Nonnegative Blind Source Separation by Sparse Component Analysis Based on Determinant Measure

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

Abstract—The problem of nonnegative blind source separation (NBSS) is addressed in this paper, where both the sources and the mixing matrix are nonnegative. Motivated by that many real-world signals are sparse, we deal with NBSS by sparse component analysis. First, a determinant-based sparseness measure, named D-measure, is introduced to gauge the temporal sparseness and spatial sparseness of signals. Based on this measure, a new NBSS model is derived and an iterative sparseness maximization (ISM) approach is proposed to solve this model. In the ISM approach, the NBSS problem can be cast into row-to-row optimizations with respect to the unmixing matrix and then the quadratic programming (QP) technique is used to optimize each row. Furthermore, we analyze the source identifiability and the computational complexity of the proposed ISM-QP method. The new method requires relatively weak conditions on the sources and the mixing matrix, has high computational efficiency, and is easy to implement. Simulation results demonstrate the effectiveness of our method.

Index Terms—Blind source separation, nonnegative sources, sparse component analysis, determinant-based sparseness measure.

I. Introduction

SINCE blind source separation (BSS) techniques can separate unknown sources only from the observed mixtures, they can be used in a wide range of practical applications [1]–[5]. Recently, nonnegative BSS (NBSS), in which both the sources and the mixing matrix are nonnegative, has attracted considerable attentions. So far, a number of methods have been developed for NBSS, such as nonnegative independent

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@yahoo.com.cn).

Y. Xiang is with School of Information Technology, Deakin University, Burwood Campus, Melbourne, VIC 3125, Australia (tel: +61 3 52272905, fax: +61 3 52272167, e-mail: yxiang@deakin.edu.au).

S. Xie is with the Faculty of Automation, Guangdong University of Technology, Guangzhou, 510006, China (tel: +86 20 87114702, fax: +86 20 87114702, e-mail: eeoshlxie@scut.edu.cn).

S. Ding is with the School of Computer Science and Engineering, the University of Aizu, Tsuruga, Ikki-machi, Aizu-Wakamatsu City, Fukushima 965-8580, Japan (tel: +81 242 37 2799, fax: +81 242 37 2734, e-mail: sding@u-aizu.ac.jp).

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).

component analysis (NICA) [6], nonnegative matrix factorization (NMF) [7], convex analysis [8], and nonnegative least-correlated component analysis (nLCA) [9]. NICA is built upon the well-known ICA approach and it has several variations such as nonlinear principal component analysis [6] and geodesic search [10]. Similar to other ICA-based methods, NICA requires that the sources are mutually independent. Other aspects of ICA and NICA including identifiability and convergence are discussed in [11]–[13].

NMF aims to decompose a given nonnegative matrix into two nonnegative factor matrices [14]. Since both the sources and the mixing matrix are nonnegative in NBSS, NMF has potential to be applied to NBSS [7], [15], [16]. By further exploiting other properties of the nonnegative mixing system, some constrained NMF methods have been developed to perform NBSS, such as the flexible component analysis based NMF (FCA-NMF) [17] and the minimum volume constrained NMF [18]–[21]. However, these methods rely on the correct usage of the optimal balance parameter[1] which is difficult to choose in practice.

Convex analysis uses the nonnegativity of the sources to achieve NBSS [8], [22]. Whilst the method in [8] directly yields the recovered sources, the vertex component analysis (VCA) method in [22] estimates the mixing matrix. VCA is often combined with the nonnegative least-square (NLS) method to retrieve the sources, which is called VCA-NLS. The methods in [8] and [22] require the *pure-source sample* assumption, which means that for each source, there exists at least one time instant at which that source dominates [8]. Besides, these methods as well as the NICA- and NMF-based methods are expensive in computation.

The nLCA-based methods utilize the least correlation between the sources [9], [23], [24], where nLCA by iterative volume maximization (nLCA-IVM) in [9] is a representative method. Similar to the minimum volume constrained NMF methods [18]–[21] which minimize the volume of the convex hull spanned by the mixing matrix, nLCA-IVM aims to maximize the volume of the convex hull spanned by the sources. The maximization is implemented by an efficient linear programming based scheme. However, nLCA-IVM can only deal with the case where the mixing matrix is square. In the over-determined scenario (i.e., there are more observations than sources), it needs to use the principal component analysis

---

[1]It is used to balance the decomposition error and the constraint [17].

(PCA) method to preprocess the observations. This may lead to loss of some useful information and thus affects source separation performance [25].

It is known [26]– [29] that many real-world signals are sparse. In this paper, we exploit the sparseness feature of the sources, together with the nonnegativity of the sources and the mixing matrix, to tackle NBSS. To achieve this, it is important to find a proper mathematical measure to gauge the sparseness of the sources. The sparseness of a single signal can be gauged by some existing sparseness measures, such as Donoho's measure [30] which is based on the L0-norm of the signal (i.e., the number of zero elements) and Hoyer's measure [31] which is based the normalized ratio of the L1-norm and L2-norm of the signal (i.e., the ratio of the absolute sum of the elements and the squared root of the quadratic sum of the elements). However, these sparseness measures do not reflect the joint sparseness of multiple sources. In order to describe the joint sparseness of the nonnegative sources, we develop a determinant-based sparseness measure, called D-measure. Based on the D-measure, we propose an iterative sparseness maximization (ISM) approach to perform NBSS, in which the NBSS problem can be cast into row-by-row optimizations with respect to the unmixing matrix and the quadratic programming (QP) can be invoked to optimize each row. For convenience, the proposed method is called ISM-QP. We also analyze the identifiability of the sources and show that our method is of high efficiency in computation.

Unlike the NICA-based methods, the proposed ISM-QP method does not restrict the sources to be mutually independent. Whilst the NMF-based methods suffer from the selection of a proper balance parameter, this problem is avoided in the proposed method. In relation to the methods based on convex analysis, our method does not need the *pure-source sample* assumption. Moreover, the new method is much more efficient in computation than the NICA-based methods, the NMF-based methods and the methods based on convex analysis. Furthermore, compared with nLCA-IVM, the proposed method can be applied to the over-determined case and simulations also show that it has better source separation performance.

The remainder of this paper is organized as follows. In Section II, we first propose the determinant-based sparseness measure, i.e., the D-measure. Based on this measure, a new NBSS model and the corresponding ISM-QP method are derived, together with analysis on source identifiability and computational cost. Section III illustrates the performance of the proposed ISM-QP method using both computer generated data and real biomedical data. Finally, conclusions are drawn in Section IV.

The following notations are used throughout the paper:

$\mathbf{x}, x_i$	column vector, the ith element of $\mathbf{x}$
$\mathbf{X}, \mathbf{x}_j, x_{ij}$	matrix, the jth row of $\mathbf{X}$, the (i,j)th entry of $\mathbf{X}$
$\mathbf{X}^T, \mathbf{X}^{-1}$	transpose of $\mathbf{X}$, inverse of $\mathbf{X}$
$\mathbf{X}^{-T}, \det(\mathbf{X})$	transposed inverse of $\mathbf{X}$, determinant of $\mathbf{X}$.

II. NBSS BASED ON SPARSE COMPONENT ANALYSIS

We consider the following instantaneous BSS mixing model with m observations and n sources:

$$\mathbf{X} = \mathbf{AS} \qquad (1)$$

where $\mathbf{X} \in \Re^{m \times K}$ is the observation matrix, $\mathbf{A} \in \Re^{m \times n}$ is the mixing matrix, $\mathbf{S} \in \Re^{n \times K}$ is the source matrix, and K denotes the number of samples. In practice, K is usually much greater than m and n, i.e., $K \gg m, n$. The corresponding unmixing model is:

$$\mathbf{Y} = \mathbf{WX} = \mathbf{WAS} \qquad (2)$$

where $\mathbf{W} \in \Re^{n \times m}$ is the unmixing matrix to be obtained and $\mathbf{Y} \in \Re^{n \times K}$ denotes the estimate of the source matrix $\mathbf{S}$. If $\mathbf{WA}$ equals the multiplication of a permutation matrix and a diagonal scaling matrix, then $\mathbf{Y}$ will be equal to $\mathbf{S}$, neglecting the permutation and scaling ambiguities.

In this paper, we make the following assumptions:

A1) $\forall j, t, \ s_{jt} \geq 0$ and $\sum_{t=1}^{K} s_{jt} = 1$.

