# Nonnegative Least-Correlated Component Analysis for Separation of Dependent Sources by Volume Maximization

Fa-Yu Wang, Chong-Yung Chi, *Senior Member*, *IEEE*, Tsung-Han Chan, and Yue Wang

**Abstract**—Although significant efforts have been made in developing nonnegative blind source separation techniques, accurate separation of positive yet dependent sources remains a challenging task. In this paper, a joint correlation function of multiple signals is proposed to reveal and confirm that the observations after nonnegative mixing would have higher joint correlation than the original unknown sources. Accordingly, a new nonnegative least-correlated component analysis ($n$LCA) method is proposed to design the unmixing matrix by minimizing the joint correlation function among the estimated nonnegative sources. In addition to a closed-form solution for unmixing two mixtures of two sources, the general algorithm of $n$LCA for the multisource case is developed based on an iterative volume maximization (IVM) principle and linear programming. The source identifiability and required conditions are discussed and proven. The proposed $n$LCA algorithm, denoted by $n$LCA-IVM, is evaluated with both simulation data and real biomedical data to demonstrate its superior performance over several existing benchmark methods.

**Index Terms**—Nonnegative blind source separation, nonnegative least-correlated component analysis, dependent sources, joint correlation function of multiple signals, iterative volume maximization.

✦

## 1 INTRODUCTION

Although signal decomposition is a popular technique widely used in many branches of science and engineering, blind source separation (BSS) deals with a real-world situation where neither the sources nor the mixing matrix is known and the only available information comes from the mixed observations [1], [2]. For instance, multiprobe biomedical imaging exploits simultaneous imaging of multiple biomarkers, where the measured pixel values often represent a composite of multiple sources independent of spatial resolution (e.g., multispectral microscopy, dual-energy X-ray imaging, dynamic functional imaging, and electroencephalogram or magnetoencephalogram (EEG/MEG) [3], [4], [5], [6], [7], [8]). Other examples include remote sensing [1], astronomical imaging [9], analytical spectroscopy [10], and telecommunications [11]. A popular approach to BSS is the independent component analysis (ICA) [12], where the sources are fundamentally assumed to be mutually and statistically independent, although this fundamental assumption may hardly be true in many real-world problems.

Blind separation of nonnegative sources, referred to herein as nonnegative blind source separation ($n$BSS), has been extensively studied recently. Many efforts aim to empirically incorporate intrinsic nonnegativity of source signals into the ICA-based principles and have made some reasonable progress in $n$BSS. For example, Oja and Plumbley [13] proposed a nonnegative independent component analysis ($n$ICA) scheme that exploits the nonnegativity assumption of statistically uncorrelated sources and guarantees the identifiability when the sources are well-grounded (i.e., for each source $s$ and any $\delta > 0$, probability $\mathrm{P_r}(s < \delta) > 0$). More recently, a stochastic nonnegative independent component analysis (SNICA) method [10] was reported that minimizes mutual information between recovered components by using a nonnegativity-constrained simulated annealing algorithm. Despite significant progress in developing various $n$BSS techniques, accurate separation of dependent sources still remains a challenging task.

Nonnegative matrix factorization (NMF) [14] is a benchmark method of decomposing a nonnegative matrix into a product of two nonnegative matrices, and has been widely applied to $n$BSS problems [15], [16], [17]. Since a multiplicative algorithm [14] primarily used to perform NMF is shown to be of slow convergence for large-scale problems, some efforts were reported to apply projected alternating least-squares method [16] or projected Newton method [18] to NMF for speeding up its convergence. Moreover, it is known that NMF may not yield unique decomposition. Some modifications of NMF have been reported for the decomposition to be unique by imposing the sparseness constraint on mixing matrix or source matrix, or both [17],

- *F.-Y. Wang and T.-H. Chan are with the Institute of Communications Engineering, National Tsing Hua University, 101, Section 2, Kuang-Fu Road, Hsinchu, Taiwan 30013, R.O.C. E-mail: wangfaa@ms16.hinet.net, tsunghan@mx.nthu.edu.tw.*
- *C.-Y. Chi is with the Department of Electrical Engineering and the Institute of Communications Engineering, National Tsing Hua University, 101, Section 2, Kuang-Fu Road, Hsinchu, Taiwan 30013, R.O.C. E-mail: cychi@ee.nthu.edu.tw.*
- *Y. Wang is with the Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, 4300 Wilson Blvd, Suite 750 Computational Bioinformatics and Bio-Imaging Lab, Arlington, VA 22203. E-mail: yuewang@vt.edu.*

[19]. Possible circumstances under which the NMF yields unique decomposition can also be found in [20].

The $n$BSS based on the pure-source sample assumption (where a pure-source sample means a data point that is fully dominated by only one source) has been investigated for separation of dependent sources in biomedical imaging [21], [22], [23] and remote sensing applications [24], [25], [26]. In [27], a neural network algorithm, searching for the edges of data scatter plot, exploits the existence of pure-source samples to iteratively update the unmixing matrix such that the sum of all the off-diagonal entries of the angular proximity matrix (analogous to the cross-correlation matrix) of the extracted edges is minimum. A recently reported $n$BSS method, convex analysis of mixtures of nonnegative sources (CAMNS) [22], based on the pure-source sample assumption, has been theoretically proven to achieve perfect separation by searching for all the extreme points of an observation-constructed polyhedral set. In remote sensing, many spectral unmixing algorithms [24], [25], [26], also based on the pure-source sample assumption, try to search for the purest observed pixels from the spectral data set and then identify the mixing matrix. The use of such spectral unmixing methods is usually followed by an inversion process, such as nonnegative least-squares method, to recover the sources. Similar ideas for finding the mixing matrix can be found in our recently reported $n$BSS method, namely $n$LCA [23], which estimates the mixing matrix by an algebraic edge search algorithm. Herein, we renamed the $n$LCA reported in [23] $n$LCA-edge search ($n$LCA-ES) so as to distinguish it from the one to be presented in this paper.

In this paper, we propose a joint correlation function of multiple signals for the design of the demixing matrix. The joint correlation function of the observations mixed by nonnegative combinations of the sources would be higher than that of the sources. Based on this idea and the nonnegativity constraint of the sources, a novel $n$LCA is proposed for which the $n$BSS problem is formulated into a problem of minimizing the joint correlation function of all the demixed sources under the nonnegativity constraint. A closed-form solution for the case of two sources with two observations is provided. For the general case of multiple sources, the proposed $n$LCA can be fulfilled by an iterative volume maximization algorithm ($n$LCA-IVM) using linear program (LP), which finds the optimal demixing matrix by maximizing the volume of a solid region formed by the demixed source vectors regardless of whether the pure-source sample assumption is valid or not. However, the source identifiability of the $n$LCA can be proven under the existence of pure-source samples. In contrast to other spectral unmixing methods, the proposed $n$LCA-IVM, which is never a pure-source sample criterion-based $n$BSS algorithm, is able to separate all the (dependent or independent) sources from the given mixtures without involving any inversion process. Comparative experimental results using synthetic data and real biomedical data are presented to demonstrate the superior performance of $n$LCA-IVM over several existing benchmark methods.

Section 2 presents the $n$BSS problem formulation and the model assumptions. In Section 3, the novel $n$LCA is presented followed by the closed-form solution of the optimum demixing matrix for the two-source case, the $n$LCA-IVM algorithm, and the source identifiability of the $n$LCA under the existence of pure-source samples. For

real-world applications, where the number of mixtures is larger than the number of sources, Section 4 presents noise reduction and rank reduction prior to the $n$BSS processing using the $n$LCA-IVM algorithm. In Section 5, simulation results are presented with synthetic mixed human face images, infrared spectral signals, and dual-energy X-ray images to evaluate the performance of the proposed $n$LCA-IVM algorithm. Then a real data analysis of fluorescence microscopy imaging and that of dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) using the proposed $n$LCA-IVM algorithm are presented to show its effectiveness. Finally, some conclusions are drawn in Section 6.

## 2   PROBLEM FORMULATION

For ease of later use, let us define the following notations:

| | |
|---|---|
| $\mathbb{R}, \mathbb{R}^N, \mathbb{R}^{M \times N}$ | Set of real numbers, $N$-dimensional vectors, $M \times N$ matrices |
| $\mathbb{R}_+, \mathbb{R}_+^N, \mathbb{R}_+^{M \times N}$ | Set of nonnegative real numbers, $N$-dimensional vectors, $M \times N$ matrices |
| $\mathbf{1}_N$ | $N \times 1$ vector with all the entries equal to unity |
| $\mathbf{I}_N$ | $N \times N$ identity matrix |
| $\mathbf{e}_i$ | Unit vector of proper dimension with the $i$th entry equal to unity |
| $\mathbf{0}$ | Zero vector of proper dimension |
| $\| \cdot \|$ | Euclidean norm of a vector or Frobenius norm of a matrix |
| $\succeq$ | Componentwise inequality |
| $\mathrm{diag}\{\alpha_1, \ldots, \alpha_N\}$ | $N \times N$ diagonal matrix with diagonal entries $\alpha_1, \ldots, \alpha_N$ |

Consider the following $M \times N$ linear mixing system:

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n], \quad n = 1, 2, \ldots, L, \tag{1}$$

where $\mathbf{x}[n] = (x_1[n], x_2[n], \ldots, x_M[n])^T$ is the $n$th *data point* of the given $M$ observations (or mixtures), $\mathbf{A} = [a_{ij}]_{M \times N} \in \mathbb{R}^{M \times N}$ is an unknown mixing matrix, $\mathbf{s}[n] = (s_1[n], s_2[n], \ldots, s_N[n])^T$ is the $n$th *source point* consisting of $N$ sources, and $L > \max\{M, N\}$ is the data length (i.e., number of pixels in each observation and each source).

Our goal herein is to design a real demixing matrix $\mathbf{W} = [w_{ij}]_{N \times M}$ such that

$$\mathbf{y}[n] = \mathbf{W}\mathbf{x}[n] = (\mathbf{W}\mathbf{A})\mathbf{s}[n] = \mathbf{P}\mathbf{s}[n], \tag{2}$$

where $\mathbf{y}[n] = (y_1[n], y_2[n], \ldots, y_N[n])^T$ is the extracted (unmixed or demixed) source point and $\mathbf{P} = \mathbf{W}\mathbf{A} \in \mathbb{R}^{N \times N}$ is a permutation matrix, meaning that the extracted source point $\mathbf{y}[n]$ is equivalent to the true source point $\mathbf{s}[n]$ up to a permutation.

For ease of the ensuing derivations, we let

$$\boldsymbol{x}_i = (x_i[1], x_i[2], \ldots, x_i[L])^T, \quad (i\text{th observation})$$

where $i = 1, 2, \ldots, M$ and

$$\boldsymbol{s}_i = (s_i[1], s_i[2], \ldots, s_i[L])^T, \quad (i\text{th source}),$$

$$\boldsymbol{y}_i = (y_i[1], y_i[2], \ldots, y_i[L])^T, \quad (i\text{th extracted source}),$$

where $i = 1, 2, \ldots, N$. Then, an alternative form of (1) is given by

$$\boldsymbol{x}_i = \sum_{j=1}^{N} a_{ij} \boldsymbol{s}_j, \quad i = 1, 2, \ldots, M. \tag{3}$$

Likewise, (2) can be alternatively expressed as

$$\boldsymbol{y}_i = \sum_{j=1}^{M} w_{ij} \boldsymbol{x}_j = \boldsymbol{s}_{t_i}, \quad i = 1, 2, \ldots, N, \tag{4}$$

where $t_i \in \{1, \ldots, N\}$, $t_i \neq t_j$ for $i \neq j$.

