A CONVEX ANALYSIS BASED CRITERION FOR BLIND SEPARATION OF NON-NEGATIVE SOURCES

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

[†]Institute Commun. Eng. & Dept. Elect. Eng.
National Tsing Hua University
Hsinchu, Taiwan
E-mail: {wkma,cychi}@ee.nthu.edu.tw

[‡]Dept. Elect., Computer, & Biomed. Eng.
Virginia Polytechnic Institute & State University
Arlington, VA 22203 USA
E-mail: yuewang@vt.edu

ABSTRACT

In this paper, we apply convex analysis to the problem of blind source separation (BSS) of non-negative signals. Under realistic assumptions applicable to many real-world problems such as multichannel biomedical imaging, we formulate a new BSS criterion that does not require statistical source independence, a fundamental assumption to many existing BSS approaches. The new criterion guarantees perfect separation (in the absence of noise), by constructing a convex set from the observations and then finding the extreme points of the convex set. Some experimental results are provided to demonstrate the efficacy of the proposed method.

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

1. INTRODUCTION

The problem of independent component analysis (ICA) or blind source separation (BSS) has received wide attention in various fields such as biomedical signal analysis, speech enhancement, wireless communications and image processing [1] [2]. Most existing ICA or BSS algorithms were developed under certain assumptions on the sources; e.g., that the sources are statistically independent and non-Gaussian. For instance, the fast fixed-point algorithm (also called the FastICA) [1] is based on the assumption that the sources are mutually independent with non-zero kurtosis (a fourth-order cumulant). The non-negative ICA (nICA) [3] is based on the assumption that sources are non-negative and mutually uncorrelated. However, this fundamental assumption, source independence, becomes problematic in many real world applications, such as biomedical imaging [4] and hyperspectral imaging [5] where sources can be correlated. Thus, many researchers exploit the non-negativity of both the sources and mixing matrix to develop their BSS algorithms. The non-negative matrix factorization (NMF) [6] can be used to decompose the observation matrix as a product of two non-negative matrices (representing the source matrix and mixing matrix). However, the solutions to the NMF may be non-unique. Moreover, a method proposed by Wang *et al.* [7] [8], called the non-negative least-correlated component analysis (nLCA), extracts original (independent or correlated) sources by minimizing the correlation coefficient between two estimated sources.

This paper presents a convex analysis approach to BSS of non-negative sources that are allowed to be dependent. The main contribution is formulation of a new BSS criterion, which can guarantee

perfect separation in the absence of noise under several mild assumptions that usually hold in biomedical imaging applications. A BSS method based on the new criterion will also be discussed.

2. SYSTEM MODEL

Consider an $M \times N$ mixing system with inputs and outputs denoted by $\mathbf{s}[n] = [\, s_1[n], \ldots, s_N[n] \,]^T$ and $\mathbf{x}[n] = [\, x_1[n], \ldots, x_M[n] \,]^T$ respectively. Assume that the noise is absent. Then the mixing signal model is given by

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

where $\mathbf{A}$ is the mixing matrix which is unknown, and L is the data length and $L \gg \max\{M, N\}$. Our BSS approach to be presented is based on the following assumptions which are motivated by magnetic resonance imaging (MRI) applications:

(A1) (Non-negative source signals) For all n, $\mathbf{s}[n] \in \mathbb{R}_+^N$ where $\mathbb{R}_+$ is the set of non-negative real numbers.

(A2) For each source index $i \in \{1, \ldots, N\}$, there exists a time index ℓ_i such that $s_i[\ell_i] > 0$ and $s_j[\ell_i] = 0, \forall j \neq i$.

(A3) (Unit row sum) For all $i = 1, \ldots, M$,

$$\sum_{j=1}^{N} a_{ij} = 1 \qquad (2)$$

where a_{ij} denotes the (i, j) entry of $\mathbf{A}$.

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

(A5) (Non-negative mixing matrix) $\mathbf{A} \in \mathbb{R}_+^{M \times N}$.

