Non-Negative Blind Source Separation using Convex Analysis

Wing-Kin (Ken) Ma
The Chinese University of Hong Kong (CUHK)

Course on Convex Optimization for Wireless Comm. and Signal Proc.
Jointly taught by Daniel P. Palomar and Wing-Kin (Ken) Ma
National Chiao Tung Univ., Hsinchu, Taiwan

December 19-21, 2008

Acknowledgement: Tsung-Han Chan, Chong-Yung Chi, and Yue Wang

Blind source separation (BSS): Problem statement

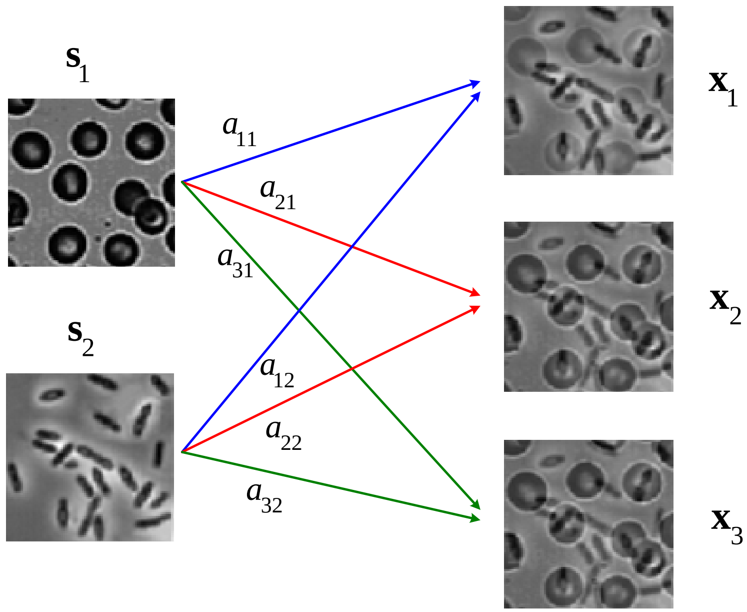
Signal model: a real-valued, N-input, M-output linear mixing model:

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

where

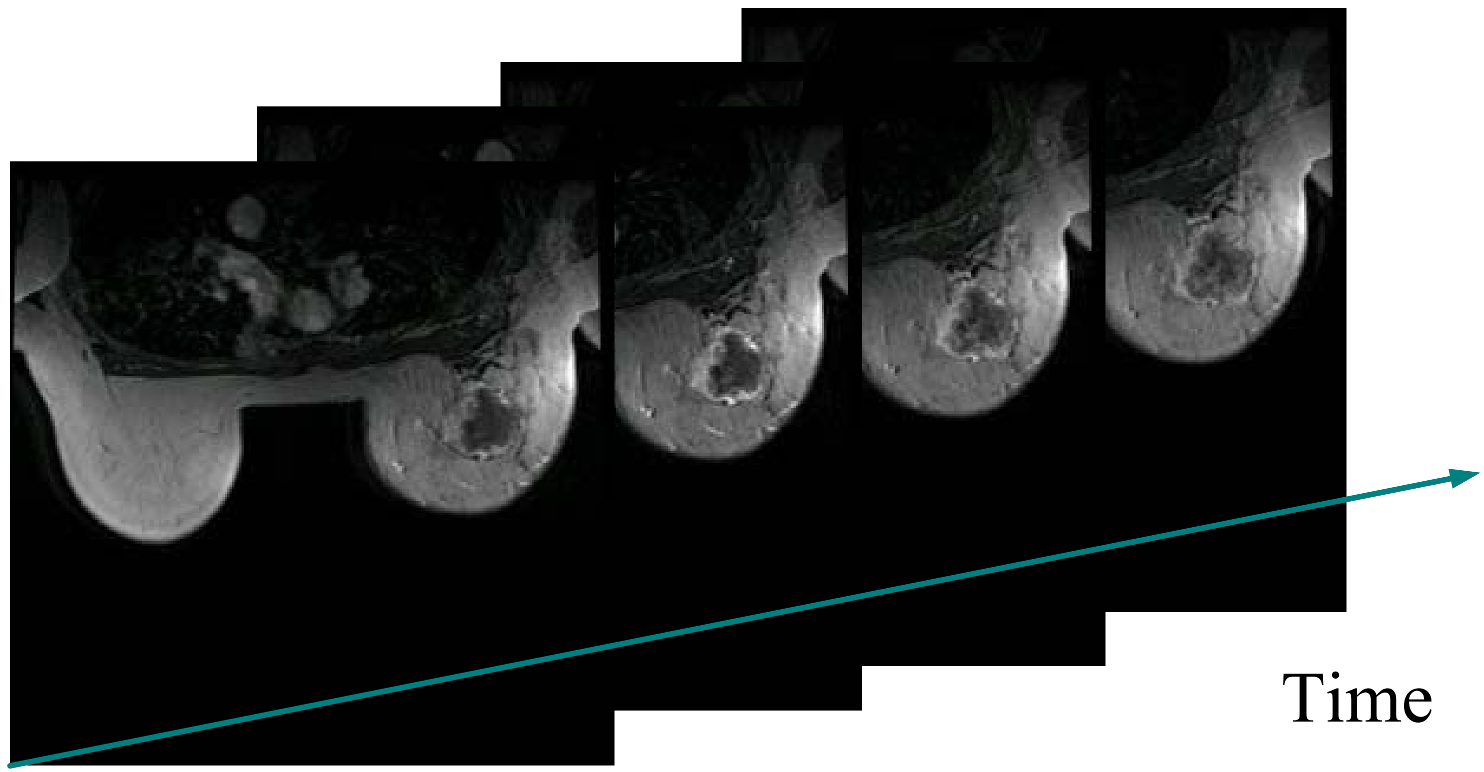
$$\boldsymbol{x}_i = \begin{bmatrix} x_i[1] \\ \vdots \\ x_i[L] \end{bmatrix}, \quad \boldsymbol{s}_i = \begin{bmatrix} s_i[1] \\ \vdots \\ s_i[L] \end{bmatrix}$$

