

A Convex Analysis Framework for Blind Separation of Non-Negative Sources

Tsung-Han Chan[†], Wing-Kin Ma^{*}, Chong-Yung Chi[‡], and Yue Wang[§]

Abstract

This paper presents a new framework for blind source separation (BSS) of non-negative source signals. The proposed framework, referred herein to as convex analysis of mixtures of non-negative sources (CAMNS), is deterministic requiring no source independence assumption, the entrenched premise in many existing (usually statistical) BSS frameworks. The development is based on a special assumption called local dominance. It is a good assumption for source signals exhibiting sparsity or high contrast, and thus is considered realistic to many real-world problems such as multichannel biomedical imaging. Under local dominance and several standard assumptions, we apply convex analysis to establish a new BSS criterion, which states that the source signals can be perfectly identified (in a blind fashion) by finding the extreme points of an observation-constructed polyhedral set. Methods for fulfilling the CAMNS criterion are also derived, using either linear programming or simplex geometry. Simulation results on several data sets are presented to demonstrate the efficacy of the proposed method over several other reported BSS methods.

Index Terms

Blind separation, Non-negative sources, Convex analysis criterion, Linear program, Simplex geometry, Convex optimization

EDICS: SSP-SSEP (Signal separation), SAM-APPL (Applications of sensor & array multichannel processing), BIO-APPL (Applications of biomedical signal processing)

This work was supported in part by the National Science Council (R.O.C.) under Grants NSC 95-2221-E-007-122 and NSC 95-2221-E-007-036, and by the US National Institutes of Health under Grants EB000830 and CA109872. This work was partly presented at the 32th IEEE International Conference on Acoustics, Speech, and Signal Processing, Honolulu, Hawaii, USA, April 15-20, 2007.

[†]Tsung-Han Chan is with Institute of Communications Engineering & Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, R.O.C. E-mail: d935608@oz.nthu.edu.tw, Tel: +886-3-5715131X34033, Fax: +886-3-5751787.

^{*}Wing-Kin Ma is the corresponding author. Address: Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. E-mail: wkma@iee.org, Tel: +852-31634350, Fax: +852-26035558.

[‡]Chong-Yung Chi is with Institute of Communications Engineering & Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, R.O.C. E-mail: cychi@ee.nthu.edu.tw, Tel: +886-3-5731156, Fax: +886-3-5751787.

[§]Yue Wang is with Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Arlington, VA 22203, USA E-mail: yuewang@vt.edu, Tel: 703-387-6056, Fax: 703-528-5543.

I. INTRODUCTION

Recently there has been much interest in blind separation of non-negative source signals [1], [2], referred herein to as non-negative blind source separation (nBSS). There are many applications where the sources to be separated are non-negative by nature; for example, in analytical chemistry [3], [4], hyperspectral imaging [5], and biomedical imaging [6]. How to cleverly exploit the non-negative signal characteristic in nBSS has been an intriguing subject, leading to numerous nBSS alternatives being proposed [7]–[11].

One major class of nBSS methods utilizes the statistical property that the sources are mutually uncorrelated or independent, supposing that the sources do satisfy that property. Methods falling in this class include second-order blind identification (SOBI) [12], fast fixed-point algorithm for independent component analysis (ICA) [13], non-negative ICA (nICA) [7], stochastic non-negative ICA (SNICA) [8], and Bayesian positive source separation (BPSS) [9], to name a few. SOBI and fast ICA were originally developed for more general blind source separation (BSS) problems where signals can be negative. The two methods can be directly applied to nBSS, but it has been reported that their separated signals may have negative values especially in the presence of finite sample effects [4], [14]. nICA takes source non-negativity into account, and is shown to provide perfect separation when the sources have non-vanishing density around zero (which is also called the well-grounded condition). SNICA uses a simulated annealing algorithm for extracting non-negative sources under the minimum mutual information criterion. BPSS also uses source non-negativity. It applies Bayesian estimation with both the sources and mixing matrix being assigned Gamma distribution priors.

Another class of nBSS methods is deterministic requiring no assumption on source independence or zero correlations. Roughly speaking, these methods explicitly exploit source non-negativity or even mixing matrix non-negativity, with an attempt to achieve some kind of least square criterion. Alternating least squares (ALS) [10], [15] deals with a sequence of least squares problems where non-negativity constraints on either the sources or mixing matrix are imposed. Non-negative matrix factorization (NMF) [11] decomposes the observation matrix as a product of two non-negative matrices, one serving as the estimate of the sources while another the mixing matrix. NMF is not a unique decomposition, which may result in indeterminacy of the sources and mixing matrix. To overcome this problem, a sparse constraint on the sources has been proposed [16].

In this paper we propose a new nBSS framework, known as convex analysis of mixtures of non-negative sources (CAMNS). Convex analysis and optimization techniques have drawn considerable attention in

signal processing, serving as powerful tools for various topics such as communications [17]–[22], array signal processing [23], and sensor networks [24]. Apart from using source non-negativity, CAMNS adopts a special deterministic assumption called *local dominance*. This assumption was initially proposed to capture the sparse characteristics of biomedical images [25], [26], but we found it a good assumption or approximation for high contrast images such as human portraits, as well. Under the local dominant assumption and some standard nBSS assumptions, we show using convex analysis that the true source signals serve as the extreme points of some observation-constructed polyhedral set. This geometrical discovery is not only surprising but important, since it provides a novel nBSS criterion that guarantees perfect blind separation. To practically realize CAMNS, we derive extreme-point finding algorithms using either linear programming or simplex geometry. As we will see by simulations, the blind separation performance of the CAMNS algorithms is promising even in the presence of strongly correlated sources.

We should stress that the proposed CAMNS framework is deterministic, but it is conceptually different from the other existing deterministic frameworks such as NMF. In particular, the idea of using convex analysis to establish nBSS criterion cannot be found in the other frameworks. Moreover, CAMNS does not require mixing matrix non-negativity while NMF requires. The closest match to CAMNS would be non-negative least correlated component analysis (nLCA) [25], [26], a concurrent development that also exploits the local dominance assumption. Nevertheless, convex analysis is not involved in nLCA.

CAMNS exploits the signal geometry arising from local dominance, using convex analysis and optimization. Such an exploration is reminiscent of that in BSS of magnitude bounded sources (BSS-MBS) [27]–[29], the latter of which utilizes signal geometry due to signal boundedness. In [27], Prieto *et al.* use a neural learning algorithm to determine the vertices of an observation space generated by MBS. Erdogan [28] proposes an infinity norm optimization problem in his endeavor of BSS-MBS, and shows that the optimal solution of that problem can result in perfect signal separation. In [29], Vrins *et al.* develop a BSS-MBS method by maximizing a range-based contrast. The range-based contrast is shown to be *discriminant*, i.e., every local maxima leads to perfect separation of one source.

The paper is organized as follows. In Section II, the problem statement is given. In Section III, we review some key concepts of convex analysis, which would be useful for understanding of the mathematical derivations that follow. The new BSS criterion for separating non-negative sources is developed in Section IV. We show how to use linear programs (LPs) to practically achieve the new criterion in Section V. Section VI studies a geometric alternative to the LP method. Finally, in Section VII, we use simulations to evaluate the performance of the proposed methods as well as some other existing nBSS methods.

II. PROBLEM STATEMENT AND ASSUMPTIONS

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 -vectors, $M \times N$ matrices
$\mathbb{R}_+, \mathbb{R}_+^N, \mathbb{R}_+^{M \times N}$	Set of non-negative real numbers, N -vectors, $M \times N$ matrices
$\mathbf{1}$	All one vector
\mathbf{I}_N	$N \times N$ identity matrix
\mathbf{e}_i	Unit vector with the i th entry being equal to 1
\succeq	Componentwise inequality
$\ \cdot\ $	Euclidean norm
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$

The scenario under consideration is that of linear instantaneous mixtures. The signal model is

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n], \quad n = 1, \dots, L \quad (1)$$

where $\mathbf{s}[n] = [s_1[n], \dots, s_N[n]]^T$ is the input or source vector sequence with N denoting the input dimension, $\mathbf{x}[n] = [x_1[n], \dots, x_M[n]]^T$ is the output or observation vector sequence with M denoting the output dimension, $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the mixing matrix describing the input-output relation, and L is the sequence (or data) length and we assume $L \gg \max\{M, N\}$. Note that (1) can be rewritten as

$$\mathbf{x}_i = \sum_{j=1}^N a_{ij} \mathbf{s}_j, \quad i = 1, \dots, M, \quad (2)$$

where a_{ij} is the (i, j) th element of \mathbf{A} , $\mathbf{s}_j = [s_j[1], \dots, s_j[L]]^T$ is a vector representing the j th source signal and $\mathbf{x}_i = [x_i[1], \dots, x_i[L]]^T$ is a vector representing the i th observed signal.

In BSS, the problem is to extract $\mathbf{s}[n]$ without information of \mathbf{A} . The BSS framework to be proposed is based on the following assumptions:

- (A1) All \mathbf{s}_j are componentwise non-negative; i.e., for each j , $\mathbf{s}_j \in \mathbb{R}_+^L$.
- (A2) Each source signal vector is *local dominant*, in the following sense: For each $i \in \{1, \dots, N\}$, there exists an (unknown) index ℓ_i such that $s_i[\ell_i] > 0$ and $s_j[\ell_i] = 0, \forall j \neq i$. (This means that for each source there is at least one n at which the source dominates.)
- (A3) The mixing matrix has unit row sum; i.e., for all $i = 1, \dots, M$,

$$\sum_{j=1}^N a_{ij} = 1. \quad (3)$$

- (A4) $M \geq N$ and \mathbf{A} is of full column rank.

Let us discuss the practicality of (A1) – (A4). Assumption (A1) is true in image analysis [5], [6] where image intensities are often represented by non-negative numbers. Assumption (A2) is special and instrumental to the development that ensues. It may be completely satisfied or serve as a good approximation when the source signals are sparse (or contain many zeros). In brain magnetic resonance imaging (MRI), for instance, the non-overlapping region of the spatial distribution of a fast perfusion and a slow perfusion source images [6] can be higher than 95%. It may also be appropriate to assume (A2) when the source signals exhibit high contrast. Assumption (A4) is a standard assumption in BSS. Assumption (A3) is automatically satisfied in MRI due to the partial volume effect [26], and in hyperspectral images due to the full additivity condition [5]. When (A3) is not satisfied, the following idea [26] can be used. Suppose that $\mathbf{x}_i^T \mathbf{1} \neq 0$ and $\mathbf{s}_j^T \mathbf{1} \neq 0$, $\forall i, j$ and consider the following normalized observation vectors:

$$\bar{\mathbf{x}}_i = \frac{\mathbf{x}_i}{\mathbf{x}_i^T \mathbf{1}} = \sum_{j=1}^N \left(\frac{a_{ij} \mathbf{s}_j^T \mathbf{1}}{\mathbf{x}_i^T \mathbf{1}} \right) \left(\frac{\mathbf{s}_j}{\mathbf{s}_j^T \mathbf{1}} \right). \quad (4)$$

By letting $\bar{a}_{ij} = a_{ij} \mathbf{s}_j^T \mathbf{1} / \mathbf{x}_i^T \mathbf{1}$ and $\bar{\mathbf{s}}_j = \mathbf{s}_j / \mathbf{s}_j^T \mathbf{1}$, we obtain a BSS problem formulation $\bar{\mathbf{x}}_i = \sum_{j=1}^N \bar{a}_{ij} \bar{\mathbf{s}}_j$ which is in the same form as the original signal model in (2). It is easy to show that the new mixing matrix, denoted by $\bar{\mathbf{A}}$, has unit row sum. In addition, the rank of $\bar{\mathbf{A}}$ is the same as that of \mathbf{A} . To show this, we notice that

$$\bar{\mathbf{A}} = \mathbf{D}_1^{-1} \mathbf{A} \mathbf{D}_2 \quad (5)$$

where $\mathbf{D}_1 = \text{diag}(\mathbf{x}_1^T \mathbf{1}, \dots, \mathbf{x}_M^T \mathbf{1})$ and $\mathbf{D}_2 = \text{diag}(\mathbf{s}_1^T \mathbf{1}, \dots, \mathbf{s}_N^T \mathbf{1})$. Since \mathbf{D}_1 and \mathbf{D}_2 are of full rank, we have $\text{rank}(\bar{\mathbf{A}}) = \text{rank}(\mathbf{A})$.