To be practical in imaging applications, the $n$LCA to be presented is based on the following general assumptions:

- ($\mathcal{A}1$) (nonnegative sources): For each $i$, $\boldsymbol{s}_i \in \mathbb{R}_+^L$.
- ($\mathcal{A}2$) (nonnegative mixing matrix): $\mathbf{A} \in \mathbb{R}_+^{M \times N}$.
- ($\mathcal{A}3$) $M \geq N$ and $\mathbf{A}$ is of full column rank, i.e., rank($\mathbf{A}$) = $N$.
- ($\mathcal{A}4$) (unit row sum): Each row sum of $\mathbf{A}$ equals one, i.e.,

$$\mathbf{A}\mathbf{1}_N = \mathbf{1}_M. \tag{5}$$

Assumptions ($\mathcal{A}1$) and ($\mathcal{A}2$) are usually satisfied in biomedical imaging and remote sensing, where the mixing process is a nonnegative combination of nonnegative signal intensities. Assumption ($\mathcal{A}3$) is an assumption generally satisfied in BSS problems. Assumption ($\mathcal{A}4$) intrinsically holds in biomedical imaging due to partial volume effect [28], in contrast to the full additivity condition on the sources (i.e., $\boldsymbol{s}^T[n]\mathbf{1}_N = 1$ for all $n$) in hyperspectral imaging [29]. However, when ($\mathcal{A}4$) is not valid, i.e., $\mathbf{A}\mathbf{1}_N \neq \mathbf{1}_M$, we can convert the mixing model (1) into the following normalized model [23]:

$$\bar{\mathbf{x}}[n] = \mathbf{D}_1^{-1}\mathbf{x}[n] = \left(\mathbf{D}_1^{-1}\mathbf{A}\mathbf{D}_2\right)\mathbf{D}_2^{-1}\mathbf{s}[n], \tag{6}$$

where $\mathbf{D}_1 = \mathrm{diag}\{\boldsymbol{x}_1^T\mathbf{1}_L, \boldsymbol{x}_2^T\mathbf{1}_L, \ldots, \boldsymbol{x}_M^T\mathbf{1}_L\}$ (which can be easily obtained from the data) and $\mathbf{D}_2 = \mathrm{diag}\{\boldsymbol{s}_1^T\mathbf{1}_L, \boldsymbol{s}_2^T\mathbf{1}_L, \ldots, \boldsymbol{s}_N^T\mathbf{1}_L\}$. By letting $\bar{\mathbf{A}} = \mathbf{D}_1^{-1}\mathbf{A}\mathbf{D}_2$ (which is unknown) and the normalized unknown source point $\bar{\mathbf{s}}[n] = \mathbf{D}_2^{-1}\mathbf{s}[n]$ (for which each normalized source $\bar{\boldsymbol{s}}_i$ is equivalent to the original source $\boldsymbol{s}_i$ except for an unknown scale factor), the resulting $n$BSS model will be

$$\bar{\mathbf{x}}[n] = \bar{\mathbf{A}}\bar{\mathbf{s}}[n], \tag{7}$$

which satisfies all the assumptions ($\mathcal{A}1$)-($\mathcal{A}4$) and thus is ready to be processed by the $n$LCA-IVM for the estimation of $\bar{\mathbf{s}}[n]$.

## 3 NONNEGATIVE LEAST-CORRELATED COMPONENT ANALYSIS

We first introduce the following three convex sets, which play an important role in the subsequent presentation of $n$LCA:

First, the *convex hull* of $N$ vectors $\{z_1, z_2, \ldots, z_N\} \subset \mathbb{R}^L$ is defined as [30]

$$\mathrm{conv}\{z_1, z_2, \ldots, z_N\} =$$

$$\left\{ z \mid z = \sum_{i=1}^{N} \alpha_i z_i, \quad \sum_{i=1}^{N} \alpha_i = 1, \alpha_i \in \mathbb{R}_+ \right\}.$$
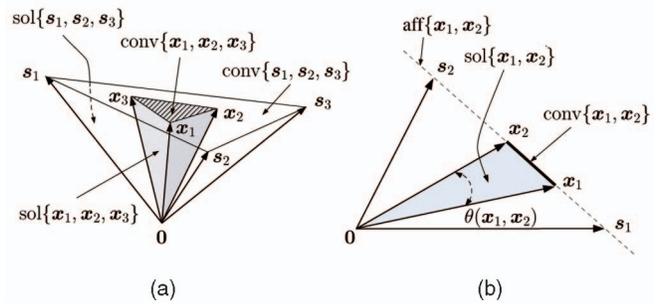


Fig. 1. Geometric illustrations of Theorem 1 for (a) $M = N = 3$ and (b) $M = N = 2$, respectively.

Second, the *affine hull* of $\{z_1, z_2, \ldots, z_N\}$ is defined as [30]

$$\mathrm{aff}\{z_1, z_2, \ldots, z_N\} =$$

$$\left\{ z \mid z = \sum_{i=1}^{N} \alpha_i z_i, \quad \sum_{i=1}^{N} \alpha_i = 1, \alpha_i \in \mathbb{R} \right\}.$$

Finally, the *solid region* formed by $\{z_1, z_2, \ldots, z_N\}$ is defined as

$$\mathrm{sol}\{z_1, z_2, \ldots, z_N\} =$$

$$\left\{ z \mid z = \sum_{i=1}^{N} \alpha_i z_i, \quad \sum_{i=1}^{N} \alpha_i \leq 1, \alpha_i \in \mathbb{R}_+ \right\}.$$

The $n$LCA to be presented below is motivated by the relationship between $\mathrm{conv}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}$ (after nonnegative mixing) and $\mathrm{conv}\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N\}$ (before nonnegative mixing), as well as the relationship between $\mathrm{sol}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}$ and $\mathrm{sol}\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N\}$ as described in the following theorem:

**Theorem 1.** *Suppose that ($\mathcal{A}2$) and ($\mathcal{A}4$) hold. Then,*

1. $\mathrm{conv}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\} \subseteq \mathrm{conv}\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N\}$ *and*
2. $\mathrm{sol}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\} \subseteq \mathrm{sol}\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N\}$.

The proof of Theorem 1 is given in Appendix A. Geometric illustrations of Theorem 1 for the cases of $M = N = 3$ and $M = N = 2$ are shown in Figs. 1a and 1b, respectively. One can see in Fig. 1a that the triangle formed by $\boldsymbol{x}_1, \boldsymbol{x}_2,$ and $\boldsymbol{x}_3$ is inside that formed by $\boldsymbol{s}_1, \boldsymbol{s}_2,$ and $\boldsymbol{s}_3$ (i.e., $\mathrm{conv}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\} \subseteq \mathrm{conv}\{\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3\}$), and the bottom-up pyramid formed by $\boldsymbol{x}_1, \boldsymbol{x}_2,$ and $\boldsymbol{x}_3$ is inside that formed by $\boldsymbol{s}_1, \boldsymbol{s}_2,$ and $\boldsymbol{s}_3$ (i.e., $\mathrm{sol}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\} \subseteq \mathrm{sol}\{\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3\}$). Similar observations can be seen in Fig. 1b including that the distance between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is smaller than that between $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, (i.e., $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \leq \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$) since $\mathrm{conv}\{\boldsymbol{x}_1, \boldsymbol{x}_2\} \subseteq \mathrm{conv}\{\boldsymbol{s}_1, \boldsymbol{s}_2\}$, and the area of $\mathrm{sol}\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ is smaller than that of $\mathrm{sol}\{\boldsymbol{s}_1, \boldsymbol{s}_2\}$ since $\mathrm{sol}\{\boldsymbol{x}_1, \boldsymbol{x}_2\} \subseteq \mathrm{sol}\{\boldsymbol{s}_1, \boldsymbol{s}_2\}$.

The conventional pairwise correlation coefficient between the observation vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, where $\boldsymbol{x}_1 \succeq \mathbf{0}$ and $\boldsymbol{x}_2 \succeq \mathbf{0}$, is widely known as

$$0 < \varrho(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{\boldsymbol{x}_1^T \boldsymbol{x}_2}{\|\boldsymbol{x}_1\| \cdot \|\boldsymbol{x}_2\|} = \cos(\theta(\boldsymbol{x}_1, \boldsymbol{x}_2)) \leq 1.$$

It is straightforward to see that $\varrho(\boldsymbol{s}_1, \boldsymbol{s}_2) \leq \varrho(\boldsymbol{x}_1, \boldsymbol{x}_2)$ since $\theta(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is never larger than $\theta(\boldsymbol{s}_1, \boldsymbol{s}_2)$ for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathrm{conv}\{\boldsymbol{s}_1, \boldsymbol{s}_2\}$ for $M = N = 2$. When $M = N > 2$, it will not be very appropriate to use all the $M(M-1)/2$ conventional pairwise correlation coefficients among all of the extracted

sources to form an effective criterion for the design of the demixing matrix. Instead, we propose a joint correlation function of $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}$, where $M \geq 2$, defined as

$$\operatorname{corr}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M) \triangleq \frac{1}{V(\operatorname{sol}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\})}, \qquad (8)$$

where

$$V(\operatorname{sol}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}) = \int_{\operatorname{sol}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}} \mathrm{d}\boldsymbol{\xi}, \qquad (9)$$

is the volume of $\operatorname{sol}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}$ and $\int_{\operatorname{sol}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}} \mathrm{d}\boldsymbol{\xi}$ is the multiple integral over $\operatorname{sol}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}$. The proposed joint correlation function of multiple signals also reflects the same behavior as the conventional pairwise correlation coefficient. Specifically, as shown in Fig. 1b, the larger the value of $\operatorname{corr}(\boldsymbol{x}_1, \boldsymbol{x}_2)$, the smaller the area of $\operatorname{sol}\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$. Since the distance from the origin to the hyperplane passing through $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ (i.e., $\operatorname{aff}(\boldsymbol{x}_1, \boldsymbol{x}_2)$) is fixed, the smaller the area of $\operatorname{sol}\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$, the smaller the value of $\theta(\boldsymbol{x}_1, \boldsymbol{x}_2)$. Hence, it can be inferred that the larger the value of $\operatorname{corr}(\boldsymbol{x}_1, \boldsymbol{x}_2)$, the larger the value of $\varrho(\boldsymbol{x}_1, \boldsymbol{x}_2)$.

Based on Theorem 1 and (9), we have $V(\operatorname{sol}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}) \leq V(\operatorname{sol}\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N\})$, which implies the following corollary:

**Corollary 1.** *Suppose that (A2) and (A4) hold. Then,*

$$\operatorname{corr}(\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N) \leq \operatorname{corr}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M).$$

Note that $\operatorname{corr}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M) \leq \operatorname{corr}(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N)$ if the demixing matrix satisfies (A2) (i.e., $\mathbf{W} \in \mathbb{R}_+^{N \times M}$) and (A4) (i.e., $\mathbf{W}\mathbf{1}_M = \mathbf{1}_N$) by Corollary 1. However, $\mathbf{W} \notin \mathbb{R}_+^{N \times M}$ in the demixing matrix design to be presented below is a real $N \times M$ matrix such that $\operatorname{corr}(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N) \leq \operatorname{corr}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M)$. If $\mathbf{W} \in \mathbb{R}^{N \times M}$ can be properly designed, $\operatorname{corr}(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N) = \operatorname{corr}(\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N)$ can be achieved.