are observation & true source vectors.



Problem: extract $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ from $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ without information of the mixing matrix $\mathbf{A} = \{a_{ij}\}$.

BSS: A biomedical imaging example



Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) assessments of breast cancer captured at different times. Courtesy to Yue Wang **[Wang *et al.* 2003].**

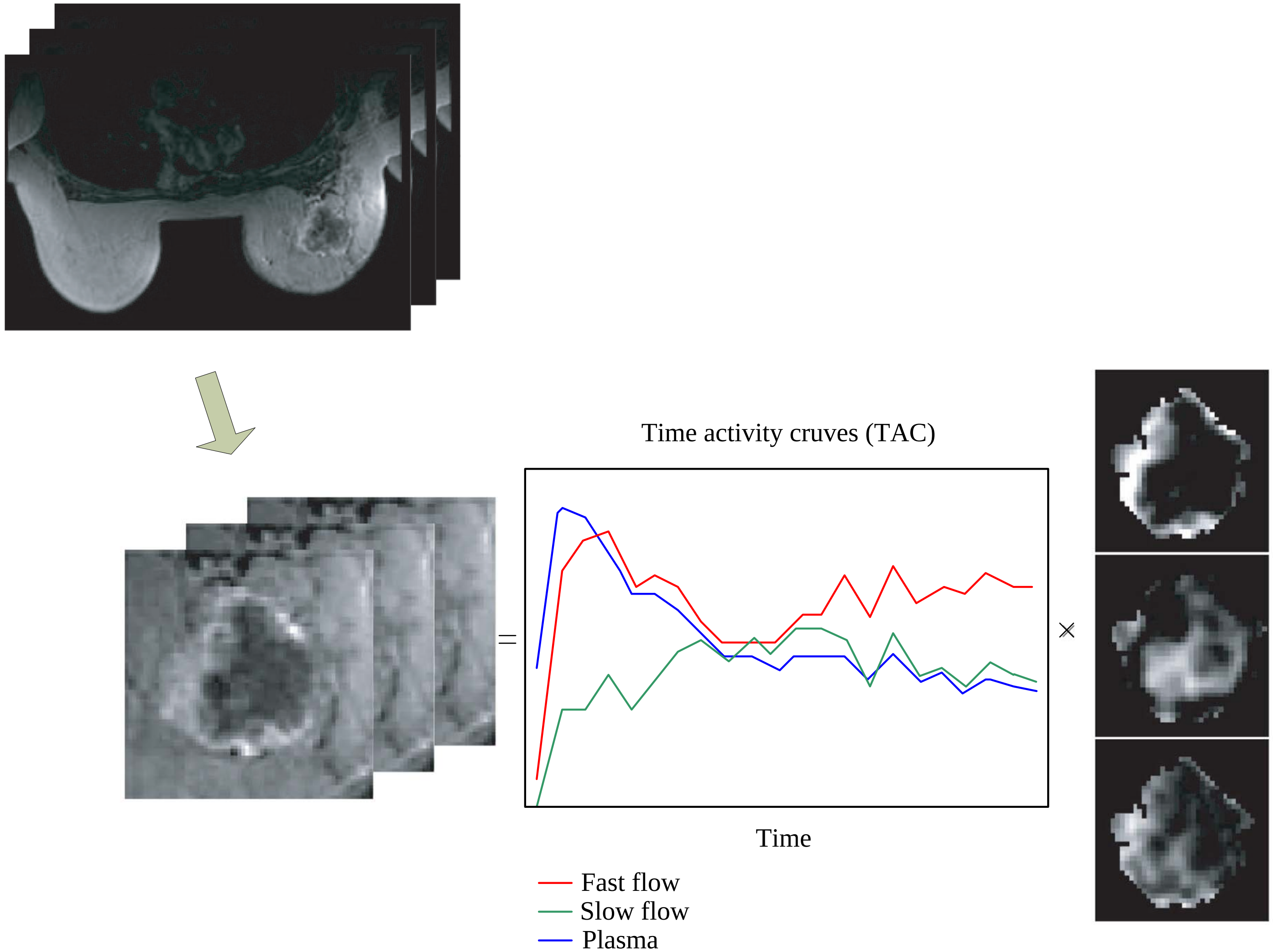


Illustration of source pattern mixing process. The signals represent a summation of vascular permeability with various diffusion rates. The goal is to separate the distribution of multiple biomarkers with the same diffusion rate.

