department of Electrical and Computer Engineering, Curtin University, Bentley, WA 6102, Australia (Tel: +61 8 92667398, fax: +61 8 92662584, e-mail: y.rong@curtin.edu.au).

K. Xie is with the Faculty of Automation, Guangdong University of Technology, Guangzhou, 510006, China (tel: +86 20 87114702, fax: +86 20 87114702, e-mail: kanxiegdut@gmail.com).

have local minima problem that results from using the alternative least square iteration optimization scheme, and thus, BSS is not a guarantee. Similarly, by using the nonnegativity of sources, the method in [35] does not depend on statistical features of the sources. Specifically, it is shown [35] that one can obtain a finite set of candidate source signals that contain the original source signals, and the latter can be identified if they are the most linearly independent (MLI) among the respective set of candidate source signals. However, this MLI assumption does not hold in many practical signals.

Recently, Chan *et al* has developed a BSS method by jointly utilizing the nonnegativity and the spatial feature of the sources [36]. With regard to this feature, it requires the pure source sampling (PSS) (or local dominance) condition, i.e., there exists one time instant at which only its sample is nonzero for each source [36]. This condition is also employed by other methods [29], [37], [38] as well, and the PSS-based methods can achieve BSS perfectly without knowing any prior statistical information of the sources. That is to say, these methods need neither large samples nor different spectra of the sources. However, they are restricted by the PSS condition, which requires that all sources are zero besides just one. In order to cope with a wider range of applications, it is important to relax this condition.

In this paper, under weaker condition, we propose a novel convex geometry (CG) based method for blind separation of nonnegative sources. In the CG-based BSS method, the accessible observation matrix is first mapped to be column-sum-to-one, such that the unaccessible source matrix is also mapped to be column-sum-to-one. As a result, the columns of the mapped observation matrix span a convex hull that is covered by the convex hull spanned by the columns of the mixing matrix. Then, the zero-sample positions of the sources can be found by searching the facets of this observed hull with existing algorithms. After that, a quadratic cost function, with respect to the rows of the unmixing matrix, is proposed and the corresponding linear constraints to these rows are derived. Finally, the unmixing matrix is estimated by minimizing the proposed cost function under given linear constraints, which is a typical convex optimization problem. To achieve BSS, the proposed CG-based method only requires a mild sparsity assumption on the sources, which is called sufficient boundary sampling (SBS) assumption. As will be shown in Section III, the SBS assumption only requires sources to have a small number of zero-samples, i.e., the sources do not need to be very sparse. Also, the SBS assumption is weaker than the PSS condition used in [36] and [37]. Besides, compared with the traditional SOS- or HOS-based BSS methods, our method does not require the sources to be independent nor require them to be uncorrelated.

The remainder of the paper is organized as follows. Section II formulates the problem of BSS and gives relevant hypotheses. The CG-based BSS method is proposed in Section III, which also provides an analysis on source identifiability. Section IV provides simulation results to compare the performance of the proposed method with that of some benchmark methods. Section V concludes the paper.

The following notations are used throughout the paper:

\mathbf{x}, x_i	Column vector, the i th element of \mathbf{x} ;
$\mathbf{X}, \mathbf{x}_j, x_{ij}$	Matrix, the j th column of \mathbf{X} , the (i, j) th entry of \mathbf{X} ;
$\mathbf{X}^T, \mathbf{X}^{-1}$	Transpose of \mathbf{X} , inverse of \mathbf{X} ;
Ω	Convex hull;
Φ_i	The i th facet of a convex hull;
Ψ_i	Index set of the points covered by Φ_i ;
$\aleph(\Psi_i)$	Number of the elements in Ψ_i ;
\Re	Real number set;
$\mathbf{0}$	All zero column vector;
$\mathbf{1}$	All one column vector.

II. PROBLEM FORMULATION

The instantaneous mixing model under consideration is as follows [10], [15], [36]:

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (1)$$

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, N denote the numbers of the observations (or outputs), the sources (or inputs) and the samples, respectively. The objective of BSS is to find a matrix $\mathbf{W} \in \Re^{m \times n}$ such that the matrix $\mathbf{Y} \in \Re^{n \times N}$ given by

$$\mathbf{Y} = \mathbf{W}^T \mathbf{X} \quad (2)$$

is an estimate of \mathbf{S} , up to row permutation and scaling. Equivalently, this means that

$$\mathbf{W}^T \mathbf{A} = \mathbf{P}\mathbf{\Lambda} \quad (3)$$

where \mathbf{P} is a permutation matrix and $\mathbf{\Lambda}$ is a diagonal scaling matrix. Here, \mathbf{W}^T is called the unmixing matrix. In order to achieve BSS, some assumptions must be made on the sources and the mixing matrix. In this paper, we assume

A-1) All sources are nonnegative, i.e., $s_{j,t} \geq 0$, where $j = 1, 2, \dots, n$ and $t = 1, 2, \dots, N$.

A-2) The sources satisfy the SBS condition which will be defined later.

A-3) $m \geq n$ and \mathbf{A} is nonnegative with full column rank.

A-1) holds in various applications (e.g., image processing, as image intensities are often represented by numbers having either positive or zero values [39]). A-2) is a mild sparsity condition on the sources. We shall show in the following section that it is weaker than the PSS or *local dominance* condition used in [36] and [37]. A-3) originates from some practical applications, such as optical spectroscopy analysis and remote sensing image interpretation, as the observations are nonnegative combinations of nonnegative signals [37], [40], [41].

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 existed nonnegative rank reduction method. We also assume that there is no zero-columns in \mathbf{X} . If this is not the case, zero-valued columns can be easily detected and removed in advance.

III. THE CG-BASED METHOD

Under A-3), it follows from (3) that any row of \mathbf{W}^T should be orthogonal to some $n - 1$ columns of \mathbf{A} . Thus, it is orthogonal to the hyper-plane spanned by these column vectors. If one can find enough *useful points* in each of the n hyper-planes spanned by $n - 1$ column vectors of \mathbf{A} , then all rows of \mathbf{W}^T can be estimated. Since only the observation matrix \mathbf{X} is known in the context of BSS, we shall search these *useful points* by analyzing \mathbf{X} . To proceed, we first introduce the related convex hulls and facets.

A. Convex hulls and facets

From (1), the t th column of \mathbf{X} can be written as

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t = \sum_{j=1}^n \mathbf{a}_j s_{jt}, \forall t. \quad (4)$$

Also, based on A-1) and A-3), it holds that $\mathbf{1}^T \mathbf{a}_j > 0, \forall j$ and $\mathbf{1}^T \mathbf{x}_t > 0, \forall t$, where $\mathbf{1}$ denotes a vector whose elements are all one. Now we map \mathbf{x}_t as follows:

$$\begin{aligned} \tilde{\mathbf{x}}_t &= \frac{\mathbf{x}_t}{\mathbf{1}^T \mathbf{x}_t} \\ &= \frac{\mathbf{A}\mathbf{s}_t}{\mathbf{1}^T \mathbf{x}_t} \\ &= \sum_{j=1}^n \frac{\mathbf{a}_j}{\mathbf{1}^T \mathbf{a}_j} \frac{\mathbf{1}^T \mathbf{a}_j s_{jt}}{\mathbf{1}^T \mathbf{x}_t} \\ &= \sum_{j=1}^n \tilde{\mathbf{a}}_j \tilde{s}_{jt} \end{aligned} \quad (5)$$

where

$$\tilde{\mathbf{a}}_j = \frac{\mathbf{a}_j}{\mathbf{1}^T \mathbf{a}_j} \quad \text{and} \quad \tilde{s}_{jt} = \frac{\mathbf{1}^T \mathbf{a}_j}{\mathbf{1}^T \mathbf{x}_t} s_{jt}.$$