(A1) and (A5) are generally true in biomedical imaging [3] where image intensities and the mixing process are represented by non-negative numbers. (A2) usually holds for sparse signals such as brain MRI in biomedical imaging, where the non-overlapping region of the spatial distribution of a fast perfusion and a slow perfusion source images [4] can be higher than 95%, thereby making (A2) an advisable assumption. (A3) can be achieved by a straightforward preprocessing [8], and it is automatically satisfied in MRI due to the partial volume effect [8]. (A4) is an assumption widely made in the BSS problem.

For notational convenience, we will focus on the following alternate signal model. Let $\boldsymbol{s}_i = [\, s_i[1], \ldots, s_i[L] \,]^T$ and $\boldsymbol{x}_i = [\, x_i[1], \ldots, x_i[L] \,]^T$. The model in (1) can be rewritten as

$$\boldsymbol{x}_i = \sum_{j=1}^{N} a_{ij} \boldsymbol{s}_j, \; i = 1, \ldots, M. \qquad (3)$$

3. NEW BSS CRITERION BY CONVEX ANALYSIS

In this section, we first review some convex analysis concepts essential to this work, and then present our main results.

3.1. Some Basic Concepts of Convex Analysis

We review two convex analysis concepts [9], [10] that will play an important role to the ensuing development. The first is the *affine hull* of a set of vectors $\{s_1, \ldots, s_N\} \subset \mathbb{R}^L$, defined as

$$\text{aff}\{s_1, \ldots, s_N\} = \left\{ x = \sum_{i=1}^{N} \theta_i s_i \;\middle|\; \theta \in \mathbb{R}^N, \sum_{i=1}^{N} \theta_i = 1 \right\}. \tag{4}$$

For $N = 2$, the affine hull is a line passing through s_1 and s_2; see Fig. 1(a). The affine hull can always be represented by an affine set:

$$\text{aff}\{s_1, \ldots, s_N\} = \left\{ x = \mathbf{C}\alpha + \mathbf{d} \;\middle|\; \alpha \in \mathbb{R}^P \right\} \tag{5}$$

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 represents the dimension of the affine set which must be less than N. For example, if $\{s_1, \ldots, s_N\}$ is a linearly independent set, then $P = N - 1$. In that case, a legitimate $(\mathbf{C}, \mathbf{d})$ can simply be obtained by $\mathbf{C} = [\, s_1 - s_N, s_2 - s_N, \ldots, s_{N-1} - s_N \,]$ and $\mathbf{d} = s_N$. In the ensuing subsections we will be faced with the problem of finding $(\mathbf{C}, \mathbf{d})$ from $\{s_1, \ldots, s_N\}$ for $P < N - 1$.

The second key concept is that of the *convex hull* of a set of vectors $\{s_1, \ldots, s_N\} \subset \mathbb{R}^L$, defined as

$$\text{conv}\{s_1, \ldots, s_N\} = \left\{ x = \sum_{i=1}^{N} \theta_i s_i \;\middle|\; \theta \in \mathbb{R}_+^N, \sum_{i=1}^{N} \theta_i = 1 \right\}. \tag{6}$$

For $N = 2$, the convex hull is a line segment between s_1 and s_2; see Fig. 1(a). Geometrically $s_1, \ldots, s_N$ would be the 'corner points' of its convex hull, defined formally as the *extreme points*. A point $x \in \text{conv}\{s_1, \ldots, s_N\}$ is an extreme point of $\text{conv}\{s_1, \ldots, s_N\}$ if it cannot be a nontrivial convex combination of $s_1, \ldots, s_N$; i.e.,

$$x \neq \sum_{i=1}^{N} \theta_i s_i \tag{7}$$

for all $\theta \in \mathbb{R}_+^N$, $\sum_{i=1}^{N} \theta_i = 1$, and $\theta \neq \mathbf{e}_i$ for any i (where $\mathbf{e}_i$ is the unit vector with the ith entry being equal to 1). The set of extreme points of $\text{conv}\{s_1, \ldots, s_N\}$ must be either the full set or a subset of $\{s_1, \ldots, s_N\}$. In addition, if $\{s_1, \ldots, s_N\}$ is a linearly independent set, then the set of extreme points of $\text{conv}\{s_1, \ldots, s_N\}$ is exactly $\{s_1, \ldots, s_N\}$.
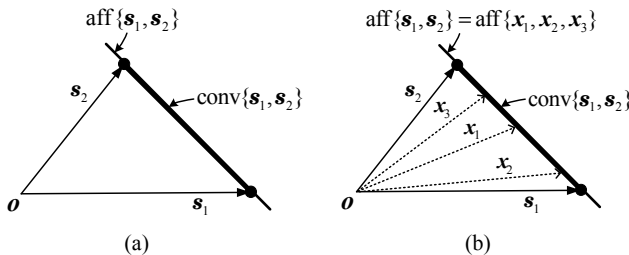

Fig. 1. Example of 2-dimensional signal space geometry for $N = 2$.