BSS techniques

- A BSS approach is based on some assumptions on the characteristics of $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ and/or $\mathbf{A}$.

- 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 random non-Gaussian & is mutually independent.
Mutual independence is a good assumption in speech & wireless commun., but not so in hyperspectral imaging.

Non-negative blind source separation (nBSS)

- 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.

- Essentially NMF deals with an optimization

$$\begin{aligned} \min_{\boldsymbol{S}\in\mathbb{R}^{L\times M},\mathbf{A}\in\mathbb{R}^{M\times N}} \quad & \|\boldsymbol{X}-\boldsymbol{S}\mathbf{A}\|_F^2 \\ \text{s.t.} \quad & \boldsymbol{S}\succeq\mathbf{0},\ \mathbf{A}\succeq\mathbf{0} \quad \text{(elementwise non-negative)} \end{aligned}$$

where $\boldsymbol{X}=[\ \boldsymbol{x}_1,\ldots,\boldsymbol{x}_N\ ]$, & $\boldsymbol{S}=[\ \boldsymbol{s}_1,\ldots,\boldsymbol{s}_M\ ]$.

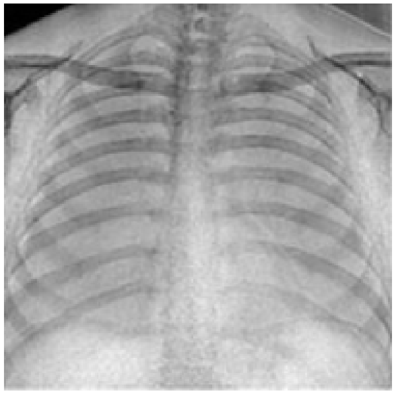
NMF may not be a unique factorization, however.

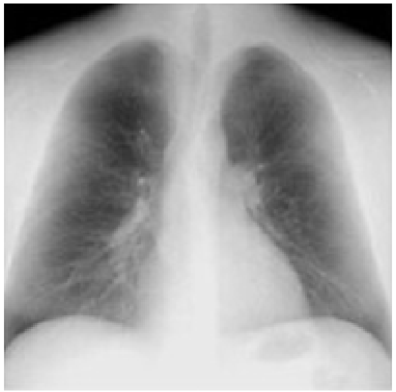
CAMNS:
Convex analysis of mixtures of non-negative sources

- CAMNS **[Chan-Ma-Chi-Wang 2008]** is a deterministic nBSS approach.

- In addition to utilizing source non-negativity, CAMNS employs a special deterministic assumption called **local dominance**.