III. SOME BASIC CONCEPTS OF CONVEX ANALYSIS

We review some convex analysis concepts that will play an important role in the ensuing development. For detailed explanations of these concepts, readers are referred to [30], [31], [32] which are excellent literatures in convex analysis.

Given a set of vectors $\{\mathbf{s}_1, \dots, \mathbf{s}_N\} \subset \mathbb{R}^L$, the *affine hull* is defined as

$$\text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\} = \left\{ \mathbf{x} = \sum_{i=1}^N \theta_i \mathbf{s}_i \mid \boldsymbol{\theta} \in \mathbb{R}^N, \sum_{i=1}^N \theta_i = 1 \right\}. \quad (6)$$

An affine hull can always be represented by an affine set:

$$\text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\} = \{ \mathbf{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \mid \boldsymbol{\alpha} \in \mathbb{R}^P \} \quad (7)$$

for some (non-unique) $\mathbf{d} \in \mathbb{R}^L$ and $\mathbf{C} \in \mathbb{R}^{L \times P}$. Here, \mathbf{C} is assumed to be of full column rank and P is the affine dimension which must be less than N . For example, if $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ is linearly independent,

then $P = N - 1$. In that case, a legitimate (\mathbf{C}, \mathbf{d}) is given by $\mathbf{C} = [\mathbf{s}_1 - \mathbf{s}_N, \mathbf{s}_2 - \mathbf{s}_N, \dots, \mathbf{s}_{N-1} - \mathbf{s}_N]$ and $\mathbf{d} = \mathbf{s}_N$. To give some insights, Fig. 1 pictorially illustrates an affine hull for $N = 3$. We can see that it is a plane passing through $\mathbf{s}_1, \mathbf{s}_2$, and \mathbf{s}_3 .

Given a set of vectors $\{\mathbf{s}_1, \dots, \mathbf{s}_N\} \subset \mathbb{R}^L$, the *convex hull* is defined as

$$\text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_N\} = \left\{ \mathbf{x} = \sum_{i=1}^N \theta_i \mathbf{s}_i \mid \boldsymbol{\theta} \in \mathbb{R}_+^N, \sum_{i=1}^N \theta_i = 1 \right\}. \quad (8)$$

For $N = 3$, a convex hull is a triangle with vertices $\mathbf{s}_1, \mathbf{s}_2$, and \mathbf{s}_3 . Geometrically $\mathbf{s}_1, \dots, \mathbf{s}_N$ would be the ‘corner points’ of its convex hull, defined formally as the *extreme points*. A point $\mathbf{x} \in \text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ is an extreme point of $\text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ if it cannot be a nontrivial convex combination of $\mathbf{s}_1, \dots, \mathbf{s}_N$; i.e.,

$$\mathbf{x} \neq \sum_{i=1}^N \theta_i \mathbf{s}_i \quad (9)$$

for all $\boldsymbol{\theta} \in \mathbb{R}_+^N$, $\sum_{i=1}^N \theta_i = 1$, and $\boldsymbol{\theta} \neq \mathbf{e}_i$ for any i . The set of extreme points of $\text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ must be either the full set or a subset of $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$. In addition, if $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ is an affinely independent set (or $\{\mathbf{s}_1 - \mathbf{s}_N, \dots, \mathbf{s}_{N-1} - \mathbf{s}_N\}$ is a linearly independent set), then the set of extreme points of $\text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ is exactly $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$. Moreover, the boundary of $\text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ is entirely constituted by all its *faces*, defined by $\text{conv}\{\mathbf{s}_{i_1}, \dots, \mathbf{s}_{i_K}\}$ where $\{i_1, \dots, i_K\} \subset \{1, \dots, N\}$, $i_k \neq i_l$ for $k \neq l$, and $K < N$. A *facet* is a face with $K = N - 1$. For example, in Fig. 1, we can see that the facets are the line segments $\text{conv}\{\mathbf{s}_1, \mathbf{s}_2\}$, $\text{conv}\{\mathbf{s}_2, \mathbf{s}_3\}$ and $\text{conv}\{\mathbf{s}_1, \mathbf{s}_3\}$.

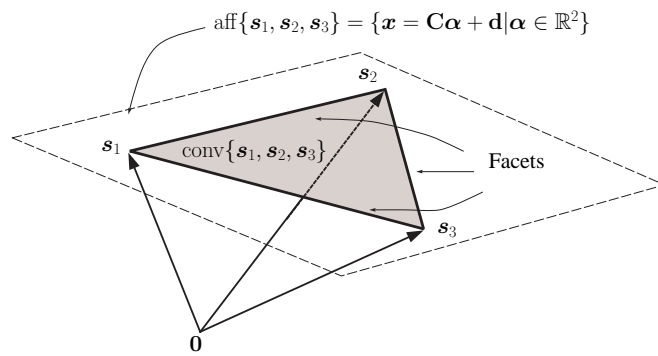


Fig. 1. Example of 3-dimensional signal space geometry for $N = 3$.

A *simplest simplex* of affine dimension $N - 1$, or $(N - 1)$ -simplex is defined as the convex hull of N affinely independent vectors $\{\mathbf{s}_1, \dots, \mathbf{s}_N\} \subset \mathbb{R}^{N-1}$. An $(N - 1)$ -simplex is formed by N extreme points,

N facets, and a number of faces. The family of these simplest simplexes has the vertex-descriptions including a point, line segment, triangle for $N = 1, 2$, and 3, respectively.

IV. NEW NBSS CRITERION BY CONVEX ANALYSIS

Now, consider the BSS problem formulation in (2) and the assumptions in (A1)-(A4). Under (A3), we see that every observation \mathbf{x}_i is an affine combination of the true source signals $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$; that is

$$\mathbf{x}_i \in \text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\} \quad (10)$$

for all $i = 1, \dots, M$. This leads to an interesting question whether the observations $\mathbf{x}_1, \dots, \mathbf{x}_M$ provide sufficient information to construct the source signal affine hull $\text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$. This is indeed possible, as described in the following lemma:

Lemma 1. *Under (A3) and (A4), we have*

$$\text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\} = \text{aff}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}. \quad (11)$$

The proof of Lemma 1 is given in Appendix A. Figure 2(a) demonstrates geometrically the validity of Lemma 1, for the special case of $N = 2$ where $\text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ is a line. Now let us focus on the characterization of $\text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$. It can easily be shown from (A2) that $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ is linearly independent. Hence, $\text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ has dimension $N - 1$ and admits a representation

$$\text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\} = \{ \mathbf{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \mid \boldsymbol{\alpha} \in \mathbb{R}^{N-1} \} \quad (12)$$

for some $(\mathbf{C}, \mathbf{d}) \in \mathbb{R}^{L \times (N-1)} \times \mathbb{R}^L$ such that $\text{rank}(\mathbf{C}) = N - 1$. Note that (\mathbf{C}, \mathbf{d}) is non-unique. Without loss of generality, we can restrict \mathbf{C} to a semi-orthogonal matrix, i.e., $\mathbf{C}^T \mathbf{C} = \mathbf{I}$. If $M = N$, it is easy to obtain (\mathbf{C}, \mathbf{d}) from the observations $\mathbf{x}_1, \dots, \mathbf{x}_M$; see the review in Section III. For the more general case of $M \geq N$, (\mathbf{C}, \mathbf{d}) may be found by solving the following minimization problem

$$(\mathbf{C}, \mathbf{d}) = \arg \min_{\substack{\tilde{\mathbf{C}}, \tilde{\mathbf{d}} \\ \tilde{\mathbf{C}}^T \tilde{\mathbf{C}} = \mathbf{I}}} \sum_{i=1}^M e_{\mathcal{A}(\tilde{\mathbf{C}}, \tilde{\mathbf{d}})}(\mathbf{x}_i) \quad (13)$$

where $e_{\mathcal{A}}(\mathbf{x})$ is the projection error of \mathbf{x} onto \mathcal{A} , defined as

$$e_{\mathcal{A}}(\mathbf{x}) = \min_{\tilde{\mathbf{x}} \in \mathcal{A}} \|\mathbf{x} - \tilde{\mathbf{x}}\|_2^2, \quad (14)$$

and

$$\mathcal{A}(\tilde{\mathbf{C}}, \tilde{\mathbf{d}}) = \{ \tilde{\mathbf{x}} = \tilde{\mathbf{C}}\boldsymbol{\alpha} + \tilde{\mathbf{d}} \mid \boldsymbol{\alpha} \in \mathbb{R}^{N-1} \} \quad (15)$$

is an affine set parameterized by $(\tilde{\mathbf{C}}, \tilde{\mathbf{d}})$. The objective of (13) is to find an $(N - 1)$ -dimensional affine set that has the minimum projection error with respect to the observations. Problem (13) can be solved analytically as stated in the following proposition:

Proposition 1. *The affine set fitting problem in (13) has a closed-form solution*

$$\mathbf{d} = \frac{1}{M} \sum_{i=1}^M \mathbf{x}_i \quad (16)$$

$$\mathbf{C} = [\mathbf{q}_1(\mathbf{U}\mathbf{U}^T), \mathbf{q}_2(\mathbf{U}\mathbf{U}^T), \dots, \mathbf{q}_{N-1}(\mathbf{U}\mathbf{U}^T)] \quad (17)$$

where $\mathbf{U} = [\mathbf{x}_1 - \mathbf{d}, \dots, \mathbf{x}_M - \mathbf{d}] \in \mathbb{R}^{L \times M}$, and the notation $\mathbf{q}_i(\mathbf{R})$ denotes the eigenvector associated with the i th principal eigenvalue of the input matrix \mathbf{R} .

The proof of Proposition 1 is given in Appendix B. We should stress that this affine set fitting provides a best affine set in terms of minimizing the projection error. Hence, in the presence of noisy data, it has an additional advantage of noise mitigation for $M > N$.

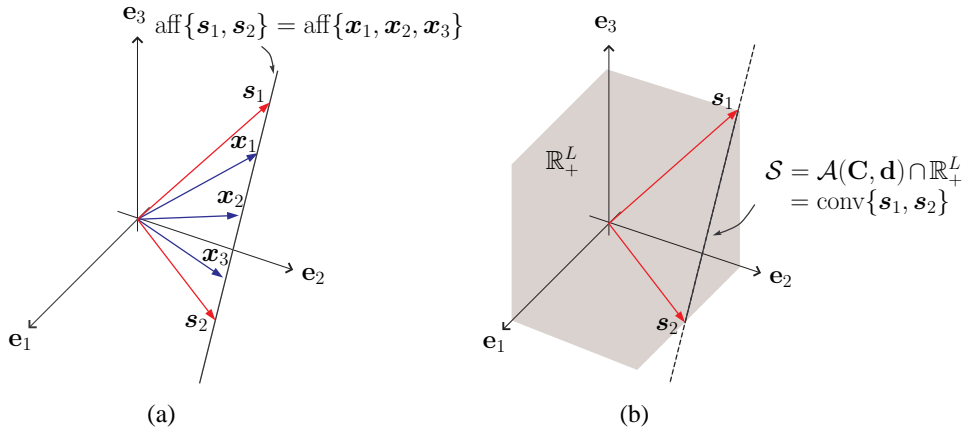


Fig. 2. Geometric illustrations of CAMNS, for the special case of $N = 2$, $M = 3$, and $L = 3$. (a) Affine set geometry indicated by Lemma 1 and Proposition 1; (b) convex hull geometry suggested in Lemma 2.