Corollary 1 is profound, which motivates the idea of designing the demixing matrix $\mathbf{W} \in \mathbb{R}^{N \times M}$ by minimizing the joint correlation function of the $N$ extracted sources $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N$, that is,

$$\min_{\mathbf{W} \in \mathbb{R}^{N \times M}} \operatorname{corr}\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N\}, \qquad (10)$$

subject to (s.t.)

$$\boldsymbol{y}_i = \sum_{j=1}^{M} w_{ij}\boldsymbol{x}_j \succeq \mathbf{0}, i = 1, 2, \ldots, N, \quad \text{(due to (A1))}, \qquad (11)$$

$$\mathbf{W}\mathbf{1}_M = \mathbf{W}\mathbf{A}\mathbf{1}_N = \mathbf{P}\mathbf{1}_N = \mathbf{1}_N. \quad \text{(due to (A4))}. \qquad (12)$$

From (8), the $n$BSS problem (10) subject to the two constraints given by (11) and (12) can be reformulated as

$$\max_{\mathbf{W} \in \mathbb{R}^{N \times M}} V(\operatorname{sol}\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N\})$$
$$\text{s.t. } \mathbf{W}\mathbf{1}_M = \mathbf{1}_N, \quad \boldsymbol{y}_i = \Sigma_{j=1}^{M} w_{ij}\boldsymbol{x}_j \succeq \mathbf{0}, \quad \forall i, \qquad (13)$$

which is a nonlinear and nonconvex optimization problem, implying that finding a closed-form solution of (13) for any $M$ and $N$ is almost formidable.
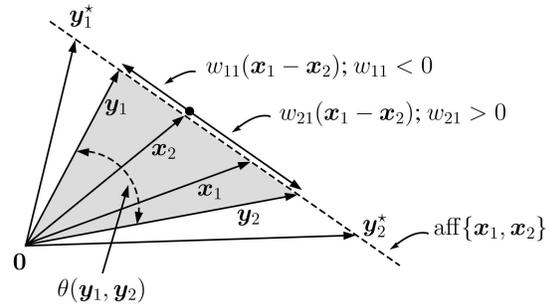


Fig. 2. Geometric illustration of the observations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ and the unmixed signals $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ for $M = N = 2$.

Next, let us present how to solve (13) for the case of $M = N = 2$ and the case of $M = N \geq 2$, respectively, followed by the associated source identifiability.

### 3.1 Case of $M = N = 2$: Closed-Form Solution

Although problem (13) is a nonconvex problem, fortunately we can find a closed-form solution for the case of $M = N = 2$, as stated in the following proposition:

**Proposition 1.** *Under (A1) to (A4), problem (13) for $M = N = 2$ has the closed-form solution $\mathbf{W}^\star$, where*

$$w_{11}^\star = \max_n \left\{ \frac{-x_2[n]}{x_1[n] - x_2[n]} \, \middle| \, x_1[n] > x_2[n], \forall n \right\}, \qquad (14a)$$

$$w_{21}^\star = \min_n \left\{ \frac{-x_2[n]}{x_1[n] - x_2[n]} \, \middle| \, x_1[n] < x_2[n], \forall n \right\}, \qquad (14b)$$

$$w_{12}^\star = 1 - w_{11}^\star, \quad w_{22}^\star = 1 - w_{21}^\star. \qquad (14c)$$

**Proof.** By (11), (12), and $M = N = 2$,

$$\boldsymbol{y}_i = w_{i1}\boldsymbol{x}_1 + w_{i2}\boldsymbol{x}_2 = w_{i1}(\boldsymbol{x}_1 - \boldsymbol{x}_2) + \boldsymbol{x}_2, \quad i = 1, 2. \qquad (15)$$

Thus, maximizing $V(\operatorname{sol}\{\boldsymbol{y}_1, \boldsymbol{y}_2\})$ (i.e., the shaded area in Fig. 2) is equivalent to increasing $|w_{11}|$ and $|w_{21}|$, where $w_{11} \leq 0$ and $w_{21} \geq 0$. Therefore, problem (13) can be decomposed into the following two subproblems:

$$w_{11}^\star = \min_w w$$
$$\text{s.t. } w \leq 0, w(\boldsymbol{x}_1 - \boldsymbol{x}_2) + \boldsymbol{x}_2 \succeq \mathbf{0}, \qquad (16a)$$

$$w_{21}^\star = \max_w w$$
$$\text{s.t. } w \geq 0, w(\boldsymbol{x}_1 - \boldsymbol{x}_2) + \boldsymbol{x}_2 \succeq \mathbf{0}. \qquad (16b)$$

Note that the inequality $w(\boldsymbol{x}_1 - \boldsymbol{x}_2) + \boldsymbol{x}_2 \succeq \mathbf{0}$ is equivalent to $w(x_1[n] - x_2[n]) + x_2[n] \geq 0$ for $n = 1, 2, \ldots, L$. Therefore, the constraint $w(\boldsymbol{x}_1 - \boldsymbol{x}_2) + \boldsymbol{x}_2 \succeq \mathbf{0}$ itself implies

$$w \geq -x_2[n]/(x_1[n] - x_2[n]), \text{ for } x_1[n] > x_2[n], \qquad (17)$$

$$w \leq -x_2[n]/(x_1[n] - x_2[n]), \text{ for } x_1[n] < x_2[n], \qquad (18)$$

which further lead to

$$\max_n \left\{ \frac{-x_2[n]}{x_1[n] - x_2[n]} \, \middle| \, x_1[n] > x_2[n] \right\} \leq w \qquad (19)$$

and

$$w \le \min_n \left\{ \frac{-x_2[n]}{x_1[n] - x_2[n]} \ \Big| \ x_1[n] < x_2[n] \right\}. \tag{20}$$

One can easily see from ($\mathcal{A}1$) and ($\mathcal{A}2$) that $x_i[n] \ge 0$ for all $i$ and $n$. As a result, the optimal solutions of (16a) and (16b) are, respectively, given by

$$w_{11}^\star = \max_n \left\{ \frac{-x_2[n]}{x_1[n] - x_2[n]} \ \Big| \ x_1[n] > x_2[n], \forall n \right\} \le 0, \tag{21}$$

$$w_{21}^\star = \min_n \left\{ \frac{-x_2[n]}{x_1[n] - x_2[n]} \ \Big| \ x_1[n] < x_2[n], \forall n \right\} \ge 0. \tag{22}$$

Then, one can obtain $w_{12}^\star = 1 - w_{11}^\star$ and $w_{22}^\star = 1 - w_{21}^\star$. Thus, the proposition has been proved. □

It is worth mentioning that the optimum solution $\mathbf{W}^\star$ given by (14) for $M = N = 2$ is exactly the same as the one obtained by the $n$LCA-ES [23]. For a more general case, where $M = N \ge 2$, we resort to the $n$LCA-IVM algorithm, to be presented next.

### 3.2 Case of $M = N \ge 2$: $n$LCA-IVM

By the use of (3) and change of integrating variables [31], the objective function of (13) can be written as

$$V(\mathrm{sol}\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}) = |\det(\mathbf{W})| \cdot V(\mathrm{sol}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}). \tag{23}$$

Since the observations $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ are given beforehand, $V(\mathrm{sol}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\})$ is thus a constant. Then, the problem (13) is equivalent to the problem:

$$\max_{\mathbf{W} \in \mathbb{R}^{N \times N}} |\det(\mathbf{W})| \tag{24}$$
$$\text{s.t. } \mathbf{W}\mathbf{1}_N = \mathbf{1}_N, \quad \boldsymbol{y}_i = \Sigma_{j=1}^N w_{ij}\boldsymbol{x}_j \succeq \mathbf{0}, \quad \forall i.$$

The $n$LCA-IVM algorithm tries to solve (24) through an iterative maximization procedure using linear program. The idea is motivated by the cofactor expansion of $\det(\mathbf{W})$ with respect to the $i$th row of $\mathbf{W}$ [32]:

$$\det(\mathbf{W}) = \sum_{j=1}^N (-1)^{i+j} w_{ij} \det(\mathcal{W}_{ij}), \tag{25}$$

where $\mathcal{W}_{ij}$ is the submatrix of $\mathbf{W}$ with the $i$th row and the $j$th column removed. One can see that, with $\mathcal{W}_{ij}$ fixed for $j = 1, 2, \ldots, N$, $\det(\mathbf{W})$ becomes a linear function of $w_{ij}$, $j = 1, 2, \ldots, N$. Let $\mathbf{w}_i^T = (w_{i1}, w_{i2}, \ldots, w_{iN})$ denote the $i$th row vector of $\mathbf{W}$. Problem (24) is then reduced to

$$\max_{\mathbf{w}_i} \left| \sum_{j=1}^N (-1)^{i+j} w_{ij} \det(\mathcal{W}_{ij}) \right| \tag{26}$$
$$\text{s.t. } \mathbf{w}_i^T \mathbf{1}_N = 1, \quad \boldsymbol{y}_i = \Sigma_{j=1}^N w_{ij}\boldsymbol{x}_j \succeq \mathbf{0}.$$

Note that the objective function in (26) is still nonconvex. Fortunately, the maximization problem (26) can be solved in a globally optimal manner by breaking it into two LPs:

$$p^\star = \max_{\mathbf{w}_i} \sum_{j=1}^N (-1)^{i+j} w_{ij} \det(\mathcal{W}_{ij}) \tag{27a}$$
$$\text{s.t. } \mathbf{w}_i^T \mathbf{1}_N = 1, \quad \boldsymbol{y}_i = \Sigma_{j=1}^N w_{ij}\boldsymbol{x}_j \succeq \mathbf{0},$$

TABLE 1
The $n$LCA $-$ IVM Algorithm

| | |
|---|---|
| **Step 1.** | Set a convergence tolerance $\delta > 0$, $i := 1$, $\mathbf{W} := \mathbf{I}_N$, and $v = |\det(\mathbf{W})|$. |
| **Step 2.** | Solve the two LPs |

$$p^\star = \max_{\mathbf{w}_i} \ \sum_{j=1}^N (-1)^{i+j} w_{ij} \det(\mathcal{W}_{ij})$$
$$\text{s.t. } \mathbf{w}_i^T \mathbf{1}_N = 1, \ \Sigma_{j=1}^M w_{ij}\boldsymbol{x}_j \succeq \mathbf{0},$$

$$q^\star = \min_{\mathbf{w}_i} \ \sum_{j=1}^N (-1)^{i+j} w_{ij} \det(\mathcal{W}_{ij})$$
$$\text{s.t. } \mathbf{w}_i^T \mathbf{1}_N = 1, \ \Sigma_{j=1}^M w_{ij}\boldsymbol{x}_j \succeq \mathbf{0},$$

| | |
|---|---|
| | to obtain their optimal solutions, denoted by $(p^\star, \bar{\mathbf{w}}_i)$ and $(q^\star, \hat{\mathbf{w}}_i)$, respectively. |
| **Step 3.** | If $|p^\star| > |q^\star|$, then update $\mathbf{w}_i := \bar{\mathbf{w}}_i$. Otherwise, update $\mathbf{w}_i := \hat{\mathbf{w}}_i$. |
| **Step 4.** | If $(i \text{ modulo } N) \ne 0$, then set $i := i + 1$ and go to **Step 2**, else |
| | If $|\max\{|p^\star|, |q^\star|\} - v|/v < \delta$, then $\mathbf{W}^\star = \mathbf{W}$ and stop. Otherwise, set $v := \max\{|p^\star|, |q^\star|\}$, $i := 1$, and then go to **Step 2**. |

$$q^\star = \min_{\mathbf{w}_i} \sum_{j=1}^N (-1)^{i+j} w_{ij} \det(\mathcal{W}_{ij}) \tag{27b}$$
$$\text{s.t. } \mathbf{w}_i^T \mathbf{1}_N = 1, \quad \boldsymbol{y}_i = \Sigma_{j=1}^N w_{ij}\boldsymbol{x}_j \succeq \mathbf{0}.$$

The optimal solution of (26), denoted by $\mathbf{w}_i^\star$, is chosen as the optimal solution of (27a) if $|p^\star| > |q^\star|$ or the optimal solution of (27b) if $|q^\star| > |p^\star|$. The $n$LCA-IVM is conducted in an iterative row-by-row manner until convergence is reached. Note that, at each iteration, all of the $N$ row vectors of $\mathbf{W}$ are updated once. The $n$LCA-IVM is summarized in Table 1.