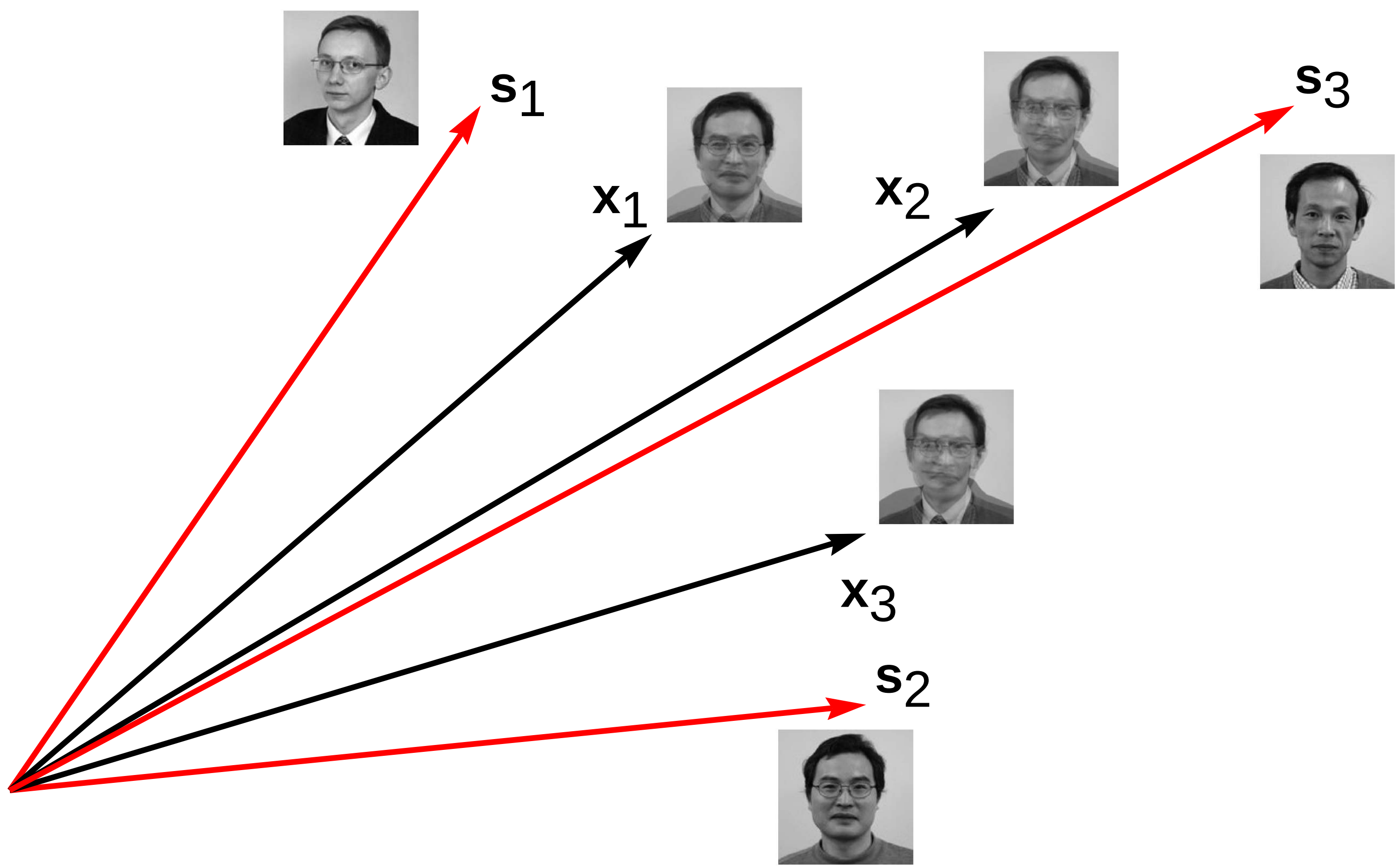
- What is local dominance? Intuitively, signals with many ‘zeros’ are likely to satisfy local dominance (math. def. available soon).





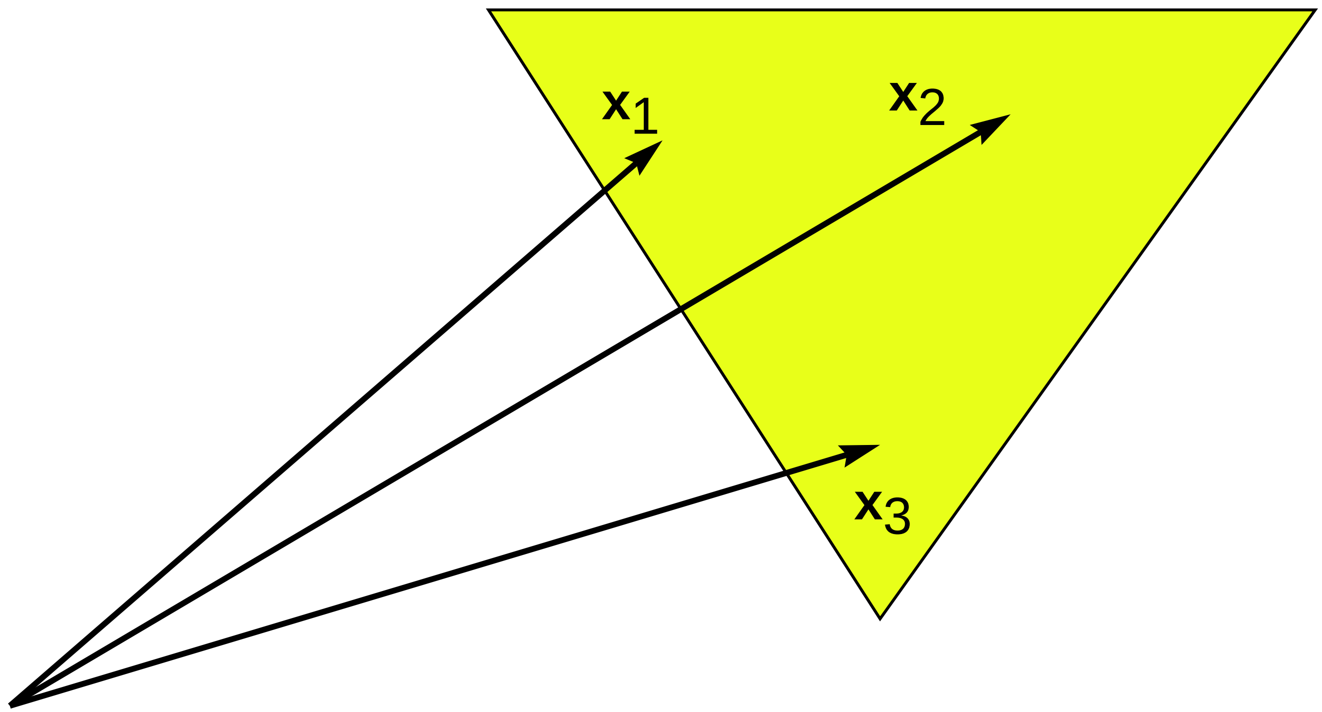
- Appears to be a good assumption for sparse or high-contrast images.