Clearly, $\mathbf{1}^T \tilde{\mathbf{a}}_j = 1, \tilde{a}_{ij} \geq 0, \mathbf{1}^T \tilde{\mathbf{x}}_t = 1, \tilde{x}_{it} \geq 0$ and $\tilde{s}_{jt} \geq 0, \forall i, j, t$. Moreover, similar to (4), one can write (5) to be

$$\tilde{\mathbf{x}}_t = \tilde{\mathbf{A}}\tilde{\mathbf{s}}_t \quad (6)$$

where $\tilde{\mathbf{A}}$ is a matrix whose j th column is $\tilde{\mathbf{a}}_j$ and $\tilde{\mathbf{s}}_t$ is a column vector whose j th element is \tilde{s}_{jt} . Similar to (1), the matrix form of (6) can be expressed as

$$\tilde{\mathbf{X}} = \tilde{\mathbf{A}}\tilde{\mathbf{S}}. \quad (7)$$

It is easy to verify that

$$\begin{aligned} \mathbf{1}^T \tilde{\mathbf{s}}_t &= \tilde{s}_{1t} + \tilde{s}_{2t} + \cdots + \tilde{s}_{nt} \\ &= \frac{\mathbf{1}^T (\mathbf{a}_1 s_{1t} + \mathbf{a}_2 s_{2t} + \cdots + \mathbf{a}_n s_{nt})}{\mathbf{1}^T \mathbf{x}_t} \\ &= \frac{\mathbf{1}^T (\mathbf{A}\mathbf{s}_t)}{\mathbf{1}^T \mathbf{x}_t} \\ &= 1. \end{aligned} \quad (8)$$

Based on (4), (6) and (8), we can see that by mapping \mathbf{x}_t , the inaccessible source vector \mathbf{s}_t has been converted to $\tilde{\mathbf{s}}_t$ whose elements are sum-to-one, where $t = 1, 2, \dots, N$.

Denote the convex hulls spanned by points $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_N$ and $\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \dots, \tilde{\mathbf{a}}_n$ by Ω and Ω^* , respectively. We have the following lemma.

Lemma 1: Ω is covered by Ω^* , i.e., $\Omega \subseteq \Omega^*$.

Proof: Since Ω is a convex hull, for any point $\mathbf{y} \in \Omega$, it can be represented as

$$\mathbf{y} = \sum_{t=1}^N \tilde{\mathbf{x}}_t \alpha_t$$

where $\sum_{t=1}^N \alpha_t = 1, \alpha_t \geq 0, \forall t$. Substituting (5) into the above equation yields

$$\begin{aligned} \mathbf{y} &= \sum_{t=1}^N \sum_{i=1}^n \tilde{\mathbf{a}}_i \tilde{s}_{it} \alpha_t \\ &= \sum_{i=1}^n \tilde{\mathbf{a}}_i \sum_{t=1}^N \tilde{s}_{it} \alpha_t \\ &= \sum_{i=1}^n \tilde{\mathbf{a}}_i \beta_i \end{aligned}$$

where $\beta_i = \sum_{t=1}^N \tilde{s}_{it} \alpha_t, \forall i$. Since $\tilde{s}_{it} \geq 0, \forall i, t$, it is in accordance with $\beta_i \geq 0, \forall i$. Furthermore, based on (8), we have

$$\begin{aligned} \sum_{i=1}^n \beta_i &= \sum_{i=1}^n \sum_{t=1}^N \tilde{s}_{it} \alpha_t \\ &= \sum_{t=1}^N \sum_{i=1}^n \tilde{s}_{it} \alpha_t \\ &= \sum_{t=1}^N \alpha_t \\ &= 1. \end{aligned}$$

Therefore, $\mathbf{y} \in \Omega^*$, which leads to $\Omega \subseteq \Omega^*$. This completes the proof. \blacksquare

Lemma 1 shows that any point in Ω is covered by Ω^* , which implies $\{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_N\} \in \Omega^*$. Note that in the case $m = n$, the hyper-planes spanned by $n - 1$ column vectors of \mathbf{A} correspond to those spanned by $n - 1$ column vectors of $\tilde{\mathbf{A}}$ one-by-one, and the hyper-planes associated with $\tilde{\mathbf{A}}$ correspond to the facets of Ω^* one-by-one. Hence, for the sake of estimating the unmixing matrix \mathbf{W}^T , searching the *useful points* from \mathbf{X} is equivalent to searching the *special points* lying in the facets of Ω^* from the mapped observation matrix $\tilde{\mathbf{X}}$. Meaning, the locations of the *useful points* in \mathbf{X} are the same as the locations of the *special points* in $\tilde{\mathbf{X}}$. As for these special points, we have another lemma.

Lemma 2: For any point from $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_N$, if it lies in a facet of Ω^* , it must lie in one facet of Ω .

Proof: We use apagoge to prove this lemma. Let \mathbf{x} be a point from $\{\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_N\}$ lying in a facet of Ω^* . Then \mathbf{x} can be represented as

$$\mathbf{x} = \sum_{i=1}^n \tilde{\mathbf{a}}_i \beta_i \quad (9)$$

where $\sum_{i=1}^n \beta_i = 1, \beta_i \geq 0, \forall i$ and $\exists k, \beta_k = 0$. Since \mathbf{x} lies in Ω , if \mathbf{x} does not lie in any facet of Ω , then there exist n linear independent vectors or points in Ω , denoted by $\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \dots, \tilde{\mathbf{y}}_n$, without loss of generality, such that

$$\mathbf{x} = \sum_{j=1}^n \tilde{\mathbf{y}}_j \gamma_j$$

where $\sum_{j=1}^n \gamma_j = 1$ and γ_j is strictly greater than 0 for all j .

Since $\tilde{\mathbf{y}}_j \in \Omega, \forall j$, it can be represented as $\tilde{\mathbf{y}}_j = \sum_{t=1}^N \tilde{\mathbf{x}}_t \lambda_{tj}$,

where $\sum_{t=1}^N \lambda_{tj} = 1, \lambda_{tj} \geq 0, \forall t, j$. Moreover, it follows from

(5) that $\tilde{\mathbf{x}}_t = \sum_{i=1}^n \tilde{\mathbf{a}}_i \tilde{s}_{it}$, which leads to

$$\begin{aligned} \mathbf{x} &= \sum_{j=1}^n \sum_{t=1}^N \tilde{\mathbf{x}}_t \lambda_{tj} \gamma_j \\ &= \sum_{j=1}^n \sum_{t=1}^N \sum_{i=1}^n \tilde{\mathbf{a}}_i \tilde{s}_{it} \lambda_{tj} \gamma_j \\ &= \sum_{i=1}^n \tilde{\mathbf{a}}_i \sum_{j=1}^n \sum_{t=1}^N \tilde{s}_{it} \lambda_{tj} \gamma_j. \end{aligned} \quad (10)$$