Recall that the source signals are non-negative. Hence, we have $\mathbf{s}_i \in \text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\} \cap \mathbb{R}_+^L$ for any i . Let us define

$$\mathcal{S} = \text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\} \cap \mathbb{R}_+^L = \mathcal{A}(\mathbf{C}, \mathbf{d}) \cap \mathbb{R}_+^L \quad (18)$$

$$= \{ \mathbf{x} \mid \mathbf{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d}, \mathbf{x} \succeq \mathbf{0}, \boldsymbol{\alpha} \in \mathbb{R}^{N-1} \} \quad (19)$$

which is a polyhedral set. We can show that

Lemma 2. *Under (A1) and (A2), we have*

$$\mathcal{S} = \text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}. \quad (20)$$

The proof of Lemma 2 is given in Appendix C. Following the simple illustrative example in Figure 2(a), in Figure 2(b) we verify geometrically that \mathcal{S} is equivalent to $\text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ for $N = 2$. From Lemma 2, we notice an important consequence that the set of extreme points of \mathcal{S} or $\text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ is $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$, since $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ is linearly independent [due to (A2)]. The extremal property of $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ can be seen in the illustration in Figure 2(b).

From the derivations above, we conclude that

Theorem 1. (nBSS criterion by CAMNS) *Under (A1) to (A4), the polyhedral set*

$$\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^L \mid \mathbf{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \succeq \mathbf{0}, \boldsymbol{\alpha} \in \mathbb{R}^{N-1}\} \quad (21)$$

where (\mathbf{C}, \mathbf{d}) is obtained from the observation set $\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ by the affine set fitting procedure in Proposition 1, has N extreme points given by the true source vectors $\mathbf{s}_1, \dots, \mathbf{s}_N$.

Proof: Theorem 1 is a direct consequence of Lemma 1, Proposition 1, Lemma 2, and the basic result that the extreme points of $\text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ are $\mathbf{s}_1, \dots, \mathbf{s}_N$. ■

The theoretical implication of Theorem 1 is profound: It suggests that the true source vectors can be perfectly identified by finding all the extreme points of \mathcal{S} . This provides new opportunities in nBSS that cannot be found in the other presently available literature to our best knowledge. Exploring these opportunities in practice are the subject of the following sections.

V. LINEAR PROGRAMMING METHOD FOR CAMNS

This section, as well as the next section are dedicated to the practical implementation of CAMNS. In this section, we propose an approach that uses linear programs (LPs) to systematically fulfil the CAMNS criterion. The next section will describe a geometric approach as an alternative to the LP.

Our problem as indicated in Theorem 1 is to find all the extreme points of the polyhedral set \mathcal{S} in (19). In the optimization literature this problem is known as vertex enumeration; see [33]–[35] and the references therein. The available extreme-point finding methods are sophisticated, requiring no assumption on the extreme points. However, the complexity of those methods would increase exponentially with the number of inequalities L (note that L is also the data length in our problem, which is often large in practice),

except for a few special cases not applicable to this work. The notable difference of the development here is that we exploit the characteristic that the extreme points s_1, \dots, s_N are linearly independent in the CAMNS problem [recall that this property is a direct consequence of (A2)]. By doing so we will establish an extreme-point finding method (for CAMNS) whose complexity is polynomial in L .

We first concentrate on identifying one extreme point from \mathcal{S} . Consider the following linear minimization problem:

$$\begin{aligned} p^* = \min_{\mathbf{s}} \quad & \mathbf{r}^T \mathbf{s} \\ \text{subject to (s.t.)} \quad & \mathbf{s} \in \mathcal{S} \end{aligned} \quad (22)$$

for some arbitrarily chosen direction $\mathbf{r} \in \mathbb{R}^L$, where p^* denotes the optimal objective value of (22). Using the polyhedral structures of \mathcal{S} in (19), problem (22) can be equivalently represented by an LP

$$\begin{aligned} p^* = \min_{\boldsymbol{\alpha}} \quad & \mathbf{r}^T (\mathbf{C}\boldsymbol{\alpha} + \mathbf{d}) \\ \text{s.t.} \quad & \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \succeq \mathbf{0} \end{aligned} \quad (23)$$

which can be solved by readily available algorithms such as the polynomial-time interior-point methods [36], [37]. Problem (23) is the problem we solve in practice, but (22) leads to important implications to extreme-point search.

A fundamental result in LP theory is that $\mathbf{r}^T \mathbf{s}$, the objective function of (22), attains the minimum at a point of the boundary of \mathcal{S} . To provide more insights, some geometric illustrations are given in Fig. 3. We can see that the solution of (22) may be uniquely given by one of the extreme points s_i [Fig. 3(a)], or it may be any point on a face [Fig. 3(b)]. The latter case poses a trouble to our task of identifying s_i , but it is arguably not a usual situation. For instance, in the demonstration in Fig. 3(b), \mathbf{r} must be normal to $s_2 - s_3$ which may be unlikely to happen for a randomly picked \mathbf{r} . With this intuition in mind, we prove in Appendix D that

Lemma 3. *Suppose that \mathbf{r} is randomly generated following a distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}_L)$. Then, with probability 1, the solution of (22) is uniquely given by s_i for some $i \in \{1, \dots, N\}$.*

The idea behind Lemma 3 is that undesired cases, such as that in Fig. 3(b) happen with probability zero.

We may find another extreme point by solving the maximization counterpart of (22)

$$\begin{aligned} q^* = \max_{\boldsymbol{\alpha}} \quad & \mathbf{r}^T (\mathbf{C}\boldsymbol{\alpha} + \mathbf{d}) \\ \text{s.t.} \quad & \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \succeq \mathbf{0}. \end{aligned} \quad (24)$$

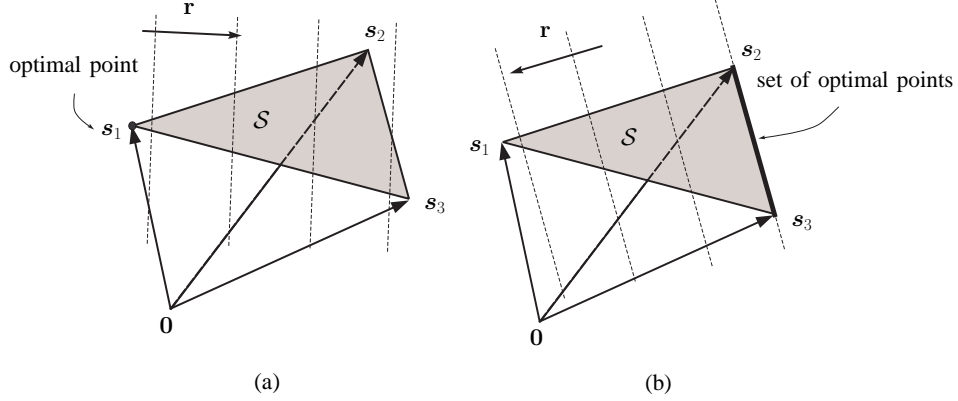


Fig. 3. Geometric interpretation of an LP.

Using the same derivations as above, we can show the following: Under the premise of Lemma 3, the solution of (24) is, with probability 1, uniquely given by an extreme point s_i different from that in (22).

Suppose that we have identified l extreme points, say, without loss of generality, $\{s_1, \dots, s_l\}$. Our interest is in refining the above LP extreme-point finding procedure such that the search space is restricted to $\{s_{l+1}, \dots, s_N\}$. To do so, consider a thin QR decomposition [38] of $[s_1, \dots, s_l]$

$$[s_1, \dots, s_l] = \mathbf{Q}_1 \mathbf{R}_1, \quad (25)$$

where $\mathbf{Q}_1 \in \mathbb{R}^{L \times l}$ is semi-unitary and $\mathbf{R}_1 \in \mathbb{R}^{l \times l}$ is upper triangular. Let

$$\mathbf{B} = \mathbf{I}_L - \mathbf{Q}_1 \mathbf{Q}_1^T. \quad (26)$$

We assume that \mathbf{r} takes the form

$$\mathbf{r} = \mathbf{B} \mathbf{w} \quad (27)$$

for some $\mathbf{w} \in \mathbb{R}^L$, and consider solving (23) and (24) with such an \mathbf{r} . Since \mathbf{r} is orthogonal to the old extreme points s_1, \dots, s_l , the intuitive expectation is that (23) and (24) should both lead to new extreme points. Interestingly, we found theoretically that expectation is not true, but close. Consider the following lemma which is proven in Appendix E:

Lemma 4. *Suppose that $\mathbf{r} = \mathbf{B} \mathbf{w}$, where $\mathbf{B} \in \mathbb{R}^{L \times L}$ is given by (26) and \mathbf{w} is randomly generated following a distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}_L)$. Then, with probability 1, at least one of the optimal solutions of (23) and (24) is a new extreme point; i.e., s_i for some $i \in \{l+1, \dots, N\}$. The certificate of finding new extreme points is indicated by $|p^*| \neq 0$ for the case of (23), and $|q^*| \neq 0$ for (24).*

By repeating the above described procedures, we can identify all the extreme points $\mathbf{s}_1, \dots, \mathbf{s}_N$. The resultant CAMNS-LP method is summarized in Table I.

TABLE I
A SUMMARY OF THE CAMNS-LP METHOD.

CAMNS-LP method	
Given	an affine set characterization 2-tuple (\mathbf{C}, \mathbf{d}) .
Step 1.	Set $l = 0$, and $\mathbf{B} = \mathbf{I}_L$.
Step 2.	Randomly generate a vector $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_L)$, and set $\mathbf{r} := \mathbf{B}\mathbf{w}$.
Step 3.	Solve the LPs $p^* = \min_{\alpha: \mathbf{C}\alpha + \mathbf{d} \geq \mathbf{0}} \mathbf{r}^T (\mathbf{C}\alpha + \mathbf{d})$ $q^* = \max_{\alpha: \mathbf{C}\alpha + \mathbf{d} \geq \mathbf{0}} \mathbf{r}^T (\mathbf{C}\alpha + \mathbf{d})$ and obtain their optimal solutions, denoted by α_1^* and α_2^* , respectively.
Step 4.	If $l = 0$ $\hat{\mathbf{S}} = [\mathbf{C}\alpha_1^* + \mathbf{d}, \mathbf{C}\alpha_2^* + \mathbf{d}]$ else $\text{If } p^* \neq 0 \text{ then } \hat{\mathbf{S}} := [\hat{\mathbf{S}} \mathbf{C}\alpha_1^* + \mathbf{d}].$ $\text{If } q^* \neq 0 \text{ then } \hat{\mathbf{S}} := [\hat{\mathbf{S}} \mathbf{C}\alpha_2^* + \mathbf{d}].$
Step 5.	Update l to be the number of columns of $\hat{\mathbf{S}}$.
Step 6.	Apply QR decomposition $\hat{\mathbf{S}} = \mathbf{Q}_1 \mathbf{R}_1,$ where $\mathbf{Q}_1 \in \mathbb{R}^{L \times l}$ and $\mathbf{R}_1 \in \mathbb{R}^{l \times l}$. Update $\mathbf{B} := \mathbf{I}_L - \mathbf{Q}_1 \mathbf{Q}_1^T$.
Step 7.	Repeat Step 2 until $l = N$.