3.2. Signal Affine Hull Characterization

With the aforedescribed convex analysis concepts, we see from the signal formulation in (3) and (A3) that

$$x_i \in \text{aff}\{s_1, \ldots, s_N\} \tag{8}$$

for all $i = 1, \ldots, M$. This leads to an interesting question whether the observations $x_1, \ldots, x_M$ provide sufficient information to determine the source signal affine hull $\text{aff}\{s_1, \ldots, s_N\}$. This is indeed possible, as described in the following lemma:

Lemma 1. *Suppose that* (A3) *and* (A4) *hold. Then,*

$$\text{aff}\{s_1, \ldots, s_N\} = \text{aff}\{x_1, \ldots, x_M\}. \tag{9}$$

The proof of Lemma 1 is omitted due to the page limit. Fig. 1(b) demonstrates geometrically the validity of Lemma 1. We will show later that given knowledge of $\text{aff}\{s_1, \ldots, s_N\}$, it is possible to perfectly recover all $s_1, \ldots, s_N$. Now let us focus on characterization of $\text{aff}\{s_1, \ldots, s_N\}$. It can easily be shown from (A2) that $\{s_1, \ldots, s_N\}$ is linearly independent. Hence, $\text{aff}\{s_1, \ldots, s_N\}$ has dimension $N - 1$ and admits a representation

$$\text{aff}\{s_1, \ldots, s_N\} = \left\{ x = \mathbf{C}\alpha + \mathbf{d} \;\middle|\; \alpha \in \mathbb{R}^{N-1} \right\} \triangleq \mathcal{A}(\mathbf{C}, \mathbf{d}) \tag{10}$$

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 be a semi-unitary matrix, i.e., $\mathbf{C}^T \mathbf{C} = \mathbf{I}$.

It is easy to obtain $(\mathbf{C}, \mathbf{d})$ from the observations $x_1, \ldots, x_M$ if $M = N$; see the example mentioned in Section 3.1. The nontrivial case lies in $M > N$, for which a method, referred to as the affine set fitting, is proposed next.

Consider the following projection problem

$$\min_{\substack{\tilde{\mathbf{C}}, \tilde{\mathbf{d}} \\ \tilde{\mathbf{C}}^T \tilde{\mathbf{C}} = \mathbf{I}}} \sum_{i=1}^{M} \| x_i - \mathcal{P}_{\mathcal{A}(\tilde{\mathbf{C}}, \tilde{\mathbf{d}})}(x_i) \|_2^2 \tag{11}$$

where $\|.\|_2$ denotes the 2-norm; and $\mathcal{P}_{\mathcal{A}}(x)$ stands for the projection of x onto the set $\mathcal{A}$, defined as

$$\mathcal{P}_{\mathcal{A}}(x) = \arg \min_{\tilde{x} \in \mathcal{A}} \| x - \tilde{x} \|_2^2. \tag{12}$$

The idea behind is to find an $(N - 1)$-dimensional affine set that has the minimum projection error with respect to the observations. At first glance, (11) is a nonconvex optimization problem, but in fact it can be solved analytically as stated in the following proposition:

Proposition 1. *The affine set fitting problem in (11) has a closed form solution*

$$\mathbf{d} = \frac{1}{M} \sum_{i=1}^{M} x_i \tag{13}$$

$$\mathbf{C} = [\, q_1(\mathbf{U}\mathbf{U}^T), q_2(\mathbf{U}\mathbf{U}^T), \ldots, q_{N-1}(\mathbf{U}\mathbf{U}^T) \,] \tag{14}$$

where $\mathbf{U} = [\, x_1 - \mathbf{d}, \ldots, x_M - \mathbf{d} \,] \in \mathbb{R}^{L \times M}$, *and* $q_i(\mathbf{R})$ *denotes the eigenvector associated with the ith principal eigenvalue of* $\mathbf{R}$.