By using the primal-dual interior-point method [33], each LP problem in (27) can be solved with a computational complexity of $\mathcal{O}(L(N-1) + (N-1)^3) \simeq \mathcal{O}(L(N-1))$, on average [22], since $L \gg N$. Because the $n$LCA-IVM algorithm solves $2N$ LP problems per iteration, one can infer that its complexity (abbreviation of computational complexity) per iteration is $\mathcal{O}(N^2L)$. However, if the number ($L$) of inequality constraints in problem (27) can be significantly reduced, the complexity can then be significantly reduced. This is possible by eliminating the redundant inequality constraints of problem (27), as presented next.

The feasible set of (27) can be equivalently written as

$$\mathcal{F} = \{\mathbf{w} \in \mathbb{R}^N \mid \mathbf{w}^T \mathbf{1}_N = 1, \mathbf{w}^T \mathbf{x}[n] \ge 0, \forall n\}, \tag{29}$$

$$= \{\mathbf{w} \in \mathbb{R}^N \mid \mathbf{w}^T \mathbf{1}_N = 1, \mathbf{w}^T \mathbf{v} \ge 0, \tag{30}$$
$$\mathbf{v} \in \mathrm{conv}\{\mathbf{x}[1], \mathbf{x}[2], \ldots, \mathbf{x}[L]\}\}.$$

The convex hull of all the data points is given by

$$\mathrm{conv}\{\mathbf{x}[1], \mathbf{x}[2], \ldots, \mathbf{x}[L]\} =$$
$$\left\{ \sum_{i=1}^r \theta_i \mathbf{x}[l_i] \ \Big| \ \sum_{i=1}^r \theta_i = 1, \theta_i \ge 0, \forall i \right\}, \tag{31}$$

TABLE 2
Complexity Order Comparison of Various BSS Methods

| Methods | $n$LCA-IVM | $n$LCA-ES | CAMNS | $n$ICA | FastICA | NMF | QNMF | $s$NMF |
|---|---|---|---|---|---|---|---|---|
| Complexity orders | $\mathcal{O}(N^2 r\eta)$ | $\mathcal{O}(N^3 L)$ | $\mathcal{O}((N-1)^2 L)$ | $\mathcal{O}(N^2 L\eta)$ | $\mathcal{O}(N^3 L\eta)$ | $\mathcal{O}(N^2 L\eta)$ | $\mathcal{O}(N^2 L\eta)$ | $\mathcal{O}(N^2 L\eta)$ |

*In the table, $N$ denotes the number of sources, $L$ denotes the data length, $\eta$ denotes the number of iterations, and $r \leq L$ denotes the number of extreme points of the convex hull (31).*

where $\mathbf{x}[l_1], \ldots, \mathbf{x}[l_r]$ are the extreme points of the convex hull and $r \leq L$ is the total number of the extreme points. Then, by (31) and (30), it follows that

$$
\begin{aligned}
\mathcal{F} = \Big\{ & \mathbf{w} \in \mathbb{R}^N \mid \mathbf{w}^T \mathbf{1}_N = 1, \\
& \sum_{i=1}^{r} \theta_i \mathbf{w}^T \mathbf{x}[l_i] \geq 0, \sum_{i=1}^{r} \theta_i = 1, \theta_i \geq 0, \forall i \Big\},
\end{aligned}
\tag{32}
$$

$$
\begin{aligned}
= \{ & \mathbf{w} \in \mathbb{R}^N \mid \mathbf{w}^T \mathbf{1}_N = 1, \\
& \mathbf{w}^T \mathbf{x}[l_i] \geq 0, i = 1, \ldots, r \},
\end{aligned}
\tag{33}
$$

where $\mathbf{x}[l_1], \ldots, \mathbf{x}[l_r]$ can be identified by using the quickhull algorithm [34], a well-known extreme point search method. Suppose that $r$ extreme points, or equivalently resultant $r$ inequality constraints in (33), are found, the computational complexity of the $n$LCA-IVM then becomes $\mathcal{O}(N^2 r\eta)$ instead of $\mathcal{O}(N^2 L\eta)$, where $\eta$ is the number of iterations involved. Due to the use of the quickhull algorithm, the computational efficiency improvement of the proposed $n$LCA-IVM can be measured by the *constraint reduction ratio*, defined as

$$
\Delta = \frac{L - r}{L}.
\tag{34}
$$

Surely, the larger the value of $\Delta$, the less the computational complexity in solving (24).

Some BSS methods, $n$LCA-ES [23], CAMNS [22], $n$ICA [13], FastICA [12], NMF [14], quasi-Newton NMF (QNMF) [18], and sparse NMF ($s$NMF) [17], can be easily verified to have complexities (for the case of $M = N$) given by $\mathcal{O}(N^3 L)$, $\mathcal{O}((N-1)^2 L)$, $\mathcal{O}(N^2 L\eta)$, $\mathcal{O}(N^3 L\eta)$, $\mathcal{O}(N^2 L\eta)$, and $\mathcal{O}(N^2 L\eta)$, respectively. The complexity order comparison for the above-mentioned BSS methods is listed in Table 2. Obviously, the proposed $n$LCA-IVM has lower complexity order than $n$ICA, FastICA, NMF, QNMF, and $s$NMF.

A geometrical illustration of how the proposed $n$LCA-IVM method works for the case of $M = N = 3$ is shown in Fig. 3. For each row vector update of $\mathbf{W}$, the unmixed source, say $y_1$, is moved to the boundary of the
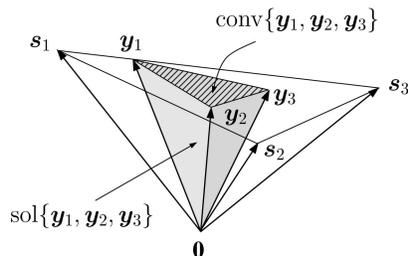


Fig. 3. Geometric interpretation of $n$LCA-IVM for $M = N = 3$, where $y_1$ is updated with $y_2$ and $y_3$ fixed, such that $V(\mathrm{sol}\{y_1, y_2, y_3\})$ increases maximally.

feasible set, while fixing $y_2$ and $y_3$, such that $V(\mathrm{sol}\{y_1, y_2, y_3\})$ increases maximally (since the LP always yields the optimal $y_1$ on the boundary of the feasible set). By repeating the procedure, the three sources will be literally identified. Similarly, for the $M = N = 2$ case as shown in Fig. 2, the $n$LCA-IVM algorithm involves only two row vector updates, implying that the global optimal solution for $\mathbf{W}$ can be obtained in only one iteration (in addition to one iteration for the convergence check). Thus, we can conclude that the proposed $n$LCA-IVM algorithm is able to provide the globally optimal solution for $M = N = 2$ as given in Proposition 1 above, but no guarantee of the global optimality for $M = N > 2$.

### 3.3  Source Identifiability of the $n$LCA

We address here the source identifiability problem of $n$LCA concerning certain conditions (assumptions) under which the $N$ extracted sources are identical to all of the $N$ true sources. In many biomedical imaging [8], [21] and remote sensing applications [24], [29], the sources meet the assumption that *pure-source samples* exist:

- ($\mathcal{A}$5) There exists at least one index set $\{l_1, l_2, \ldots, l_N\}$ such that $\mathbf{s}[l_i] = s_i[l_i]\mathbf{e}_i, i = 1, 2, \ldots, N$.

Note that the corresponding data points $\mathbf{x}[n]$ for $n \in \{l_1, l_2, \ldots, l_N\}$ become

$$
\mathbf{x}[l_i] = \mathbf{a}_i s_i[l_i],
\tag{35}
$$

where $\mathbf{a}_i$ is the $i$th column vector of $\mathbf{A}$. Namely, the data point $\mathbf{x}[l_i]$ is only contributed from source $s_i$. For investigating the source identifiability of the $n$LCA, the following lemma is needed:

**Lemma 1.** *Suppose that ($\mathcal{A}$2), ($\mathcal{A}$3), and ($\mathcal{A}$4) hold and $M = N$. Then, $|\det(\mathbf{A})| \leq 1$, and the equality holds if and only if $\mathbf{A}$ is a permutation matrix.*

The proof of Lemma 1 is given in Appendix B. Under the existence of pure-source samples, the source identifiability of the proposed $n$LCA is established in the following theorem:

**Theorem 2. (Source identifiability of $n$LCA).** *Suppose that ($\mathcal{A}$1)-($\mathcal{A}$5) hold. Then,*

$$
\mathbf{W}^\star \mathbf{A} = \mathbf{P},
\tag{36}
$$

*where $\mathbf{W}^\star$ is the optimal solution of problem (24) and $\mathbf{P}$ is an $N \times N$ permutation matrix.*

**Proof.** Substituting $\mathbf{W}^\star$ into the constraints of problem (24) yields

$$
\mathbf{W}^\star \mathbf{1}_N = \mathbf{1}_N,
\tag{37a}
$$

$$
\mathbf{W}^\star \mathbf{x}[n] \succeq \mathbf{0}, \quad n = 1, 2, \ldots, L.
\tag{37b}
$$

By (1), (37b) can also be expressed as

$$\mathbf{W}^{\star} \sum_{i=1}^{N} s_i[n]\mathbf{a}_i = \sum_{i=1}^{N} s_i[n](\mathbf{W}^{\star}\mathbf{a}_i) \succeq \mathbf{0}, \quad \forall n, \qquad (38)$$

which can be further reduced to the following $N$ inequalities:

$$\mathbf{W}^{\star}\mathbf{a}_i \succeq \mathbf{0}, i = 1, 2, \ldots, N. \text{ (by } (\mathcal{A}5)). \qquad (39)$$

By letting $\mathbf{B} = \mathbf{W}^{\star}\mathbf{A}$, it can be inferred from (37a), ($\mathcal{A}4$), and (39) that the feasible set of $\mathbf{B}$ is given by

$$\mathcal{B} = \left\{ \mathbf{B} \mid \mathbf{B}\mathbf{1}_N = \mathbf{1}_N, \mathbf{B} \in \mathbb{R}_+^{N \times N} \right\}.$$

Now, we prove this theorem by contradiction. Suppose that $\mathbf{B} \in \mathcal{B}$ but $\mathbf{B} \notin \mathcal{P}_N$, where $\mathcal{P}_N$ denotes the set of $N \times N$ permutation matrices. Then one can infer, by Lemma 1, that

$$1 > |\det(\mathbf{B})| = |\det(\mathbf{W}^{\star}\mathbf{A})| = |\det(\mathbf{W}^{\star})| \cdot |\det(\mathbf{A})|. \quad (40)$$

Furthermore, it can be inferred from (40) that

$$|\det(\mathbf{W}^{\star})| < \frac{1}{|\det(\mathbf{A})|} = |\det(\mathbf{A}^{-1})| = |\det(\mathbf{W}')|, \qquad (41)$$

where $\mathbf{W}' = \mathbf{A}^{-1}$. It can be easily shown that $\mathbf{W}'$ satisfies (37a) and (39). Therefore, (41) contradicts with the optimality of (24). Thus, the theorem has been proved. □

Though the source identifiability is proven under ($\mathcal{A}5$), we will show with experimental results (Section 5) that the proposed $n$LCA-IVM is still able to achieve accurate source estimation even if ($\mathcal{A}5$) is not perfectly satisfied. Because the proposed $n$LCA-IVM above only considers the case of $M = N$ without additive noise contamination, we next present the noise suppression and dimension reduction methods for the case of $M > N$ to obtain the best $N$ nonnegative rank and noise-reduced observations prior to employing the proposed $n$LCA-IVM.