The CAMNS-LP method in Table I is not only systematically straightforward to apply, it is also efficient due to the maturity of convex optimization algorithms. Using a primal-dual interior-point method, each LP problem [or the problem in (22) or (24)] can be solved with a worst-case complexity of $\mathcal{O}(L^{0.5}(L(N-1) + (N-1)^3)) \simeq \mathcal{O}(L^{1.5}(N-1))$ for $L \gg N$ [37]. Since the algorithm solves $2(N-1)$ LP problems in the worst case, we infer that its worst-case complexity is $\mathcal{O}(L^{1.5}(N-1)^2)$.

Based on Theorem 1, Lemma 3, Lemma 4, and the above complexity discussion, we assert that

Proposition 2. *Under (A1)-(A4), the CAMNS-LP method in Table I finds all the true source vectors*

s_1, \dots, s_N with probability 1. It does so with a worst-case complexity of $\mathcal{O}(L^{1.5}(N-1)^2)$.

Some further discussions on the algorithm complexity are now in order.

- 1) In the complexity result in Proposition 1, the factor $L^{0.5}$ is the theoretical worst-case number of iterations for an interior-point optimization algorithm to solve an LP. In many LP applications the number of iterations is usually found to grow much slower than the worst-case and seem like a constant. Hence, in practice CAMNS-LP works more like $\mathcal{O}(L(N-1)^2)$ on average.
- 2) It is interesting to compare the complexity of CAMNS-LP and some benchmarked BSS methods. nICA [7], NMF [11], Erdogan's BSS-MBS algorithm [28] can be verified to have complexities given by $\mathcal{O}(N^2L\eta)$, $\mathcal{O}(N^3L^2\eta)$, and $\mathcal{O}(N^2L\eta)$, respectively, where η is the number of iterations for each respective algorithm. To put the comparison into context, we rewrite the CAMNS-LP complexity as $\mathcal{O}((N-1)^2L\eta)$ where η is the number of iterations again. Clearly CAMNS-LP has a competitive complexity order compared to those methods.

We have provided a practical implementation of CAMNS-LP at <http://www.ee.cuhk.edu.hk/~wkma/CAMNS/CAMNS.htm>. The source codes were written in MATLAB, and are based on a reliable convex optimization software SeDuMi [36]. Readers are encouraged to test the codes and give us some feedback.

VI. GEOMETRIC METHOD FOR CAMNS

The CAMNS-LP method developed in the last section effectively uses numerical optimization to achieve the CAMNS criterion. In this section we develop analytical or semi-analytical alternatives to achieving CAMNS, that are simple and more efficient than CAMNS-LP. The methods to be proposed are for the cases of two and three sources only, where the relatively simple geometrical structures in the two cases are utilized.

To facilitate the development, in Section VI-A we provide an alternate form of the CAMNS criterion. Then, Sections VI-B and VI-C describe the geometric algorithms for two and three sources, respectively.

A. An Alternate n BSS Criterion

Let us consider the pre-image of \mathcal{S} under the mapping $s = \mathbf{C}\alpha + \mathbf{d}$, denoted by

$$\begin{aligned} \mathcal{F} &= \{ \alpha \in \mathbb{R}^{N-1} \mid \mathbf{C}\alpha + \mathbf{d} \succeq \mathbf{0} \} \\ &= \{ \alpha \in \mathbb{R}^{N-1} \mid \mathbf{c}_n^T \alpha + d_n \geq 0, \ n = 1, \dots, L \} \end{aligned} \quad (28)$$

where \mathbf{c}_n^T is the n th row of \mathbf{C} . There is a direct correspondence between the extreme points of \mathcal{S} and \mathcal{F} , as described in the following lemma:

Lemma 5. *The polyhedral set \mathcal{F} in (28) is equivalent to an $(N-1)$ -simplex*

$$\mathcal{F} = \text{conv}\{\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N\} \quad (29)$$

where each $\boldsymbol{\alpha}_i \in \mathbb{R}^{N-1}$ satisfies

$$\mathbf{C}\boldsymbol{\alpha}_i + \mathbf{d} = \mathbf{s}_i. \quad (30)$$

The proof of Lemma 5 is given in Appendix F. Hence, as an alternative to our previously proposed approach where we find $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ by identifying the extreme points of \mathcal{S} , we can achieve perfect blind separation by identifying the extreme points of \mathcal{F} . In this alternative, we are aided by a useful convex analysis result presented as follows:

Lemma 6. (Extreme point validation for polyhedra [31]) *For a polyhedral set in form of (28), a point $\boldsymbol{\alpha} \in \mathcal{F}$ is an extreme point of \mathcal{F} if and only if the following collection of vectors*

$$\mathcal{C}_{\boldsymbol{\alpha}} = \{ \mathbf{c}_n \in \mathbb{R}^{N-1} \mid \mathbf{c}_n^T \boldsymbol{\alpha} = -d_n, n = 1, \dots, L \} \quad (31)$$

contains $N - 1$ linearly independent vectors.

In summary, an alternate version of the BSS criterion in Theorem 1 is given as follows

Theorem 2. (Alternate nBSS criterion by CAMNS) *Under (A1) to (A4), the set*

$$\mathcal{F} = \{ \boldsymbol{\alpha} \in \mathbb{R}^{N-1} \mid \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \succeq \mathbf{0} \} \quad (32)$$

where (\mathbf{C}, \mathbf{d}) are obtained from the observations $\mathbf{x}_1, \dots, \mathbf{x}_M$ via the affine set fitting solution in Proposition 1, has N extreme points $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N$. Each extreme point corresponds to a source signal through the relationship $\mathbf{C}\boldsymbol{\alpha}_i + \mathbf{d} = \mathbf{s}_i$. The extreme points of \mathcal{F} may be validated by the procedure in Lemma 6.

Proof: Theorem 2 directly follows from Theorem 1, Lemma 5, and Lemma 6. ■

Based on Theorem 2, we propose geometric based methods for the cases of two and three sources in the following two subsections.

B. Two Sources: A Simple Closed-Form Solution

For the case of $N = 2$, the set \mathcal{F} is simply

$$\mathcal{F} = \{\alpha \in \mathbb{R} \mid c_n \alpha + d_n \geq 0, \ n = 1, \dots, L\}. \quad (33)$$

By Lemma 6, α is extremal if and only if

$$\alpha = -d_n/c_n, \quad c_n \neq 0, \quad \alpha \in \mathcal{F} \quad (34)$$

for some $n \in \{1, \dots, L\}$. From (33) we see that $\alpha \in \mathcal{F}$ implies the following two conditions:

$$\alpha \geq -d_n/c_n, \text{ for all } n \text{ such that } c_n > 0, \quad (35)$$

$$\alpha \leq -d_n/c_n, \text{ for all } n \text{ such that } c_n < 0. \quad (36)$$

We therefore conclude from (34), (35), and (36) that the extreme points are given by

$$\alpha_1 = \min\{-d_n/c_n \mid c_n < 0, \ n = 1, 2, \dots, L\}, \quad (37)$$

$$\alpha_2 = \max\{-d_n/c_n \mid c_n > 0, \ n = 1, 2, \dots, L\} \quad (38)$$

which are simple closed-form expressions.

The complexity of the 2-source geometric method in (37)-(38) is $\mathcal{O}(L)$.

C. Three Sources: A Semi-Analytic Solution

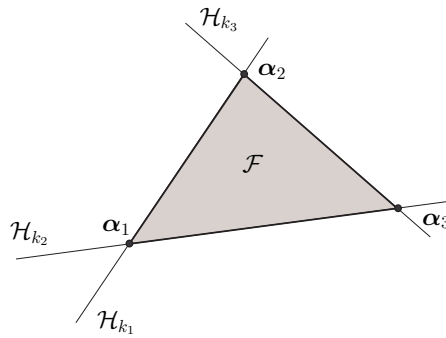


Fig. 4. Geometric illustration of \mathcal{F} for $N = 3$.

For $N = 3$, the set \mathcal{F} is a triangle on \mathbb{R}^2 . By exploiting the relatively simple geometry of \mathcal{F} , we can locate the extreme points very effectively. Figure 4 shows the geometry of \mathcal{F} in this case. We see that the

boundary of \mathcal{F} is entirely constituted by three facets $\text{conv}\{\alpha_1, \alpha_2\}$, $\text{conv}\{\alpha_1, \alpha_3\}$, and $\text{conv}\{\alpha_2, \alpha_3\}$. They may be represented by the following polyhedral expression

$$\text{conv}\{\alpha_1, \alpha_2\} = \mathcal{F} \cap \mathcal{H}_{k_1}, \quad (39a)$$

$$\text{conv}\{\alpha_1, \alpha_3\} = \mathcal{F} \cap \mathcal{H}_{k_2}, \quad (39b)$$

$$\text{conv}\{\alpha_2, \alpha_3\} = \mathcal{F} \cap \mathcal{H}_{k_3}, \quad (39c)$$

for some $k_1, k_2, k_3 \in \{1, \dots, L\}$, where

$$\mathcal{H}_k = \{\alpha \in \mathbb{R}^2 \mid \mathbf{c}_k^T \alpha = -d_k\}. \quad (40)$$

Suppose that (k_1, k_2, k_3) is known. Then, by Lemma 6, the three extreme points are given by the closed form

$$\alpha_1 = - \begin{bmatrix} \mathbf{c}_{k_1}^T \\ \mathbf{c}_{k_2}^T \end{bmatrix}^{-1} \begin{bmatrix} d_{k_1} \\ d_{k_2} \end{bmatrix}, \quad (41a)$$

$$\alpha_2 = - \begin{bmatrix} \mathbf{c}_{k_1}^T \\ \mathbf{c}_{k_3}^T \end{bmatrix}^{-1} \begin{bmatrix} d_{k_1} \\ d_{k_3} \end{bmatrix}, \quad (41b)$$

$$\alpha_3 = - \begin{bmatrix} \mathbf{c}_{k_2}^T \\ \mathbf{c}_{k_3}^T \end{bmatrix}^{-1} \begin{bmatrix} d_{k_2} \\ d_{k_3} \end{bmatrix}. \quad (41c)$$

A geometric interpretation of (41) is as follows: an extreme point is the intersection of any two of the facet-forming lines \mathcal{H}_{k_1} , \mathcal{H}_{k_2} , and \mathcal{H}_{k_3} . This can also be seen in Figure 4.

Inspired by the fact that finding all the facets is equivalent to finding all the extreme points, we propose an extreme-point finding heuristic in Table II. The idea behind is illustrated in Figure 5. In the first stage [Figure 5(a)], we move from some interior point α' over a direction \mathbf{r} until we reach the boundary. This process helps identify one facet line, say \mathcal{H}_{k_1} . Similarly, by travelling over the direction opposite to \mathbf{r} , we may locate another facet line, say \mathcal{H}_{k_2} . The intersection of \mathcal{H}_{k_1} and \mathcal{H}_{k_2} results in finding an extreme point α_1 . In the second stage [Figure 5(b)], we use the same idea to locate the last facet line \mathcal{H}_{k_3} , by travelling over direction $\alpha' - \alpha_1$.

The proposed algorithm requires an interior point α' as the starting point for facet search. Sometimes, the problem nature allows us to determine such a point easily. For instance, if the mixing matrix is

TABLE II
A SUMMARY OF GEOMETRIC EXTREME-POINT FINDING ALGORITHM.

Geometric Extreme-Point Finding Algorithm for $N = 3$

Given an affine set characterization 2-tuple (\mathbf{C}, \mathbf{d}) , and a vector α' interior to \mathcal{F} .

Step 1. Randomly generate a direction $\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$.

