Wing-Kin (Ken) Ma

Dept. Electronic Eng.
The Chinese University of Hong Kong
Hong Kong

Institute Commun. Eng.
National Tsing Hua University
Taiwan


Acknowledgement to Tsung-Han Chan & Prof. Chong-Yung Chi.
1. Introduction to blind source separation (BSS) & non-negative BSS (nBSS)

2. Convex analysis of mixtures of non-negative sources (CAMNS)
   - Theory: a new nBSS criterion by CAMNS
   - Practical implementation of CAMNS: Systematic LP method

3. Simulation results & Conclusion
Outline

1. Introduction to blind source separation (BSS) & non-negative BSS (nBSS)

2. Convex analysis of mixtures of non-negative sources (CAMNS)
   - Theory: a new nBSS criterion by CAMNS
   - Practical implementation of CAMNS: Systematic LP method

3. Simulation results & Conclusion
Outline

1. Introduction to blind source separation (BSS) & non-negative BSS (nBSS)

2. Convex analysis of mixtures of non-negative sources (CAMNS)
   - Theory: a new nBSS criterion by CAMNS
   - Practical implementation of CAMNS: Systematic LP method

3. Simulation results & Conclusion
Blind source separation (BSS): Problem statement

**Signal model:** consider a real-valued, $N$-input, $M$-output linear mixing model:

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

where

$$\mathbf{x}_i = \begin{bmatrix} x_i[1] \\ \vdots \\ x_i[L] \end{bmatrix}, \quad \mathbf{s}_i = \begin{bmatrix} s_i[1] \\ \vdots \\ s_i[L] \end{bmatrix}$$

are observation & true source vectors.

**Problem:** extract $\{s_1, \ldots, s_N\}$ from $\{x_1, \ldots, x_M\}$ without information of the mixing matrix $\mathbf{A} = \{a_{ij}\}$. 

Blind Separation of Non-negative Sources using Convex Analysis: Theory and Methods

Wing-Kin Ma
BSS: A biomedical imaging example

**Figure:** Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) assessments of breast cancer captured at different times. Courtesy to Yue Wang [Wang et al. 2003]
Figure: Illustration of source pattern mixing process. The signals represent a summation of vascular permeability with various diffusion rates. The goal is to separate the distribution of multiple biomarkers with the same diffusion rate.
A BSS approach is based on some assumptions on the characteristics of \( \{s_1, \ldots, s_N\} \) and/or \( A \).

There are two aspects in developing a BSS approach:
- criterion established from the assumptions made, &
- optimization methods for fulfilling the criterion.

The suitability of the assumptions (\& the approach as a result) depends much on the applications under consideration.

Example:

**Independent component analysis (ICA)**, a well-known BSS technique, typically assumes that each \( s_i[n] \) is non-Gaussian random & is mutually independent of one other.

Mutual independence is a good assumption in speech \& wireless commun., but not so in hyperspectral imaging.
In some applications source signals are non-negative by nature; imaging.

nBSS approaches exploit the signal non-negativity characteristic (plus some additional assumptions).

**Applications:** biomedical imaging, hyperspectral imaging, & analytical chemistry.

**Some existing nBSS approaches:** non-negative ICA (nICA) [Plumbley 2003], & non-negative matrix factorization (NMF) [Lee-Seung 1999].

nICA is a statistical approach adopting the mutual independence assumption.

NMF is a deterministic approach that may cope with correlated sources. It may not be a unique factorization, however.
Outline

1. Introduction to blind source separation (BSS) & non-negative BSS (nBSS)

2. Convex analysis of mixtures of non-negative sources (CAMNS)
   - Theory: a new nBSS criterion by CAMNS
   - Practical implementation of CAMNS: Systematic LP method

3. Simulation results & Conclusion
CAMNS: Convex analysis of mixtures of non-negative sources

- CAMNS is a deterministic nBSS approach.
- In addition to utilizing source non-negativity, CAMNS employs a special deterministic assumption called local dominance.
- What is local dominance? Intuitively, signals with many ‘zeros’ are likely to satisfy local dominance (math. def. available soon).