A2) $m \geq n, \ \forall i, j, \ a_{ij} \geq 0$, and $\mathbf{A}$ is of full column rank.

Assumption A1) is made by taking advantage of the scaling ambiguity in NBSS and Assumption A2) is a common assumption widely used in NBSS [9], [29]. Based on these assumptions, we shall develop the ISM-QP method by exploiting the sparseness feature of the sources. We start with the proposition of a new sparseness measure.

A. Sparseness measure

There are several existing measures which can assess the sparseness of signals. The measure in [30] uses the L0-norm of the signal and the one in [31] uses the normalized ratio of the L1-norm and L2-norm of the signal. These measures can reflect the temporal sparseness of a single signal [32] but rarely refer to the spatial or cross sparseness of multiple signals, which is more important for solving NBSS. On the other hand, the widely used determinant constraint has shown some implicit connections with sparseness. In spectral unmixing for remote sensing image interpretation, the determinant-based method [33] produces very similar results to those yielded by the sparseness based method [29]. So there is possibility that the sparseness of nonnegative signals can be measured by a determinant associated with the signals.

We first recall the well-known Fischer Inequality [34].

Lemma 1 (Fischer Inequality): For the matrices $\mathbf{U}_{11} \in \Re^{I \times I}$, $\mathbf{U}_{12} \in \Re^{I \times J}$ and $\mathbf{U}_{22} \in \Re^{J \times J}$, if the matrix $\mathbf{U} = \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{12}^T & \mathbf{U}_{22} \end{pmatrix}$ is positive definite, it holds that

$$\det(\mathbf{U}) \leq \det(\mathbf{U}_{11}) \det(\mathbf{U}_{22}) \qquad (3)$$

where the equation holds if and only if all of the entries of $\mathbf{U}_{12}$ are zero.

Let $\mathbf{V} \in \Re^{I \times K}$, $K > I > 1$ be a nonnegative matrix whose rows satisfy sum-to-one. We define the determinant-based sparseness measure, or the D-measure as follows:

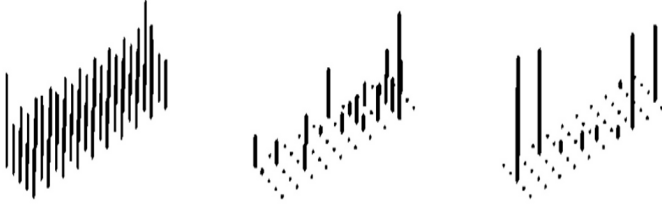$$D(\mathbf{V}) = \det\left(\mathbf{VV}^T\right). \qquad (4)$$

Fig. 1: Illustration of various degrees of sparseness. The D-measure values corresponding to the three matrices (from left to right) are 0.1, 0.5, 0.9, respectively.

Based on Lemma 1, we have the following proposition for the D-measure.

Proposition 1: It holds that $0 \leq D(\mathbf{V}) \leq 1$, where $D(\mathbf{V}) = 0$ if all entries of $\mathbf{V}$ are equal, and $D(\mathbf{V}) = 1$ if and only if the following two conditions satisfy at the same time:

c1) For $\forall i \in \{1, 2, \cdots, m\}$, there is only one nonzero element in $\mathbf{v}_i$.

c2) For $\forall i, j \in \{1, 2, \cdots, m\}$ and $i \neq j$, it holds that $\mathbf{v}_i$ and $\mathbf{v}_j$ are orthogonal.

Proof: See Appendix A. ∎

Proposition 1 shows that the proposed D-measure is well bounded and its value interpolates smoothly between the two extremes 0 and 1. If the sources are non-sparse, then $D(\mathbf{V})$ is close to 0. In contrast, $D(\mathbf{V})$ approaches to 1 if and only if the sources are of sufficient temporal sparseness and spatial sparseness. Fig. 1 illustrates the sparseness degrees of three different matrices gauged by the D-measure. It can be seen that the sparser the matrix is, the larger value the D-measure gives.

B. NBSS model

From Assumptions A1) and A2), the signal matrix $\mathbf{S}$ and the mixing matrix $\mathbf{A}$ in (1) are nonnegative. Hence, the observation matrix $\mathbf{X}$ is constructed only by additive mixing operations. Consequently, $\mathbf{X}$ is less sparse than $\mathbf{S}$, and then the sparseness feature of the sources can be utilized in NBSS. On the other hand, since the sources are nonnegative, the recovered sources should also be nonnegative, or the matrix $\mathbf{Y}$ in (2) should be nonnegative. Furthermore, due to the inherent scaling ambiguity in NBSS, the sources are assumed to be sum-to-one. It is expected that the sum-to-one property is also remained in the recovered signals, i.e., $\mathbf{Y}$ is row-sum-to-one. This can be achieved if the observation matrix $\mathbf{X}$ is normalized to be row-sum-to-one in advance and later the unmixing matrix $\mathbf{W}$ is constrained to be row-sum-to-one. For the sake of simplicity, we assume that $\mathbf{X}$ has been normalized to be row-sum-to-one in the sequel. Therefore, based on the D-measure defined in (4), we propose a new NBSS model as follows:

$$Maximize : \det \left(\mathbf{Y}\mathbf{Y}^T \right) \tag{5}$$
$$\text{s.t.} \begin{cases} y_{it} = \sum_{j=1}^{m} w_{ij} x_{jt} \geq 0, \forall i, t \\ \sum_{j=1}^{m} w_{ij} = 1, \forall i. \end{cases}$$

Substituting (2) into (5), it follows

$$Maximize : \det \left(\mathbf{W}\mathbf{X}\mathbf{X}^T \mathbf{W}^T \right) \tag{6}$$
$$\text{s.t.} \begin{cases} \sum_{j=1}^{m} w_{ij} x_{jt} \geq 0, \forall i, t \\ \sum_{j=1}^{m} w_{ij} = 1, \forall i. \end{cases}$$

We would like to note that maximizing the determinant in (6) with respect to the unmixing matrix $\mathbf{W}$ under the given conditions could be equivalent to minimizing the determinant of the mixing matrix under some other conditions. The latter approach is used in the minimum volume constrained NMF [18]– [21]. However, in the minimum volume constrained NMF, one needs to optimize both a decomposition error and a determinant. Thus one obstacle is how to choose an optimal parameter to balance the two terms. This problem is avoided in our method as the cost function in (6) has only a determinant term. Moreover, this simpler cost function will result in a more efficient BSS algorithm.

It can be seen that the constraints to the cost function (6) are related to the rows of $\mathbf{W}$, denoted by $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_n$. Also, since the determinant in the cost function can be expanded with respect to $\mathbf{w}_i$, $i = 1, 2, \cdots, n$, the optimization for (6) may become easier by splitting it into several sub-optimization problems, each of which is related to a row of $\mathbf{W}$. As will be shown later, these sub-optimization problems can be solved by a typical quadratic programming method. The detailed algorithm derivation will be shown next.

C. Algorithm derivation

Let $\tilde{\mathbf{Y}} = \mathbf{W}\mathbf{X}\mathbf{X}^T \mathbf{W}^T$ and $\tilde{\mathbf{X}} = \mathbf{X}\mathbf{X}^T$. It holds that

$$\tilde{\mathbf{Y}} = \begin{bmatrix} \mathbf{w}_1 \tilde{\mathbf{X}} \mathbf{w}_1^T & \cdots & \mathbf{w}_1 \tilde{\mathbf{X}} \mathbf{w}_i^T & \cdots & \mathbf{w}_1 \tilde{\mathbf{X}} \mathbf{w}_n^T \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{w}_i \tilde{\mathbf{X}} \mathbf{w}_1^T & \cdots & \mathbf{w}_i \tilde{\mathbf{X}} \mathbf{w}_i^T & \cdots & \mathbf{w}_i \tilde{\mathbf{X}} \mathbf{w}_n^T \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{w}_n \tilde{\mathbf{X}} \mathbf{w}_1^T & \cdots & \mathbf{w}_n \tilde{\mathbf{X}} \mathbf{w}_i^T & \cdots & \mathbf{w}_n \tilde{\mathbf{X}} \mathbf{w}_n^T \end{bmatrix}. \tag{7}$$

From (7), one can see that $\mathbf{w}_i$ only appears in the ith row and the ith column of $\tilde{\mathbf{Y}}$. Using the cofactor expansion with respect to the ith row of $\tilde{\mathbf{Y}}$, it gives

$$\det \left(\tilde{\mathbf{Y}} \right) = \sum_{j=1}^{n} (-1)^{i+j} \det \left(\tilde{\mathbf{Y}}_{ij} \right) \mathbf{w}_i \tilde{\mathbf{X}} \mathbf{w}_j^T \tag{8}$$

where $\tilde{\mathbf{Y}}_{ij}$ denotes a $(n-1) \times (n-1)$ sub-matrix of $\tilde{\mathbf{Y}}$ with the ith row and the jth column removed.