Recall that $\exists k, \beta_k = 0$. Then, by comparing (9) and (10), one can conclude that there exists k , such that $\sum_{j=1}^n \sum_{t=1}^N \tilde{s}_{kt} \lambda_{tj} \gamma_j = 0$. Since $\tilde{\mathbf{S}}$ and $\lambda_{tj}, \forall t, j$ are nonnegative

and $\gamma_j > 0, \forall j$, there must exist k such that $\sum_{t=1}^N \tilde{s}_{kt} \lambda_{tj} = 0, \forall j$. That is to say, all elements in the k th row of the matrix $\left[\sum_{t=1}^N \tilde{s}_t \lambda_{t1}, \sum_{t=1}^N \tilde{s}_t \lambda_{t2}, \dots, \sum_{t=1}^N \tilde{s}_t \lambda_{tn} \right]$ are zeros. Therefore,

$$\begin{aligned} &\det([\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \dots, \tilde{\mathbf{y}}_n]) \\ &= \det(\tilde{\mathbf{A}}) \det\left(\left[\sum_{t=1}^N \tilde{s}_t \lambda_{t1}, \dots, \sum_{t=1}^N \tilde{s}_t \lambda_{tn}\right]\right) \\ &= 0. \end{aligned} \quad (11)$$

On the other hand, $\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \dots, \tilde{\mathbf{y}}_n$ are linear independent. So, it must hold that $\det([\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \dots, \tilde{\mathbf{y}}_n]) \neq 0$, which contradicts (11). This completes the proof. ■

Lemma 2 shows that the implicit *special points* of $\tilde{\mathbf{X}}$ can be found through searching the facets of Ω which is known. In general, the number of the facets of Ω is more than n . In order to estimate the unmixing matrix which has n rows, one needs to select n proper facets. The determination of these facets often depends on the conditions on the sources. The SBS condition mentioned in A-2) will be presented in the next subsection.

B. SBS condition

We start with the following definition.

Definition 1: For the nonnegative source matrix $\mathbf{S} \in \mathbb{R}^{n \times N}$, if each source has at least $n - 1$ zero-samples and the

sub-matrix composed of the samples of all other sources corresponding to those zero-sample instants is full row rank, then \mathbf{S} is said to satisfy the *divergent boundary sampling* condition.

This definition describes a class of sources which have both zero-samples and some kind of nonzero-samples. Due to the *divergent boundary sampling* condition of the sources, the available convex hull Ω has a special relationship with the unknown convex hull Ω^* .

Proposition 1: If the nonnegative source matrix $\mathbf{S} \in \mathbb{R}^{n \times N}$ satisfies the *divergent boundary sampling* condition, then Ω contains n special facets such that in each facet, all points from $\tilde{\mathbf{X}}$ lie in a facet of Ω^* .

Proof: Since \mathbf{S} satisfies the *divergent boundary sampling* condition, then $\tilde{\mathbf{S}}$ satisfies this condition. Thus, for $\forall i \in \{1, 2, \dots, n\}$, there exists a $n \times (n - 1)$ sub-matrix in $\tilde{\mathbf{S}}$, such that the i th row consists of zero-elements and the remaining rows form a full rank $(n - 1) \times (n - 1)$ square matrix. Denote the facet spanned by the columns of $\tilde{\mathbf{X}}$ corresponding to this sub-matrix by $\tilde{\Phi}_i$. Because both $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{S}}$ are column-sum-to-one, one can see that for $\tilde{\mathbf{X}}$, any point covered by $\tilde{\Phi}_i, \forall i$ corresponds to a boundary sample column of $\tilde{\mathbf{S}}$ and the i th element of this column equals zero. Thus, these points lie in a facet of Ω^* , and $\tilde{\Phi}_1, \tilde{\Phi}_2, \dots, \tilde{\Phi}_n$ are the eligible facets. This completes the proof. ■

Without loss of generality, throughout this paper, we denote the n special facets of Ω by $\Phi_1, \Phi_2, \dots, \Phi_n$ and the other facets of Ω by $\Phi_{n+1}, \Phi_{n+2}, \dots$. Let $\Psi_i, \forall i$ be the index set of the columns in $\tilde{\mathbf{X}}$ which are covered by $\tilde{\Phi}_i$. In order to find these special facets, we present the SBS condition as follows.

Definition 2: For the source matrix $\mathbf{S} \in \mathbb{R}^{n \times N}$ satisfying the *divergent boundary sampling* condition, if $\forall i \leq n, j > n$ and $\aleph(\Psi_i) > \aleph(\Psi_j)$, then \mathbf{S} is said to satisfy the SBS condition.

The following proposition shows that the SBS condition is less restrictive on the sources than the PSS condition utilized in [36] and [37].

Proposition 2: The SBS condition is equivalent to the PSS condition in the case of $n = 2$, but weaker than the latter for $n > 2$.

Proof: If $n = 2$, i.e., there are only two sources, the SBS condition means that for each source, there exists one time instant at which only its sample value is zero. This is the same as PSS condition which requires that for each source, there exists one time instant at which only its sample value is nonzero.

For $n > 2$, if \mathbf{S} satisfies the PSS condition, there exists the identity matrix in $\tilde{\mathbf{S}}$. Thus \mathbf{S} must satisfy *divergent boundary sampling* condition and Ω has only n facets where each facet covers at least $n - 1$ points in $\tilde{\mathbf{X}}$, i.e., $\aleph(\Psi_i) \geq n - 1, \forall i \leq n$ and $\aleph(\Psi_i) = 0, \forall i > n$. Meaning, \mathbf{S} must also satisfy the SBS condition. On the other hand, the SBS condition does not have restriction on the exact locations of the points in the facets, i.e., none of the points needs to be at a vertex of Ω^* . In contrast, the PSS condition requires that some points¹ must appear at all the

¹Actually, these points correspond to the pure source sample columns in the source matrix \mathbf{S} .

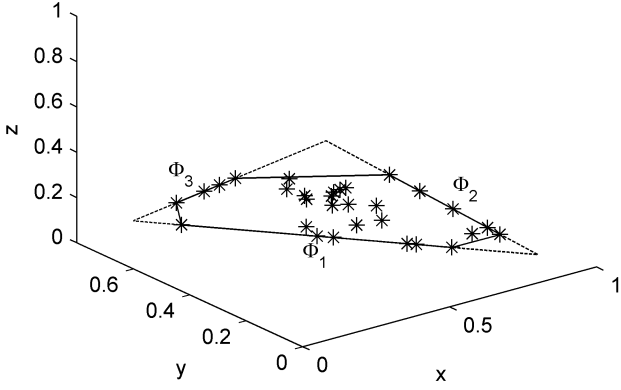


Fig. 1. An illustration of $\tilde{\mathbf{X}}$ in the case of $n = 3$, which shows that \mathbf{S} satisfies the SBS condition but violates the PSS condition, where the symbol ‘*’ represents the columns of $\tilde{\mathbf{X}}$, the triangle bounded by the dashed lines corresponds to Ω^* , the hull bounded by the solid lines corresponds to Ω , and x, y, z denote the three row variables of $\tilde{\mathbf{X}}$, respectively.

vertices of Ω^* . This means that \mathbf{S} satisfying the SBS condition is not guaranteed to satisfy the PSS condition. Therefore, the SBS condition is less restrictive than the PSS condition. This completes the proof. ■

Fig. 1 is an illustration of $\tilde{\mathbf{X}}$ in the case of $n = 3$. It can be seen that \mathbf{S} satisfies the *divergent boundary sampling* condition, as in Φ_1, Φ_2 and Φ_3 , the number of the covered different points are all more than $n - 1$. Moreover, it satisfies the SBS condition, as Φ_1, Φ_2 and Φ_3 all cover at least 4 points in $\tilde{\mathbf{X}}$, more than the number of points covered by any other facet. However, since no point in $\tilde{\mathbf{X}}$ appears at any vertex of Ω^* , then \mathbf{S} does not satisfy the PSS condition.