- Appears to be a good assumption for sparse or high-contrast images.
An intuitive illustration of how CAMNS works

How can we extract \( \{s_1, \ldots, s_N\} \) from \( \{x_1, \ldots, x_M\} \) without knowing \( \{a_{ij}\} \)?
An intuitive illustration of how CAMNS works

Based on some assumptions (e.g., signal non-negativity & local dominance) & by convex analysis, we use \( \{x_1, \ldots, x_M\} \) to construct a polyhedral set (yellow).
We show that the ‘corners’ (formally speaking, extreme points) of this polyhedral set are exactly \( \{ s_1, \ldots, s_N \} \) (rather surprisingly).
An intuitive illustration of how CAMNS works

By using linear programs, we can locate the ‘corners’ of the polyhedral set effectively. As a result perfect separation will be achieved.
A quick review of some convex analysis concepts

### Affine Hull

Affine hull of a set of vectors \( \{s_1, \ldots, s_N\} \subset \mathbb{R}^L \):

\[
\text{aff}\{s_1, \ldots, s_N\} = \left\{ x = \sum_{i=1}^{N} \theta_i s_i \mid \theta \in \mathbb{R}^N, \sum_{i=1}^{N} \theta_i = 1 \right\}.
\]

- An affine hull can always be represented by:

\[
\text{aff}\{s_1, \ldots, s_N\} = \left\{ x = C\alpha + d \mid \alpha \in \mathbb{R}^P \right\}
\]

for some (non-unique) \( d \in \mathbb{R}^L \) and \( C \in \mathbb{R}^{L \times P} \), where \( P \leq N - 1 \) is the affine dimension.

- If \( \{s_1, \ldots, s_N\} \) is affine independent (or \( \{s_1 - s_N, \ldots, s_{N-1} - s_N\} \) is linearly independent) then \( P = N - 1 \).
**Convex Hull**

Convex hull of a set of vectors \( \{ s_1, \ldots, s_N \} \subset \mathbb{R}^L \):

\[
\text{conv}\{s_1, \ldots, s_N\} = \left\{ x = \sum_{i=1}^{N} \theta_i s_i \mid \theta \in \mathbb{R}_+^N, \sum_{i=1}^{N} \theta_i = 1 \right\}
\]

- A point \( x \in \text{conv}\{s_1, \ldots, s_N\} \) is an **extreme point** of \( \text{conv}\{s_1, \ldots, s_N\} \) if \( x \) is not any nontrivial convex combination of \( \{ s_1, \ldots, s_N \} \).

- If \( \{ s_1, \ldots, s_N \} \) is affine independent then \( \{ s_1, \ldots, s_N \} \) is the set of all extreme points of its convex hull.
aff\{s_1, s_2, s_3\} = \{x = C\alpha + d | \alpha \in \mathbb{R}^2\}

**Figure:** Example of 3-dimensional signal space geometry with $N = 3$. In this example, $\text{aff}\{s_1, s_2, s_3\}$ is a plane passing through $s_1, s_2, s_3$, & $\text{conv}\{s_1, s_2, s_3\}$ is a triangle with corners (extreme points) $s_1, s_2, s_3$. 
The assumptions in CAMNS

Recall the model $x_i = \sum_{j=1}^{M} a_{ij} s_j$. Our assumptions:

(A1) **Source non-negativity**: For each $j$, $s_j \in \mathbb{R}_+^L$.

(A2) **Local dominance**: For each $i \in \{1, \ldots, N\}$, there exists an (unknown) index $\ell_i$ such that $s_i[\ell_i] > 0$ and $s_j[\ell_i] = 0$, $\forall j \neq i$. (Reasonable assumption for sparse or high-contrast signals).

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

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

(Already satisfied in MRI and hyperspectral imaging, can be relaxed).

(A4) $M \geq N$ and $A$ is of full column rank. (Standard BSS assumption)
Since $\sum_{j=1}^{N} a_{ij} = 1$ \textit{[(A3)]}, we have for each observation
\[ x_i = \sum_{j=1}^{N} a_{ij} s_j \in \text{aff}\{s_1, \ldots, s_N\} \]
This implies
\[ \text{aff}\{s_1, \ldots, s_N\} \supseteq \text{aff}\{x_1, \ldots, x_M\}. \]
In fact, we can show that

\textbf{Lemma 1}

Under \textit{(A3)} and \textit{(A4)}, \[ \text{aff}\{s_1, \ldots, s_N\} = \text{aff}\{x_1, \ldots, x_M\}. \]
Consider the representation

\[
\text{aff}\{s_1, \ldots, s_N\} = \text{aff}\{x_1, \ldots, x_N\} = \left\{ x = C\alpha + d \mid \alpha \in \mathbb{R}^{N-1} \right\} \triangleq \mathcal{A}(C, d)
\]

for some \((C, d) \in \mathbb{R}^{L \times (N-1)} \times \mathbb{R}^L\) with \(\text{rank}(C) = N - 1\).