Note that $\tilde{\mathbf{Y}}_{ij}$ does not contain $\mathbf{w}_i$ if $j = i$. However, it still contains $\mathbf{w}_i$ or $\mathbf{w}_i^T$ if $j \neq i$. In order to extract $\mathbf{w}_i$ completely, the scenario of $j \neq i$ is further analyzed under the following two cases.

1) If $j < i$, $\tilde{\mathbf{Y}}_{ij}$ contains $\mathbf{w}_i$ in its $(i-1)$th column, which corresponds to the ith column of $\tilde{\mathbf{Y}}$, i.e., the column $\left[\mathbf{w}_1 \tilde{\mathbf{X}} \mathbf{w}_i^T, \cdots, \mathbf{w}_{i-1} \tilde{\mathbf{X}} \mathbf{w}_i^T, \mathbf{w}_{i+1} \tilde{\mathbf{X}} \mathbf{w}_i^T, \cdots, \mathbf{w}_n \tilde{\mathbf{X}} \mathbf{w}_i^T \right]^T$.

Expanding $\det\left(\tilde{\mathbf{Y}}_{ij}\right)$ with respect to this column, we obtain

$$
\begin{aligned}
\det\left(\tilde{\mathbf{Y}}_{ij}\right) = &\sum_{t=1}^{i-1}(-1)^{t+i-1}\det\left(\tilde{\mathbf{Y}}_{ij,t(i-1)}\right)\mathbf{w}_t\tilde{\mathbf{X}}\mathbf{w}_i^T \\
&+\sum_{t=i}^{n-1}(-1)^{t+i-1}\det\left(\tilde{\mathbf{Y}}_{ij,t(i-1)}\right)\mathbf{w}_{t+1}\tilde{\mathbf{X}}\mathbf{w}_i^T
\end{aligned} \tag{9}
$$

where $\tilde{\mathbf{Y}}_{ij,t(i-1)}$ denotes a $(n-2)\times(n-2)$ sub-matrix of $\tilde{\mathbf{Y}}_{ij}$ with the tth row and the $(i-1)$th column removed. $\tilde{\mathbf{Y}}_{ij,t(i-1)}$ does not contain $\mathbf{w}_i$.

2) If $j > i$, $\tilde{\mathbf{Y}}_{ij}$ contains $\mathbf{w}_i$ in its ith column, which is also $\left[\mathbf{w}_1\tilde{\mathbf{X}}\mathbf{w}_i^T,\cdots,\mathbf{w}_{i-1}\tilde{\mathbf{X}}\mathbf{w}_i^T,\mathbf{w}_{i+1}\tilde{\mathbf{X}}\mathbf{w}_i^T,\cdots,\mathbf{w}_n\tilde{\mathbf{X}}\mathbf{w}_i^T\right]^T$. Then, by expanding $\det\left(\tilde{\mathbf{Y}}_{ij}\right)$ with respect to this column, it yields

$$
\begin{aligned}
\det\left(\tilde{\mathbf{Y}}_{ij}\right) = &\sum_{t=1}^{i-1}(-1)^{t+i}\det\left(\tilde{\mathbf{Y}}_{ij,ti}\right)\mathbf{w}_t\tilde{\mathbf{X}}\mathbf{w}_i^T \\
&+\sum_{t=i}^{n-1}(-1)^{t+i}\det\left(\tilde{\mathbf{Y}}_{ij,ti}\right)\mathbf{w}_{t+1}\tilde{\mathbf{X}}\mathbf{w}_i^T
\end{aligned} \tag{10}
$$

where $\tilde{\mathbf{Y}}_{ij,ti}$ denotes a $(n-2)\times(n-2)$ sub-matrix of $\tilde{\mathbf{Y}}_{ij}$ with the tth row and the ith column removed, and it does not contain $\mathbf{w}_i$.

From (9) and (10), we can rewrite (8) as

$$
\begin{aligned}
&\det\left(\tilde{\mathbf{Y}}\right) \\
&=\sum_{j=1}^{i-1}(-1)^{i+j}\mathbf{w}_i\tilde{\mathbf{X}}\mathbf{w}_j^T \\
&\quad\left[\sum_{t=1}^{i-1}(-1)^{t+i-1}\det\left(\tilde{\mathbf{Y}}_{ij,t(i-1)}\right)\mathbf{w}_t\tilde{\mathbf{X}}\mathbf{w}_i^T\right.\\
&\qquad\left.+\sum_{t=i}^{n-1}(-1)^{t+i-1}\det\left(\tilde{\mathbf{Y}}_{ij,t(i-1)}\right)\mathbf{w}_{t+1}\tilde{\mathbf{X}}\mathbf{w}_i^T\right] \\
&\quad+(-1)^{i+i}\det\left(\tilde{\mathbf{Y}}_{ii}\right)\mathbf{w}_i\tilde{\mathbf{X}}\mathbf{w}_i^T \\
&\quad+\sum_{j=i+1}^{n}(-1)^{i+j}\mathbf{w}_i\tilde{\mathbf{X}}\mathbf{w}_j^T\left[\sum_{t=1}^{i-1}(-1)^{t+i}\det\left(\tilde{\mathbf{Y}}_{ij,ti}\right)\mathbf{w}_t\tilde{\mathbf{X}}\mathbf{w}_i^T\right.\\
&\qquad\left.+\sum_{t=i}^{n-1}(-1)^{t+i}\det\left(\tilde{\mathbf{Y}}_{ij,ti}\right)\mathbf{w}_{t+1}\tilde{\mathbf{X}}\mathbf{w}_i^T\right] \\
&=\mathbf{w}_i\tilde{\mathbf{X}}\sum_{j=1}^{i-1}(-1)^{i+j}\mathbf{w}_j^T\left[\sum_{t=1}^{i-1}(-1)^{t+i-1}\det\left(\tilde{\mathbf{Y}}_{ij,t(i-1)}\right)\mathbf{w}_t\right.\\
&\qquad\left.+\sum_{t=i}^{n-1}(-1)^{t+i-1}\det\left(\tilde{\mathbf{Y}}_{ij,t(i-1)}\right)\mathbf{w}_{t+1}\right]\tilde{\mathbf{X}}\mathbf{w}_i^T \\
&\quad+\mathbf{w}_i(-1)^{i+i}\det\left(\tilde{\mathbf{Y}}_{ii}\right)\tilde{\mathbf{X}}\mathbf{w}_i^T \\
&\quad+\mathbf{w}_i\tilde{\mathbf{X}}\sum_{j=i+1}^{n}(-1)^{i+j}\mathbf{w}_j^T\left[\sum_{t=1}^{i-1}(-1)^{t+i}\det\left(\tilde{\mathbf{Y}}_{ij,ti}\right)\mathbf{w}_t\right.\\
&\qquad\left.+\sum_{t=i}^{n-1}(-1)^{t+i}\det\left(\tilde{\mathbf{Y}}_{ij,ti}\right)\mathbf{w}_{t+1}\right]\tilde{\mathbf{X}}\mathbf{w}_i^T \\
&=\mathbf{w}_i\mathbf{C}\mathbf{w}_i^T
\end{aligned} \tag{11}
$$

where

$$
\mathbf{C} = \mathbf{C}_1 + \mathbf{C}_2 + \mathbf{C}_3 \tag{12}
$$

with

$$
\begin{cases}
\mathbf{C}_1 = \tilde{\mathbf{X}}\displaystyle\sum_{j=1}^{i-1}(-1)^{i+j}\mathbf{w}_j^T\left[\sum_{t=1}^{i-1}(-1)^{t+i-1}\det\left(\tilde{\mathbf{Y}}_{ij,t(i-1)}\right)\mathbf{w}_t\right.\\
\qquad\left.+\displaystyle\sum_{t=i}^{n-1}(-1)^{t+i-1}\det\left(\tilde{\mathbf{Y}}_{ij,t(i-1)}\right)\mathbf{w}_{t+1}\right]\tilde{\mathbf{X}} \\
\mathbf{C}_2 = (-1)^{i+i}\det\left(\tilde{\mathbf{Y}}_{ii}\right)\tilde{\mathbf{X}} \\
\mathbf{C}_3 = \tilde{\mathbf{X}}\displaystyle\sum_{j=i+1}^{n}(-1)^{i+j}\mathbf{w}_j^T\left[\sum_{t=1}^{i-1}(-1)^{t+i}\det\left(\tilde{\mathbf{Y}}_{ij,ti}\right)\mathbf{w}_t\right.\\
\qquad\left.+\displaystyle\sum_{t=i}^{n-1}(-1)^{t+i}\det\left(\tilde{\mathbf{Y}}_{ij,ti}\right)\mathbf{w}_{t+1}\right]\tilde{\mathbf{X}}.
\end{cases}
$$