In the next subsection, we shall derive the CG-based BSS algorithm and analyze the identifiability of the sources.

C. CG-based algorithm and source identifiability analysis

Given the observation matrix \mathbf{X} , the mapped observation matrix $\tilde{\mathbf{X}}$ and the convex hull Ω spanned by all columns of $\tilde{\mathbf{X}}$ are fixed. The facets of Ω can be obtained by using the convex hull based MATLAB toolbox, where the invoked function is `convhulln()` and the parameters depend only on the columns of $\tilde{\mathbf{X}}$ [42]. Besides, it is known [43] that if a facet is spanned by n -dimensional linear independent vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1}$, then any point \mathbf{v}^* lies in this facet if and only if

$$\det([\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1}, \mathbf{v}^*]) = 0. \quad (12)$$

Therefore, for each facet of Ω , one can use the classical determinant based method in (12) to find all the points in $\tilde{\mathbf{X}}$ which are covered by this facet.

As we previously mentioned, Ω usually has more than n facets and the n special facets of Ω , denoted by $\Phi_1, \Phi_2, \dots, \Phi_n$, need to be identified. Based on A-2) together with the definition of the SBS condition, the special facets $\Phi_1, \Phi_2, \dots, \Phi_n$ are the facets that cover a larger number of points in $\tilde{\mathbf{X}}$, i.e., the first n facets with the largest number of

points. Hence, by counting the number of points in each facet, we can find $\Phi_1, \Phi_2, \dots, \Phi_n$ and thus find the corresponding $\Psi_1, \Psi_2, \dots, \Psi_n$, where $\Psi_i, i = 1, 2, \dots, n$ are the index set of the columns in $\tilde{\mathbf{X}}$ which are covered by Φ_i . It is known from the subsection III.B, the columns of $\tilde{\mathbf{X}}$ which are indicated by $\Psi_1, \Psi_2, \dots, \Psi_n$ are those *special points* lying in the facets of Ω^* . In other words, the *special points* in $\tilde{\mathbf{X}}$ are $\tilde{\mathbf{x}}_t, t \in \Psi_1 \cup \Psi_2 \cup \dots \cup \Psi_n$. As a result, the *useful points* in \mathbf{X} which can be used to estimate the unmixing matrix \mathbf{W}^T are identified as $\mathbf{x}_t, t \in \Psi_1 \cup \Psi_2 \cup \dots \cup \Psi_n$.

For each $i \in \{1, 2, \dots, n\}$, the i th row of \mathbf{W}^T should be as orthogonal to all the points of \mathbf{X} , whose indices belong to Ψ_i , as possible. Thus it is logical to estimate the i th row of \mathbf{W}^T by minimizing the quadratic function $\sum_{t \in \Psi_i} \mathbf{w}_i^T \mathbf{x}_t \mathbf{x}_t^T \mathbf{w}_i$. In order to avoid trivial solution, we constrain \mathbf{w}_i to be nonzero by a linear constraint $\mathbf{w}_i^T \mathbf{X} \mathbf{1} = 1$, which means that the i th estimated source is normalized to be sum-to-one. Based on these discussions, we propose to estimate the i th row of \mathbf{W}^T through the following constrained optimization:

$$\begin{cases} \text{Minimize } \sum_{t \in \Psi_i} \mathbf{w}_i^T \mathbf{x}_t \mathbf{x}_t^T \mathbf{w}_i \\ \text{subject to } \mathbf{w}_i^T \mathbf{X} \mathbf{1} = 1 \end{cases} \quad (13)$$

where $i = 1, 2, \dots, n$.

Clearly, the above constrained optimization is a typical convex quadratic programming (QP) problem. The following theorem shows that the optimal solutions of (13) with respect to all i lead to perfect source separation.

Theorem 1: Under A-1) to A-3), it holds that

$$\hat{\mathbf{W}}^T \mathbf{A} = \mathbf{P} \mathbf{A} \quad (14)$$

where $\hat{\mathbf{W}}^T$ is a matrix which is formed by using the n optimal solutions of (13) as its rows.

Proof: See Appendix A. ■

It is worth mentioning that based on A-3), it holds that $\Phi_i \not\subseteq \Phi_j, \forall i \neq j$, resulting in $\Psi_i \not\subseteq \Psi_j, \forall i \neq j$. Thus, estimating any row of \mathbf{W}^T from the optimization in (13) is irrelevant to the estimation of the other rows of \mathbf{W}^T , which avoids the accumulation of estimation errors. The optimization in (13) can be implemented by using existing softwares and the optimization toolbox in MATLAB is used in this paper.

In summary, the proposed CG-based algorithm is formulated as follows.

- Step 1: Map \mathbf{X} into $\tilde{\mathbf{X}}$ by using (5).
- Step 2: Calculate² the facets of the convex hull Ω spanned by the columns of $\tilde{\mathbf{X}}$.
- Step 3: For each facet of Ω , use (12) to find all the points in $\tilde{\mathbf{X}}$ which are covered by this facet.
- Step 4: Obtain the index set $\Psi_1, \Psi_2, \dots, \Psi_n$, which corresponds to the first n facets with the largest number of points.
- Step 5: Obtain $\hat{\mathbf{w}}_i, i = 1, 2, \dots, n$ by optimizing (13). Then the unmixing matrix and the sources are estimated as $\hat{\mathbf{W}}^T = [\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, \dots, \hat{\mathbf{w}}_n]^T$ and $\hat{\mathbf{S}} = \hat{\mathbf{W}}^T \mathbf{X}$.

²One can invoke the function `convhulln()` in Matlab toolbox.

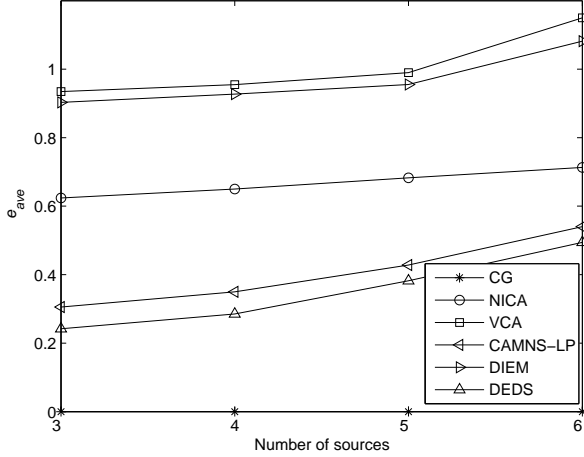


Fig. 2. e_{ave} versus the number of sources, where noise is absent.