Let us consider determining the source affine set parameters \((C, d)\) from \(\{x_1, \ldots, x_M\}\).

The solution is simple for \(M = N\):

\[
d = x_N, \quad C = [x_1 - x_N, \ldots, x_{N-1} - x_N]
\]

For \(M > N\), we use an affine set fitting solution.
Affine set fitting problem:

\[(C, d) = \arg \min_{\tilde{C}, \tilde{d}} \sum_{i=1}^{M} e_{A(\tilde{C}, \tilde{d})}(x_i) \]  
\[ \tilde{C}^T \tilde{C} = I \]  

where \( e_A(x) = \min_{\tilde{x} \in A} \| x - \tilde{x} \|_2^2 \) is the projection error of \( x \) onto \( A \), & \( A(C, d) = \{ x = C\alpha + d \mid \alpha \in \mathbb{R}^{N-1} \} \).

**Proposition 1**

Problem (*) has a closed-form solution

\[ d = \frac{1}{M} \sum_{i=1}^{M} x_i, \quad C = [ q_1(UU^T), q_2(UU^T), \ldots, q_{N-1}(UU^T) ] \]

where \( U = [ x_1 - d, \ldots, x_M - d ] \in \mathbb{R}^{L \times M} \), and \( q_i(R) \) denotes the eigenvector associated with the \( i \)th principal eigenvalue of \( R \).
Be reminded that $s_i \in \mathbb{R}^L_+$. Hence, it is true that

$$s_i \in \text{aff}\{s_1, \ldots, s_N\} \cap \mathbb{R}^L_+ = A(C, d) \cap \mathbb{R}^L_+ \triangleq S$$

The following lemma arises from local dominance (A2):

**Lemma 2**

Under (A1) and (A2),

$$S = \text{conv}\{s_1, \ldots, s_N\}$$

Moreover, the set of all its extreme points is $\{s_1, \ldots, s_N\}$. 
Proof of Lemma 2:

\( \text{aff}\{s_1, \ldots, s_N\} \cap \mathbb{R}_+^L \subseteq \text{conv}\{s_1, \ldots, s_N\} : \)

- Every \( z \in \text{aff}\{s_1, \ldots, s_N\} \cap \mathbb{R}_+^L \) may be expressed as

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

- Due to (A2), for each \( i \) \( \exists \ell_i \) such that \( z[\ell_i] = \theta_i s_i[\ell_i] \geq 0 \).
- Since \( s_i[\ell_i] > 0, \theta_i \geq 0 \). \( \implies z \in \text{conv}\{s_1, \ldots, s_N\} \).

\( \text{aff}\{s_1, \ldots, s_N\} \cap \mathbb{R}_+^L \supseteq \text{conv}\{s_1, \ldots, s_N\} : \)

- Every \( z \in \text{conv}\{s_1, \ldots, s_N\} \) may be expressed as

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

which already implies \( z \in \text{aff}\{s_1, \ldots, s_N\} \).

- \( s_i \succeq 0 \ \forall i, \ \theta \succeq 0 \implies z \succeq 0. \)
From the above results, we obtain an nBSS criterion that is based on convex analysis & cannot be found in the other BSS literature to our best knowledge:

**Theorem 1 (nBSS criterion by CAMNS)**

Under (A1) to (A4), the polyhedral set

\[ S = \{ x \in \mathbb{R}^L \mid x = C\alpha + d \succeq 0, \ \alpha \in \mathbb{R}^{N-1} \} \]

where \((C, d)\) is obtained from the observation set \(\{x_1, \ldots, x_M\}\) by the affine set fitting procedure in Proposition 1, has \(N\) extreme points given by the true source vectors \(s_1, \ldots, s_N\).
Practical realization of CAMNS

- CAMNS boils down to finding all the extreme points of an observation-constructed polyhedral set.
- In the optimization context this is known as vertex enumeration.
- In CAMNS, there is one important problem structure that we can take full advantage of; that is,

Property (implied by (A2))
The extreme points $s_1, \ldots, s_N$ (or the true source vectors) are linear independent.