If $\forall j \neq i$, $\mathbf{w}_j$ is known, which results in that $\forall t$, $\tilde{\mathbf{Y}}_{ij,t(i-1)}$ and $\tilde{\mathbf{Y}}_{ij,ti}$ are known. On the other hand, $\tilde{\mathbf{Y}}_{ii}$ does not contain $\mathbf{w}_i$ and $\tilde{\mathbf{X}}$ can be calculated from the observation $\mathbf{X}$. Thus, $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{C}_3$ are independent of $\mathbf{w}_i$. Therefore, $\det\left(\tilde{\mathbf{Y}}\right)$, i.e., the cost function in (5) or (6), is a quadratic function with respect to $\mathbf{w}_i$. To utilize this property, we optimize (5) or (6) by using an alternative iteration updating scheme, i.e., optimizing one row of $\mathbf{W}$ while fixing the rest rows, which has been verified to be an efficient scheme [9], [35]. With regard to optimizing the ith row $\mathbf{w}_i$ of $\mathbf{W}$, the model (6) is simplified as follows:

$$
Maximize : \mathbf{w}_i\mathbf{C}\mathbf{w}_i^T \tag{13}
$$

$$
\text{s.t.}\begin{cases}
\displaystyle\sum_{j=1}^{m} w_{ij}x_{jt} \geq 0, \ \forall t \in \{1,2,\cdots,K\} \\
\displaystyle\sum_{j=1}^{m} w_{ij} = 1
\end{cases}
$$

where $\mathbf{C}$ is a matrix given in (12), which is independent of $\mathbf{w}_i$, and $i = 1,2,\cdots,n$.

Based on the above analysis, the initial optimization in (6) is broken into a series of QP problems which can be solved easily by using MATLAB function *quadprog()*. However, if the sample number K is very large, the optimization may be time-consuming. To improve computational efficiency, one can find in prior the extreme points of the convex hull spanned by the observations and then replace the initial inequality constraints by the constraints only with respect to these points [9]. Consequently, (13) can be simplified to

$$
Maximize : \mathbf{w}_i\mathbf{C}\mathbf{w}_i^T \tag{14}
$$

$$
\text{s.t.}\begin{cases}
\displaystyle\sum_{j=1}^{m} w_{ij}v_{jl} \geq 0, \ \forall l \in \{1,2,\cdots,L\} \\
\displaystyle\sum_{j=1}^{m} w_{ij} = 1
\end{cases}
$$

where L denotes the number of the extreme points of the convex hull spanned by the observations and $i = 1,2,\cdots,n$. Since L is often much smaller than K, the problem dimension can be reduced significantly. Also, the existing quickhull algorithm in [36] can be used to find these extreme points.

Finally, we formulate the proposed ISM-QP algorithm as follows:

1) *Preprocessing:* Normalize the sum of each row of $\mathbf{X}$ to one. Use the quickhull algorithm in [36] to find the extreme points $\mathbf{v}_1,\mathbf{v}_2,\cdots,\mathbf{v}_L$ of the convex hull spanned by $\mathbf{X}$.

2) *Initialization:* Set a random initial matrix for $\mathbf{W}$ and then normalize the sum of each of its rows to one. Let $i = 1$.

3) *Updating:* compute the matrix $\mathbf{C}$ by (12), obtain the optimal solution $\mathbf{w}_i^*$ of (14) by using MATLAB function *quadprog()*, update $\mathbf{w}_i$ by $\mathbf{w}_i^*$, and set $i = i+1$. If $i > n$, reset $i = mod(i, n)$.

4) *Stop:* If a given stop criterion is satisfied, the algorithm stops; otherwise, go to Step 3).

Remark 1: For the proposed algorithm, the stop criterion in Step 4) is not necessarily fixed. One can use the maximum iteration number [37] or the convergence tolerance [9]. In this paper, the first criterion is utilized.

Remark 2: With respect to solving the involved QP problem, we use the MATLAB function *quadprog()* with the variable $\mathbf{x} = \mathbf{w}_i^T$ and the coefficients f, A, b, H, Aeq, Beq given below:

$$f = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{m \times 1}, \ A = -\begin{bmatrix} v_{11} & \cdots & v_{1L} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mL} \end{bmatrix}^T,$$

$$b = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{L \times 1}, \ H = -\mathbf{C}, \ Aeq = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}^T_{m \times 1}, \ Beq = 1.$$

D. Computational complexity and source identifiability

As far as the computational complexity of the algorithm is concerned, by using a reflective Newton method in [38], each QP problem in (14) can be solved with an approximate computational complexity of $O(mL)$ in the best case and $O(m^2L)$ in the worst case. Practically, by using the *subroutine improve* method, the complexity of the QP problem is approximately $O(kmL)$, where $1 \leq k < m$ is often small [38]. Therefore, the computational complexity of the proposed ISM-QP is approximately $O(kmnL)$ in each iteration. For the sake of comparison, the computational complexities of some benchmark BSS algorithms are also shown here, including NICA [6], FCA-NMF [17], nLCA-IVM [9], and VCA-NLS [22]. In each iteration, their approximate complexities are (for the case of $m = n$) $O(m^2K)$, $O(m^2K)$, $O(m^2L)$, $O(m^2K)$, respectively. Note that in the over-determined scenario, additional computation cost is needed by nLCA-IVM to reduce the dimension of the observations from m to n.

Furthermore, taking the optimization result into account, one can see that (14) is a typical quadratic programming problem and the result can be obtained by invoking the existing software package. Since this result corresponds to only one row of the unmixing matrix, the original optimization problem in (6) with respect to the whole unmixing matrix is considered directly for the source identifiability analysis. Like most existing methods, it needs some conditions so that the optimal solution of (6) corresponds to the actual unmixing matrix, which implies the perfect recovery of the sources. One basic condition is that the permutation indeterminacy of the sources can be ignored. Based on this common condition in BSS problem, we have the following theorem.

TABLE I: ρ_{ave} AND T_{ave} OF DIFFERENT ALGORITHMS, WHERE SOURCES ARE FINGERPRINT IMAGES

	ISM-QP	NICA	FCA-NMF	nLCA-IVM	VCA-NLS
ρ_{ave}	1.0000	0.7875	0.9180	1.0000	1.0000
T_{ave}	7.6734	311.7219	331.5047	8.3484	248.4125

Theorem 1 (Source Identifiability): If there exists a $n \times n$ submatrix $\hat{\mathbf{S}}$ satisfying $D(\hat{\mathbf{S}}) = 1$, where $\hat{\mathbf{S}}$ is normalized to be row-sum-to-one, then it holds that

$$\mathbf{W}^* \mathbf{A} = \mathbf{P} \tag{15}$$

where $\mathbf{W}^*$ is the optimal solution of (6), $\mathbf{P}$ is a permutation matrix.

Proof: See Appendix B. ∎

III. EXPERIMENTAL RESULTS

In this section, both computer generated data and real biomedical data are used to test the proposed ISM-QP algorithm, and the results are compared with some existing benchmark methods, including NICA, FCA-NMF, nLCA-IVM, and VCA-NLS. For NICA and nLCA-IVM, the so-called nonnegative principle component analysis is used to preprocess the observations in the over-determined case.

