BLIND SEPARATION OF NON-NEGATIVE SOURCES BY CONVEX ANALYSIS: EFFECTIVE METHOD USING LINEAR PROGRAMMING

Tsung-Han Chan†, Wing-Kin Ma†, Chong-Yung Chi†, and Yue Wang‡*

†Institute Commun. Eng.
National Tsinghua University
Hsinchu, Taiwan
E-mail: cychi@ee.nthu.edu.tw

*Dept. Electronic Eng.,
Chinese Univ. Hong Kong
Shatin, N.T., Hong Kong
E-mail: wkma@ieee.org

‡ Dept. Elect., Computer, & Biomed. Eng.
Virginia Tech. Institute & State Univ.
Arlington, VA22303, USA
E-mail: yuewang@vt.edu

ABSTRACT

We recently reported a criterion for blind separation of non-negative sources, using a new concept called convex analysis for mixtures of non-negative sources (CAMNS). Under some assumptions that are considered realistic for sparse or high-contrast signals, the criterion is that the true source signals can be perfectly recovered by finding the extreme points of some observation-constructed convex set. In our last work we also developed methods for fulfilling the CAMNS criterion, but only for two to three sources. In this paper we propose a systematic linear programming (LP) based method that is applicable to any number of sources. The proposed method has two advantages. First, its dependence on LP means that the method does not suffer from local minima. Second, the maturity of LP solvers enables efficient implementation of the proposed method in practice. Simulation results are provided to demonstrate the efficacy of the proposed method.

Index Terms— Blind separation, Non-negative sources, Convex analysis criterion, Linear program

1. INTRODUCTION

The problem of blind separation of non-negative sources, or non-negative blind source separation (nBSS), has received wide attention in a variety of fields, such as analytical chemistry [1], hyperspectral imaging [2], and biomedical imaging [3]. Existing methods for solving the nBSS problem usually adopt the statistical assumption that the sources are mutually uncorrelated or independent; e.g., non-negative independent component analysis (nICA) [4] which assumes uncorrelated sources, and Bayesian positive source separation (BPSS) [5] which assumes independent sources. Recently, some nBSS approaches requiring no assumption on source independence or zero correlations have emerged. One such nBSS approach is the non-negative matrix factorization (NMF) [6]. It decomposes the observation matrix as a product of two non-negative matrices, one serving as the estimate of the sources while the other the mixing matrix. NMF, however, may be a non-unique decomposition and some remedies have been suggested [7]. Here we are interested in another deterministic approach proposed by us recently, called CAMNS [8, 9]. CAMNS adopts a deterministic assumption called *local dominance*. This assumption was initially proposed to capture the sparse characteristics of biomedical images [10], but we found it a good as-

sumption or approximation for high contrast images as well; e.g., the human face separation example in Section 5. Under the local dominant assumption and some standard nBSS assumptions, we proved that the true source signals can be perfectly recovered by finding the extreme points of an observation-constructed polyhedral set. We also developed extreme-point search methods for CAMNS in our last published work [8]. However, those previously proposed methods can handle up to three sources only.

In this paper we propose an extreme-point search method that fulfils the CAMNS criterion for any number of sources. The idea is to use LP to systematically locate all the extreme pints (which are the true sources). As we will elaborate upon, the proposed LP-based method does not suffer from local minima and can be implemented efficiently. Our simulation results will show that this CAMNS-LP method has promising separation performance.

2. SYSTEM MODEL

Consider a noise-free linear mixing signal model

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

where $\mathbf{s}[n] = [\ s_1[n], \ldots, s_N[n]\ ]^T$ is the source vector, $\mathbf{x}[n] = [\ x_1[n], \ldots, x_M[n]\ ]^T$ is the observation vector, $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the unknown mixing matrix, and L is the data length and we assume $L \gg \max\{M, N\}$. Note that (1) can be alternatively expressed as

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

where a_{ij} is the (i,j)th element of $\mathbf{A}$, $\boldsymbol{s}_j = [\ s_j[1], \ldots, s_j[L]\ ]^T$ is the jth source signal vector and $\boldsymbol{x}_i = [\ x_i[1], \ldots, x_i[L]\ ]^T$ is the ith observed signal vector. The CAMNS criterion to be presented is based on the following assumptions:

(A1) All $\boldsymbol{s}_j$ are componentwise non-negative; i.e., for each j, $\boldsymbol{s}_j \in \mathbb{R}_+^L$ (a set of non-negative real L-vectors).