The proof of Proposition 1 is given in Section 7. We should add that the affine set fitting solution in Proposition 1 works for $M = N$ as well. Moreover, in the presence of noise the affine set fitting would serve as a favorable way for noise reduction, since it provides a best affine set that achieves the minimum data fitting error.

3.3. Signal Convex Hull, Extreme Points, and Perfect Blind Separation

Recall that we are dealing with non-negative source signals. Hence,

$$\boldsymbol{s}_i \in \text{aff}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} \cap \mathbb{R}_+^L \triangleq \mathcal{S} \tag{15}$$

for any i. Now, we can show that

Lemma 2. *Suppose that* (A1) *and* (A2) *hold. Then,*

$$\mathcal{S} = \text{conv}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} \tag{16}$$

and the set of all its extreme points is $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$.

The proof of Lemma 2 is given in Section 7. Lemma 2 gives the important implication that perfect blind separation can be done by identifying the extreme points of the signal convex set $\mathcal{S}$.

Given the observations, we only know from (10) and (15) that $\mathcal{S}$ is of the form

$$\begin{aligned}\mathcal{S} &= \left\{ \boldsymbol{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \mid \boldsymbol{\alpha} \in \mathbb{R}^{N-1} \right\} \cap \mathbb{R}_+^L \\ &= \left\{ \boldsymbol{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \mid \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \succeq \mathbf{0}, \boldsymbol{\alpha} \in \mathbb{R}^{N-1} \right\} \end{aligned} \tag{17}$$

for some $\mathbf{C}$ and $\mathbf{d}$ estimated from the observations (see Proposition 1), and $\succeq$ is the elementwise inequality. Let $\mathcal{F}$ be the image of $\mathcal{S}$ under the mapping from $\boldsymbol{x}$ to $\boldsymbol{\alpha}$, i.e.,

$$\begin{aligned}\mathcal{F} &= \left\{ \boldsymbol{\alpha} \in \mathbb{R}^{N-1} \mid \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \succeq \mathbf{0} \right\} \\ &= \left\{ \boldsymbol{\alpha} \in \mathbb{R}^{N-1} \mid \mathbf{c}_n^T \boldsymbol{\alpha} + d_n \geq 0, \ n = 1, \ldots, L \right\} \end{aligned} \tag{18}$$

where $\mathbf{c}_n^T$ is the nth row of $\mathbf{C}$. We prove the correspondence between the extreme points of $\mathcal{S}$ and $\mathcal{F}$ in the following lemma:

Lemma 3. *The set* $\mathcal{F}$ *given by (18) is equivalent to*

$$\mathcal{F} = \text{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\} \tag{19}$$

where each $\boldsymbol{\alpha}_i \in \mathbb{R}^{N-1}$ *is such that* $\mathbf{C}\boldsymbol{\alpha}_i + \mathbf{d} = \boldsymbol{s}_i$. *It has N extreme points, given by* $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N$.

The proof of Lemma 3 is given in Section 7. More importantly, in convex analysis there is a powerful result that can help us identify the extreme points of $\mathcal{F}$:

Lemma 4. (Extreme point validation for polyhedra [10]) *A point* $\boldsymbol{\alpha} \in \mathcal{F}$ *is an extreme point of* $\mathcal{F}$ *if and only if the following collection of vectors*

$$\mathcal{C}_{\boldsymbol{\alpha}} = \left\{ \mathbf{c}_n \in \mathbb{R}^{N-1} \mid \mathbf{c}_n^T \boldsymbol{\alpha} = -d_n, \ n = 1, \ldots, L \right\} \tag{20}$$

contains $N-1$ linearly independent vectors.