Let $\check{\mathbf{S}}$ denote the recovered source matrix. The cross-correlation coefficient (C-Coef) ρ between $\mathbf{S}$ and $\check{\mathbf{S}}$, defined as [9]

$$\rho = \frac{1}{n} \max_{\pi_i \in \Pi_n, c_i \in \{1, -1\}} \sum_{i=1}^n \frac{(\mathbf{s}_i - \mathbf{q}(\mathbf{s}_i))(c_i \check{\mathbf{s}}_{\pi_i} - c_i \mathbf{q}(\check{\mathbf{s}}_{\pi_i}))^T}{\|\mathbf{s}_i - \mathbf{q}(\mathbf{s}_i)\| \cdot \|c_i \check{\mathbf{s}}_{\pi_i} - c_i \mathbf{q}(\check{\mathbf{s}}_{\pi_i})\|}$$

is utilized to evaluate the source separation performance of the tested algorithms, where $\mathbf{s}_i$ denotes the ith source, $\mathbf{q}(\mathbf{s}_i)$ is a K-dimension vector composed of the mean of $\mathbf{s}_i$, $\Pi_n = \{\pi = (\pi_1, \pi_2, \cdots, \pi_n) | \pi_i \in \{1, 2, \cdots, n\}, \pi_i \neq \pi_j, \forall i \neq j\}$ is the set of all the permutations of $\{1, 2, \cdots, n\}$, and c_i is the sign (or the polarity) ambiguity between the recovered source $\check{\mathbf{s}}_{\pi_i}$ and the true source $\mathbf{s}_i$. Clearly, $\rho \in [0, 1]$ and the larger the value, the better the source separation performance.

A. Human fingerprint image separation

Fingerprints are the special features of individuals, often used as evidence in judicial cases. However, they may be touched by other people. As a result, only some mixed fingerprints can be collected directly. It is appealing to recover the source fingerprints by a BSS method. In the simulation, we consider the case of $m = 5$ and $n = 4$. The algorithms are tested using four fingerprint images (640×480, see Fig. 2(a)) from DB1 dataset of fingerprint verification competition 2004 (see *http://biometrics.cse.msu.edu/fvc04db/index.html*). For each algorithm, we carry out 50 independent runs to compute the average C-Coef (ρ_{ave}) index and the average CPU-time (T_{ave}). In each run, a mixing matrix with uniform distribution on [0, 1] is randomly generated to mix the fingerprint images.

Table I shows ρ_{ave} and T_{ave} resulted from ISM-QP, NICA, FCA-NMF, nLCA-IVM, and VCA-NLS. As far as the ρ_{ave}
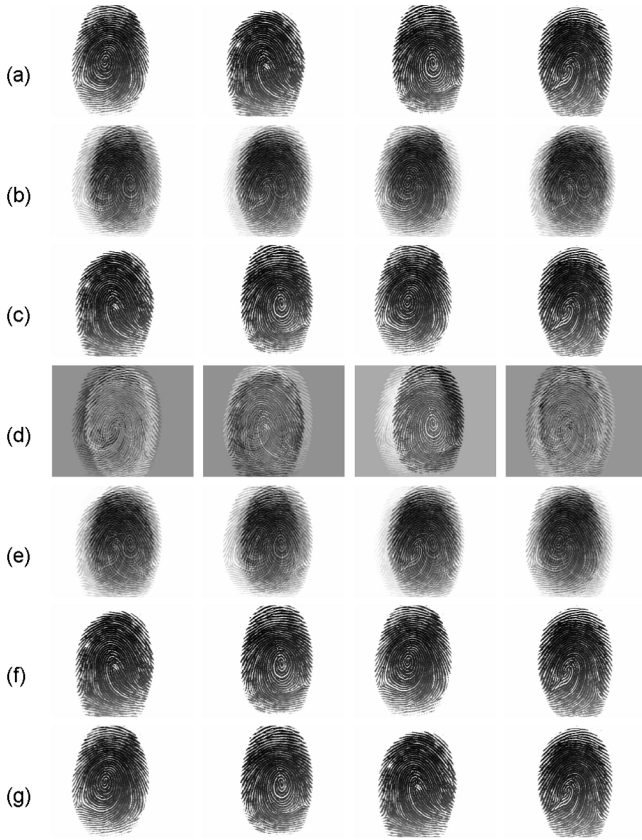
Fig. 2: Performance comparison in separating human fingerprint images. (a) Source images; (b) Mixed images; (c) Recovered images by ISM-QP; (d) Recovered images by NICA; (e) Recovered images by FCA-NMF; (f) Recovered images by nLCA-IVM; (g) Recovered images by VCA-NLS.

index is concerned, nLCA-IVM and VCA-NLS perform the same as the proposed ISM-QP. The reasons are that there exist *pure-source samples* in the source images and the identifiability condition in Theorem 1 is satisfied. The ρ_{ave} index of NICA is small, because the sources are highly correlated. Regarding computational costs, the T_{ave} index of ISM-QP is the smallest and is much smaller than those of NICA, FCA-NMF, and VCA-NLS. Fig. 2(b) shows the mixtures in one random experiment for reference (for the convenience of presentation, only four mixtures are displayed here), and Fig. 2(c)-(g) show the corresponding recovered sources by the compared algorithms. From Fig. 2 (d), one can see that the results of NICA are seriously affected by the dependence among the sources.

B. Recovery of natural images under noise

In this simulation, we consider a mixing system with $m = 5$ and $n = 3$ in the presence of additive Gaussian noise. Three 128×128 natural images are used as sources, which can be downloaded from the website *http://www.bsp.brain.riken.jp/ICALAB*. Firstly, the results of 50 random runs are presented, where the signal to noise ratio

TABLE II: ρ_{ave} AND T_{ave} OF DIFFERENT ALGORITHMS, WHERE SOURCES ARE NATURAL IMAGES (SNR=25dB)

	ISM-QP	NICA	FCA-NMF	nLCA-IVM	VCA-NLS
ρ_{ave}	0.9810	0.9585	0.7644	0.9661	0.9521
T_{ave}	1.0641	12.3672	15.0563	1.1969	7.2203

$(\text{SNR})^2$ is kept at 25dB. In each run, a mixing matrix with uniform distribution on [0, 1] is randomly generated. It can be seen from Table II that the ρ_{ave} value of ISM-QP is the best. In contrast, the ρ_{ave} value of FCA-NMF is quite low as FCA-NMF does not necessarily generate the desired solution. Moreover, ISM-QP is also the most efficient as its T_{ave} value is the smallest.

Secondly, in real-world scenarios, some pixels of images may be corrupted locally by noise. So the performance of the algorithms is assessed versus different percentages of corrupted pixels. For each percentage, we carry out 50 independent runs using different randomly selected corrupted pixels. In all 50 runs, the same mixing matrix is used, which is randomly generated with uniform distribution on [0, 1]. Fig. 3 shows the ρ_{ave} values corresponding to 20%, 40%, 60% and 80% corrupted pixels, where SNR=25dB. Since the ρ_{ave} values of FCA-NMF are quite small, we only show the results obtained by ISM-QP, NICA, nLCA-IVM and VCA-NLS for better visual comparisons. One can see from Fig. 3 that the proposed ISM-QP algorithm performs the best. The reason of ISM-QP outperforming nLCA-IVM is that the latter employs PCA to preprocess data, resulting in loss of some useful information. For VCA-NLS, the results change significantly versus the percentages of corrupted pixels because the vertex search algorithm is sensitive to the number of corrupted pixels.

Furthermore, we compare the algorithms against different SNRs when the percentage of the corrupted pixels is fixed at 25%. For each SNR level, 50 independent runs with different mixing matrix which are distributed uniformly on [0, 1] are performed. Fig. 4 shows the ρ_{ave} values of ISM-QP, NICA, nLCA-IVM and VCA-NLS but those of FCA-NMF are not shown here due to the same reason mentioned above. We can see that if noise is weak, ISM-QP and nLCA-IVM performs similarly but they outperform the other compared algorithms. However, when noise becomes stronger, the performance of nLCA-IVM deteriorates more significantly. The reason is that more useful information is lost in this case due to the usage of PCA in preprocessing data. In addition, NICA seems to perform well in low SNR situations. This is because NICA mainly exploits the non-Gaussianity of the sources [6], [10] and thus it is robust to Gaussian noise. However, it performs much worse than ISM-QP at high SNRs.