(A2) Each source signal vector is *locally dominant*, the definition of which is as follows: For each $i \in \{1, \ldots, N\}$, there exists an (unknown) index ℓ_i such that $s_i[\ell_i] > 0$ and $s_j[\ell_i] = 0$, $\forall j \neq i$.

(A3) The mixing matrix has unit row sum; i.e., for all $i = 1, \ldots, M$,

$$\sum_{j=1}^{N} a_{ij} = 1. \tag{3}$$

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

Assumptions (A1) and (A4) are standard in nBSS [4]. Assumption (A2) is special and instrumental to CAMNS. For high-contrast sources or sparse sources which contain many zeros, (A2) may be completely satisfied or serve as a good approximation. Assumption (A3) is automatically satisfied in MRI due to the partial volume effect [10], and in hyperspectral images due to the full additivity condition [2]. When (A3) is not satisfied, the normalization procedure in [10] can be used to enforce (A3).

3. CONVEX ANALYSIS OF MIXTURES OF NON-NEGATIVE SOURCES: THEORY

The purpose of this section is to provide a concise, self-contained description to CAMNS [8,9]. Of particular significance is the nBSS criterion derived from CAMNS, which will be stated in Theorem 2.

3.1. Some Basic Concepts of Convex Analysis

Before proceeding to describing CAMNS, it is useful to review several basic results in convex analysis [11]. Given a set of vectors $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} \subset \mathbb{R}^L$ (a set of real L-vectors), the *affine hull* is defined as

$$\mathrm{aff}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} = \left\{ \boldsymbol{x} = \sum_{i=1}^{N} \theta_i \boldsymbol{s}_i \ \middle|\ \boldsymbol{\theta} \in \mathbb{R}^N, \mathbf{1}^T \boldsymbol{\theta} = 1 \right\}, \quad (4)$$

where $\boldsymbol{\theta} = [\,\theta_1, \ldots, \theta_N\,]^T$ and $\mathbf{1}$ is an all-one vector. An affine hull can be represented by a polyhedral set, in form of

$$\mathrm{aff}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} = \left\{ \boldsymbol{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \ \middle|\ \boldsymbol{\alpha} \in \mathbb{R}^P \right\} \quad (5)$$

for some (non-unique) $\mathbf{d} \in \mathbb{R}^L$ and full column rank $\mathbf{C} \in \mathbb{R}^{L \times P}$, where P is the affine dimension which must be less than N.

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

$$\mathrm{conv}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} = \left\{ \boldsymbol{x} = \sum_{i=1}^{N} \theta_i \boldsymbol{s}_i \ \middle|\ \boldsymbol{\theta} \in \mathbb{R}^N_+, \mathbf{1}^T \boldsymbol{\theta} = 1 \right\}. \quad (6)$$

A point $\boldsymbol{x} \in \mathrm{conv}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ is an extreme point of $\mathrm{conv}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ if $\boldsymbol{x}$ cannot be a nontrivial convex combination of $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N$, (more specifically, $\boldsymbol{x} \neq \sum_{i=1}^{N} \theta_i \boldsymbol{s}_i$ 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, where $\mathbf{e}_i$ is an $N \times 1$ unit vector with the ith entry equal to 1).

A situation particularly relevant to this work is when $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ is linearly independent. In this situation, the affine dimension of $\mathrm{aff}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ is $P = N - 1$. Moreover, $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ is the set of extreme points of $\mathrm{conv}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$. To provide some insights into the concepts above, Fig. 1 shows the geometry of an affine hull and convex hull for $N = 3$.

3.2. New nBSS Criterion by CAMNS

Let us turn our attention back to the nBSS problem stated in Section 2, with the convex analysis concepts incorporated. From (A2), it can be shown that the true source vector set $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ is linearly independent. Based on the affine hull concepts described above, the source affine hull $\mathrm{aff}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ can be represented by

$$\mathrm{aff}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} = \left\{ \boldsymbol{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \ \middle|\ \boldsymbol{\alpha} \in \mathbb{R}^{N-1} \right\} \triangleq \mathcal{A}(\mathbf{C}, \mathbf{d}) \quad (7)$$

$$\mathrm{aff}\{\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3\} = \{\boldsymbol{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} | \boldsymbol{\alpha} \in \mathbb{R}^2\}$$
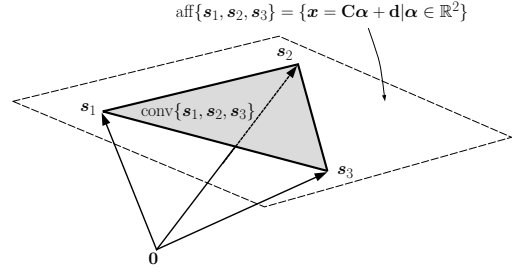
Fig. 1. Example of 3-dimensional signal space geometry for $N = 3$.