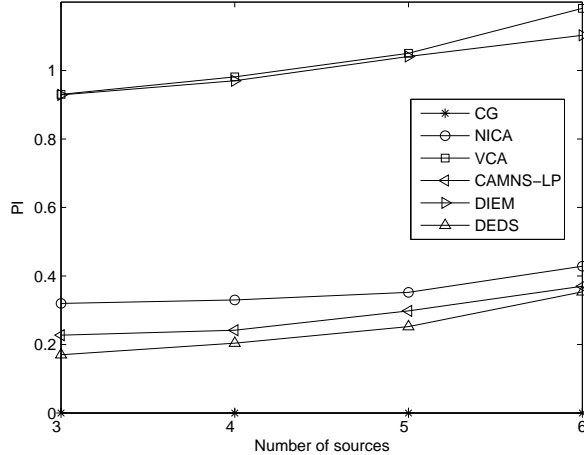


Fig. 3. PI versus the number of sources, where noise is absent.

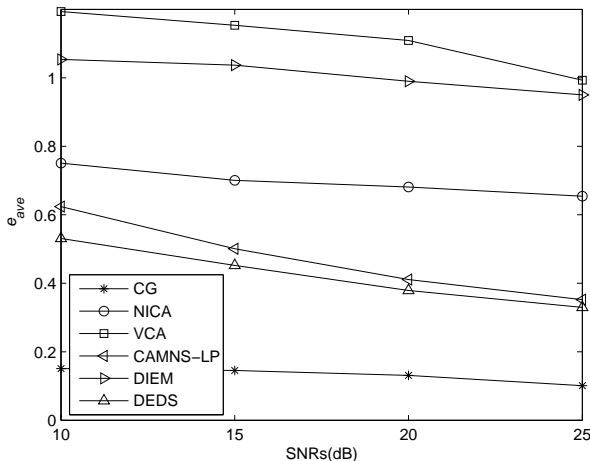


Fig. 4. e_{ave} versus SNRs, where the number of sources is three.

IV. SIMULATION RESULTS

In this section, we provide simulation examples to illustrate the performance of the proposed CG-based algorithm, in comparison with the NICA algorithm [7], the DIEM algorithm [12], the DEDS algorithm [35], the CAMNS-LP algorithm [36], and the VCA algorithm [37]. Based on [36], the source separation performance is measured by the mean of the sum square error (M-SSE) index defined as follows:

$$e(\mathbf{S}, \hat{\mathbf{S}}) = \frac{1}{n} \min_{\pi \in \Pi_n} \sum_{i=1}^n \|\mathbf{s}^i - \hat{\mathbf{s}}^{\pi_i}\|^2 \quad (15)$$

where \mathbf{s}^i is the i th row of the source matrix \mathbf{S} , $\hat{\mathbf{s}}^i$ is the i th row of the estimated source matrix $\hat{\mathbf{S}}$, $\boldsymbol{\pi} = [\pi_1, \pi_2, \dots, \pi_n]^T$, and $\Pi_n = \{\boldsymbol{\pi} \in \mathbb{R}^{n \times 1} \mid \pi_i \in \{1, 2, \dots, n\}, \pi_i \neq \pi_j, \forall i \neq j\}$ is the set of all permutations of $\{1, 2, \dots, n\}$. Here, the L2-norms of \mathbf{s}^i and $\hat{\mathbf{s}}^i$, $\forall i$ are normalized to be one. The optimization in (15) aims to find the best match between the original sources and the estimated sources, which can be solved by the Hungarian algorithm in [44]. In addition, the classical BSS performance index (PI) for the global matrix $\mathbf{G} = \hat{\mathbf{W}}^T \mathbf{A}$ is also used to measure the source separation performance. This index is defined as [12]:

$$\text{PI}(\mathbf{G}) = \frac{1}{n(n-1)} \left[\sum_{i \in \mathcal{I}_n} \left(\sum_{j \in \mathcal{I}_n} \frac{|g_{ij}|^2}{\max_l |g_{il}|^2} - 1 \right) + \sum_{j \in \mathcal{I}_n} \left(\sum_{i \in \mathcal{I}_n} \frac{|g_{ij}|^2}{\max_l |g_{lj}|^2} - 1 \right) \right]$$

where \mathcal{I}_n denotes the index set $\{1, 2, \dots, n\}$. Clearly, for the M-SSE index e and the PI, the smaller they are, the better the algorithm performs.

Example 1: Computer generated nonnegative signals with 5000 samples are first used as source signals. These sources satisfy the SBS condition but violate the PSS condition. Besides, the sources are mutually correlated and higher-order dependent. For each compared algorithm, 1000 independent runs are carried out to compute the average M-SSE index e_{ave} and the PI. In each run, the sources are mixed by a randomly generated mixing matrix.

Firstly, we assume that noise is absent. Fig. 2 shows the e_{ave} values obtained by the CG, NICA, VCA, CAMNS-LP, DIEM and DEDS algorithms under different number of sources. One can see that the proposed CG-based algorithm yields perfect source separation regardless of the source number, thanks to the fact that the required SBS condition is satisfied. The DEDS and CAMNS-LP algorithms are the second and third best performers, respectively, among all algorithms compared herein and they do not achieve perfect source separation. The reason is that the MLI (resp. PSS) condition required by the DEDS (resp. CAMNS-LP) algorithm does not hold. Because of the violation of the PSS condition, the performance of the VCA algorithm is not satisfactory. As for the NICA and DIEM algorithms, they fail as the source signals are neither independent nor mutually uncorrelated. Fig. 3 shows the PI values of these algorithms versus the number of sources.

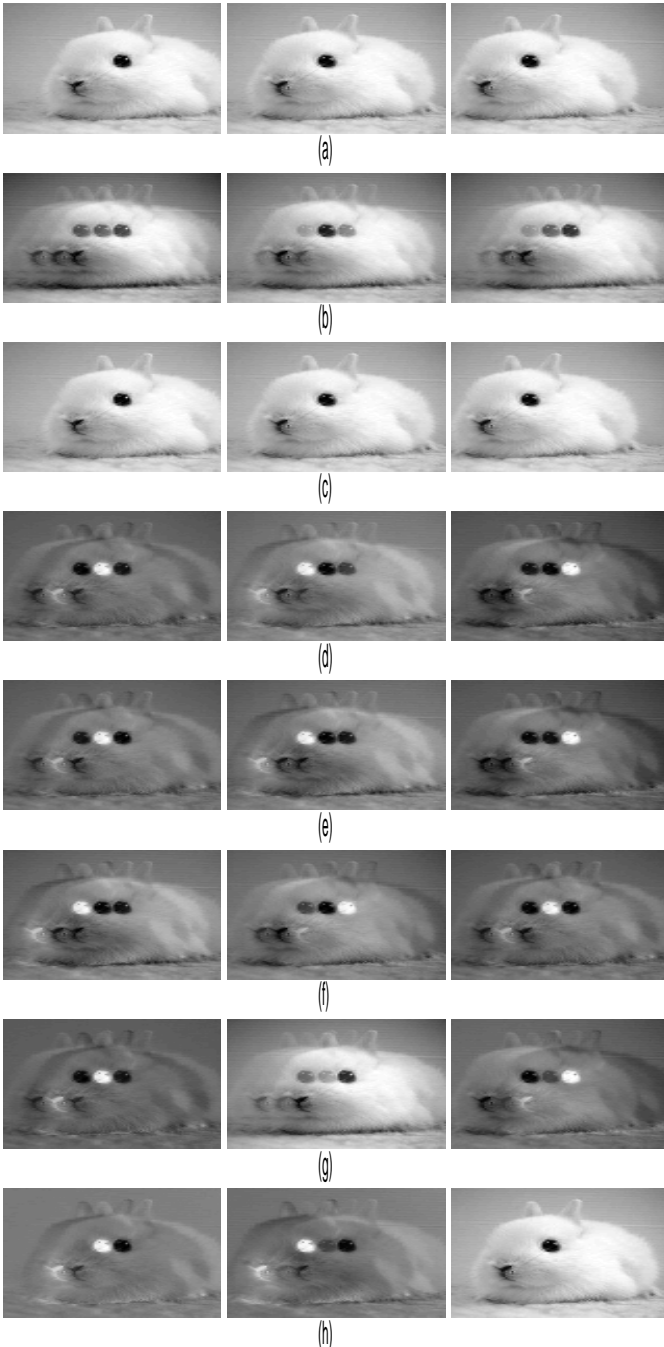


Fig. 5. (a) Original images; (b) Ghosting images; (c) Images recovered by the CG-based algorithm; (d) Images recovered by NICA; (e) Images recovered by VCA. (f) Images recovered by CAMNS-LP; (g) Images recovered by DIEM; (h) Images recovered by DEDS.