Step 2. Locate a boundary point

$$\psi_1 = \alpha' + t_1 \mathbf{r}$$

where

$$\begin{aligned} t_1 &= \sup\{t \mid \alpha' + t\mathbf{r} \in \mathcal{F}\} \\ &= \min\{-(\mathbf{c}_n^T \alpha' + d_n)/\mathbf{c}_n^T \mathbf{r} \mid \mathbf{c}_n^T \mathbf{r} < 0, n = 1, \dots, L\}. \end{aligned}$$

Step 3. Find the index set

$$\mathcal{K}_1 = \{n \mid \mathbf{c}_n^T \psi_1 = -d_n, n = 1, \dots, L\}$$

If $\{\mathbf{c}_k \mid k \in \mathcal{K}_1\}$ contains 2 linearly independent vectors (i.e., ψ_1 is an extreme point and there is an indeterminacy that $\mathcal{H}_k \neq \mathcal{H}_l$ for some $k, l \in \mathcal{K}_1$), then go to **Step 1**.

Step 4. Locate a boundary point

$$\psi_2 = \alpha' + t_2 \mathbf{r}$$

where

$$\begin{aligned} t_2 &= \inf\{t \mid \alpha' + t\mathbf{r} \in \mathcal{F}\} \\ &= \max\{-(\mathbf{c}_n^T \alpha' + d_n)/\mathbf{c}_n^T \mathbf{r} \mid \mathbf{c}_n^T \mathbf{r} > 0, n = 1, \dots, L\}. \end{aligned}$$

Step 5. Find the index set

$$\mathcal{K}_2 = \{n \mid \mathbf{c}_n^T \psi_2 = -d_n, n = 1, \dots, L\}$$

If $\{\mathbf{c}_k \mid k \in \mathcal{K}_2\}$ contains 2 linearly independent vectors, then go to **Step 1**.

Step 6. Determine

$$\alpha_1 = - \begin{bmatrix} \mathbf{c}_{k_1}^T \\ \mathbf{c}_{k_2}^T \end{bmatrix}^{-1} \begin{bmatrix} d_{k_1} \\ d_{k_2} \end{bmatrix},$$

for an arbitrary $k_1 \in \mathcal{K}_1$ and $k_2 \in \mathcal{K}_2$.

Step 7. Set $\mathbf{r} = \alpha' - \alpha_1$, and locate a boundary point

$$\psi_3 = \alpha' + t_3 \mathbf{r}$$

where

$$\begin{aligned} t_3 &= \sup\{t \mid \alpha' + t\mathbf{r} \in \mathcal{F}\} \\ &= \min\{-(\mathbf{c}_n^T \alpha' + d_n)/\mathbf{c}_n^T \mathbf{r} \mid \mathbf{c}_n^T \mathbf{r} < 0, n = 1, \dots, L\}. \end{aligned}$$

Step 8. Find the index set

$$\mathcal{K}_3 = \{n \mid \mathbf{c}_n^T \psi_3 = -d_n, n = 1, \dots, L\}.$$

Step 9. Determine

$$\alpha_2 = - \begin{bmatrix} \mathbf{c}_{k_1}^T \\ \mathbf{c}_{k_3}^T \end{bmatrix}^{-1} \begin{bmatrix} d_{k_1} \\ d_{k_3} \end{bmatrix}, \quad \alpha_3 = - \begin{bmatrix} \mathbf{c}_{k_2}^T \\ \mathbf{c}_{k_3}^T \end{bmatrix}^{-1} \begin{bmatrix} d_{k_2} \\ d_{k_3} \end{bmatrix},$$

for an arbitrary $k_1 \in \mathcal{K}_1$, $k_2 \in \mathcal{K}_2$, and $k_3 \in \mathcal{K}_3$.

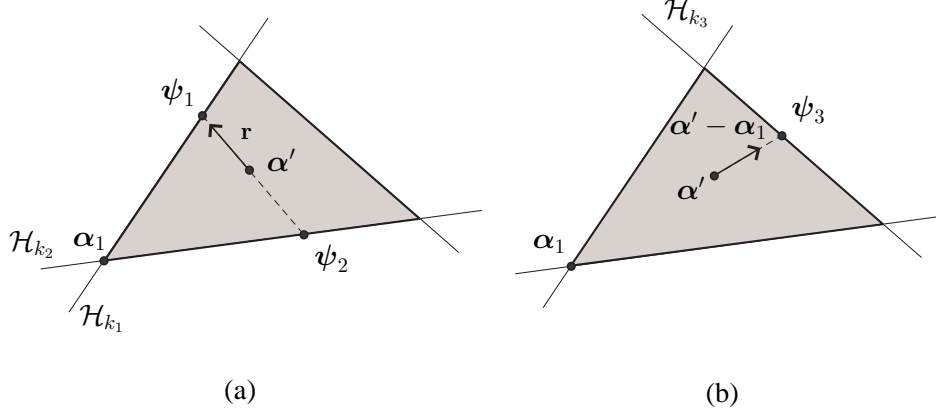


Fig. 5. Illustration of the operations of the geometric extreme-point finding algorithm for $N = 3$.

componentwise non-negative or $a_{ij} \geq 0$ for all i, j , then it can be verified that $\alpha' = \mathbf{0}$ is interior to \mathcal{F} . When an interior point is not known, we can find one numerically by solving the LP

$$\begin{aligned} \max_{\alpha, \beta} \quad & \beta \\ \text{s.t.} \quad & \mathbf{c}_n^T \alpha + d_n \geq \beta, \quad n = 1, \dots, L \end{aligned} \quad (42)$$

(This is known as the phase I method in optimization [30].)

The complexity of the 3-source geometric heuristic in Table II is $\mathcal{O}(L)$.

VII. SIMULATIONS

To demonstrate the efficacy of the CAMNS-based algorithms, four simulation results are presented here. Section VII-A is an X-ray image example where our task is to distinguish bone structures from soft tissue. Section VII-B considers a benchmarked problem [2] in which the sources are faces of three different persons. Section VII-C focuses on a challenging scenario reminiscent of ghosting effects in photography. Section VII-D uses Monte Carlo simulation to evaluate the performance of CAMNS-based algorithms under noisy condition. For performance comparison, we also test three existing nBSS algorithms, namely non-negative least-correlated component analysis (nLCA) [26], non-negative matrix factorization (NMF) [11], non-negative independent component analysis (nICA) [7], and Ergodan's BSS-MBS algorithm [28].

The performance measure used in this paper is described as follows. Let $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_N]$ be the true multi-source signal matrix, and $\hat{\mathbf{S}} = [\hat{\mathbf{s}}_1, \dots, \hat{\mathbf{s}}_N]$ be the multi-source output of a BSS algorithm. It is well known that a BSS algorithm is inherently subject to permutation and scaling ambiguities. We

propose a sum square error (SSE) measure for \mathbf{S} and $\hat{\mathbf{S}}$ [39], [40], given as follows:

$$e(\mathbf{S}, \hat{\mathbf{S}}) = \min_{\boldsymbol{\pi} \in \Pi_N} \sum_{i=1}^N \left\| \mathbf{s}_i - \frac{\|\mathbf{s}_i\|}{\|\hat{\mathbf{s}}_{\pi_i}\|} \hat{\mathbf{s}}_{\pi_i} \right\|^2 \quad (43)$$

where $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N)$, and $\Pi_N = \{\boldsymbol{\pi} \in \mathbb{R}^N \mid \pi_i \in \{1, 2, \dots, N\}, \pi_i \neq \pi_j \text{ for } i \neq j\}$ is the set of all permutations of $\{1, 2, \dots, N\}$. The optimization of (43) is to adjust the permutation $\boldsymbol{\pi}$ such that the best match between true and estimated signals is yielded, while the factor $\|\mathbf{s}_i\|/\|\hat{\mathbf{s}}_{\pi_i}\|$ is to fix the scaling ambiguity. Problem (43) is the optimal assignment problem which can be efficiently solved by Hungarian algorithm¹ [41].

A. Example of $N = M = 2$: Dual-energy Chest X-ray Imaging

Dual-energy chest x-ray imaging is clinically used for detecting calcified granuloma, a symptom of lung nodules [42]. The diagnostic images are acquired from two stacked detectors separated by a copper filter along which x-rays at two different energies are passed. For visualizing the symptom of calcified granuloma, it is necessary to separate bone structures and soft tissue from the diagnostic images.

In this simulation we have two 164×164 source images, one representing bone structure and another soft tissue. The two images can be found in [43] and they are displayed in Figure 6(a). Each image is represented by a source vector $\mathbf{s}_i \in \mathbb{R}^L$, by scanning the image vertically from top left to bottom right (thereby $L = 164^2 = 26896$). We found that the two source signals satisfy the local dominant assumption [or (A2)] perfectly, by numerical inspection. The observation vectors, or the diagnostic images are synthetically generated using a mixing matrix

$$\mathbf{A} = \begin{bmatrix} 0.55 & 0.45 \\ 0.63 & 0.37 \end{bmatrix}. \quad (44)$$

The mixed images are shown in Figure 6(b). The separated images of the various nBSS methods are illustrated in Figure 6(c)-(h). By visual inspection, the CAMNS-based methods and nLCA appear to yield the best separation among the various methods. We also see that nICA and Erdogan's algorithm provide reasonably good performance. In Table III the various methods are quantitatively compared, using the SSE in (43). The table suggests that the CAMNS-based methods, along side with nLCA achieve perfect separation.

¹A Matlab implementation is available at <http://si.utia.cas.cz/Tichavsky.html>.

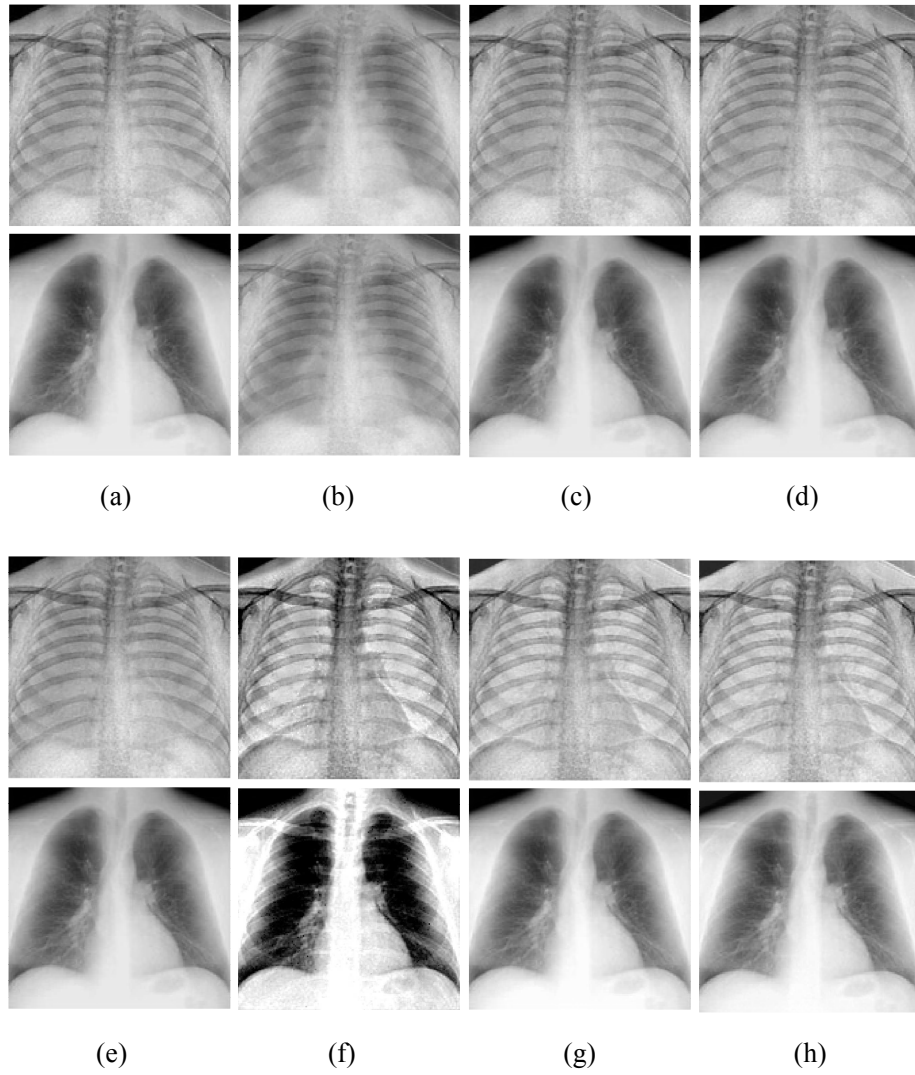


Fig. 6. Dual-energy chest x-ray imaging: (a) the sources, (b) the observations, and the extracted sources obtained by (c) CAMNS-LP method, (d) CAMNS-geometric method, (e) nLCA, (f) NMF, (g) nICA and (h) Erdogan's algorithm.