for some $(\mathbf{C}, \mathbf{d}) \in \mathbb{R}^{L \times (N-1)} \times \mathbb{R}^L$ such that $\mathrm{rank}(\mathbf{C}) = N - 1$. In addition, from (2) and assumption (A3), each observation $\boldsymbol{x}_i$ is seen to be an affine combination of $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$; i.e.,

$$\boldsymbol{x}_i \in \mathcal{A}(\mathbf{C}, \mathbf{d}) \quad (8)$$

for all $i = 1, \ldots, M$. The first key ingredient of CAMNS is identification of the source affine hull parameters $(\mathbf{C}, \mathbf{d})$ from the observations $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$. Consider the following theorem:

Theorem 1. (Source affine set construction [8]) *Under* (A2) *to* (A4)*, the observation affine hull is identical to the source affine hull:*

$$\mathcal{A}(\mathbf{C}, \mathbf{d}) = \mathrm{aff}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}. \quad (9)$$

Moreover, $(\mathbf{C}, \mathbf{d})$ *can be obtained from* $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ *by the following closed-form solution*

$$\mathbf{d} = \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{x}_i, \quad (10)$$

$$\mathbf{C} = [\, \boldsymbol{q}_1(\mathbf{U}\mathbf{U}^T), \boldsymbol{q}_2(\mathbf{U}\mathbf{U}^T), \ldots, \boldsymbol{q}_{N-1}(\mathbf{U}\mathbf{U}^T)\,], \quad (11)$$

where $\mathbf{U} = [\, \boldsymbol{x}_1 - \mathbf{d}, \ldots, \boldsymbol{x}_M - \mathbf{d}\,] \in \mathbb{R}^{L \times M}$*, and* $\boldsymbol{q}_i(\mathbf{R})$ *denotes the eigenvector associated with the ith principal eigenvalue of* $\mathbf{R}$*.*

We should add that the above source affine set closed-form solution is based on an optimization that finds an affine set that yields the best fitting with respect to the observations [8].

Recall that the source signals are non-negative. Hence, we have $\boldsymbol{s}_i \in \mathrm{aff}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} \cap \mathbb{R}^L_+$ for any i. Let us define

$$\mathcal{S} = \mathrm{aff}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} \cap \mathbb{R}^L_+ = \mathcal{A}(\mathbf{C}, \mathbf{d}) \cap \mathbb{R}^L_+ \quad (12)$$

$$= \{\boldsymbol{x} \mid \boldsymbol{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d}, \ \boldsymbol{x} \succeq \mathbf{0}, \ \boldsymbol{\alpha} \in \mathbb{R}^{N-1}\} \quad (13)$$

(where $\succeq$ is the componentwise inequality), which can be seen to be a polyhedral set. The second important ingredient of CAMNS, leading to a new nBSS criterion, is as follows:

Theorem 2. (CAMNS criterion [8]) *Under* (A1) *and* (A2)*, the set* $\mathcal{S}$ *in* (13) *is also the source convex hull; that is,*

$$\mathcal{S} = \mathrm{conv}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}. \quad (14)$$

Moreover, $\mathcal{S}$ *has N extreme points given by the true source vectors* $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N$*.*

The theoretical implication of Theorem 2 is profound: It suggests that the true source vectors can be perfectly identified by finding all the extreme points of $\mathcal{S}$. Hence, if we are able to develop realizable methods for fulfilling the CAMNS criterion we can achieve perfect blind separation in practice. There are simple methods of doing this when N equals 2 or 3; see [8]. The next section considers the situation where N is arbitrary.

4. LINEAR PROGRAMMING METHOD FOR CAMNS

We now describe the main contribution of this paper, namely the systematic LP-based method for fulfilling the CAMNS criterion.