An intuitive illustration of how CAMNS works



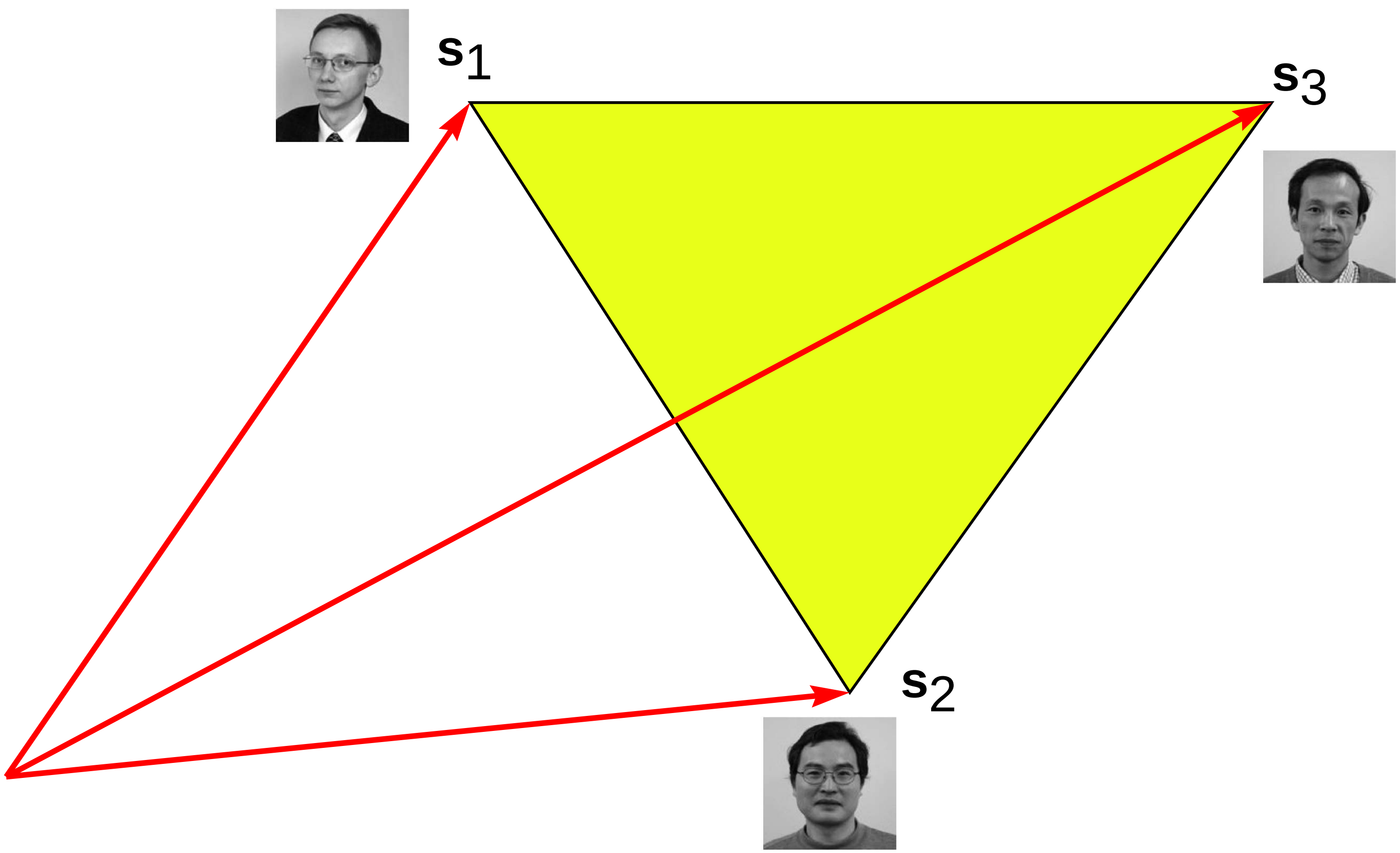
How can we extract $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ from $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ without knowing $\{a_{ij}\}$?