Similarly, our algorithm results in perfect source separation and significantly outperforms the other algorithms.

Secondly, we consider the noisy case and assume that the number of sources is $n = 3$. Fig. 4 shows the e_{ave} values of the compared algorithms versus different SNR levels ranging from 10dB to 25dB. As expected, with the rise of SNR, the e_{ave} values of these algorithms decrease. One can also see that the proposed CG algorithm performs much better than the other algorithms at all of the SNR levels considered.

Example 2: The second simulation tests the ability of the

concerned algorithms in reducing image ghosting. Fig. 5(a) shows the three source images, where the second and the third images are the shifted versions of the first image. Clearly, these images are dependent (or correlated) and they do not satisfy the PSS and MLI conditions. The ghosting images are created by merging these source images using the following randomly generated mixing matrix:

$$\mathbf{A} = \begin{bmatrix} 0.8518 & 0.6499 & 0.5666 \\ 0.3342 & 0.6843 & 0.1629 \\ 0.8413 & 0.5801 & 0.2190 \end{bmatrix}.$$

Fig. 5(b) shows the three ghosting images, and Fig. 5(c)-(h) show the ghosting-reduced images by the CG, NICA, VCA, CAMNS-LP, DIEM and DEDS algorithms, respectively. By visual comparison, it can be seen that the CG-based algorithm almost perfectly recover the source images. However, on the other hand, the NICA, VCA, CAMNS-LP and DIEM algorithms fail to reduce the ghosting effects. While the DEDS algorithm successfully extract the second source image, it fails to recover the other two source images.

Example 3: In this example, different natural images shown in Fig. 6(a) are utilized as sources to assess the performance of the CG, NICA, VCA, CAMNS-LP, DIEM and DEDS algorithms. We compute the average M-SSE index e_{ave} based on 1000 independent runs and in each run, a random mixing matrix is used to mix up the source images. The e_{ave} values of the compared algorithms are 0.0002, 1.2309, 1.8275, 1.4018, 1.4686 and 1.3868, respectively. It can be seen that the CG-based algorithm yields very high separation accuracy. In contrast, since the source images are neither independent (or uncorrelated) nor satisfy the PSS and MLI conditions, the other algorithms fail to achieve blind source separation.

Example 4: We compare the performance of the CG, NICA, VCA, CAMNS-LP, DIEM and DEDS algorithms in separating mixed images polluted by additive noise. The noise is measured by the signal to noise ratio (SNR) defined as $SNR = 10\log_{10} (\|\mathbf{x}\|^2 / \|\Delta\mathbf{x}\|^2)$, where \mathbf{x} and $\Delta\mathbf{x}$ denote the signal and the noise, respectively. In this simulation, the source images are the same as those used in Example 3, which are shown in Fig. 6(a). The same mixing matrix \mathbf{A} utilized in Example 2 is employed to mix the sources in the presence of noise, where $SNR=25dB$. Fig. 6(b) shows the three mixtures of the sources, and Fig. 6(c)-(h) show the recoveries obtained by the CG, NICA, VCA, CAMNS-LP, DIEM and DEDS algorithms, respectively. Clearly, the CG-based algorithm achieves satisfactory source separation performance, whilst the other algorithms are unsuccessful in separating the source images from their mixtures.

V. CONCLUSION

In this paper, a CG-based BSS method is proposed to separate nonnegative sources. This method first maps the observation matrix, such that the source matrix is normalized to be column-sum-to-one. Subsequently, the zero-samples of the sources are uncovered by searching the facets of the convex hull comprising the mapped observations. Based on these zero-samples, a quadratic cost function with linear constraint is proposed. The estimate of the unmixing matrix

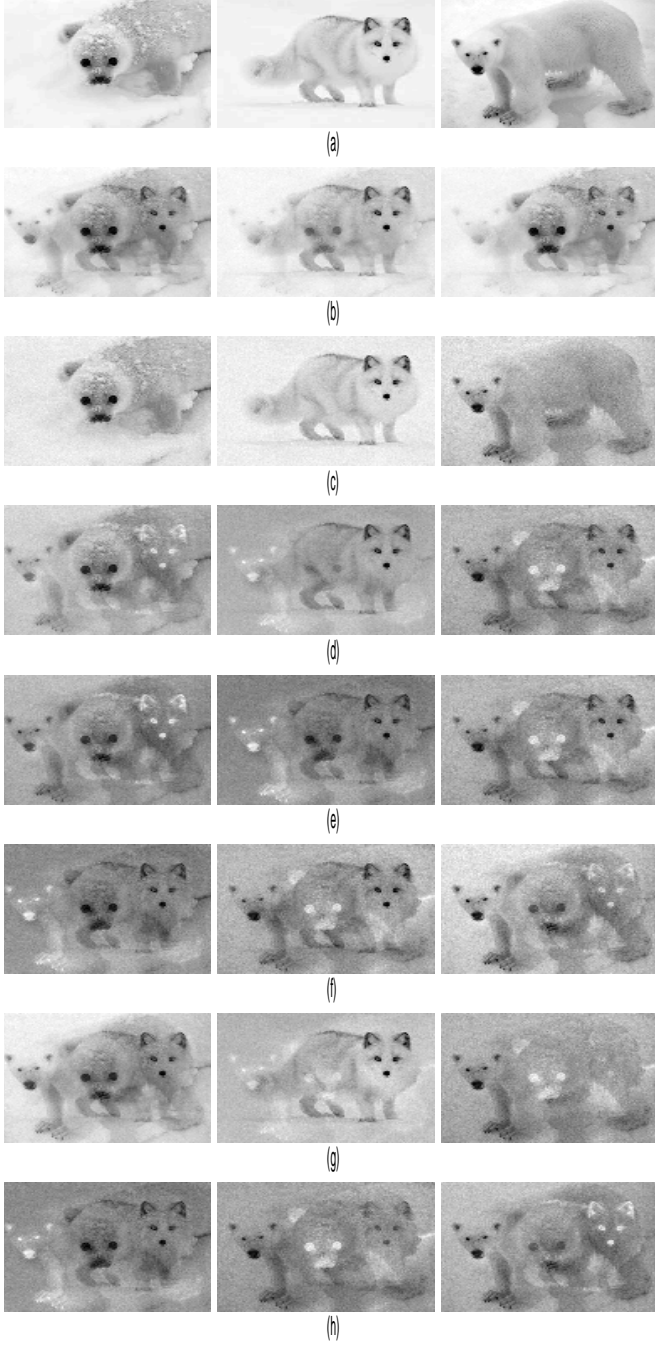


Fig. 6. (a) Original natural images; (b) Noisy mixtures of images, where SNR=25dB; (c) Images recovered by the proposed algorithm; (d) Images recovered by NICA; (e) Images recovered by VCA; (f) Images recovered by CAMNS-LP; (g) Images recovered by DIEM; (h) Images recovered by DEFS.

can be obtained by solving this convex optimization problem. Compared with the BSS methods specifically developed for nonnegative sources, the CG-based method needs a weaker sparsity condition. In relation to conventional BSS methods, the proposed method does not rely on the independence or un-correlation assumption. The superior performance of our method is demonstrated by simulation examples.