Summarizing the results developed above, we have the following new blind separation criterion for non-negative sources:

Theorem 1. (New BSS criterion) *Suppose that* (A1) *to* (A4) *hold. The set*

$$\mathcal{F} = \left\{ \boldsymbol{\alpha} \in \mathbb{R}^{N-1} \mid \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \succeq \mathbf{0} \right\} \tag{21}$$

where $(\mathbf{C}, \mathbf{d})$ *are obtained from the observations* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$ *via the affine set fitting solution in Proposition 1, has N extreme points* $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N$. *Each extreme point corresponds to a source signal through* $\mathbf{C}\boldsymbol{\alpha}_i + \mathbf{d} = \boldsymbol{s}_i$. *The extreme points of* $\mathcal{F}$ *can be identified by using the validation procedure in Lemma 4.*

4. BSS METHOD BASED ON THE NEW CRITERION

Based on the BSS criterion in Theorem 1, a BSS method which works by identifying the extreme points of $\mathcal{F}$ is proposed in this section. Let us focus on the case of $N = 2$ and the case of $N = 3$, respectively. By (A1) to (A5), one can see that $\mathbf{d} = \sum_{i=1}^{M} \boldsymbol{x}_i / M \succeq \mathbf{0}$.

Case A: $N = 2$. It can be shown that there is a closed form solution

$$\alpha_1 = \min_n \{-d_n/c_n \mid c_n < 0, \ n = 1, 2, ..., L\},$$
$$\alpha_2 = \max_n \{-d_n/c_n \mid c_n > 0, \ n = 1, 2, ..., L\}.$$

Case B: $N = 3$. In Fig. 2, we can see that there are three faces forming the boundary of $\mathcal{F}$, and that identifying any 2 faces will be sufficient to determine one extreme point. An extreme point search algorithm is then as follows.

Step 1. Given an arbitrary vector $\mathbf{r}_1 \in \mathbb{R}^2$ and setting $\mathbf{r}_2 = -\mathbf{r}_1$, find the indices associated with two of the faces by

$$k_i = \arg \min_n \{-d_n/\mathbf{c}_n^T \mathbf{r}_i \mid \mathbf{c}_n^T \mathbf{r}_i < 0, \ n = 1, ..., L\} \tag{22}$$
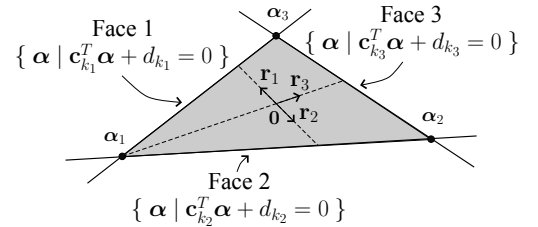
for $i = 1, 2$.

Step 2. Obtain the first extreme point

$$\boldsymbol{\alpha}_1 = - \begin{bmatrix} \mathbf{c}_{k_1}^T \\ \mathbf{c}_{k_2}^T \end{bmatrix}^{-1} \begin{bmatrix} d_{k_1} \\ d_{k_2} \end{bmatrix} \in \mathbb{R}^2.$$

Step 3. Set $\mathbf{r}_3 = -\boldsymbol{\alpha}_1$. Find the third face by (22) for $i = 3$.

Step 4. Obtain the other two extreme points

$$\boldsymbol{\alpha}_2 = - \begin{bmatrix} \mathbf{c}_{k_2}^T \\ \mathbf{c}_{k_3}^T \end{bmatrix}^{-1} \begin{bmatrix} d_{k_2} \\ d_{k_3} \end{bmatrix}, \boldsymbol{\alpha}_3 = - \begin{bmatrix} \mathbf{c}_{k_1}^T \\ \mathbf{c}_{k_3}^T \end{bmatrix}^{-1} \begin{bmatrix} d_{k_1} \\ d_{k_3} \end{bmatrix}.$$

Fig. 2. A geometry of finding extreme points for $N = 3$.