An intuitive illustration of how CAMNS works (cont'd)



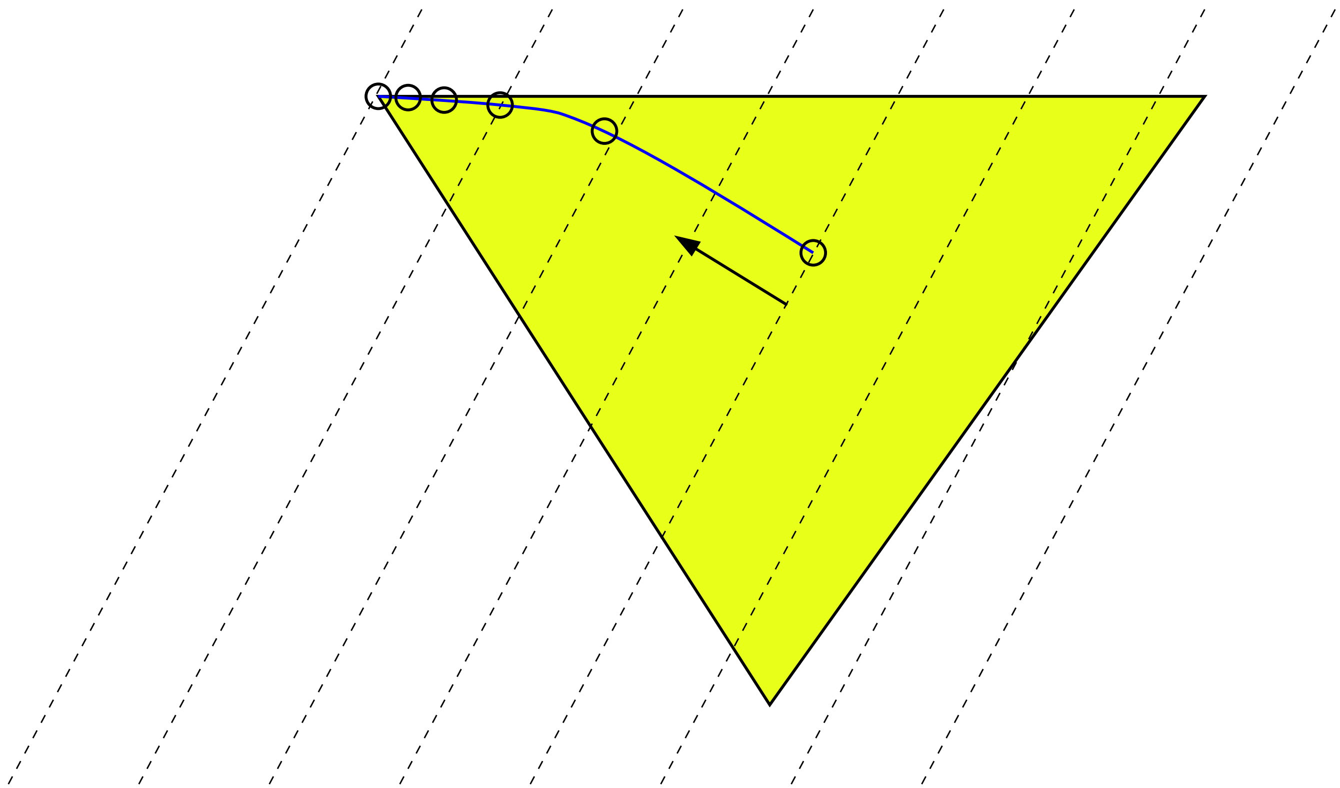
Based on some assumptions (e.g., signal non-negativity & local dominance) & by convex analysis, we use $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ to construct a polyhedral set.

An intuitive illustration of how CAMNS works (cont'd)



We show that the 'corners' (formally speaking, extreme points) of this polyhedral set are exactly $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ (rather surprisingly).

An intuitive illustration of how CAMNS works (cont'd)



Using LP, we can locate the 'corners' of the polyhedral set effectively. As a result perfect separation can be achieved.

A quick review of some convex analysis concepts

Affine hull of a given set of vectors $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} \subset \mathbb{R}^L$:

$$\text{aff}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} = \left\{ \boldsymbol{x} = \sum_{i=1}^{N} \theta_i \boldsymbol{s}_i \;\middle|\; \boldsymbol{\theta} \in \mathbb{R}^N, \sum_{i=1}^{N} \theta_i = 1 \right\}.$$

- An affine hull can always be represented by

$$\text{aff}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} = \left\{ \boldsymbol{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \;\middle|\; \boldsymbol{\alpha} \in \mathbb{R}^P \right\}$$