- By exploiting this property, we can locate all the extreme points by solving a sequence of LPs ($\approx 2N$ LPs at worst).
- Hence, this special vertex enumeration problem can be solved in polynomial time in both $L \& N$. 
Consider the following LP

$$p^* = \min_{s} r^T s$$

s.t. $s \in S$

(†)

for an arbitrary $r \in \mathbb{R}^L$. From basic LP theory, the solution of (†) is

- one of the extreme points of $S$ (that is, one of the $s_i$), or
- any point on a face of $S$ (look rather unlikely, intuitively).
We can prove that getting a non-extreme-pt. solution is very unlikely:

**Lemma 3**

Suppose that $\mathbf{r}$ is randomly generated following $\mathcal{N}(0, \mathbf{I}_L)$. Then, with probability 1, the solution of

$$p^* = \min_{\mathbf{s}} \mathbf{r}^T \mathbf{s}$$

s.t. $\mathbf{s} \in \mathcal{S}$

is uniquely given by $s_i$ for some $i \in \{1, \ldots, N\}$.
Suppose that we have found $l$ extreme point, say, $\{s_1, \ldots, s_l\}$.

We can find the other extreme points, by using the linear independence of $\{s_1, \ldots, s_N\}$ to ‘annihilate’ the old extreme points.

**Lemma 4**

Suppose $\mathbf{r} = \mathbf{Bw}$, where $\mathbf{w} \sim \mathcal{N}(0, \mathbf{I}_{L-l})$, & $\mathbf{B} \in \mathbb{R}^{L \times (L-l)}$ is such that

$$\mathbf{B}[s_1, \ldots, s_l] = 0 \quad \mathbf{B}^T \mathbf{B} = \mathbf{I}_{L-l}$$

Then, with probability 1, at least one of the LPs

$$p^* = \min_{s \in S} \mathbf{r}^T s \quad q^* = \max_{s \in S} \mathbf{r}^T s$$

finds a new extreme point; i.e., $s_i$ for some $i \in \{l + 1, \ldots, N\}$. The 1st LP finds a new extreme pt. if $|p^*| \neq 0$; the 2nd LP finds a new extreme pt. if $|q^*| \neq 0$. 

Wing-Kin Ma

Blind Separation of Non-negative Sources using Convex Analysis: Theory and
Remarks on alternatives of implementing CAMNS

- We have another theorem that converts $S \subset \mathbb{R}^L$ to another polyhedral set on $\mathbb{R}^{(N-1)}$, denoted by $\mathcal{F}$ here.

- The set $\mathcal{F}$ not only has a smaller vector dim. (note that $L \gg N$), but is also a simplex with extreme points related to those of $S$ in a one-to-one manner.

- For $N = 2$, $\mathcal{F}$ is a line segment on $\mathbb{R}$ and there is a closed form for locating its extreme points.

- For $N = 3$, $\mathcal{F}$ is a triangle on $\mathbb{R}^2$ and there is also a simple way for locating its extreme points.
Outline

1. Introduction to blind source separation (BSS) & non-negative BSS (nBSS)

2. Convex analysis of mixtures of non-negative sources (CAMNS)
   - Theory: a new nBSS criterion by CAMNS
   - Practical implementation of CAMNS: Systematic LP method

3. Simulation results & Conclusion
Simulation example 1: dual energy chest X-Ray

Original sources
Simulation example 1: dual energy chest X-Ray

Observations
Simulation example 1: dual energy chest X-Ray

Separated sources by CAMNS
Simulation example 1: dual energy chest X-Ray

Separated sources by \textbf{nICA} (a benchmarked nBSS method)
Simulation example 1: dual energy chest X-Ray

Separated sources by **NMF** (yet another benchmarked nBSS method)
Simulation example 2: human faces