## 4 RANK REDUCTION AND NOISE SUPPRESSION BY PRINCIPAL COMPONENT ANALYSIS

As $M > N$, the rank-reducing approximation, so-called the principal component analysis (PCA) [35], can provide $M$ approximations (of rank equal to $N$) to the given $M$ observations. Then $N$ rank-reduced approximated observations with the smallest approximation errors are considered to be processed by the proposed $n$LCA-IVM. Now let us consider the noisy observations

$$\boldsymbol{x}_i = \sum_{j=1}^{N} a_{ij}\boldsymbol{s}_j + \boldsymbol{\eta}_i, \quad i = 1, 2, \ldots, M, \qquad (42)$$

where $\boldsymbol{\eta}_i \in \mathbb{R}^L$ is the additive noise vector. Another matrix expression of (42) is given by

$$\mathbf{X} = \mathbf{S}\mathbf{A}^T + \boldsymbol{\Psi}, \qquad (43)$$

where $\mathbf{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M) \in \mathbb{R}^{L \times M}$ is the observation matrix, $\mathbf{S} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N) \in \mathbb{R}^{L \times N}$ is the source matrix, and $\boldsymbol{\Psi} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_M) \in \mathbb{R}^{L \times M}$ is the noise matrix.

It is generally true that $\text{rank}(\mathbf{S}) = N$ because the sources $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N$ are linearly independent. It can be easily shown by the rank inequality of the product of matrices [36] that $\text{rank}(\mathbf{S}\mathbf{A}^T) = N < M$, which implies $N < \text{rank}(\mathbf{X}) \leq M$, as $\boldsymbol{\Psi} \neq \mathbf{0}$. Note that $\mathbf{X}$ is usually of full column rank as $\boldsymbol{\Psi} \neq \mathbf{0}$. The optimum rank-reducing approximation to the data matrix, denoted by $\mathbf{X}^{\star}$, can be realized by minimizing the following approximation error:

$$J = \|\mathbf{X} - \mathbf{X}^{\star}\|^2 = \sum_{i=1}^{M} \|\boldsymbol{x}_i - \boldsymbol{x}_i^{\star}\|^2, \qquad (44)$$

subject to the constraint of $\text{rank}(\mathbf{X}^{\star}) = N$. Each column of $\mathbf{X}^{\star}$ takes the following linear form:

$$\boldsymbol{x}_i^{\star} = \boldsymbol{\mu} + \mathbf{V}\boldsymbol{\lambda}_i, \quad i = 1, 2, \ldots, M, \qquad (45)$$

where $\boldsymbol{\mu}$ is an $L \times 1$ vector, $\mathbf{V}$ is an $L \times (N-1)$ semiunitary matrix (i.e., $\mathbf{V}^T\mathbf{V} = \mathbf{I}_{N-1}$), and $\boldsymbol{\lambda}_i$ are $(N-1) \times 1$ vectors. It is straightforward to prove that the optimum $\boldsymbol{\mu}$, $\boldsymbol{\lambda}_i$, and $\mathbf{V}$ are given by $\boldsymbol{\mu}^{\star} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{x}_i, \boldsymbol{\lambda}_i^{\star} = (\mathbf{V}^{\star})^T(\boldsymbol{x}_i - \boldsymbol{\mu}^{\star})$, and $\mathbf{V}^{\star}$ is a matrix consisting of $(N-1)$ principal left singular vectors of the matrix $(\boldsymbol{x}_1 - \boldsymbol{\mu}^{\star}, \boldsymbol{x}_2 - \boldsymbol{\mu}^{\star}, \ldots, \boldsymbol{x}_M - \boldsymbol{\mu}^{\star})$. Note that, in the presence of additive noise, an additional advantage is noise mitigation as $M > N$ besides rank reduction.

After obtaining the optimum $\boldsymbol{x}_i^{\star} = \boldsymbol{\mu}^{\star} + \mathbf{V}^{\star}\boldsymbol{\lambda}_i^{\star}$, $i = 1, 2, \ldots, M$, we select $N$ noise-reduced (and rank-reduced) observations $\boldsymbol{x}_i^{\star} \in \mathbb{R}^L$ with the smallest approximation errors $\|\boldsymbol{x}_i^{\star} - \boldsymbol{x}_i\|^2$, say $\boldsymbol{x}_1^{\star}, \ldots, \boldsymbol{x}_N^{\star}$ without loss of generality. To maintain the nonnegativity of the selected $\boldsymbol{x}_i^{\star}$, let $\bar{\boldsymbol{x}}_i = (\bar{x}_{i1}, \ldots, \bar{x}_{iL})^T$ be the $i$th noise-reduced observation to be processed by the proposed $n$LCA-IVM, where

$$\bar{x}_{ij} = \max\{x_{ij}^{\star}, 0\}, \quad j = 1, 2, \ldots, L \qquad (46)$$

in which $x_{ij}^{\star}$ denotes the $j$th component of $\boldsymbol{x}_i^{\star}$.

## 5 EXPERIMENTAL RESULTS

In this section, five sets of experimental results are presented to demonstrate the efficacy of the proposed $n$LCA-IVM (after removing redundant constraints of (27) using the quickhull algorithm [34]). In the first three experiments (for scenarios of different numbers of sources and existence/nonexistence of pure-source samples as in Experiment 1, sparse sources as in Experiment 2, and long data length as in Experiment 3), simulation data over 50 Monte Carlo runs are used to evaluate the performance and complexity of the $n$LCA-IVM, and seven existing BSS algorithms $n$LCA-ES [23], CAMNS [22], $n$ICA [13], FastICA [12], NMF [14], QNMF [18], and $s$NMF [17]. At each run, all the entries of the mixing matrix were randomly generated with a uniform distribution over $[0, 1]$ followed by normalization of each row sum equal to unity. The last two experiments are real data applications of the proposed $n$LCA-IVM to fluorescence microscopy and DCE-MRI images.

The parameter settings for each algorithm in the first three experiments are as follows: For $n$LCA-IVM and FastICA, the convergence tolerances were set to $\delta = 10^{-5}$ and $\delta = 10^{-4}$, respectively. For $n$ICA, NMF, QNMF, and $s$NMF, the convergence rule was the maximum number of

iterations set to 2,000. The other two algorithms, $n$LCA-ES and CAMNS, themselves are not iterative algorithms (without involving convergence). For $s$NMF [17], the regulation parameter controlling the sparseness constraint was set to 0.6. Other parameter settings for QNMF are as follows: At each iteration, the source matrix $\mathbf{S}$ is updated using the fixed-point algorithm with regulation parameter $\lambda = \lambda_0 e^{-\tau \kappa}$, where $\lambda_0 = 20$, $\tau = 0.02$, and $\kappa$ denotes the iteration number, and the mixing matrix $\mathbf{A}$ is updated using the quasi-Newton algorithm [18]. On the other hand, in the last two experiments, the convergence tolerance was still set to $\delta = 10^{-5}$ for the proposed $n$LCA-IVM.

Let $\mathbf{q}(s_i) = (\mathbf{1}_L^T s_i / L) \mathbf{1}_L$ and let $\hat{\mathbf{S}} = (\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_N)$ denote the extracted source matrix. The cross-correlation coefficient $\rho$ between $\hat{\mathbf{S}}$ and $\mathbf{S}$, defined as

$$\rho = \frac{1}{N} \max_{\substack{\pi \in \Pi_N \\ c_i \in \{1, -1\}}} \sum_{i=1}^{N} \frac{(s_i - \mathbf{q}(s_i))^T (c_i \hat{s}_{\pi_i} - c_i \mathbf{q}(\hat{s}_{\pi_i}))}{\|s_i - \mathbf{q}(s_i)\| \cdot \|\hat{s}_{\pi_i} - \mathbf{q}(\hat{s}_{\pi_i})\|},$$

($0 \leq \rho \leq 1$) is used as the measure for performance evaluation and comparison of the algorithms under test, where $\Pi_N = \{\pi = (\pi_1, \ldots, \pi_N) \mid \pi_i \in \{1, \ldots, N\}, \pi_i \neq \pi_j, \forall i \neq j\}$ is the set of all the permutations of $\{1, 2, \ldots, N\}$ and $c_i$ is the sign (or the polarity) ambiguity between the extracted source $\hat{s}_{\pi_i}$ and the true source $s_i$. The above performance measure itself is an optimization problem to find the optimal $\pi$ and $c_i$ such that $\rho$ is maximum, and it can be efficiently solved by Kuhn-Munkres algorithm [37]. Note that the larger the value of $\rho$, the better the performance of the associated algorithm under evaluation.

For the computational complexity comparison of the proposed $n$LCA-IVM (including the use of the quickhull algorithm) and the other seven algorithms, the computation time (seconds) of each BSS algorithm (implemented with Matlab 7.0) when executed using a Toshiba Satellite A100 laptop personal computer (equipped with Intel Centrino Duo Processor T2050 CPU 1.6 GHz, 1 GB memory and Microsoft windows XP home edition version 2002 with service pack 2) was measured as the corresponding computational complexity index.

## 5.1 Experiment 1: Human Face Image Separation

In this experiment, two test scenarios were considered. In the first scenario, six $128 \times 128$ human face images (i.e., $L = 16{,}384$) (shown in Fig. 4a) taken from [38] were used as the source signals to generate six noise-free observations (i.e., $M = 6$) and 20 noisy observations (i.e., $M = 20$) with the same signal-to-noise ratio ($\text{SNR}_i = 25$ dB, $i = 1, 2, \ldots, 20$) where $\text{SNR}_i = \|x_i - \eta_i\|^2 / (L\sigma_i^2)$ and $\sigma_i^2$ is the variance of independent and identically distributed zero-mean Gaussian noise components in $\eta_i$. In the second scenario, the same simulation was performed except that only the first four human face images in Fig. 4a were used. For the noise-free case, apart from the mixtures of the original sources, we also considered the case by artificially adding a set of pure-source samples $\{\mathbf{s}[l_i] = s_i[l_i]\mathbf{e}_i, i = 1, 2, \ldots, N\}$ to the original sources in order to verify the source identifiability of the $n$LCA. For the noisy case, we applied the various BSS methods to process the noise-reduced observations, i.e., $\bar{x}_i$, $i = 1, 2, \ldots, N$, given by (46).



Fig. 4. Human face images in the absence of noise: (a) the sources, (b) the observations, and the extracted sources obtained by (c) $n$LCA-IVM, (d) $n$LCA-ES, (e) CAMNS, (f) $n$ICA, (g) FastICA, (h) NMF, (i) QNMF, and (j) $s$NMF.