C. Real-world fluorescence microscopy image analysis

In this experiment, the proposed ISM-QP algorithm is used to analyze the real-world fluorescence microscopy images, which are often collected by using an optical sensor array [39].

---

[2]SNR is defined as $10\log_{10} \frac{\|\mathbf{x}\|}{\|\mathbf{x}-\mathbf{y}\|}$, where $\mathbf{x}$ and $\mathbf{y}$ denote the original signal and the signal polluted by noise, respectively.
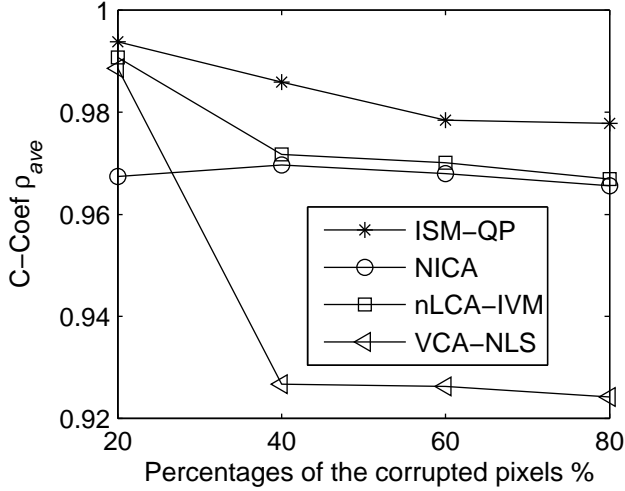
Fig. 3: ρ_{ave} of different algorithms versus the percentages of corrupted pixels (SNR=25dB).
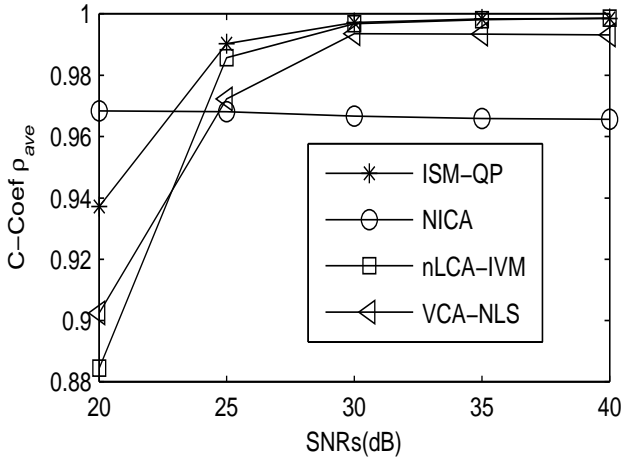

Fig. 4: ρ_{ave} of different algorithms versus SNRs, where the percentage of corrupted pixels is 25%.
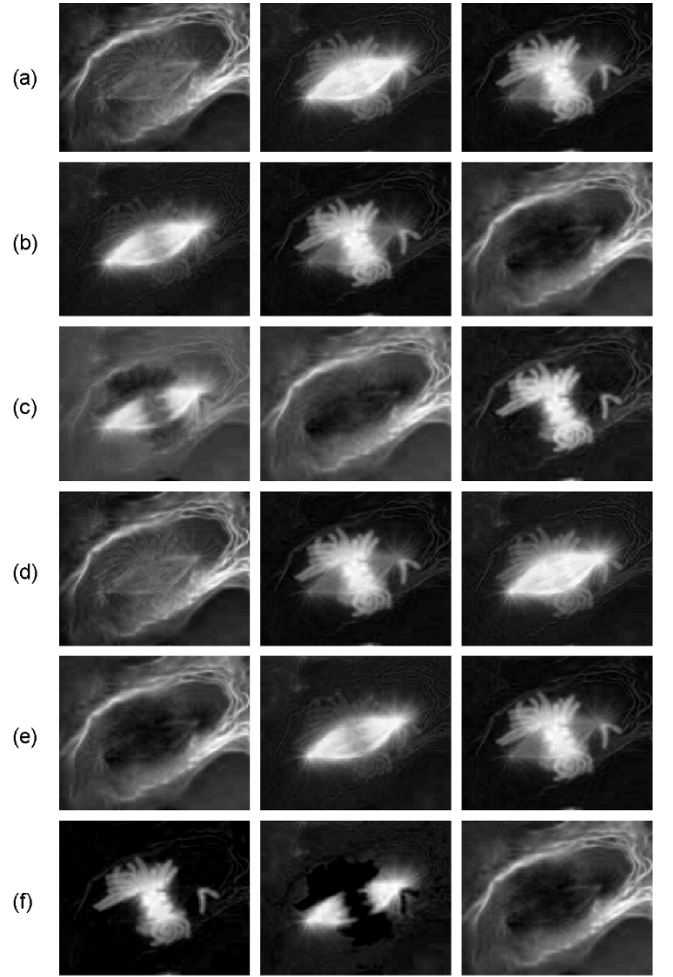

Fig. 5: Analysis of real-world newt lung cell images using NBSS technique. (a) Collected images; (b) Unmixed images by ISM-QP; (c) Unmixed images by NICA; (d) Unmixed images by FCA-NMF; (e) Unmixed images by nLCA-IVM; (f) Unmixed images by VCA-NLS.

Due to the limitation of sensor resolution, the collected images often suffer from the spectral-overlap problem, which leads to information leak-through from one spectral channel to another. NBSS has the potential to find individual maps associated with specific biomarkers from such images. Specifically, three newt lung cell images (200×191) obtained from *http://publications.nigms.nih.gov/insidethecell/chapter1.html* are analyzed in this experiment, which are shown in Fig. 5(a).

Fig. 5(b) shows the unmixed images by using ISM-QP. By visually comparing the images in Fig. 5(b) with those in Fig. 5(a), it can be seen that the shape of the unmixed rope chromosomes by ISM-QP is much clearer than that in the original collected images, and similar result can also be found from the separation of the spindle fibers. Since the source images are unknown, it is difficult to use a numerical index to measure the separation precision of ISM-QP. Instead, we give the unmixed images by the NICA, FCA-NMF, nLCA-IVM and VCA-NLS algorithms for reference, which are shown in Fig. 5(c)-(f), respectively. We can see that the result from our

algorithm is comparable to those results obtained by the other algorithms. In addition, the corresponding CPU-times of ISM-QP, NICA, FCA-NMF, nLCA-IVM and VCA-NLS are 0.4188, 33.4500, 22.8234, 1.1094, 17.9219, respectively. This confirms again that ISM-QP is more efficient than the other competing algorithms.

IV. CONCLUSION

In this paper, a determinant-based sparseness measure called *D*-measure is proposed and the sparseness of the sources is exploited for NBSS. In contrast to the traditional single-signal-based measure, the proposed signal-matrix-based *D*-measure can reflect both the temporal and spatial sparseness of signals. This joint sparseness is beneficial for solving the BSS problem, which has been analyzed in [26] and further verified in this paper. Based on the *D*-measure, a new NBSS model is derived and the corresponding source identifiability is also analyzed. A remarkable advantage of the new NBSS model is that the optimization can be cast into a series of

QP problems, which can be easily solved. Finally, an efficient ISM-QP algorithm is developed to perform NBSS. When the identifiability condition is satisfied, the estimation precisions of ISM-QP are extremely high (see the results in Section III.A). Even if the identifiability condition is violated, the results are still robust (see Section III.B). The experimental results also show the advantages of the proposed ISM-QP algorithm over the benchmark algorithms NICA, FCA-NMF, nLCA-IVM and VCA-NLS.