B. Example of $N = M = 3$: Human Face Separation

Three 128×128 human face images, taken from the benchmarks in [2], are used to generate three observations. The mixing matrix is

$$\mathbf{A} = \begin{bmatrix} 0.20 & 0.62 & 0.18 \\ 0.35 & 0.37 & 0.28 \\ 0.40 & 0.40 & 0.20 \end{bmatrix}. \quad (45)$$

In this example, the local dominant assumption is not perfectly satisfied. To shed some light into this, we propose a measure called the *local dominance proximity factor* (LDPF) of the i th source, defined as follows:

$$\kappa_i = \max_{n=1,\dots,L} \frac{s_i[n]}{\sum_{j \neq i} s_j[n]}. \quad (46)$$

When $\kappa_i = \infty$, we have the i th source satisfying the local dominant assumption perfectly. The values of κ_i 's in this example are shown in Table IV, where we see that the LDPFs of the three sources are strong but not infinite.

Figure 7 shows the separated images of the various nBSS methods. We see that the CAMNS-based methods and nLCA provide good separation, despite the fact that the local dominance assumption is not perfectly satisfied. This result indicates that the CAMNS-based methods have some robustness against violation of local dominance. Moreover, nICA and Erdogan's algorithm work poorly due to the violation of the assumption of uncorrelated sources. The SSE performance of the various methods is given in Table III, where we have two observations. First, the CAMNS-LP method yields the best performance among all the methods under test. Second, the performance of the LP method is better than that of the CAMNS-geometric method. The latter suggests that the LP method is more robust than the geometric method, when local dominance is not exactly satisfied. This result will be further confirmed in the Monte Carlo simulation in Section VII-D.

C. Example of $M = N = 4$: Ghosting Effect

We take a 285×285 Lena image from [2] as one source and then shift it diagonally to create three more sources; see Figure 8(a). Apparently, these sources are strongly correlated. Even worse, their LDPFs, shown in Table IV are not too satisfactory compared to the previous two examples. The mixing matrix is

$$\mathbf{A} = \begin{bmatrix} 0.02 & 0.37 & 0.31 & 0.30 \\ 0.31 & 0.21 & 0.26 & 0.22 \\ 0.05 & 0.38 & 0.28 & 0.29 \\ 0.33 & 0.23 & 0.21 & 0.23 \end{bmatrix}. \quad (47)$$

Figure 8(b) displays the observations, where the mixing effect is reminiscent of the ghosting effect in analog televisions. The image separation results are illustrated in Figure 8(c)-(g). Clearly, only the CAMNS-LP method and nLCA provide sufficiently good mitigation of the "ghosts". This result once again suggests that the CAMNS-based methods (as well as nLCA) is not too sensitive to the effect of local dominance violation. Regarding the comparison of the CAMNS-LP method and nLCA, one may

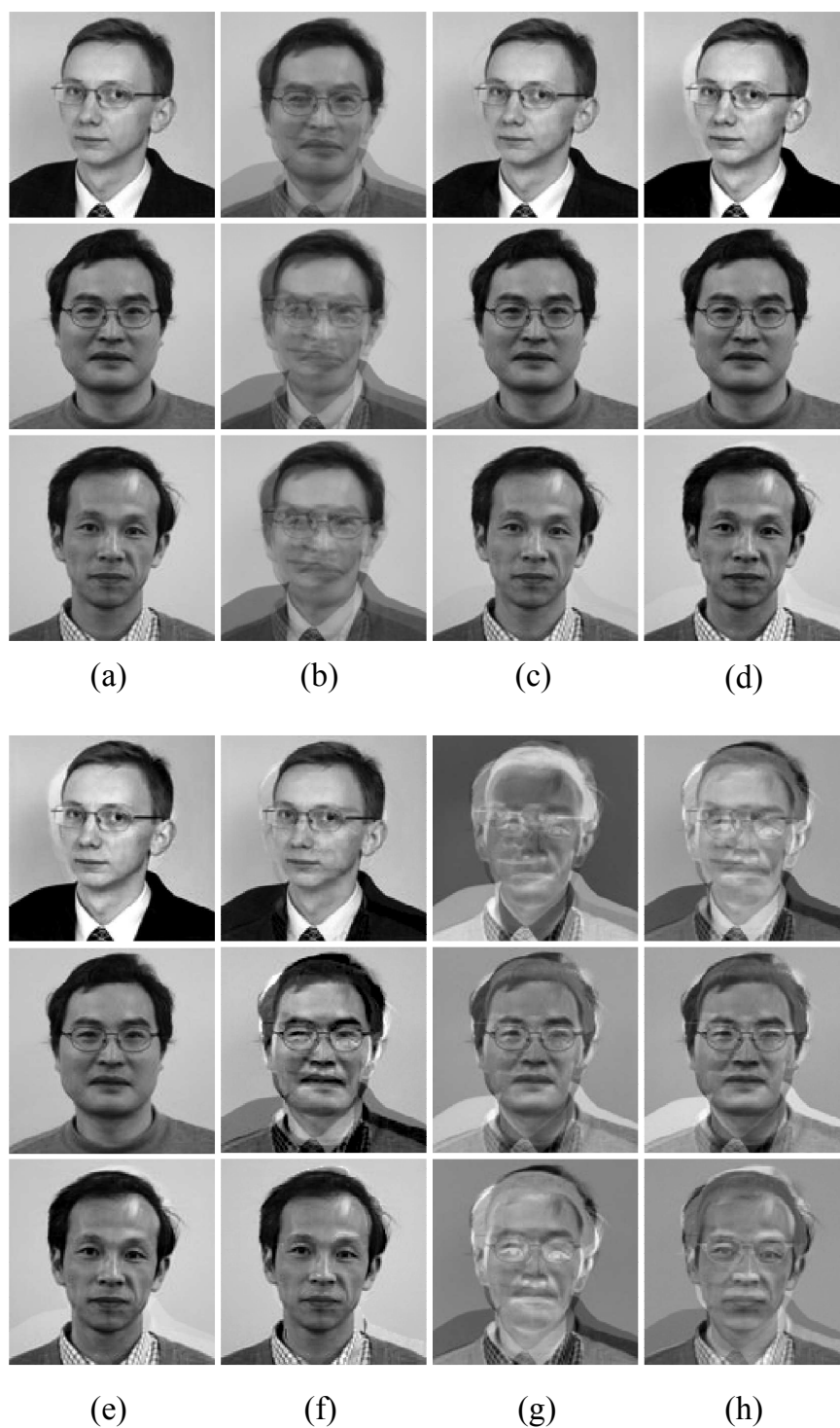


Fig. 7. Human face separation: (a) the sources, (b) the observations, and the extracted sources obtained by (c) CAMNS-LP method, (d) CAMNS-geometric method, (e) nLCA, (f) NMF, (g) nICA and (h) Erdogan's algorithm.

TABLE III
LOCAL DOMINANCE PROXIMITY FACTORS IN THE THREE SCENARIOS.

	κ_i			
	source 1	source 2	source 3	source 4
Dual-energy X-ray	∞	∞	-	-
Human face separation	15.625	6.172	10.000	-
Ghosting reduction	2.133	2.385	2.384	2.080

TABLE IV
THE SSEs OF THE VARIOUS NBSS METHODS IN THE THREE SCENARIOS.

	SSE $e(\mathbf{S}, \hat{\mathbf{S}})$ (in dB)					
	CAMNS-LP	CAMNS-Geometric	nLCA	NMF	nICA	Erdogan's algorithm
Dual-energy X-ray	-252.2154	-247.2876	-259.0132	30.8372	24.4208	23.8371
Human face separation	9.4991	18.2349	19.6589	24.6158	38.5425	32.5437
Ghosting reduction	20.7535	-	31.3767	38.6202	41.8963	39.1264

find that the nLCA separated images have some ghosting residuals, upon very careful visual inspection. As for the proposed method, we argue that the residuals are harder to notice. Moreover, our numerical inspection found a problem that the nLCA signal outputs have negative values sometimes. For this reason, we see in Table III that the SSE of nLCA is about 10dB larger than that of the CAMNS-LP method.

D. Example of $M = 6, N = 3$: Noisy Environment

We use Monte Carlo simulation to test the performance of the various methods when noise is present. The three face images in Figure 7(a) were used to generate six noisy observations. The noise is independently and identically distributed (i.i.d.), following a Gaussian distribution with zero mean and variance σ^2 . To maintain non-negativity of the observations in the simulation, we force the negative noisy observations to zero. We performed 100 independent runs. At each run the mixing matrix was i.i.d. uniformly generated on $[0,1]$ and then each row is normalized to 1 to maintain (A3). The average errors $e(\mathbf{S}, \hat{\mathbf{S}})$ for different SNRs (defined here as $\text{SNR} = \sum_{i=1}^N \|\mathbf{s}_i\|^2 / LN\sigma_i^2$) are shown in Figure 9. One can see that the CAMNS-LP method performs better than the other methods.