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

$$p^\star = \min_{\boldsymbol{s}} \ \mathbf{r}^T \boldsymbol{s}$$
$$\text{subject to (s.t.)} \quad \boldsymbol{s} \in \mathcal{S} \tag{15}$$

for some arbitrarily chosen direction $\mathbf{r} \in \mathbb{R}^L$, where $p^\star$ denotes the optimal objective value of (15). Using the polyhedral structure of $\mathcal{S}$ in (13), problem (15) can be equivalently represented by an LP

$$p^\star = \min_{\boldsymbol{\alpha}} \ \mathbf{r}^T (\mathbf{C}\boldsymbol{\alpha} + \mathbf{d})$$
$$\text{s.t.} \quad \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \succeq \mathbf{0}. \tag{16}$$

A fundamental result in LP theory is that $\mathbf{r}^T \boldsymbol{s}$, the objective function of (15), attains the minimum at a point of the boundary of $\mathcal{S}$. To provide more insights, some geometric illustrations are given in Fig. 2. We can see that the solution of (15) may be uniquely given by one of the extreme points $\boldsymbol{s}_i$ [Fig. 2(a)], or it may be any point on a face [Fig. 2(b)]. The latter case poses a trouble to our task of identifying $\boldsymbol{s}_i$, but it is arguably not a usual situation. For instance, in the demonstration in Fig. 2(b), $\mathbf{r}$ must be normal to $\boldsymbol{s}_2 - \boldsymbol{s}_3$ which may be unlikely to happen for a randomly picked $\mathbf{r}$. With this intuition in mind, we prove in the Appendix that

Lemma 1. *Suppose that* $\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_L)$ *(i.e., $\mathbf{r}$ being Gaussian distributed with zero mean and covariance matrix equal to $L \times L$ identity matrix). Then, with probability 1, the solution of (15) is uniquely given by $\boldsymbol{s}_i$ for some $i \in \{1, ..., N\}$.*

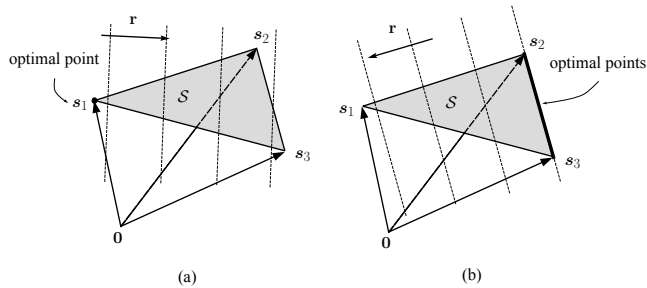The idea behind Lemma 1 is to show that undesired cases, such as that in Fig. 2(b) happen with probability zero.

Fig. 2. Geometric interpretation of an LP.

We may find another extreme point by solving the maximization counterpart of (15)

$$q^\star = \max_{\boldsymbol{\alpha}} \ \mathbf{r}^T (\mathbf{C}\boldsymbol{\alpha} + \mathbf{d})$$
$$\text{s.t.} \quad \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \succeq \mathbf{0}. \tag{17}$$

Using the same idea as above, we can show the following: Under the premise of Lemma 1, the solution of (17) is, with probability 1, uniquely given by an extreme point $\boldsymbol{s}_i$ different from that in (15).

Suppose that we have identified l extreme points, say, without loss of generality, $\{\boldsymbol{s}_1, ..., \boldsymbol{s}_l\}$. Our interest is to refine the above LP extreme-point finding procedure such that the search space is restricted to $\{\boldsymbol{s}_{l+1}, ..., \boldsymbol{s}_N\}$. To do so, consider a thin QR decomposition [13] of $[\boldsymbol{s}_1, ..., \boldsymbol{s}_l]$

$$[\boldsymbol{s}_1, ..., \boldsymbol{s}_l] = \mathbf{Q}_1 \mathbf{R}_1, \tag{18}$$

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

$$\mathbf{B} = \mathbf{I}_L - \mathbf{Q}_1 \mathbf{Q}_1^T. \tag{19}$$

We assume that $\mathbf{r}$ takes the form

$$\mathbf{r} = \mathbf{B}\boldsymbol{w} \tag{20}$$

for some $\boldsymbol{w} \in \mathbb{R}^L$, and consider solving (16) and (17) with such an $\mathbf{r}$. Since $\mathbf{r}$ is orthogonal to the old extreme points $\boldsymbol{s}_1, ..., \boldsymbol{s}_l$, the intuitive expectation is that (16) and (17) should both lead to new extreme points. Interestingly, we found theoretically that expectation is not true, but close. Consider the following lemma:

Lemma 2. *Suppose that* $\mathbf{r} = \mathbf{B}\boldsymbol{w}$*, where* $\mathbf{B} \in \mathbb{R}^{L \times L}$ *satisfies (19) and* $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_L)$*. Then, with probability 1, at least one of the optimal solutions of (16) and (17) is a new extreme point; i.e., $\boldsymbol{s}_i$ for some $i \in \{l + 1, ..., N\}$. The certificate of finding new extreme points is indicated by $|p^\star| \neq 0$ for (16), and $|q^\star| \neq 0$ for (17).*

Lemma 2 is proven using the same concept as that in the Appendix. We omit the proof due to lack of space here, and its details will be given in [9]. By repeating the above described procedures, we can identify all the extreme points $\boldsymbol{s}_1, ..., \boldsymbol{s}_N$. The resultant CAMNS-LP method is summarized in the following steps:

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 $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_L)$, and set $\mathbf{r} := \mathbf{B}\boldsymbol{w}$.

Step 3. Solve the LPs (16) and (17), and obtain their optimal solutions, denoted by $\boldsymbol{\alpha}_1^\star$ and $\boldsymbol{\alpha}_2^\star$, respectively.

Step 4. If $l = 0$, then $\hat{\mathbf{S}} = [\ \mathbf{C}\boldsymbol{\alpha}_1^\star + \mathbf{d}, \ \mathbf{C}\boldsymbol{\alpha}_2^\star + \mathbf{d}\ ]$,
else
If $|p^\star| \neq 0$, then $\hat{\mathbf{S}} := [\ \hat{\mathbf{S}} \quad \mathbf{C}\boldsymbol{\alpha}_1^\star + \mathbf{d}\ ]$,
If $|q^\star| \neq 0$, then $\hat{\mathbf{S}} := [\ \hat{\mathbf{S}} \quad \mathbf{C}\boldsymbol{\alpha}_2^\star + \mathbf{d}\ ]$.

Step 5. Update l as 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** to **Step 6** until $l = N$.

Let us consider the implementation issues of the above proposed method, which depends on those of LPs. It is well known that LPs do not suffer from local minima. Moreover, the LPs we encounter [(16) or (17)] can be solved effectively by interior-point algorithms, 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 $L \gg N$ [12]. 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 2, Lemma 1, Lemma 2, and the above discussion, we assert that

Proposition 1. *Under (A1)-(A4), the CAMNS-LP method finds all the true source vectors $\boldsymbol{s}_1, ..., \boldsymbol{s}_N$ with probability 1. It does so with a worst-case complexity of $\mathcal{O}(L^{1.5}(N-1)^2)$.*
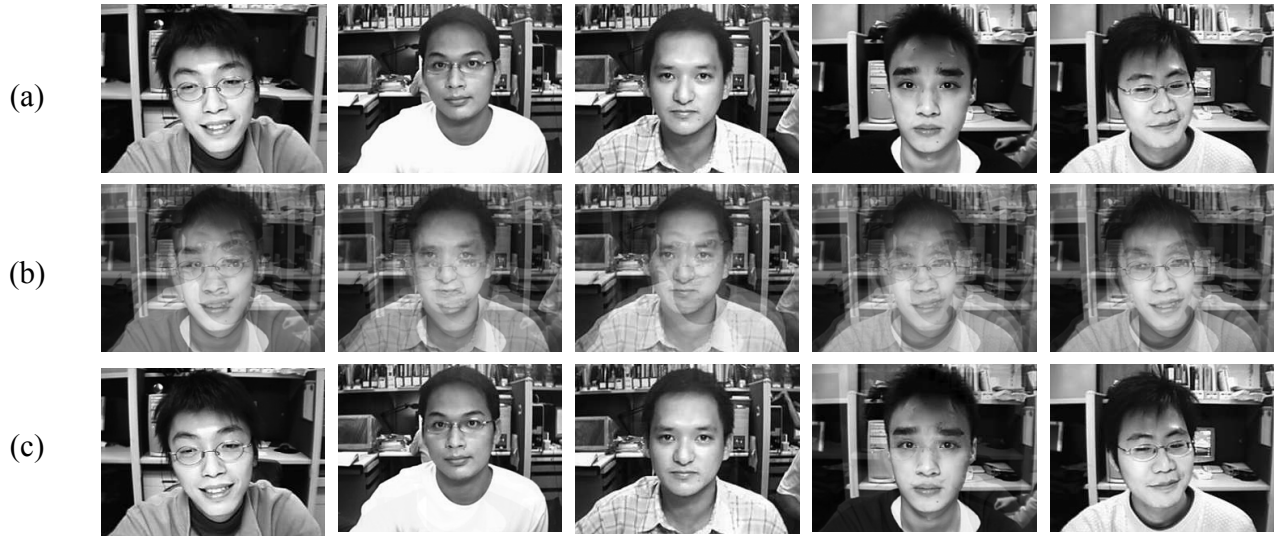
Fig. 3. Human face images: (a) the sources, (b) the observations, and (c) the extracted sources obtained by CAMNS-LP.