APPENDIX A
PROOF OF PROPOSITION 1

Proof. As we all know, $\mathbf{V}\mathbf{V}^T$ is a positive semi-definite matrix. This means $\det\left(\mathbf{V}\mathbf{V}^T\right) \geq 0$, or equivalently $D(\mathbf{V}) \geq 0$ according to (4). In the case of $D(\mathbf{V}) > 0$, i.e., $\det\left(\mathbf{V}\mathbf{V}^T\right) \geq 0$, $\mathbf{V}\mathbf{V}^T$ is positive definite, then it follows from Lemma 1 that

$$
\begin{aligned}
&\det\left(\mathbf{V}\mathbf{V}^T\right) \\
&\leq \det\left(\mathbf{v}_1\mathbf{v}_1^T\right) \det\left(\begin{bmatrix} \mathbf{v}_2\mathbf{v}_2^T & \cdots & \mathbf{v}_2\mathbf{v}_I^T \\ \vdots & \ddots & \vdots \\ \mathbf{v}_I\mathbf{v}_2^T & \cdots & \mathbf{v}_I\mathbf{v}_I^T \end{bmatrix}\right) \\
&\leq \det\left(\mathbf{v}_1\mathbf{v}_1^T\right) \det\left(\mathbf{v}_2\mathbf{v}_2^T\right) \cdots \det\left(\mathbf{v}_I\mathbf{v}_I^T\right) \\
&= \|\mathbf{v}_1\|_2^2 \|\mathbf{v}_2\|_2^2 \cdots \|\mathbf{v}_I\|_2^2 \\
&\leq \|\mathbf{v}_1\|_1^2 \|\mathbf{v}_2\|_1^2 \cdots \|\mathbf{v}_I\|_1^2 \\
&= 1.
\end{aligned}
\tag{16}
$$

Thus we obtain $0 \leq D(\mathbf{V}) \leq 1$.

If all entries of $\mathbf{V}$ are equal, $\mathbf{V}$ is not of full row rank. As a result, $\mathbf{V}\mathbf{V}^T$ is not of full rank, leading to $D(\mathbf{V}) = 0$. As for the equation $D(\mathbf{V}) = 1$, the proof is decomposed into two parts.

i) Sufficiency: From the condition c1), one can obtain $\mathbf{v}_i\mathbf{v}_i^T = 1, \forall i$. Also, from the condition c2), it holds that $\mathbf{v}_i\mathbf{v}_j^T = 0, \forall i \neq j$. Since $\mathbf{V}\mathbf{V}^T$ becomes the identity matrix in this case, we have $D(\mathbf{V}) = \det\left(\mathbf{V}\mathbf{V}^T\right) = 1$.

ii) Necessity: If $D(\mathbf{V}) = 1$, then $\mathbf{V}\mathbf{V}^T$ is positive definite and all inequalities in (16) should degenerate into equations. The degeneration of the third inequality implies that there is only one nonzero element in $\mathbf{v}_i, \forall i$, which is the condition c1). At the same time, based on Lemma 1, the degenerations of the first two inequalities mean that $\forall i \neq j, \mathbf{v}_i\mathbf{v}_j^T = 0$, i.e., $\mathbf{v}_i$ and $\mathbf{v}_j$ are orthogonal, which is the condition c2). This completes the proof.

APPENDIX B
PROOF OF THEOREM 1

Proof. It is shown in Assumption A2) that $m \geq n$. So we prove theorem by considering the cases of $m = n$ and $m > n$ separately.

1) In the case of $m = n$, based on the features of determinant, we have

$$
\begin{aligned}
&\det\left(\mathbf{W}\mathbf{X}\mathbf{X}^T\mathbf{W}^T\right) \\
&= \det(\mathbf{W}) \det\left(\mathbf{X}\mathbf{X}^T\right) \det\left(\mathbf{W}^T\right) \\
&= \det^2(\mathbf{W}) \det\left(\mathbf{X}\mathbf{X}^T\right).
\end{aligned}
\tag{17}
$$

Because $\mathbf{X}$ is known and $\det\left(\mathbf{X}\mathbf{X}^T\right) > 0$, maximizing $\det\left(\mathbf{W}\mathbf{X}\mathbf{X}^T\mathbf{W}^T\right)$ is equivalent to maximizing $\det^2(\mathbf{W})$ or $|\det(\mathbf{W})|$. Under this circumstance, (6) degenerates into Wang's optimization (see (24) in [9]). Then, this theorem can be proved by directly using the results of Theorem 2 in [9].

2) In the case of $m > n$, since $x_{jt} = \sum_{k=1}^{n} a_{jk}s_{kt}, \forall j, t$ and $\sum_{j=1}^{m} w_{ij}x_{jt} \geq 0, \forall i, t$, it holds that

$$
\sum_{j=1}^{m} w_{ij} \sum_{k=1}^{n} a_{jk}s_{kt} = \sum_{j=1}^{m}\sum_{k=1}^{n} s_{kt}w_{ij}a_{jk} \geq 0, \forall i, t.
\tag{18}
$$

So, for the sub-matrix $\hat{\mathbf{S}}$, it also holds that

$$
\sum_{j=1}^{m}\sum_{k=1}^{n} \hat{s}_{kt}w_{ij}a_{jk} \geq 0, \forall i, t.
\tag{19}
$$

Given that $\hat{\mathbf{S}}$ satisfies $D(\hat{\mathbf{S}}) = 1$, then $\hat{\mathbf{S}}$ can be represented by $\hat{\mathbf{S}} = \hat{\mathbf{L}}\hat{\mathbf{P}}$, where $\hat{\mathbf{L}}$ and $\hat{\mathbf{P}}$ denote a diagonal matrix and a permutation matrix, respectively. Therefore, (19) can be simplified to

$$
\sum_{j=1}^{m} w_{ij}a_{jk} \geq 0, \forall i, k.
\tag{20}
$$

Considering that $\sum_{j=1}^{m} w_{ij} = 1, \forall i$ and $\sum_{k=1}^{n} a_{jk} = 1, \forall j$, it follows

$$
\sum_{k=1}^{n}\sum_{j=1}^{m} w_{ij}a_{jk} = 1, \forall i.
\tag{21}
$$

Moreover, since $\hat{\mathbf{S}} = \hat{\mathbf{L}}\hat{\mathbf{P}}$, the sub-matrix $\hat{\mathbf{X}}$ corresponding to $\hat{\mathbf{S}}$ of $\mathbf{X}$ can be accordingly represented as

$$
\hat{\mathbf{X}} = \mathbf{A}\hat{\mathbf{S}} = \mathbf{A}\hat{\mathbf{L}}\hat{\mathbf{P}}.
\tag{22}
$$

On the other hand, based on the principle of diagonal reduction in [40], for the observation matrix $\mathbf{X}$, there exist nonsingular square matrices $\mathbf{B}$ and $\mathbf{C}$ such that

$$
\mathbf{B}\mathbf{X}\mathbf{C} = \begin{bmatrix} \mathbf{I}_{n\times n} & \mathbf{0}_{n\times(T-n)} \\ \mathbf{0}_{(m-n)\times n} & \mathbf{0}_{(m-n)\times(T-n)} \end{bmatrix}.
\tag{23}
$$

Hence, $\mathbf{X}$ can be decomposed as

$$
\mathbf{X} = \mathbf{G}\mathbf{H}
\tag{24}
$$

where $\mathbf{G}$ is a $m \times n$ matrix consisting of the first n column vectors of $\mathbf{B}^{-1}$, and $\mathbf{H}$ is a $n \times K$ matrix consisting of the first n row vectors of $\mathbf{C}^{-1}$. Then, $\hat{\mathbf{X}}$ can also be expressed as

$$
\hat{\mathbf{X}} = \mathbf{G}\hat{\mathbf{H}}
\tag{25}
$$

where $\hat{\mathbf{H}}$ is the sub-matrix (corresponding to $\hat{\mathbf{X}}$) of $\mathbf{H}$. Note that $\hat{\mathbf{X}}$ is of full column rank. So, $\hat{\mathbf{H}}$ is nonsingular. Based on (22) and (25), it follows

$$\mathbf{G} = \mathbf{A}\hat{\mathbf{L}}\hat{\mathbf{P}}\hat{\mathbf{H}}^{-1}. \tag{26}$$

Substituting (24) and (26) into the cost function in (6), we obtain

$$\begin{aligned}
&\det\left(\mathbf{W}\mathbf{X}\mathbf{X}^T\mathbf{W}^T\right)\\
&= \det\left(\mathbf{W}\mathbf{G}\mathbf{H}\mathbf{H}^T\mathbf{G}^T\mathbf{W}^T\right)\\
&= \det\left(\mathbf{W}\mathbf{A}\hat{\mathbf{L}}\hat{\mathbf{P}}\hat{\mathbf{H}}^{-1}\mathbf{H}\mathbf{H}^T\hat{\mathbf{H}}^{-T}\hat{\mathbf{P}}^T\hat{\mathbf{L}}^T\mathbf{A}^T\mathbf{W}^T\right)\\
&= \det(\mathbf{W}\mathbf{A})\det\left(\hat{\mathbf{L}}\hat{\mathbf{P}}\hat{\mathbf{H}}^{-1}\mathbf{H}\mathbf{H}^T\hat{\mathbf{H}}^{-T}\hat{\mathbf{P}}^T\hat{\mathbf{L}}^T\right)\det\left(\mathbf{A}^T\mathbf{W}^T\right)\\
&= \det{}^2(\mathbf{W}\mathbf{A})\det\left(\hat{\mathbf{L}}\hat{\mathbf{P}}\hat{\mathbf{H}}^{-1}\mathbf{H}\mathbf{H}^T\hat{\mathbf{H}}^{-T}\hat{\mathbf{P}}^T\hat{\mathbf{L}}^T\right).
\end{aligned} \tag{27}$$