The values of averaged $\rho$, denoted by $\rho_{\text{ave}}$, for $N = 6$ in this experiment are listed in Table 3. It is to be mentioned that the original six sources do not have pure-source samples. From the table, one can observe that, for both the noise-free and noisy cases, the proposed $n$LCA-IVM (with maximum $\rho_{\text{ave}}$) outperforms all the other algorithms. For the case that pure-source samples were artificially added in the original six sources, $n$LCA-IVM, $n$LCA-ES, and CAMNS achieve perfect separation (with $\rho_{\text{ave}} = 1$), and hence, outperform all of the other algorithms. These results indicate that the proposed $n$LCA-IVM is less sensitive to the existence of pure-source samples than $n$LCA-ES and CAMNS. In addition, it obtains the global optimum solution for each realization when six pure-source samples are artificially made existent in the original sources, thus justifying the source identifiability of the $n$LCA as stated in Theorem 2 (despite no guarantee for the global optimum solution when $N > 2$). One typical realization among the 50 independent runs for the noiseless case is shown in Fig. 4, where all of the separated images have been properly ordered for ease of illustrative comparison among all the BSS algorithms under test. One can see in Fig. 4 that the separated images obtained by the $n$LCA-IVM match the

TABLE 3
Averaged Cross-Correlation Coefficients ($\rho_{\text{ave}}$) between Sources and Their Estimates, Averaged Computation Time ($T_{\text{ave}}$), Averaged Number of Iterations ($\eta_{\text{ave}}$) of Various BSS Methods over 50 Monte Carlo Runs for $N = 6$ in Experiment 1

| Testing cases | Exp. 1 (Human face image separation) $L = 16384$ |
|---|---|---|---|---|---|---|---|---|---|
| | $M = N = 6$, SNR=$\infty$ (No pure-source sample) | | | $M = 20$, $N = 6$, SNR=25 dB (No pure-source sample) | | | $M = N = 6$, SNR=$\infty$ (Pure-source samples added) | | |
| Method | $\rho_{\text{ave}}$ | $T_{\text{ave}}$ (sec) | $\eta_{\text{ave}}$ | $\rho_{\text{ave}}$ | $T_{\text{ave}}$ (sec) | $\eta_{\text{ave}}$ | $\rho_{\text{ave}}$ | $T_{\text{ave}}$ (sec) | $\eta_{\text{ave}}$ |
| $n$LCA-IVM | 0.9952 | 22.95 | 3.92 | 0.9213 | 15.60 | 5.48 | 1 | 3.94 | 2.26 |
| $n$LCA-ES | 0.7661 | 0.05 | — | 0.6383 | 0.06 | — | 1 | 0.06 | — |
| CAMNS | 0.8973 | 13.66 | — | 0.7685 | 9.54 | — | 1 | 12.23 | — |
| $n$ICA | 0.4694 | 54.19 | 2000 | 0.4590 | 54.62 | 2000 | 0.5032 | 54.82 | 2000 |
| FastICA | 0.4357 | 52.80 | 83.68 | 0.3889 | 58.99 | 77.76 | 0.4050 | 61.78 | 81.22 |
| NMF | 0.6156 | 58.98 | 2000 | 0.6959 | 59.27 | 2000 | 0.6307 | 60.07 | 2000 |
| QNMF | 0.5614 | 77.14 | 2000 | 0.5465 | 76.44 | 2000 | 0.5093 | 74.64 | 2000 |
| $s$NMF | 0.5968 | 93.84 | 2000 | 0.7079 | 94.69 | 2000 | 0.6017 | 95.11 | 2000 |

original sources better than those obtained by the other algorithms.

Again, from Table 3, one can observe that the computational complexity in terms of the averaged computation time, denoted by $T_{\text{ave}}$, of the proposed $n$LCA-IVM is lower than all the other iterative algorithms, i.e., $n$ICA, FastICA, NMF, QNMF, and $s$NMF. An interesting observation is that $T_{\text{ave}}$ (3.94 sec) associated with the proposed $n$LCA-IVM when pure-source samples exist is significantly smaller than that ($T_{\text{ave}} = 22.95$ sec) without pure-source samples. This is because more redundant inequality constraints of problem (24) can be effectively removed by the quickhull algorithm. On the other hand, $T_{\text{ave}}$s associated with the two non-iterative algorithms, $n$LCA-ES and CAMNS, are similar for all the three cases. Moreover, the averaged constraint reduction ratios ($0 < \Delta \leq 1$) (see (34)) are 0.9385 for the noise-free case, 0.9504 for the noisy case, and 0.9853 for the case with artificially added pure-source samples, indicating that the computational efficiency with the use of the quickhull algorithm prior to the use of the $n$LCA-IVM is remarkably high in this scenario.

We also show the corresponding simulation results for $N = 4$ in Table 4, where the first four human face images in Fig. 4a were used. From this table, similar conclusions about the performance and computational complexity comparison of all the BSS algorithms under test can be drawn as from Table 3. Here, the averaged constraint reduction ratios $\Delta$ are 0.9887 for the noise-free case, 0.9864 for the noisy case,

and 0.9943 for the case with artificially added pure-source samples, indicating that the computational efficiency with the use of the quickhull algorithm prior to the use of the $n$LCA-IVM is also high in this scenario.

## 5.2 Experiment 2: Infrared Spectra Decomposition

In the field of analytical chemistry, infrared spectra are often used to identify and quantify a solvent in liquid or gas state by its spectral signature. When multiple solvents are mixed, the measured infrared spectrum is a mixture of the spectra of these solvents. In this experiment, we used the gas phase Fourier-transform infrared spectra of three pure materials (i.e., toluene, acetone, and dichloromethane) with $2\,\text{cm}^{-1}$ resolution from a public quantitative infrared database (i.e., http://webbook.nist.gov/chemistry/quant-ir/) to generate three mixtures, where $L = 3{,}526$. The values of $\rho_{\text{ave}}, T_{\text{ave}}$, and $\eta_{\text{ave}}$ of the BSS algorithms under test are shown in Table 5. One can observe from Table 5 that the proposed $n$LCA-IVM performs almost as well as $n$LCA-ES and CAMNS (where $\rho_{\text{ave}} \simeq 1$ does not justify the source identifiability of the former due to no pure-source samples), and its performance is superior to that of $n$ICA, FastICA, NMF, QNMF, and $s$NMF, although no pure-source sample exists in the three original sources of this experiment. As for the computational complexity, the best four algorithms are $n$LCA-ES, FastICA, CAMNS, and $n$LCA-IVM. One typical realization among the 50 independent runs for this experiment is shown in Fig. 5, where all the separated spectra (intensity versus wave number ($\text{cm}^{-1}$)) have been

TABLE 4
Averaged Cross-Correlation Coefficients ($\rho_{\text{ave}}$) between Sources and Their Estimates, Averaged Computation Time ($T_{\text{ave}}$), Averaged Number of Iterations ($\eta_{\text{ave}}$) of Various BSS Methods over 50 Monte Carlo Runs for $N = 4$ in Experiment 1

| Testing cases | Exp. 1 (Human face image separation) $L = 16384$ |
|---|---|---|---|---|---|---|---|---|---|
| | $M = N = 4$, SNR=$\infty$ (No pure-source sample) | | | $M = 20$, $N = 4$, SNR=25 dB (No pure-source sample) | | | $M = N = 4$, SNR=$\infty$ (Pure-source samples added) | | |
| Method | $\rho_{\text{ave}}$ | $T_{\text{ave}}$ (sec) | $\eta_{\text{ave}}$ | $\rho_{\text{ave}}$ | $T_{\text{ave}}$ (sec) | $\eta_{\text{ave}}$ | $\rho_{\text{ave}}$ | $T_{\text{ave}}$ (sec) | $\eta_{\text{ave}}$ |
| $n$LCA-IVM | 0.9993 | 2.11 | 3.1 | 0.9107 | 3.74 | 3.74 | 1 | 1.07 | 3.10 |
| $n$LCA-ES | 0.8876 | 0.04 | — | 0.7892 | 0.04 | — | 1 | 0.04 | — |
| CAMNS | 0.9785 | 7.47 | — | 0.8570 | 5.88 | — | 1 | 9.20 | — |
| $n$ICA | 0.4684 | 32.51 | 2000 | 0.5056 | 32.68 | 2000 | 0.4974 | 32.13 | 2000 |
| FastICA | 0.3821 | 23.77 | 25.22 | 0.3830 | 24.20 | 30.02 | 0.3937 | 23.32 | 26.74 |
| NMF | 0.6655 | 34.29 | 2000 | 0.8537 | 33.78 | 2000 | 0.6714 | 33.29 | 2000 |
| QNMF | 0.6778 | 46.10 | 2000 | 0.7918 | 46.29 | 2000 | 0.6820 | 46.56 | 2000 |
| $s$NMF | 0.6524 | 110.05 | 2000 | 0.8304 | 65.36 | 2000 | 0.6552 | 62.27 | 2000 |

TABLE 5
Averaged Cross-Correlation Coefficients ($\rho_{\text{ave}}$) between Sources and Their Estimates, Averaged Computation Time ($T_{\text{ave}}$),
Averaged Number of Iterations ($\eta_{\text{ave}}$) of Various BSS Methods over 50 Monte Carlo Runs in Experiments 2 and 3

| Testing cases | Exp. 2 (Infrared spectra decomposition) $M = N = 3$, $L = 3526$, SNR=$\infty$ (No pure-source sample) | | | Exp. 3 (Dual-energy X-ray image decomposition) $M = N = 2$, $L = 26896$, SNR=$\infty$ (Pure-source samples exist) | | |
|---|---|---|---|---|---|---|
| Methods | $\rho_{\text{ave}}$ | $T_{\text{ave}}$ (sec) | $\eta_{\text{ave}}$ | $\rho_{\text{ave}}$ | $T_{\text{ave}}$ (sec) | $\eta_{\text{ave}}$ |
| $n$LCA-IVM | $\simeq 1$ | 2.15 | 3.50 | 1 | 0.46 | 2 |
| $n$LCA-ES | 0.9999 | 0.01 | — | 1 | 0.06 | — |
| CAMNS | 0.9994 | 1.59 | — | 1 | 3.07 | — |
| $n$ICA | 0.5184 | 7.53 | 2000 | 0.6660 | 23.68 | 2000 |
| FastICA | 0.9221 | 0.52 | 9.40 | 0.3451 | 16.73 | 5.76 |
| NMF | 0.8654 | 4.89 | 2000 | 0.8201 | 22.15 | 2000 |
| QNMF | 0.8545 | 9.47 | 2000 | 0.9448 | 28.19 | 2000 |
| $s$NMF | 0.8841 | 10.11 | 2000 | 0.8044 | 48.10 | 2000 |

properly scaled and ordered for ease of illustrative comparison among all the BSS algorithms under test. One can see in this figure that the separated source spectra obtained from $n$LCA-IVM, $n$LCA-ES, and CAMNS match the original source spectra much better than those obtained by the other algorithms.



Fig. 5. Infrared spectra of acetone (left plot), dichloromethane (middle plot), and toluene (right plot): (a) the sources, (b) the observations and the extracted sources obtained by (c) $n$LCA-IVM, (d) $n$LCA-ES, (e) CAMNS, (f) $n$ICA, (g) FastICA, (h) NMF, (i) QNMF, and (j) $s$NMF.

## 5.3 Experiment 3: Dual-Energy X-Ray Image Decomposition

Accurate detection of lung nodules (the early sign of lung cancers) using dual-energy chest X-ray imaging is an important diagnostic task. However, the presence of ribs or clavicles overlapped with soft tissue presents significant challenge to detect subtle nodules. Effective separation of bone and soft tissues in dual-energy chest X-ray imaging is highly desirable. In this experiment, we applied the $n$LCA-IVM to process two mixed images of soft and bone tissues ($L = 26,896$) taken from [39] as shown in Fig. 6a. The values of $\rho_{\text{ave}}$ of all the BSS algorithms under test are also given in Table 5. It is to be mentioned that two pure-source samples exist in the original two sources. It can be seen from the table that $n$LCA-IVM, $n$LCA-ES, and CAMNS all perform perfectly (i.e., $\rho_{\text{ave}} = 1$), justifying the source identifiability of $n$LCA (as stated in Theorem 2) and the global optimum solution achieved by the proposed $n$LCA-IVM (see Proposition 1). On the other hand, the value of $T_{\text{ave}}$ (0.46 second) associated with the proposed $n$LCA-IVM is larger than that (0.06 second) associated with



Fig. 6. Dual energy X-ray images: (a) the sources, (b) the observations and the extracted sources obtained by (c) $n$LCA-IVM, (d) $n$LCA-ES, (e) CAMNS, (f) $n$ICA, (g) FastICA, (h) NMF, (i) QNMF, and (j) $s$NMF.