5. EXPERIMENTS

In this experiment, we perform fifty independent runs (i.e., use fixed sources to generate the mixtures with random mixing matrices) with the proposed method and three existing algorithms, nLCA [8], nICA [3], and NMF [6] for performance comparison. Assume that $\hat{\boldsymbol{s}}_j$ is the estimated source of $\boldsymbol{s}_j$ (with permutation ambiguity artificially removed). The mean-removed correlation coefficient between the true sources $\boldsymbol{s}_j$ and the extracted sources $\hat{\boldsymbol{s}}_j$ defined as [8]

$$\tilde{\rho} = \frac{1}{N} \sum_{j=1}^{N} \frac{|(\boldsymbol{s}_j - \frac{1}{L}\boldsymbol{s}_j^T \mathbf{1})^T (\hat{\boldsymbol{s}}_j - \frac{1}{L}\hat{\boldsymbol{s}}_j^T \mathbf{1})|}{\| \boldsymbol{s}_j - \frac{1}{L}\boldsymbol{s}_j^T \mathbf{1} \|_2 \cdot \| \hat{\boldsymbol{s}}_j - \frac{1}{L}\hat{\boldsymbol{s}}_j^T \mathbf{1} \|_2}$$

was calculated at each independent run as a performance index where $\mathbf{1}$ denotes a column vector with all the entries equal to unity.

Note that $0 \leq \tilde{\rho} \leq 1$ and the larger $\tilde{\rho}$, the better the performance of the algorithm under test. Three correlated human face image sources ($N = 3$) taken from the benchmarks in [2] were used to generate three noise-free observations and fifteen noisy observations. Note that we add zero-mean Gaussian noise with variance σ_i^2 to produce noisy observations with the same signal-to-noise ratio (SNR=SNR$_i$=25dB) where SNR$_i = \parallel x_i \parallel_2^2 / L\sigma_i^2$. To maintain non-negativity of the observations in the simulation, we force the negative noisy observations to zero. The averaged values of $\tilde{\rho}$ over the performed fifty independent runs are displayed in Table 1. From this table, one can see that proposed method performs better than the other three existing algorithms (i.e., maximum averaged $\tilde{\rho}$).

Method ($N = 3$)	Proposed Method	nLCA	nICA	NMF
SNR=∞ $M = 3$	0.9872	0.9811	0.7519	0.7406
SNR=25 dB $M = 15$	0.9336	0.9116	0.7311	0.7153

Table 1. The performance (averaged mean-removed correlation coefficient $\tilde{\rho}$) of the proposed method, nLCA, nICA and NMF for the human face image simulation.

6. CONCLUSIONS AND FUTURE WORKS

We have developed a new blind separation criterion for non-negative sources by a convex analysis approach. This approach is deterministic and geometry-based, requiring no source independence assumption. Based on the new criterion (Theorem 1), a BSS method (for $N = 2$ and $N = 3$) has been proposed. Some experimental results reveal that the proposed BSS method outperforms nLCA, nICA, and NMF. As future directions, we shall consider extension to the case of $N > 3$ and investigate relaxations of the assumptions made in the paper.

7. APPENDIX

Proof of Proposition 1. From (12), it can be shown that

$$\mathcal{P}_{\mathcal{A}(\tilde{\mathbf{C}},\tilde{\mathbf{d}})}(x_i) = \arg \min_{\substack{\tilde{x}=\tilde{\mathbf{C}}\tilde{\alpha}+\tilde{\mathbf{d}} \\ \tilde{\alpha}\in\mathbb{R}^{N-1}}} \|x - \tilde{x}\|_2^2 = \mathbf{P}_{\tilde{c}}x + \mathbf{P}_{\tilde{c}}^{\perp}\tilde{\mathbf{d}} \quad (23)$$

where $\mathbf{P}_{\tilde{c}} = \tilde{\mathbf{C}}\tilde{\mathbf{C}}^T$ and $\mathbf{P}_{\tilde{c}}^{\perp} = \mathbf{I} - \mathbf{P}_{\tilde{c}}$. By substituting (23) into (11), the problem becomes

$$\min_{\substack{\tilde{\mathbf{C}},\tilde{\mathbf{d}} \\ \tilde{\mathbf{C}}^T\tilde{\mathbf{C}}=\mathbf{I}}} \sum_{i=1}^{M} \|\mathbf{P}_{\tilde{c}}^{\perp}(x_i - \tilde{\mathbf{d}})\|_2^2 \triangleq \min_{\substack{\tilde{\mathbf{C}},\tilde{\mathbf{d}} \\ \tilde{\mathbf{C}}^T\tilde{\mathbf{C}}=\mathbf{I}}} J(\tilde{\mathbf{C}},\tilde{\mathbf{d}}). \quad (24)$$