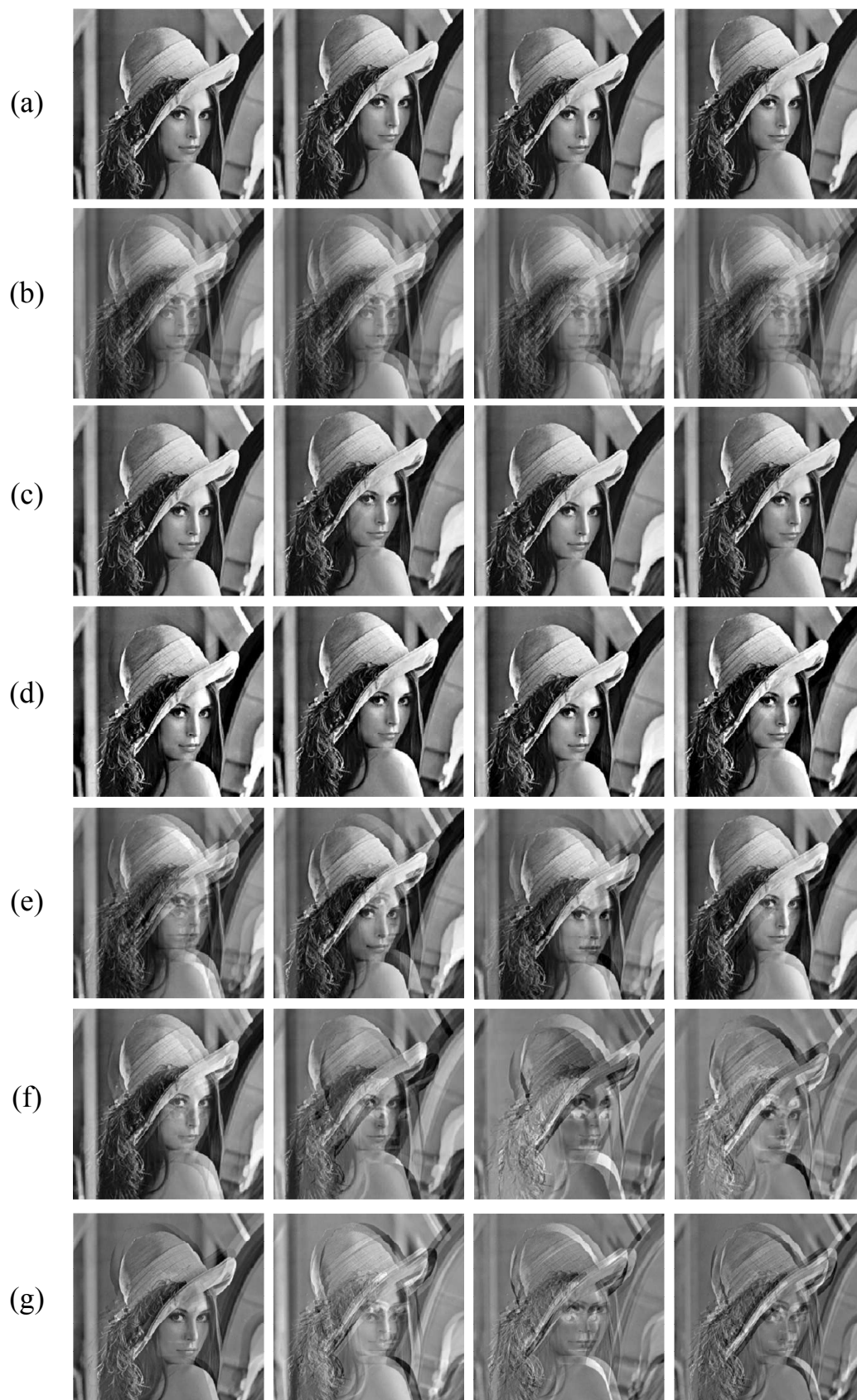


Fig. 8. Ghosting reduction: (a) the sources, (b) the observations and the extracted sources obtained by (c) CAMNS-LP method, (d) nLCA, (e) NMF, (f) nICA and (g) Erdogan's algorithm.

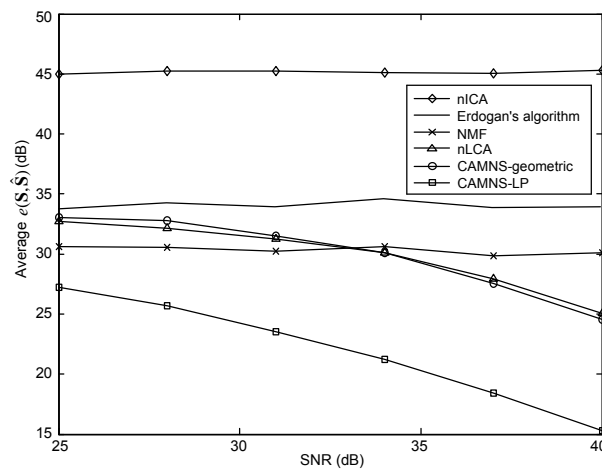


Fig. 9. Performance evaluation of the CAMNS-based methods, nLCA, NMF, nICA and Erdogan's method for the human face images experiment under noisy condition.

VIII. CONCLUSION

We have developed a convex analysis based framework for non-negative blind source separation. The core of the framework is a new nBSS criterion, which guarantees perfect separation under some assumptions [see (A1)-(A4)] that are realistic in many applications such as multichannel biomedical imaging. To practically realize this result, we have proposed a systematic LP-based method for fulfilling the criterion. We should mention a side benefit that the LP method deals with linear optimization that can be solved efficiently and does not suffer from local minima. Moreover, we have used simplex geometry to establish a computationally very cheap alternative to the LP method. Our current development has led to two simple geometric algorithms, for two and three sources respectively. Future direction should consider extension of the geometric approach to four sources and beyond. We anticipate that the extension would be increasingly complex in a combinatorial manner. By contrast, the comparatively more expensive LP method does not have such a trouble *per se*, and is applicable to any number of sources.

We have also performed extensive simulations to evaluate the separation performance of the CAMNS-based methods, under several scenarios such as x-ray, human portraits, and ghosting. The results indicate that the LP method offers the best performance among the various methods under test.

ACKNOWLEDGMENT

We would like to thank the anonymous reviewers and the associate editor in charge for their thoughtful comments, which have improved this paper.

APPENDIX

A. Proof of Lemma 1

Any $\mathbf{x} \in \text{aff}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ can be represented by

$$\mathbf{x} = \sum_{i=1}^M \theta_i \mathbf{x}_i, \quad (48)$$

where $\boldsymbol{\theta} \in \mathbb{R}^M$, $\boldsymbol{\theta}^T \mathbf{1} = 1$. Substituting (2) into (48), we get

$$\mathbf{x} = \sum_{j=1}^N \beta_j \mathbf{s}_j, \quad (49)$$

where $\beta_j = \sum_{i=1}^M \theta_i a_{ij}$ for $j = 1, \dots, N$, or equivalently

$$\boldsymbol{\beta} = \mathbf{A}^T \boldsymbol{\theta}. \quad (50)$$

Since \mathbf{A} has unit row sum [(A3)], we have

$$\boldsymbol{\beta}^T \mathbf{1} = \boldsymbol{\theta}^T (\mathbf{A} \mathbf{1}) = \boldsymbol{\theta}^T \mathbf{1} = 1. \quad (51)$$

This implies that $\boldsymbol{\beta}^T \mathbf{1} = 1$, and as a result it follows from (49) that $\mathbf{x} \in \text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$.

On the other hand, any $\mathbf{x} \in \text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ can be represented by (49) for $\boldsymbol{\beta}^T \mathbf{1} = 1$. Since \mathbf{A} has full column rank [(A4)], there always exist a $\boldsymbol{\theta}$ such that (50) holds. Substituting (50) into (49) yields (48). Since (51) implies that $\boldsymbol{\theta}^T \mathbf{1} = 1$, we conclude that $\mathbf{x} \in \text{aff}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$.

B. Proof of Proposition 1

As a basic result in least squares, each projection error in (13)

$$e_{\mathcal{A}(\tilde{\mathbf{C}}, \tilde{\mathbf{d}})}(\mathbf{x}_i) = \min_{\boldsymbol{\alpha} \in \mathbb{R}^{N-1}} \|\tilde{\mathbf{C}} \boldsymbol{\alpha} + \tilde{\mathbf{d}} - \mathbf{x}_i\|_2^2 \quad (52)$$

has a closed form

$$e_{\mathcal{A}(\tilde{\mathbf{C}}, \tilde{\mathbf{d}})}(\mathbf{x}_i) = (\mathbf{x}_i - \tilde{\mathbf{d}})^T \mathbf{P}_{\tilde{\mathbf{C}}}^\perp (\mathbf{x}_i - \tilde{\mathbf{d}}) \quad (53)$$

where $\mathbf{P}_{\tilde{\mathbf{C}}}^\perp$ is the orthogonal complement projection of $\tilde{\mathbf{C}}$. Using (53), we can therefore rewrite the affine set fitting problem [in (13)] as

$$\min_{\tilde{\mathbf{C}}^T \tilde{\mathbf{C}} = \mathbf{I}} \left\{ \min_{\tilde{\mathbf{d}}} J(\tilde{\mathbf{C}}, \tilde{\mathbf{d}}) \right\} \quad (54)$$

where

$$J(\tilde{\mathbf{C}}, \tilde{\mathbf{d}}) = \sum_{i=1}^M (\mathbf{x}_i - \tilde{\mathbf{d}})^T \mathbf{P}_{\tilde{\mathbf{C}}}^\perp (\mathbf{x}_i - \tilde{\mathbf{d}}) \quad (55)$$

The inner minimization problem in (54) is an unconstrained convex quadratic program, and it can be easily verified that $\mathbf{d} = \frac{1}{M} \sum_{i=1}^M \mathbf{x}_i$ is an optimal solution to the inner minimization problem. By substituting this optimal \mathbf{d} into (54) and by letting $\mathbf{U} = [\mathbf{x}_1 - \mathbf{d}, \dots, \mathbf{x}_M - \mathbf{d}]$, problem (54) can be reduced to

$$\min_{\tilde{\mathbf{C}}^T \tilde{\mathbf{C}} = \mathbf{I}_{N-1}} \text{Trace}\{\mathbf{U}^T \mathbf{P}_{\tilde{\mathbf{C}}}^\perp \mathbf{U}\} \quad (56)$$

When $\tilde{\mathbf{C}}^T \tilde{\mathbf{C}} = \mathbf{I}_{N-1}$, the projection matrix $\mathbf{P}_{\tilde{\mathbf{C}}}^\perp$ can be simplified to $\mathbf{I}_L - \tilde{\mathbf{C}} \tilde{\mathbf{C}}^T$. Subsequently (56) can be further reduced to

$$\max_{\tilde{\mathbf{C}}^T \tilde{\mathbf{C}} = \mathbf{I}_{N-1}} \text{Trace}\{\mathbf{U}^T \tilde{\mathbf{C}} \tilde{\mathbf{C}}^T \mathbf{U}\}. \quad (57)$$

An optimal solution of (57) is known to be the $N - 1$ principal eigenvector matrix [44].

C. Proof of Lemma 2

Assume that $\mathbf{z} \in \text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\} \cap \mathbb{R}_+^L$:

$$\mathbf{z} = \sum_{i=1}^N \theta_i \mathbf{s}_i \succeq \mathbf{0}, \quad \mathbf{1}^T \boldsymbol{\theta} = 1.$$

From (A2), it follows that $z[\ell_i] = \theta_i s_i[\ell_i] \geq 0, \forall i$. Since $s_i[\ell_i] > 0$, we must have $\theta_i \geq 0, \forall i$. Therefore, \mathbf{z} lies in $\text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$. On the other hand, assume that $\mathbf{z} \in \text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$, i.e.,

$$\mathbf{z} = \sum_{i=1}^N \theta_i \mathbf{s}_i, \quad \mathbf{1}^T \boldsymbol{\theta} = 1, \quad \boldsymbol{\theta} \succeq \mathbf{0}$$

implying that $\mathbf{z} \in \text{aff}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$. From (A1), we have $\mathbf{s}_i \succeq \mathbf{0} \forall i$ and subsequently $\mathbf{z} \succeq \mathbf{0}$. This completes the proof for (20).