$n$LCA-ES, while much smaller than that associated with each of CAMNS, $n$ICA, FastICA, NMF, QNMF, and $s$NMF. One typical realization is also shown in Fig. 6.

The above experimental results using synthetic data indicate that $n$LCA-IVM, $n$LCA-ES, and CAMNS may attain the same performance (by $\rho_{ave}$) for the noise-free case as ($\mathcal{A}5$) is true, and show that $n$LCA-IVM and CAMNS perform better than $n$LCA-ES because the nonnegativity of the demixed sources is guaranteed by the former but not by the latter. Meanwhile, in contrast to $n$LCA-ES and CAMNS, the proposed $n$LCA-IVM is less sensitive to the existence of pure-source samples (i.e., ($\mathcal{A}5$)) because it is developed by volume maximization of the solid region constructed by the separated sources, neither relying on ($\mathcal{A}5$) nor involving search of data points contributed from pure-source samples. Moreover, both $n$LCA-IVM and CAMNS outperform $n$ICA and FastICA for which the source independence or uncorrelatedness assumption is required, but the sources are neither statistically independent nor uncorrelated in these experiments. The proposed $n$LCA-IVM not only outperforms NMF, QNMF, and $s$NMF (perhaps due to inherent local optimality issue of NMF-based algorithms) but also has lower computational complexity ($T_{ave}$).

## 5.4 Experiment 4: Analyzing Fluorescence Microscopy Signals

Fluorescence microscopy uses an optical sensor array (e.g., CCD camera) to produce multispectral images in which the targets of a specimen are labeled with different fluorescence probes [6]. In an attempt to dissect the multispectral images, the ability of identifying spectral biomarkers is limited due to the spectral-overlapped problem among the probes, leading to information leak-through from one spectral channel to another. To resolve such problems by separating the fluorescence microscopy into individual maps associated with specific biomarkers can be generally formulated as an $n$BSS problem. In the investigation of the specimen labeled by multiple fluorescent probes, two or more emissions are often overlapped in the measured images. This is called crosstalk, whose separation can be treated as an $n$BSS problem. In this experiment, we employed the $n$LCA-IVM algorithm to analyze a set of dividing newt lung cell images taken from http://publications.nigms.nih.gov/insidethecell/chapter1.html. In Fig. 7a, the observed images of intermediate filaments, spindle fibers, and chromosomes are displayed from top to bottom, respectively. Because of wide emission spectra of the fluorescence probe, the spectral biomarkers are indeed overlapped as shown in Fig. 7a. The regions of interest (ROI) of the observed images are shown in Fig. 7b, and the obtained results are shown in Fig. 7c, where the unmixed spectral images of intermediate filaments, chromosomes, and spindle fibers are visually much clearer than the ROIs of the observed images before unmixing. Specifically, the rope shape of the extracted chromosomes and the spindle shape of the extracted spindle fibers can be clearly observed, and these results exhibit a good agreement with biological expectation.

## 5.5 Experiment 5: Contrast Agent Perfusion Image Extraction

In DCE-MRI, various molecular weight contrast agents are used to assess tumor vascular permeability and quantify cellular and molecular abnormalities in blood vessel walls
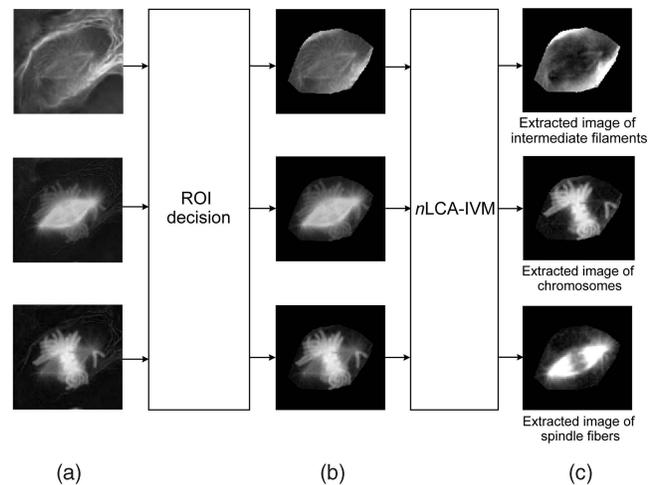


Fig. 7. Fluorescence microscopy images: (a) the measured newt lung cell images, (b) the ROI images of (a), and (c) the unmixed images of intermediate filaments (top), chromosomes (middle), and spindle fibers (bottom) obtained by the proposed $n$LCA-IVM.

[40]. The DCE-MRI can distinguish vascular heterogeneity and elucidate features that characterize angiogenic blood vessels from their normal counterparts, and has potential utility in measuring the efficacy of angiogenesis inhibitors in cancer treatment in vivo [4], [8]. Although DCE-MRI can provide a meaningful estimation of vascular permeability when a tumor is homogeneous, many malignant tumors show markedly heterogeneous areas of permeability, i.e., the observed DCE-MRI images are usually the mixtures of two permeability distributions associated with fast perfusion (tumor-induced blood vessels) and slow perfusion (normal blood vessels). In this experiment, we employed the proposed $n$LCA-IVM to separate the vascular permeability images associated with fast and slow perfusion in breast cancer study. A temporal image sequence containing 19 breast magnetic resonance images serves as the observations, as shown in Fig. 8a. The ROIs (i.e., the location of the breast cancer) were first determined by human-data interaction, which are shown in Fig. 8b. Then the PCA was employed to obtain the two noise-suppressed images displayed in Fig. 8c, which were then processed by the proposed $n$LCA-IVM. The two separated source images are shown in Fig. 8d, corresponding to the permeability images of fast perfusion (top) and slow perfusion (bottom), respectively. Note that the bright peripheral area in the extracted permeability image of fast perfusion indicates the distribution of angiogenesis of the breast cancer tumor, whereas the hypoxia in the inner core of the extracted permeability image of slow perfusion indicates the distribution of blood vessels of the normal tissue. Finally, the two $19 \times 1$ column vectors of the mixing matrix, called the time activity curves (TACs), as shown in Fig. 8e, were estimated using a nonnegative least-squares estimator [41] with the given observation matrix $\mathbf{X}$ and the source matrix estimate $\hat{\mathbf{S}}$ (see (43)). The descending curve (dashed line) represents the fast perfusion, showing the fast contrast agent washout in the breast tumor over the data acquisition time [8], and the ascending curve (solid line) denotes the slow perfusion, showing the contrast agent accumulation in the normal tissue. These results reveal the perfusion characteristics of different microvascular physiologies [4],
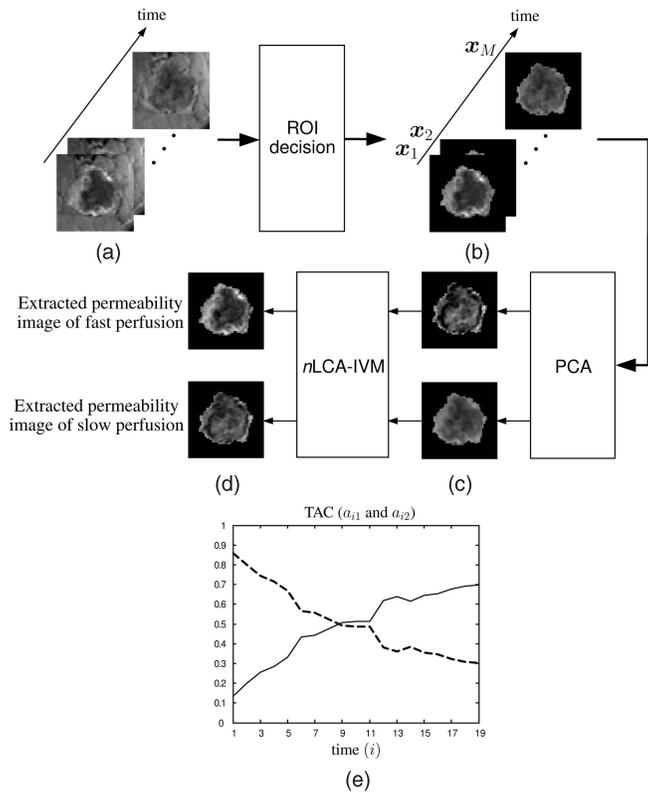
Fig. 8. DCE-MRI images where $M = 19$: (a) the measured breast MR images, (b) the ROI images of (a), (c) the outputs of PCA, (d) the extracted permeability images of fast perfusion (top) and slow perfusion (bottom), and (e) the associated TACs of fast perfusion (dashed line) and slow perfusion (solid line) obtained by the proposed $n$LCA-IVM.

as well as $a_{i1} > 0$, $a_{i2} > 0$, and $a_{i1} + a_{i2} \simeq 1$ for all $i$, which are consistent with ($\mathcal{A}$2) and ($\mathcal{A}$4).

## 6   CONCLUSIONS

We have presented a new $n$LCA method for the blind separation of nonnegative and potentially correlated sources with a given set of mixed observations. The optimum unmixing matrix is obtained by minimizing the joint correlation of all the separated source signals subject to the nonnegativity constraint. A closed-form solution for the unmixing matrix of the two-source case was provided. The $n$LCA method is implemented by an iterative volume maximization algorithm (i.e., $n$LCA-IVM) that uses computationally efficient LP solvers. We also proved the source identifiability of the $n$LCA method under the existence of pure-source samples (i.e., ($\mathcal{A}$5)), and furthermore, it never relies on ($\mathcal{A}$5) and thus exhibits good performance even when ($\mathcal{A}$5) is not satisfied perfectly. Comparative experimental results using simulated data were presented to demonstrate the superior performance and computational efficiency of the $n$LCA-IVM over several existing benchmark methods. Evaluation of the proposed $n$LCA-IVM using real biomedical data was also presented, which provides some insights into the underlying biomedical relevant characteristics and physiologies. Due to a variety of potential biomedical imaging applications, nontrivial modifications of this algorithm may be needed for specific applications, and they are left as future research along with other applications, where $n$BSS is needed.

## APPENDIX A

## PROOF OF THEOREM 1

1. By (3), ($\mathcal{A}$2), and ($\mathcal{A}$4), for any $y \in \mathrm{conv}\{x_1, \ldots, x_M\}$,

$$
\begin{aligned}
y &= \sum_{i=1}^{M} \alpha_i x_i, \quad \left(\Sigma_{i=1}^{M}\alpha_i = 1, \alpha_i \geq 0\right) \\
&= \sum_{i=1}^{M} \alpha_i \sum_{j=1}^{N} a_{ij} s_j, \quad (\text{by } (3)),
\end{aligned}
\tag{47}
$$

where $a_{ij} \geq 0$ and $\Sigma_{j=1}^{N} a_{ij} = 1$. By letting $\gamma_j = \Sigma_{i=1}^{M} \alpha_i a_{ij}$, (47) becomes

$$
y = \sum_{j=1}^{N} \gamma_j s_j \in \mathrm{conv}\{s_1, s_2, \ldots, s_N\}
\tag{48}
$$

because $\gamma_j \geq 0$ and $\Sigma_{j=1}^{N} \gamma_j = 1$. Thus, we can conclude that $\mathrm{conv}\{x_1, x_2, \ldots, x_M\} \subseteq \mathrm{conv}\{s_1, s_2, \ldots, s_N\}$.

2. Again, by (3), ($\mathcal{A}$2), and ($\mathcal{A}$4), for any $y \in \mathrm{sol}\{x_1, x_2, \ldots, x_M\}$, $y$ given by (47) still holds except for $\Sigma_{i=1}^{M}\alpha_i \leq 1$, $\alpha_i \geq 0$, and so

$$
y = \sum_{j=1}^{N} \gamma_j s_j \in \mathrm{sol}\{s_1, s_2, \ldots, s_N\}
\tag{49}
$$

because $\gamma_j \geq 0$ and $\Sigma_{j=1}^{N} \gamma_j \leq 1$. Thus, $\mathrm{sol}\{x_1, x_2, \ldots, x_M\} \subseteq \mathrm{sol}\{s_1, s_2, \ldots, s_N\}$.