Original sources
Simulation example 2: human faces

Observations
Simulation example 2: human faces

Separated sources by **CAMNS**
Simulation example 2: human faces

Separated sources by nICA
Simulation example 2: human faces

Separated sources by NMF
Simulation example 3: ghosting

Original sources
Simulation example 3: ghosting

Observations
Simulation example 3: ghosting

Separated sources by CAMNS
Simulation example 3: ghosting

Separated sources by \textit{nICA}
Simulation example 3: ghosting

Separated sources by **NMF**
Simulation example 4: five of my students

Original sources
Simulation example 4: five of my students

Observations
Simulation example 4: five of my students

Separated sources by CAMNS
Simulation example 5: Monte Carlo performance for $N = 3$

Average sum squared errors of the sources with respect to SNRs.
We have used convex analysis to develop a new approach for nBSS, called CAMNS.

CAMNS guarantees perfect separation of the true sources, by determining the extreme points of an observation constructed polyhedral set (under several assumptions).

A systematic LP-based method has been proposed to determine the true sources effectively in practice. Its complexity is polynomial (specifically, $O(L^{1.5}(N - 1)^2))$).

A number of simulation results indicate that CAMNS performs very well even in the presence of dependent sources.

Thank you!

We are going to release the source codes online soon. Please check out http://www.ee.cuhk.edu.hk/~wkma
Conclusion

- We have used convex analysis to develop a new approach for nBSS, called CAMNS.

- CAMNS guarantees perfect separation of the true sources, by determining the extreme points of an observation constructed polyhedral set (under several assumptions).

- A systematic LP-based method has been proposed to determine the true sources effectively in practice. Its complexity is polynomial (specifically, $O(L^{1.5}(N - 1)^2))$).

- A number of simulation results indicate that CAMNS performs very well even in the presence of dependent sources.

Thank you!

We are going to release the source codes online soon. Please check out http://www.ee.cuhk.edu.hk/~wkma
Conclusion

- We have used convex analysis to develop a new approach for nBSS, called CAMNS.

- CAMNS guarantees perfect separation of the true sources, by determining the extreme points of an observation constructed polyhedral set (under several assumptions).

- A systematic LP-based method has been proposed to determine the true sources effectively in practice. Its complexity is polynomial (specifically, $O(L^{1.5}(N - 1)^2)$).

- A number of simulation results indicate that CAMNS performs very well even in the presence of dependent sources.

Thank you!

We are going to release the source codes online soon. Please check out http://www.ee.cuhk.edu.hk/~wkma
Conclusion

- We have used convex analysis to develop a new approach for nBSS, called CAMNS.

- **CAMNS guarantees perfect separation of the true sources**, by determining the extreme points of an observation constructed polyhedral set (under several assumptions).

- **A systematic LP-based method has been proposed to determine the true sources effectively in practice.** Its complexity is polynomial (specifically, $O(L^{1.5}(N - 1)^2)$).

- A number of simulation results indicate that CAMNS performs very well even in the presence of dependent sources.

Thank you!

We are going to release the source codes online soon. Please check out [http://www.ee.cuhk.edu.hk/~wkma](http://www.ee.cuhk.edu.hk/~wkma)
Conclusion

- We have used convex analysis to develop a new approach for nBSS, called CAMNS.

- CAMNS guarantees perfect separation of the true sources, by determining the extreme points of an observation constructed polyhedral set (under several assumptions).

- A systematic LP-based method has been proposed to determine the true sources effectively in practice. Its complexity is polynomial (specifically, $O(L^{1.5}(N - 1)^2)$).

- A number of simulation results indicate that CAMNS performs very well even in the presence of dependent sources.

Thank you!

We are going to release the source codes online soon. Please check out http://www.ee.cuhk.edu.hk/~wkma
References


Appendix: Relaxation of (A3)

The unit row sum assumption (A3) may be relaxed.

Suppose that $x_i^T 1 \neq 0$ (where 1 is an all-one vector) for all $i$.

Consider a normalized version of $x_i$:

$$
\bar{x}_i = \frac{x_i}{x_i^T 1} = \sum_{j=1}^{N} \left( \frac{a_{ij} s_j^T 1}{x_i^T 1} \right) \left( \frac{s_j}{s_j^T 1} \right).
$$

One can show that $(\bar{a}_{ij})$ satisfies (A3).