D. Proof of Lemma 3

Any point in $\mathcal{S} = \text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ can be equivalently represented by $\mathbf{s} = \sum_{i=1}^N \theta_i \mathbf{s}_i$, where $\boldsymbol{\theta} \succeq \mathbf{0}$ and $\boldsymbol{\theta}^T \mathbf{1} = 1$. Applying this result to (22), problem (22) can be reformulated as

$$\begin{aligned} \min_{\boldsymbol{\theta} \in \mathbb{R}^N} \quad & \sum_{i=1}^N \theta_i \rho_i \\ \text{s.t.} \quad & \boldsymbol{\theta}^T \mathbf{1} = 1, \quad \boldsymbol{\theta} \succeq \mathbf{0}. \end{aligned} \quad (58)$$

where $\rho_i = \mathbf{r}^T \mathbf{s}_i$. We assume without loss of generality that $\rho_1 < \rho_2 \leq \dots \leq \rho_N$. If $\rho_1 < \rho_2 < \dots < \rho_N$, then it is easy to verify that the optimal solution to (58) is uniquely given by $\boldsymbol{\theta}^* = \mathbf{e}_1$. In its counterpart in (22), this translates into $\mathbf{s}^* = \mathbf{s}_1$. But when $\rho_1 = \rho_2 = \dots = \rho_P$ and $\rho_P < \rho_{P+1} \leq \dots \leq \rho_N$ for some P , the solution of (58) is not unique. In essence, the latter case can be shown to have a solution set

$$\Theta = \{\boldsymbol{\theta} \mid \boldsymbol{\theta}^T \mathbf{1} = 1, \quad \boldsymbol{\theta} \succeq \mathbf{0}, \quad \theta_{P+1} = \dots = \theta_N = 0\}. \quad (59)$$

We now prove that the non-unique solution case happens with probability zero. Suppose that $\rho_i = \rho_j$ for some $i \neq j$, which means that

$$(\mathbf{s}_i - \mathbf{s}_j)^T \mathbf{r} = 0. \quad (60)$$

Let $v = (\mathbf{s}_i - \mathbf{s}_j)^T \mathbf{r}$. Apparently, v follows a distribution $\mathcal{N}(0, \|\mathbf{s}_i - \mathbf{s}_j\|^2)$. Since $\mathbf{s}_i \neq \mathbf{s}_j$, the probability $\Pr[\rho_i = \rho_j] = \Pr[v = 0]$ is of measure zero. This in turn implies that $\rho_1 < \rho_2 < \dots < \rho_N$ holds with probability 1.

E. Proof of Lemma 4

The approach to proving Lemma 4 is similar to that in Lemma 3. Let

$$\rho_i = \mathbf{r}^T \mathbf{s}_i = (\mathbf{B}\mathbf{w})^T \mathbf{s}_i \quad (61)$$

for which we have $\rho_i = 0$ for $i = 1, \dots, l$. It can be shown that

$$\rho_{l+1} < \rho_{l+2} < \dots < \rho_N \quad (62)$$

holds with probability 1, as long as $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ is linearly independent. Problems (22) and (24) are respectively equivalent to

$$\begin{aligned} p^* &= \min_{\boldsymbol{\theta} \in \mathbb{R}^N} \sum_{i=l+1}^N \theta_i \rho_i \\ \text{s.t. } \quad &\boldsymbol{\theta} \succeq \mathbf{0}, \quad \boldsymbol{\theta}^T \mathbf{1} = 1, \end{aligned} \quad (63)$$

$$\begin{aligned} q^* &= \max_{\boldsymbol{\theta} \in \mathbb{R}^N} \sum_{i=l+1}^N \theta_i \rho_i \\ \text{s.t. } \quad &\boldsymbol{\theta} \succeq \mathbf{0}, \quad \boldsymbol{\theta}^T \mathbf{1} = 1. \end{aligned} \quad (64)$$

Assuming (62), we have three distinct cases to consider: (C1) $\rho_{l+1} < 0$, $\rho_N < 0$, (C2) $\rho_{l+1} < 0$, $\rho_N > 0$, and (C3) $\rho_{l+1} > 0$, $\rho_N > 0$.

For (C2), we can see the following: Problem (63) has a unique optimal solution $\boldsymbol{\theta}^* = \mathbf{e}_{l+1}$ [and $\mathbf{s}^* = \mathbf{s}_{l+1}$ in its counterpart in (22)], attaining an optimal value $p^* = \rho_{l+1} < 0$. Problem (64) has a unique optimal solution $\boldsymbol{\theta}^* = \mathbf{e}_N$ [and $\mathbf{s}^* = \mathbf{s}_N$ in its counterpart in (24)], attaining an optimal value $q^* = \rho_N > 0$. In other words, both (63) and (64) lead to finding of new extreme points. For (C1), it is still true that (63) finds a new extreme point with $p^* < 0$. However, problem (64) is shown to have a solution set

$$\Theta = \{\boldsymbol{\theta} \mid \boldsymbol{\theta}^T \mathbf{1} = 1, \boldsymbol{\theta} \succeq \mathbf{0}, \theta_{l+1} = \dots = \theta_N = 0\}. \quad (65)$$

which contains convex combinations of the old extreme points, and the optimal value is $q^* = 0$. A similar condition happens with (C3), where (64) finds a new extreme point with $q^* > 0$ while (63) does not with $p^* = 0$.

F. Proof of Lemma 5

Equation (28) can also be expressed as

$$\mathcal{F} = \{ \boldsymbol{\alpha} \in \mathbb{R}^{N-1} \mid \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \in \text{conv}\{\mathbf{s}_1, \dots, \mathbf{s}_N\} \}.$$

Thus, every $\boldsymbol{\alpha} \in \mathcal{F}$ satisfies

$$\mathbf{C}\boldsymbol{\alpha} + \mathbf{d} = \sum_{i=1}^N \theta_i \mathbf{s}_i \quad (66)$$

for some $\boldsymbol{\theta} \succeq \mathbf{0}$, $\boldsymbol{\theta}^T \mathbf{1} = 1$. Since \mathbf{C} has full column rank, (66) can be re-expressed as

$$\boldsymbol{\alpha} = \sum_{i=1}^N \theta_i \boldsymbol{\alpha}_i, \quad (67)$$

where $\boldsymbol{\alpha}_i = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T (\mathbf{s}_i - \mathbf{d})$ (or $\mathbf{C}\boldsymbol{\alpha}_i + \mathbf{d} = \mathbf{s}_i$). Equation (67) implies that $\mathcal{F} = \text{conv}\{\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N\}$. Now, assume that $\{\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N\}$ are affinely dependent, i.e., there must exist an extreme point $\boldsymbol{\alpha}_N = \sum_{i=1}^{N-1} \gamma_i \boldsymbol{\alpha}_i$ where $\sum_{i=1}^{N-1} \gamma_i = 1$. One then has $\mathbf{s}_N = \mathbf{C}\boldsymbol{\alpha}_N + \mathbf{d} = \sum_{i=1}^{N-1} \gamma_i \mathbf{s}_i$ where $\sum_{i=1}^{N-1} \gamma_i = 1$ which implies $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ are affinely dependent (contradiction). Thus, the set \mathcal{F} is an $(N-1)$ -simplex.

REFERENCES

- [1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: John Wiley, 2001.
- [2] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. John Wiley and Sons, Inc., 2002.
- [3] E. R. Malinowski, *Factor Analysis in Chemistry*. New York: John Wiley, 2002.
- [4] D. Nuzillard and J.-M. Nuzillard, "Application of blind source separation to 1-D and 2-D nuclear magnetic resonance spectroscopy," *IEEE Trans. Signal Process. Lett.*, vol. 5, no. 8, pp. 209–211, Feb. 1998.
- [5] J. M. P. Nascimento and J. M. B. Dias, "Does independent component analysis play a role in unmixing hyperspectral data?" *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 1, pp. 175–187, Jan. 2005.
- [6] Y. Wang, J. Xuan, R. Srikanchana, and P. L. Choyke, "Modeling and reconstruction of mixed functional and molecular patterns," *International Journal of Biomedical Imaging*, vol. 2006, pp. 1–9, 2006.
- [7] M. D. Plumbley, "Algorithms for non-negative independent component analysis," *IEEE Trans. Neural Netw.*, vol. 14, no. 3, pp. 534–543, 2003.
- [8] 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.
- [9] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, "Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4133–4145, Nov. 2006.

- [10] R. Tauler and B. Kowalski, "Multivariate curve resolution applied to spectral data from multiple runs of an industrial process," *Anal. Chem.*, vol. 65, pp. 2040–2047, 1993.
- [11] D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, Oct. 1999.
- [12] A. Belouchrani, K. Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.
- [13] A. Hyvärinen and E. Oja, "A fixed-point algorithm for independent component analysis," *Neur. Comput.*, vol. 9, pp. 1483–1492, 1997.
- [14] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Netw.*, vol. 13, no. 6, pp. 1450–1464, Nov. 2002.
- [15] C. Lawson and R. J. Hanson, *Solving Least-Squares Problems*. New Jersey: Prentice-Hall, 1974.
- [16] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.
- [17] W.-K. Ma, T. N. Davidson, K. M. Wong, Z.-Q. Luo, and P. C. Ching, "Quasi-maximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA," *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 912–922, April 2002.
- [18] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, 2003.
- [19] Z.-Q. Luo, T. N. Davidson, G. B. Giannakis, and K. M. Wong, "Transceiver optimization for block-based multiple access through ISI channels," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1037–1052, 2004.
- [20] Y. Ding, T. N. Davidson, Z.-Q. Luo, and K. M. Wong, "Minimum BER block precoders for zero-forcing equalization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2410–2423, 2003.
- [21] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, 2006.
- [22] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, 2006.
- [23] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, 2003.
- [24] P. Biswas, T.-C. Lian, T.-C. Wang, and Y. Ye, "Semidefinite programming based algorithms for sensor network localization," *ACM Trans. Sensor Networks*, vol. 2, no. 2, pp. 188–220, 2006.
- [25] F.-Y. Wang, Y. Wang, T.-H. Chan, and C.-Y. Chi, "Blind separation of multichannel biomedical image patterns by non-negative least-correlated component analysis," in *Lecture Notes in Bioinformatics (Proc. PRIB'06)*, Springer-Verlag, vol. 4146, Berlin, Dec. 9-14, 2006, pp. 151–162.
- [26] 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 *IEEE International Workshop on Machine Learning for Signal Processing (MLSP'06)*, Maynooth, Ireland, Sept. 6-8, 2006, pp. 73–78.
- [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] A. T. Erdogan, "A simple geometric blind source separation method for bound magnitude sources," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 438–449, 2006.

- [29] F. Vrins, J. A. Lee, and M. Verleysen, "A minimum-range approach to blind extraction of bounded sources," *IEEE Trans. Neural Netw.*, vol. 18, no. 3, pp. 809–822, 2006.
- [30] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge Univ. Press, 2004.
- [31] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Athena Scientific, 2003.
- [32] B. Grünbaum, *Convex Polytopes*. Springer, 2003.
- [33] M. E. Dyer, "The complexity of vertex enumeration methods," *Mathematics of Operations Research*, vol. 8, no. 3, pp. 381–402, 1983.
- [34] K. G. Murty and S.-J. Chung, "Extreme point enumeration," College of Engineering, University of Michigan, Technical Report 92-21, 1992, available online: <http://deepblue.lib.umich.edu/handle/2027.42/6731>.
- [35] K. Fukuda, T. M. Liebling, and F. Margot, "Analysis of backtrack algorithms for listing all vertices and all faces of a convex polyhedron," *Computational Geometry: Theory and Applications*, vol. 8, no. 1, pp. 1–12, 1997.
- [36] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, 1999.
- [37] I. J. Lustig, R. E. Marsten, and D. F. Shanno, "Interior point methods for linear programming: Computational state of the art," *ORSA Journal on Computing*, vol. 6, no. 1, pp. 1–14, 1994.
- [38] G. H. Golub and C. F. V. Loan, *Matrix Computations*. The Johns Hopkins University Press, 1996.
- [39] P. Tichavský and Z. Koldovský, "Optimal pairing of signal components separated by blind techniques," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 119–122, 2004.
- [40] J. R. Hoffman and R. P. S. Mahler, "Multitarget miss distance via optimal assignment," *IEEE Trans. System, Man, and Cybernetics*, vol. 34, no. 3, pp. 327–336, May 2004.
- [41] H. W. Kuhn, "The Hungarian method for the assignment method," *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, 1955.
- [42] S. G. Armato, "Enhanced visualization and quantification of lung cancers and other diseases of the chest," *Experimental Lung Res.*, vol. 30, no. 30, pp. 72–77, 2004.
- [43] 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. [Online]. Available: <http://suzukilab.uchicago.edu/research.htm>
- [44] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1985.