## APPENDIX B

## PROOF OF LEMMA 1

By Hadamard's inequality [36],

$$
|\det(\mathbf{A})| \leq \prod_{i=1}^{N} \left( \sum_{j=1}^{N} |a_{ij}|^2 \right)^{1/2},
\tag{50}
$$

and the equality holds if and only if all the rows of $\mathbf{A}$ are orthogonal. By the inequality (see [42, Theorem 2.33]),

$$
\left( \sum_{j=1}^{N} b_j^2 \right)^{1/2} \leq \sum_{j=1}^{N} |b_j|,
\tag{51}
$$

and the equality holds if and only if $\mathbf{b} = (b_1, b_2, \ldots, b_N)^T = b_l \mathbf{e}_l$ for some $l$ and $b_l \neq 0$. By ($\mathcal{A}$2)-($\mathcal{A}$4), (50), and (51), we have

$$
|\det(\mathbf{A})| \leq \prod_{i=1}^{N} \left( \sum_{j=1}^{N} a_{ij}^2 \right)^{1/2} \leq \prod_{i=1}^{N} \left( \sum_{j=1}^{N} a_{ij} \right) = 1,
\tag{52}
$$

where the first equality holds if and only if all the rows of $\mathbf{A} = (a_1, \ldots, a_N)^T$ are orthogonal, and the second equality holds if and only if $a_i = \mathbf{e}_l$ (since $\sum_{j=1}^{N} a_{ij} = 1$) for some $l$. Thus, it can be easily inferred that $|\det(\mathbf{A})| = 1$ if and only if $\mathbf{A}$ is a permutation matrix.

## REFERENCES

[1] N. Keshava and J. Mustard, "Spectral Unmixing," *IEEE Signal Processing Magazine,* vol. 19, no. 1, pp. 44-57, Jan. 2002.

[2] M. Mjolsness and D. DeCoste, "Machine Learning for Science: State of the Art and Future Prospects," *Science,* vol. 293, pp. 2051-2055, 2001.

[3] H.R. Herschman, "Molecular Imaging: Looking at Problems, Seeing Solutions," *Science,* vol. 302, pp. 605-608, 2003.

[4] D.M. McDonald and P.L. Choyke, "Imaging of Angiogenesis: From Microscope to Clinic," *Nature Medicine,* vol. 9, no. 6, pp. 713-725, 2003.

[5] M.E. Dickinson, G. Bearman, S. Tilie, R. Lansford, and S.E. Fraser, "Multi-Spectral Imaging and Linear Unmixing Add a Whole New Dimension to Laser Scanning Fluorescence Microscopy," *Biotechniques,* vol. 31, no. 6, pp. 1272-1278, Dec. 2001.

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

[7] A. Rabinovich, S. Agarwal, S. Krajewski, J.C. Reed, J.H. Price, and S. Belongie, "Accuracy of Unsupervised Spectral Decomposition for Densitometry of Histological Sections," *IEEE Trans. Medical Imaging,* to appear.

[8] Y. Wang, J. Xuan, R. Srikanchana, and P.L. Choyke, "Modeling and Reconstruction of Mixed Functional and Molecular Patterns," *Int'l J. Biomedical Imaging,* vol. 2006, pp. 1-9, 2006.

[9] D. Nuzillard and A. Bijaoui, "Blind Source Separation and Analysis of Multispectral Astronomical Images," *Astronomy and Astrophysics Supplement Series,* vol. 147, pp. 129-138, 2000.

[10] 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, 2006.

[11] S. Haykin, *Neural Networks: A Comprehensive Foundation.* second ed. Prentice-Hall, 2005.

[12] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis.* John Wiley, 2001.

[13] E. Oja and M. Plumbley, "Blind Separation of Positive Sources by Globally Convergent Gradient Search," *Neural Computation,* vol. 16, pp. 1811-1825, 2004.

[14] D. Lee and H.S. Seung, "Learning the Parts of Objects by Non-Negative Matrix Factorization," *Nature,* vol. 401, pp. 788-791, Oct. 1999.

[15] J.S. Lee, D.D. Lee, S. Choi, K.S. Park, and D.S. Lee, "Non-Negative Matrix Factorization of Dynamic Images in Nuclear Medicine," *IEEE Nuclear Science Symp. Conf. Record,* vol. 4, pp. 2027-2030, 2001.

[16] 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, 2007.

[17] W. Liu, N. Zheng, and X. Lu, "Nonnegative Matrix Factorization for Visual Coding," *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing,* pp. 293-296, Apr. 2003.

[18] R. Zdunek and A. Cichocki, "Nonnegative Matrix Factorization with Constrained Second-Order Optimization," *Signal Processing,* vol. 87, no. 8, pp. 1904-1916, 2007.

[19] P. Hoyer, "Nonnegative Matrix Factorization with Sparseness Constraints," *J. Machine Learning Research,* vol. 5, pp. 1457-1469, 2004.

[20] H. Laurberg, M.G. Christensen, M.D. Plumbley, L.K. Hansen, and S.H. Jensen, "Theorems on Positive Data: On the Uniqueness of NMF," *Computational Intelligence and Neuroscience,* vol. 2008, pp. 1-9, 2008.

[21] J. Mansfield, K. Gossage, C. Hoyt, and R. Levenson, "Autofluorescence Removal, Multiplexing, and Automated Analysis Methods for *In-Vivo* Fluorescence Imaging," *J. Biomedical Optics,* vol. 10, no. 4, pp. 1-9, 2005.

[22] 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.

[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," *Proc. IEEE Int'l Workshop Machine Learning for Signal Processing,* pp. 73-78, Sept. 2006.

[24] M.E. Winter, "N-Findr: An Algorithm for Fast Autonomous Spectral End-Member Determination in Hyperspectral Data," *Proc. SPIE Conf. Imaging Spectrometry,* pp. 266-275, Oct. 1999.

[25] 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.

[26] C.I. Chang, C.-C. Wu, W.-M. Liu, and Y.-C. Quyang, "A New Growing Method for Simplex-Based Endmember Extraction Algorithm," *IEEE Trans. Geoscience and Remote Sensing,* vol. 44, no. 10, pp. 2804-2819, Oct. 2006.

[27] A. Prieto, C.G. Puntonet, and B. Prieto, "A Neural Learning Algorithm for Blind Separation of Sources Based on Geometric Properties," *Signal Processing,* vol. 64, pp. 315-331, 1998.

[28] P. Santago and H.D. Gage, "Statistical Models of Partial Volume Effect," *IEEE Trans. Image Processing,* vol. 4, no. 11, pp. 1531-1540, Nov. 1995.

[29] D. Stein, S. Beaven, L. Hoff, E. Winter, A. Schaum, and A. Stocker, "Anomaly Detection from Hyperspectral Imagery," *IEEE Signal Processing Magazine,* vol. 19, no. 1, pp. 58-69, Jan. 2002.

[30] M.S. Bazaraa, H.D. Sherall, and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms,* second ed. John Wiley, 1993.

[31] S. Lang, *Undergraduate Analysis,* second ed. Springer-Verlag, 1983.

[32] S.H. Friedberg, A.J. Insel, and L.E. Spence, *Linear Algebra,* third ed. Prentice-Hall, 1997.

[33] I.J. Lustig, R.E. Marsten, and D.F. Shanno, "Interior Point Methods for Linear Programming: Computational State of the Art," *ORSA J. Computing,* vol. 6, no. 1, pp. 1-14, 1994.

[34] C.B. Barber, D.P. Dobkin, and H. Huhdanpaa, "The Quickhull Algorithm for Convex Hulls," *ACM Trans. Math. Software,* vol. 22, pp. 469-483, Dec. 1996.

[35] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning.* Springer, 2001.

[36] R.A. Horn and C.R. Johnson, *Matrix Analysis.* Cambridge Univ. Press, 1985.

[37] 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.

[38] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing.* John Wiley, 2002.

[39] K. Suzuki, R. Engelmann, H. MacMahon, and K. Doi, "Virtual Dual-Energy Radiography: Improved Chest Radiographs by Means of Rib Suppression Based on a Massive Training Artificial Neural Network (Mtann)," *Radiology,* vol. 238, http://suzukilab.uchicago.edu/research.htm, 2006.

[40] A.R. Padhani and J.E. Husband, "Dynamic Contrast-Enhanced MRI Studies in Oncology with an Emphasis on Quantification, Validation and Human Studies," *Clinical Radiology,* vol. 56, no. 8, pp. 607-620, 2001.

[41] T.F. Coleman and Y. Li, "A Reflective Newton Method for Minimizing a Quadratic Function Subject to Bounds on Some of the Variables," *SIAM J. Optimization,* vol. 6, no. 4, pp. 1040-1058, 1996.

[42] C.-Y. Chi, C.-C. Feng, C.-H. Chen, and C.-Y. Chen, *Blind Equalization and System Identification.* Springer-Verlag, 2006.

**Fa-Yu Wang** received the BS and MS degrees in power mechanical engineering and electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 1993 and 1999, respectively. From 1996 to 2005, he was with Chunghwa Telecom. Currently, he is working toward the PhD degree at the Institute of Communications Engineering, National Tsing Hua University. His research interests focus on signal processing, pattern analysis, and machine intelligence.

**Chong-Yung Chi** received the PhD degree in electrical engineering from the University of Southern California in 1983. From 1983 to 1988, he was with the Jet Propulsion Laboratory, Pasadena, California. He has been a professor in the Department of Electrical Engineering since 1989 and the Institute of Communications Engineering (ICE) since 1999 (also the Chairman of ICE for 2002-2005) at National Tsing Hua University, Hsinchu, Taiwan. Currently, he is an associate editor for the *IEEE Signal Processing Letters* and the *IEEE Transactions on Circuits and Systems I*, and a member of the IEEE Signal Processing Committee on Signal Processing Theory and Methods. He coauthored a technical book, *Blind Equalization and System Identification* (Springer, 2006) and has published more than 150 technical papers. His research interests include signal processing for wireless communications, convex analysis and optimization for blind source separation, biomedical imaging, and hyperspectral imaging. He is a senior member of the IEEE.

**Tsung-Han Chan** received the BS degree from the Department of Electrical Engineering, Yuan Ze University, Taoyuan, Taiwan, in 2004 and the PhD degree from the Institute of Communications Engineering, National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2009. Currently, he is a postdoctoral research fellow with the Institute of Communications Engineering, NTHU. In 2008, he was a visiting doctoral graduate research assistant with Virginia Polytechnic Institute and State University, Arlington. His research interests are in signal processing, convex optimization, and pattern analysis, with a recent emphasis on dynamic medical imaging and remote sensing applications.

**Yue Wang** received the BS and MS degrees in electrical and computer engineering from Shanghai Jiao Tong University in 1984 and 1987, respectively, and the PhD degree in electrical engineering from the University of Maryland Graduate School in 1995. In 1996, he was a postdoctoral fellow at Georgetown University School of Medicine. From 1996 to 2003, he was an assistant and later an associate professor of electrical engineering at the Catholic University of America. In 2003, he joined Virginia Polytechnic Institute and State University and is currently the endowed Grant A. Dove Professor of electrical and computer engineering. He became an elected fellow of the American Institute for Medical and Biological Engineering (AIMBE) and was named an ISI highly cited researcher by Thomson Scientific in 2004. His research interests focus on statistical pattern recognition, machine learning, signal and image processing, with applications to computational bioinformatics and biomedical imaging.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.