APPENDIX A PROOF OF THEOREM 1

Firstly, we consider the cost function in (13) and denote its value by F_i . Let

$$q_{ik} = \sum_{j=1}^m w_{ji} a_{jk} \quad (16)$$

and

$$d_{kk} = \sum_{t=1}^N s_{kt} \quad (17)$$

where $i, k \in \{1, 2, \dots, n\}$.

Thus, from the cost function in (13) and the equation (16), it follows

$$\begin{aligned} F_i &= \sum_{t \in \Psi_i} \left(\sum_{j=1}^m w_{ji} x_{jt} \right)^2 \\ &= \sum_{t \in \Psi_i} \left(\sum_{j=1}^m w_{ji} \sum_{k=1}^n a_{jk} s_{kt} \right)^2 \\ &= \sum_{t \in \Psi_i} \left(\sum_{k=1}^n \sum_{j=1}^m w_{ji} a_{jk} s_{kt} \right)^2 \\ &= \sum_{t \in \Psi_i} \left(\sum_{k=1}^n q_{ik} s_{kt} \right)^2 \\ &= \sum_{t \in \Psi_i} (q_{i1} s_{1t} + q_{i2} s_{2t} + \dots + q_{in} s_{nt})^2. \end{aligned}$$

Clearly, the global minimum of this cost function is zero, and $F_i = 0$ is equivalent to

$$q_{i1} s_{1t} + q_{i2} s_{2t} + \dots + q_{in} s_{nt} = 0, \quad \forall t \in \Psi_i. \quad (18)$$

Based on the analysis about Ψ_i in Section III, one can see that $\exists j, s_{jt} = 0, \forall t \in \Psi_i$. Hence, (18) can be replaced by

$$\begin{cases} \sum_{k=1}^{j-1} q_{ik} s_{kt} + \sum_{k=j+1}^n q_{ik} s_{kt} = 0, \forall t \in \Psi_i \\ q_{ij} = c_i \end{cases} \quad (19)$$

where $c_i \in \mathfrak{R}$ is a constant.

Suppose that $\bar{\mathbf{V}}_i$ is the $\aleph(\Psi_i)$ -column sub-matrix of \mathbf{S} , where the column indices are from Ψ_i . Denote \mathbf{V}_i to be $\bar{\mathbf{V}}_i$ with the j th row removed, and let

$$\mathbf{h}_i = [q_{i1}, \dots, q_{i(j-1)}, q_{i(j+1)}, \dots, q_{in}]^T. \quad (20)$$

Then the first equation in (19) can be written as

$$\mathbf{h}_i^T \mathbf{V}_i = \mathbf{0}^T. \quad (21)$$

According to A-2), \mathbf{V}_i is full row rank. Thus, it results from (21) that

$$\mathbf{h}_i^T = \mathbf{0}^T \mathbf{V}_i^T (\mathbf{V}_i \mathbf{V}_i^T)^{-1}$$

yielding $\mathbf{h}_i = \mathbf{0}$. Further considering (20) gives $q_{ik} = 0, \forall k \neq j$. Therefore, the solution of (18) is

$$q_{ik} = \begin{cases} c_i, & k = j \\ 0, & k \neq j \end{cases} \quad (22)$$

where $j \in \{1, 2, \dots, n\}$.

Secondly, it is easy to see from (1) that

$$x_{jt} = \sum_{k=1}^n a_{jk} s_{kt} \quad \forall j, t. \quad (23)$$

Then, based on the equations (16), (17) and (23), the equation constraint in (13) results in

$$\begin{aligned} \sum_{k=1}^n q_{ik} d_{kk} &= \sum_{k=1}^n \sum_{j=1}^m w_{ji} a_{jk} \sum_{t=1}^N s_{kt} \\ &= \sum_{j=1}^m w_{ji} \sum_{k=1}^n a_{jk} \sum_{t=1}^N s_{kt} \\ &= \sum_{j=1}^m w_{ji} \sum_{t=1}^N x_{jt} \\ &= 1. \end{aligned} \quad (24)$$

Considering A-1), it is obvious that $d_{kk} > 0, \forall k$. Consequently, it follows from (22) and (24) that the global optimal solution of (13) satisfies

$$q_{ik} = \begin{cases} 1/d_{kk}, & k = j \\ 0, & k \neq j \end{cases} \quad (25)$$

where $j \in \{1, 2, \dots, n\}$. From (25), it is clear that the column vector $\mathbf{q}_i = [q_{i1}, q_{i2}, \dots, q_{in}]^T$ has only one nonzero-element.

Furthermore, based on A-3), it holds that $\Psi_i \not\subseteq \Psi_k, \forall i \neq k$. Hence, for different $i \in \{1, 2, \dots, n\}$, the optima of (13) are different from each other. That is to say, for different i , the corresponding j in (25) is different. As a result, for all i, k satisfying $i \neq k$, we have $\mathbf{q}_i \neq \mathbf{0}$ and \mathbf{q}_i is orthogonal to \mathbf{q}_k . Thus, $\mathbf{Q}^T \mathbf{D}$ is a permutation matrix, where $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n]$ and \mathbf{D} is a diagonal matrix whose diagonal entries are $d_{11}, d_{22}, \dots, d_{nn}$. Hence, $\hat{\mathbf{W}}^T \mathbf{A} \mathbf{D} = \mathbf{Q}^T \mathbf{D}$ is a permutation matrix. Consequently, $\hat{\mathbf{W}}^T \mathbf{A}$ can be expressed as

$$\hat{\mathbf{W}}^T \mathbf{A} = \hat{\mathbf{W}}^T \mathbf{A} \mathbf{D} \mathbf{D}^{-1} = \mathbf{P} \mathbf{\Lambda}$$

where $\mathbf{P} = \hat{\mathbf{W}} \mathbf{A} \mathbf{D}$ is a permutation matrix and $\mathbf{\Lambda} = \mathbf{D}^{-1}$ is a diagonal scaling matrix. This completes the proof.