Since $\det\left(\hat{\mathbf{L}}\hat{\mathbf{P}}\hat{\mathbf{H}}^{-1}\mathbf{H}\mathbf{H}^T\hat{\mathbf{H}}^{-T}\hat{\mathbf{P}}^T\hat{\mathbf{L}}^T\right) > 0$ is independent of $\mathbf{W}$, maximizing $\det\left(\mathbf{W}\mathbf{X}\mathbf{X}^T\mathbf{W}^T\right)$ is meant by maximizing $\det^2(\mathbf{W}\mathbf{A})$, i.e., $|\det(\mathbf{W}\mathbf{A})|$. Based on (20), (21) and the Lemma 1 in [9], it holds that $|\det(\mathbf{W}\mathbf{A})| \leq 1$, where the equation holds if and only if $\mathbf{W}\mathbf{A}$ is a permutation matrix. Note that if $\mathbf{W}\mathbf{A}$ is a permutation matrix, the equality and inequality conditions in (6) hold. Therefore, the optimal solution $\mathbf{W}^*$ of (6) satisfies $\mathbf{W}^*\mathbf{A} = \mathbf{P}$. This completes the proof.

REFERENCES

[1] N. Jiang and D. Farina, "Covariance and time-scale methods for blind separation of delayed sources," *IEEE Trans. on Biomedical Engineering*, vol. 58, no. 3, pp. 550-556, Mar. 2010.
[2] Y. Xiang, S. K. Ng, and V. K. Nguyen, "Blind separation of mutually correlated sources using precoders," *IEEE Trans. on Neural Networks*, vol. 21, no. 1, pp. 82-90, Jan. 2010.
[3] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Non-Negative Matrix and Tensor Factorization: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*. Oxford, U.K.: Wiley-Blackwell, 2009.
[4] 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.
[5] Z. Yang, S. Ding, and S. Xie, "Blind source separation by fully nonnegative constrained iterative volume maximization," *IEEE Signal Processing Letters*, vol. 17, no. 9, pp. 799-802, Sep. 2010.
[6] M. D. Plumbley and E. Oja, "A 'nonnegative PCA' algorithm for independent component analysis," *IEEE Trans. on Neural Networks*, vol. 15, no. 1, pp. 66-76, Jan. 2004.
[7] 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.
[8] 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. on Signal Processing*, vol. 56, no. 10, pp. 5120-5134, Oct. 2008.
[9] 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. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 875-888, May 2010.
[10] M. D. Plumbley, "Algorithms for nonnegative independent component analysis," *IEEE Trans. on Neural Networks*, vol. 14, no. 3, pp. 534-543, Mar. 2003.
[11] V. Zarzoso, P. Comon, and R. Phlypo, "A contrast function for independent component analysis without permutation ambiguity," *IEEE Trans. on Neural Networks*, vol. 21, no. 5, pp. 863-868, May 2010.
[12] V. Zarzoso and P. Comon, "Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size," *IEEE Trans. on Neural Networks*, vol. 21, no. 2, pp. 248-261, Feb. 2010.
[13] M. D. Plumbley, "Conditions for nonnegative independent component analysis," *IEEE Signal Processing Letters*, vol. 9, no. 6, pp. 177-180, Jun. 2002.
[14] 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.
[15] R. Badeau, N. Bertin, and E. Vincent, "Stability analysis of multiplicative update algorithms and application to nonnegative matrix factorization," *IEEE Trans. on Neural Networks*, vol. 21, no. 12, pp. 1869-1881, Dec. 2010.
[16] R. Schachtner, G. Pöppel, and E. W. Lang, "A nonnegative blind source separation model for binary test data," *IEEE Trans. on Circuits and Systems-I:Regular Papers*, vol. 57, no. 7, pp. 1439-1448, Jul. 2010.
[17] A. Cichocki, A. H. Phan, R. Zdunek, and L. Q. Zhang, "Flexible component analysis for sparse, smooth, nonnegative coding or representation," *Lecture Notes in Computer Science*, vol. 4984, no. 1, pp. 811-820, 2008.
[18] M. Arngren, M. N. Schmidt, and J. Larsen, "Bayesian nonnegative matrix factorization with volume prior for unmixing of hyperspectral images," in *Proc. 2009 IEEE Workshop on Machine Learning for Signal Processing*, 2009.
[19] R. Schachtner, G. Pöppel, A. M. Tomé, and E. W. Lang, "Minimum determinant constraint for non-negative matrix factorization," *Lecture Notes in Computer Science*, vol. 5441 pp.106-113, 2009.
[20] R. Schachtner, G. Pöppel, and E. W. Lang, "Towards unique solutions of non-negative matrix factorization problems by a determinant criterion," *Digital Signal Processing*, vol. 21, no. 4, pp. 528-534, Jul. 2011.
[21] 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.
[22] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898-910, Apr. 2005.
[23] 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. 2006 IEEE Workshop on Machine Learning for Signal Processing*, pp. 73-78, 2006.
[24] S. A. Astakhov, H. Stogbauer, A. Kraskov, and P. Grassberger, "Monte Carlo algorithm for least dependent non-negative mixture decomposition," *Analytical Chemistry*, vol. 78, no. 5, pp. 1620-1627, May 2006.
[25] 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. on Signal Processing*, vol. 54, no. 11, pp. 4133-4145, Nov. 2006.
[26] J. Bobin, J. L. Starck, Y. Moudden, and M. J. Fadili, "Blind source separation: The sparsity revolution," *Advances in Imaging and Electron Physics*, vol. 152, no. 1, pp. 221-302, Jan. 2008.
[27] Z. Xu, X. Chang, F. Xu, H. Zhang, "L1/2 regularization: a thresholding representation theory and a fast solver," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 23, no. 7, pp. 1013-1024, July 2012.
[28] Y. Li, A. Cichocki, S. Amari, S. Xie, and C. Guan, "Equivalence probability and sparsity of two sparse solutions in sparse representation," *IEEE Trans. on Neural Networks*, vol. 19, no. 12, pp. 66-76, Dec. 2008.
[29] Z. Yang, G. Zhou, S. Xie, et al, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 1112-1125, Apr, 2011.
[30] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *PNAS*, vol. 100, no. 5, pp. 2197-2202, Mar. 2003.
[31] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, no. 1, pp. 1457-469, Jan. 2004.
[32] B. Willmore and D. J. Tolhurst, "Characterizing the sparseness of neural codes," *Network Computation in Neural Systems*, vol. 12, no. 3, pp. 255-270, Mar. 2001.
[33] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765-777, Mar. 2007.
[34] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
[35] D. D. Lee and H. S. Seung, "Algorithms for nonnegative matrix factorization," *Advances in Neural Information Processing Systems*, vol. 13, pp. 556-562, 2001.
[36] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hull," *ACM Trans. on Mathematical Software*, vol. 22, no. 4, pp. 469-483, Dec. 1996.

[37] M. W. Berry, M. Browne, and A. N. langville, "Algorithms and applications for approximate nonnegative matrix factoriazation," *Computational Statistics and Data Analysis*, vol. 52, no. 1, pp. 155-173, Sept. 2007.

[38] T. F. Coleman and Y. Li, "A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables," *SIAM Journal on Optimization*, vol. 6, no. 4, pp. 1040-1058, Apr. 1996.

[39] K. Suhling and D. Stephens, *Cell Imaging: Methods Express*. Bloxham: Scion Publishing, 2005.

[40] C. R. Rao and S. K. Mitra, *Generallized Inverse of Matrices and Its Applications*. New York: John Wiley and Sons, 1971.