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

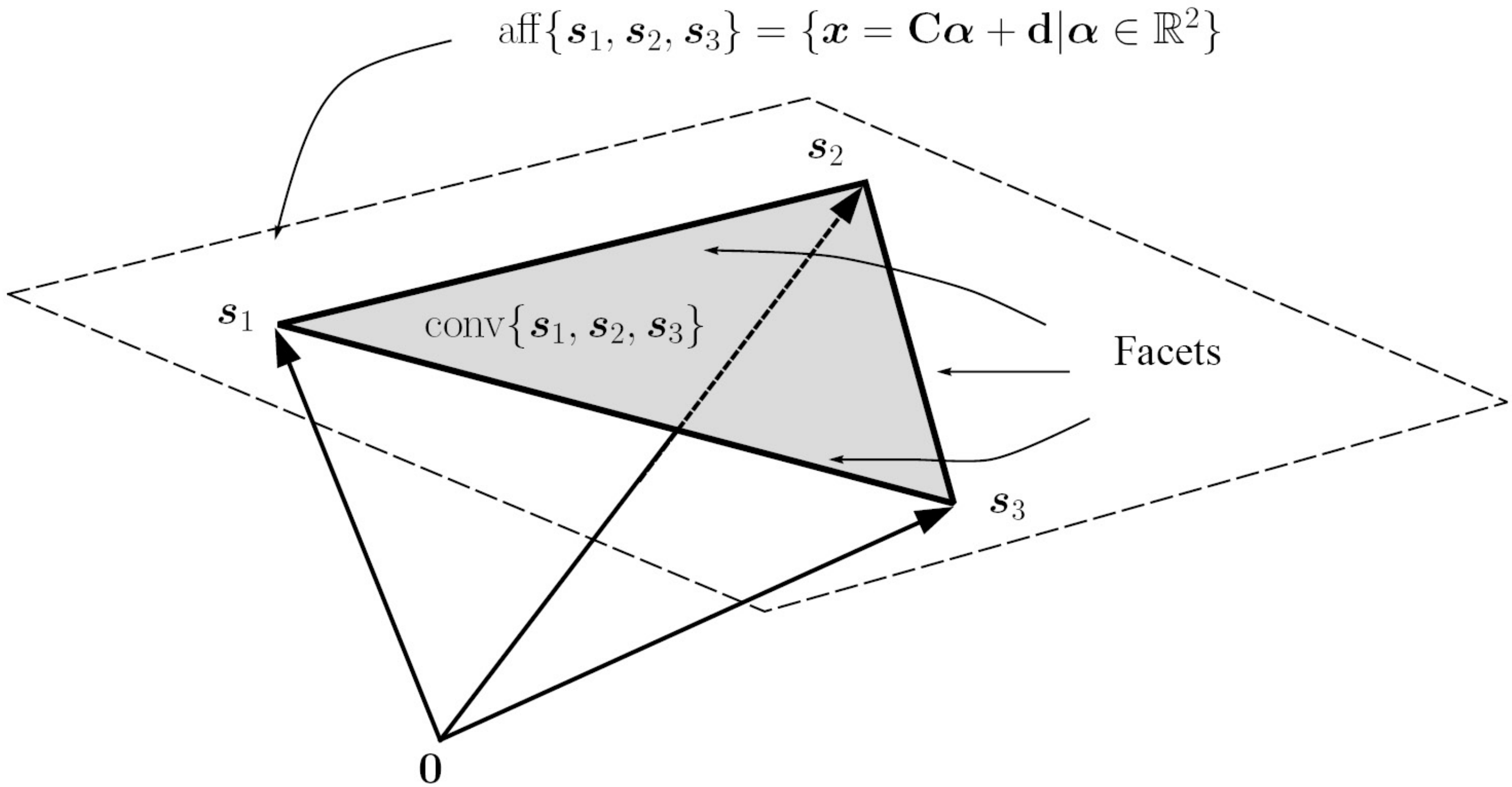
- If $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ is affine independent (or $\{\boldsymbol{s}_1 - \boldsymbol{s}_N, \ldots, \boldsymbol{s}_{N-1} - \boldsymbol{s}_N\}$ is linearly independent) then $P = N - 1$.

Convex hull of a given set of vectors $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} \subset \mathbb{R}^L$:

$$\text{conv}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} = \left\{ \boldsymbol{x} = \sum_{i=1}^{N} \theta_i \boldsymbol{s}_i \;\middle|\; \boldsymbol{\theta} \in \mathbb{R}_+^N, \sum_{i=1}^{N} \theta_i = 1 \right\}$$

- A point $\boldsymbol{x} \in \text{conv}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ is an **extreme point** of $\text{conv}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ if $\boldsymbol{x}$ is not any nontrivial convex combination of $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$.

- If $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ is affine independent then $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ is the set of all extreme points of its convex hull.



Example of 3-dimensional signal space geometry with $N = 3$. In this example, $\mathrm{aff}\{\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3\}$ is a plane passing through $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3$, & $\mathrm{conv}\{\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3\}$ is a triangle with corners (extreme points) $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3$.

The assumptions in CAMNS

Recall the model $\boldsymbol{x}_i = \sum_{j=1}^{M} a_{ij}\boldsymbol{s}_j$. Our assumptions:

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

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

(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, can be relaxed).

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

How to enforce (A3), if it does not hold

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

Suppose that $\boldsymbol{x}_i^T \mathbf{1} \neq 0$ (where $\mathbf{1}$ is an all-one vector) for all i.