REFERENCES

- [1] W. Bian and X. Chen, "Smoothing neural network for constrained non-Lipschitz optimization with applications," *IEEE Trans. Neural Networks and Learning Systems*, vol. 23, no. 3, pp. 399-411, Mar. 2012.
- [2] S. Xie, L. Yang, J.-M. Yang, G. Zhou, and Y. Xiang, "Time-frequency approach to underdetermined blind source separation," *IEEE Trans. Neural Networks and Learning Systems*, vol. 23, no. 2, pp. 306-316, Feb. 2012.
- [3] M. I. Stanciu, S. Azou, and A. Şerbănescu, "On the blind estimation of chip time of time-hopping signals through minimization of a multimodal cost function," *IEEE Trans. Signal Processing*, vol. 59, no. 2, pp. 842-847, Feb. 2011.
- [4] C. Gouy-Pailler, M. Congedo, C. Brunner, C. Jutten, and G. Pfurtscheller, "Nonstationary brain source separation for multiclass motor imagery," *IEEE Trans. Biomedical Engineering*, vol. 57, no. 2, pp. 469-478, Feb. 2010.
- [5] M. D. Plumbley and E. Oja, "A 'nonnegative pca' algorithm for independent component analysis," *IEEE Trans. Neural Networks*, vol. 15, no. 1, pp. 66-76, Jan. 2004.
- [6] Y. Xiang, V. K. Nguyen, and N. Gu, "Blind equalization of nonirreducible systems using CM criterion," *IEEE Trans. Circuits and Systems II: Express Briefs*, vol. 53, no. 8, pp. 758-762, Aug. 2006.
- [7] E. Oja and M. D. Plumbley, "Blind separation of positive sources by globally convergent gradient search," *Neural Computation*, vol. 16, no. 9, pp. 1811-1825, Sept. 2004.
- [8] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, no. 3, pp. 626-634, May 1999.
- [9] J. Vía, D. P. Palomar, L. Vielva, and I. Santamaría, "Quaternion ICA from second-order statistics," *IEEE Trans. Signal Processing*, vol. 59, no. 4, pp. 1586-1600, Apr. 2011.
- [10] X.-L. Li and T. Adali, "Independent component analysis by entropy bound minimization," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5151-5164, Oct. 2010.
- [11] 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 Processing*, vol. 53, no. 12, pp. 4429-4441, Dec. 2005.
- [12] G. Chabriel and J. Barrere, "A direct algorithm for nonorthogonal approximate joint diagonalization," *IEEE Trans. Signal Processing*, vol. 60, no. 1, pp. 39-47, Jan. 2012.
- [13] K. Abed-Meraim, Y. Xiang, J. H. Manton, and Y. Hua, "Blind source separation using second-order cyclostationary statistics," *IEEE Trans. Signal Processing*, vol. 49, no. 4, pp. 694-701, Apr. 2001.
- [14] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation using second-order statistics," *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 434-444, Feb. 1997.
- [15] A. T. Erdogan, "A simple geometric blind source separation method for bounded magnitude sources," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 438-449, Feb. 2006.
- [16] S. Cruces, "Bounded component analysis of linear mixtures: a criterion of minimum convex perimeter," *IEEE Trans. Signal Processing*, vol. 58, no. 4, pp. 2141-2154, Apr. 2010.
- [17] M. D. Iordache, J. M., Bioucas-Dias, and A. Plaza, "Sparse unmixing of Hyperspectral data," *IEEE Trans. Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2014-2039, Jun. 2011.
- [18] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 51-60, Mar. 2002.
- [19] Y. Xiang, D. Peng, Y. Xiang, and S. Guo, "Novel Z-domain precoding method for blind separation of spatially correlated signals," *IEEE Trans. Neural Networks and Learning Systems*, vol. 24, no. 1, pp. 94-105, Jan. 2013.
- [20] Y. Xiang, S. K. Ng, and V. K. Nguyen, "Blind separation of mutually correlated sources using precoders," *IEEE Trans. Neural Networks*, vol. 21, no. 1, pp. 82-90, Jan. 2010.
- [21] Z. He, A. Cichocki, S. Xie, and K. Choi, "Detecting the number of clusters in n-way probabilistic clustering," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2006-2021, Nov. 2010.
- [22] Y. Li, A. Cichocki, and S. Amari, "Analysis of sparse representation and blind source separation," *Neural Computation* vol. 16, no. 6, pp. 1193-1234, Jun. 2004.
- [23] Z. He, S. Xie, S. Ding, and A. Cichocki, "Convolutional blind source separation in frequency domain based on sparse representation," *IEEE Trans. Audio, Speech and Language Processing*, vol. 15, no. 5, pp. 1551-1563, July 2007.
- [24] P. Georgiev, F. Theis, and A. Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Trans. Neural Networks*, vol. 16, no. 4, pp. 992-996, Jul. 2005.
- [25] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353-2362, Nov. 2001.
- [26] S. Kim and C. Yoo, "Underdetermined blind source separation based on subspace representation," *IEEE Trans. Signal Processing*, vol. 57, no. 7, pp. 2604-2614, Jul. 2009.
- [27] N. Guan, D. Tao, Z. Luo, B. Yuan, "Online nonnegative matrix factorization with robust stochastic approximation," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 23, no. 7, pp. 1087-1099, July 2012.
- [28] G. Zhou, S. Xie, Z. Yang, J. Yang, and Z. He, "Minimum-volume-constrained nonnegative matrix factorization: enhanced ability of learning parts," *IEEE Trans. on Neural Networks*, vol. 22, no. 10, pp. 1626-1637, Oct. 2011.

- [29] 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. Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354-379, Apr. 2012.
- [30] A. Cichocki, R. Zdunek, and S. Amari, "Nonnegative matrix and tensor factorization," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 142-145, Jan. 2008.
- [31] Z. He, S. Xie, R. Zdunek, G. Zhou, and A. Cichocki, "Symmetric nonnegative matrix factorization: Algorithms and applications to probabilistic clustering," *IEEE Trans. Neural Networks*, vol. 22, no. 12, pp. 2117-2131, Dec. 2011.
- [32] D. D. Lee and H. S. Seung, "Learning the parts of objects by nonnegative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788-791, 1999.
- [33] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics and Data Analysis*, vol. 52, no. 1, pp. 155-173, Sept. 2007.
- [34] G. Zhou, A. Cichocki, and S. Xie, "Fast nonnegative matrix/tensor factorization based on low-rank approximation," *IEEE Trans. Signal Processing*, vol. 60, no. 6, pp. 2928-2940, June 2012.
- [35] W. Naanaa and J.-M. Nuzillard, "A geometric approach to blind separation of nonnegative and dependent source signals," *Signal Processing*, vol. 92, no. 11, pp. 2775-2784, Nov. 2012.
- [36] 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 Processing*, vol. 56, no. 10, pp. 5120-5134, Oct. 2008.
- [37] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898-910, Apr. 2005.
- [38] Z. Yang, Y. Xiang, Y. Rong, and S. Xie, "Projection-pursuit-based method for blind separation of nonnegative sources," *IEEE Trans. Neural Networks and Learning Systems*, vol. 24, no. 1, pp. 47-57, Jan. 2013.
- [39] Z. Yang, G. Zhou, S. Xie, S. Ding, J. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Trans. Image Processing*, vol. 20, no. 4, pp. 1112-1125, Apr. 2011.
- [40] S. Moussaoui, D. Brie, A. M. Djafari, and C. Carteret, "Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4133-4145, Nov. 2006.
- [41] R. Ricci, M. Ditzler, and L. Nestor, "Discovering the Beer-Lambert law," *Journal of Chemical Education*, vol. 71, pp. 983-985, Nov. 1994.
- [42] C. B. Barber, D.P. Dobkin, and H. T. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Trans. Mathematical Software*, vol. 22, no. 4, pp. 469-483, Dec. 1996.
- [43] G. Strang, *Linear Algebra and Its Application*, 3rd ed. Cambridge, MA: MIT Press, 1988.
- [44] P. Tichavský and Z. Koldovský, "Optimal pairing of signal components separated by blind techniques," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 119-122, Feb. 2004.