Fixing $\tilde{\mathbf{C}}$, $\min_{\tilde{\mathbf{d}}} J(\tilde{\mathbf{C}},\tilde{\mathbf{d}})$ is a convex optimization problem. By solving $\nabla_{\tilde{\mathbf{d}}} J(\tilde{\mathbf{C}},\tilde{\mathbf{d}}) = 0$ for $\tilde{\mathbf{d}}$ with a given $\tilde{\mathbf{C}}$, we obtain an optimal solution $\mathbf{d}$ given by (13). By substituting (13) into (24), the problem is reduced to

$$\max_{\tilde{\mathbf{C}}^T\tilde{\mathbf{C}}=\mathbf{I}} \sum_{i=1}^{M} \|\mathbf{P}_c(x_i - \mathbf{d})\|_2^2 = \max_{\tilde{\mathbf{C}}^T\tilde{\mathbf{C}}=\mathbf{I}} \text{Trace}\{\tilde{\mathbf{C}}^T\mathbf{U}\mathbf{U}^T\tilde{\mathbf{C}}\}.$$

which is the problem of finding the $N - 1$ principal eigenvalues of $\mathbf{U}\mathbf{U}^T$. Hence, we obtain an optimum solution $\mathbf{C}$ given by (14).

Proof of Lemma 2. Assume that $z \in \text{aff}\{s_1, ..., s_N\} \cap \mathbb{R}_+^L$:

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

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, z lies in $\text{conv}\{s_1, ..., s_N\}$. On the other hand, assume that $z \in \text{conv}\{s_1, ..., s_N\}$, i.e.,

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

implying that $z \in \text{aff}\{s_1, ..., s_N\}$. From (A1), we have $s_i \succeq \mathbf{0} \, \forall i$ and subsequently $z \succeq \mathbf{0}$. This completes the proof for (16). Moreover, the fact that $s_1, ..., s_N$ are linearly independent (according to (A2)) implies that $\{s_1, ..., s_N\}$ is the set of all the extreme points of $\text{conv}\{s_1, ..., s_N\}$ (cf., Sec. 3.1).

Proof of Lemma 3. Equation (18) can also be expressed as

$$\mathcal{F} = \left\{ \alpha \in \mathbb{R}^{N-1} \mid \mathbf{C}\alpha + \mathbf{d} \in \text{conv}\{s_1, ..., s_N\} \right\}.$$

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

$$\mathbf{C}\alpha + \mathbf{d} = \sum_{i=1}^{N} \theta_i s_i. \quad (25)$$

for some $\theta \succeq \mathbf{0}$, $\theta^T\mathbf{1} = 1$. Now, due to the one-to-one mapping $s_i = \mathbf{C}\alpha_i + \mathbf{d} \in \mathcal{S}$ (between $\mathcal{S}$ and $\mathcal{F}$), one can infer that (25) is equivalent to

$$\alpha = \sum_{i=1}^{N} \theta_i \alpha_i, \quad \theta^T\mathbf{1} = 1, \quad \theta \succeq \mathbf{0} \quad (26)$$

meaning that $\mathcal{F} = \text{conv}\{\alpha_1, ..., \alpha_N\}$. Now, suppose that α_1 is not an extreme point, i.e., $\alpha_1 = \sum_{i=1}^{N} \theta_i \alpha_i$, $\theta \neq e_1$. Then one can easily show that $s_1 = \mathbf{C}\alpha_1 + \mathbf{d} = \sum_{i=1}^{N} \theta_i s_i$, $\theta \neq e_1$, i.e., s_1 is not an extreme point (contradiction). So, all α_i must be extreme points of $\mathcal{F}$.

8. REFERENCES

[1] A. Hyvarinen, 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. Oja and M. Plumbley, "Blind separation of positive sources by globally convergent gradient search," *Neural Computation*, vol. 16, pp. 1811-1825, 2004.

[4] 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.

[5] 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.

[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] 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," Lecture Notes in Bioinformatics (Proc. PRIB'06), LNBI 4146, pp. 151-162, Springer-Verlag, Berlin, 2006.

[8] 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 (MLSP'06)*, Maynooth, Ireland, Sept. 6-8, 2006, pp. 73-78.

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

[10] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Athena Scientific, 2003.