5. SIMULATIONS AND CONCLUSIONS

We synthetically generated 5 mixtures [Fig. 3(b)] from 5 human face images [Fig. 3(a)]. The extracted sources obtained by CAMNS-LP is displayed in Fig. 3(c). One can see that the extracted sources are very similar to the original, with separation residuals only being slightly noticeable in the 4th image.

Interested readers can visit `http://www.ee.cuhk.edu.hk/~wkma/CAMNS/CAMNS.htm`, where comparisons with nICA [4] and NMF [6] can be found. The website also provides the MAT-LAB source codes of CAMNS-LP.

In conclusion, we have presented a systematic LP-based method for realizing the nBSS criterion by CAMNS. The proposed method uses LPs to find the true source signals, the process of which is immune to local minima. Moreover, the method is efficient in the sense that its worst-case complexity is of the order of $L^{1.5}$, where L is the data length. Simulation results give a good validation of the blind separability of the proposed method.

6. APPENDIX: PROOF OF LEMMA 1

Any point in $\mathcal{S} = \mathrm{conv}\{s_1, ..., s_N\}$ can be equivalently represented by $s = \sum_{i=1}^{N} \theta_i s_i$, where $\theta \succeq 0$ and $\theta^T \mathbf{1} = 1$. Then problem (15) can be reformulated as

$$\min_{\theta \in \mathbb{R}^N} \quad \sum_{i=1}^{N} \theta_i \rho_i$$
$$\text{s.t.} \quad \theta^T \mathbf{1} = 1, \; \theta \succeq 0, \tag{21}$$

where $\rho_i = \mathbf{r}^T s_i$. We assume without loss of generality that $\rho_1 < \rho_2 \leq \cdots \leq \rho_N$. If $\rho_1 < \rho_2 < \cdots < \rho_N$, then it is easy to verify that the optimal solution to (21) is uniquely given by $\theta^\star = \mathbf{e}_1$. In its counterpart in (15), this translates into $s^\star = s_1$. But when $\rho_1 = \rho_2 = \cdots = \rho_P$ and $\rho_P < \rho_{P+1} \leq \cdots \leq \rho_N$ for some P, the solution of (21) is not unique. In the latter case, the solution set is $\Theta = \{\theta \mid \theta^T \mathbf{1} = 1, \; \theta \succeq 0, \; \theta_{P+1} = \cdots = \theta_N = 0\}$.

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

$$(s_i - s_j)^T \mathbf{r} = 0. \tag{22}$$

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

7. REFERENCES

[1] E. R. Malinowski, *Factor Analysis in Chemistry*. New York: John Wiley, 2002.

[2] J. M. P. Nascimento, J. M. Bioucas 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.

[3] 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.

[4] M. D. Plumbley and E. Oja, "A non-negative PCA algorithm for independent component analysis," *IEEE Trans. Neural Computation*, vol. 16, no. 1, pp. 1811-1825, 2004.

[5] 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, 2006.

[6] D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788-791, Oct. 1999.

[7] P. O. Hoyer and A. Hyvärinen, "A multi-layer sparse coding network learns contour coding from natural images," *Vision Research*, vol. 42, no. 12, pp. 1593-1605, 2002.

[8] T.-H. Chan, W.-K. Ma, C.-Y. Chi and Y. Wang, " A convex analysis based criterion for blind separation of non-negative sources," *IEEE ICASSP-2007*, Honolulu, Hawaii, April 15-20, 2007, pp. 961-964.

[9] T.-H. Chan, W.-K. Ma, C.-Y. Chi and Y. Wang, "A convex analysis framework for blind separation of non-negative sources," submitted to *IEEE Trans. Signal Process.*, July 2007. Revised, Dec. 2007.

[10] 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," *IEEE International Workshop on Machine Learning for Signal Processing*, Maynooth, Ireland, Sept. 6-8, 2006, pp. 73-78.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, 2004.

[12] 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.

[13] G. H. Golub and C. F. V. Loan, *Matrix Computations*. The Johns Hoplkins University Press, 1996.