Consider a normalized version of $\boldsymbol{x}_i$:

$$\bar{\boldsymbol{x}}_i = \frac{\boldsymbol{x}_i}{\boldsymbol{x}_i^T \mathbf{1}} = \sum_{j=1}^{N} \underbrace{\left(\frac{a_{ij} \boldsymbol{s}_j^T \mathbf{1}}{\boldsymbol{x}_i^T \mathbf{1}} \right)}_{\triangleq \bar{a}_{ij}} \underbrace{\left(\frac{\boldsymbol{s}_j}{\boldsymbol{s}_j^T \mathbf{1}} \right)}_{\triangleq \bar{\boldsymbol{s}}_j}.$$

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

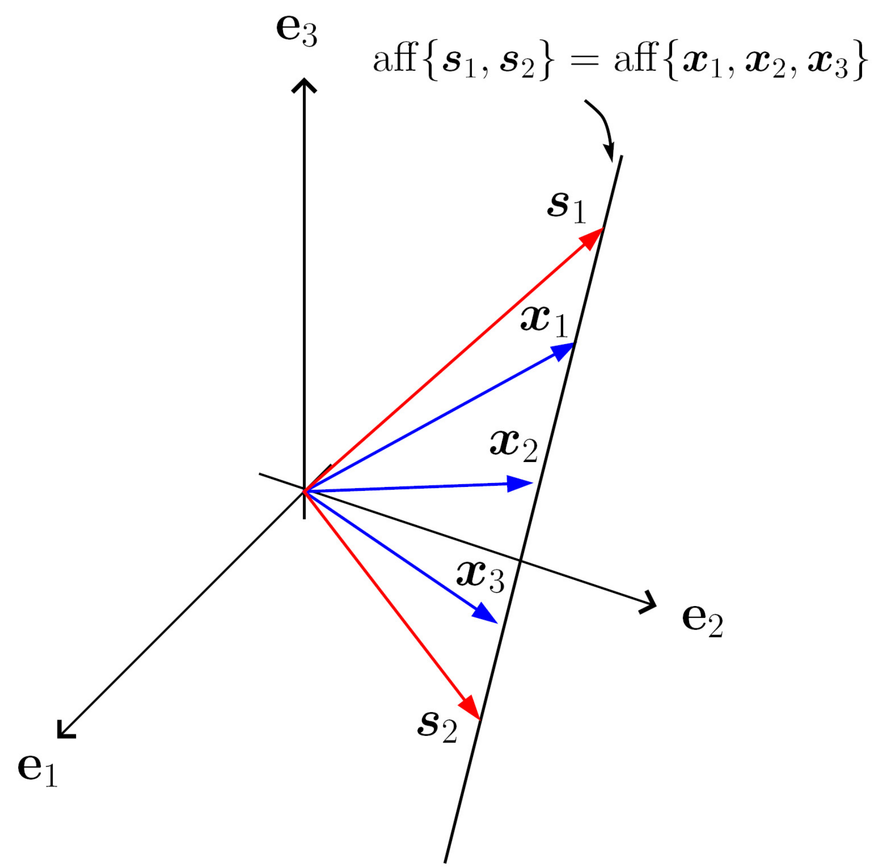
CAMNS

Since $\sum_{j=1}^{N} a_{ij} = 1$ [(A3)], we have for each observation

$$\boldsymbol{x}_i = \sum_{j=1}^{N} a_{ij}\boldsymbol{s}_j \in \mathrm{aff}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$$

This implies

$$\mathrm{aff}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} \supseteq \mathrm{aff}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}.$$



In fact, we can show that

Lemma. *Under* (A3) *and* (A4), $\mathrm{aff}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} = \mathrm{aff}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$.

- Consider the representation

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

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

- Let us consider determining the source affine set parameters $(\mathbf{C}, \mathbf{d})$ from $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$.

- The solution is simple for $M = N$:

$$
\mathbf{d} = \boldsymbol{x}_N, \qquad \mathbf{C} = [\, \boldsymbol{x}_1 - \boldsymbol{x}_N, \ldots, \boldsymbol{x}_{N-1} - \boldsymbol{x}_N \,]
$$

- For $M > N$, we use an affine set fitting solution.

Affine set fitting problem:

$$(\mathbf{C}, \mathbf{d}) = \arg \min_{\substack{\tilde{\mathbf{C}}, \tilde{\mathbf{d}} \\ \tilde{\mathbf{C}}^T \tilde{\mathbf{C}} = \mathbf{I}}} \sum_{i=1}^{M} \underbrace{\min_{\tilde{\boldsymbol{x}} \in \mathcal{A}(\tilde{\mathbf{C}}, \tilde{\mathbf{d}})} \|\tilde{\boldsymbol{x}} - \boldsymbol{x}_i\|_2^2}_{\text{proj. error of } \boldsymbol{x}_i \text{ onto } \mathcal{A}(\tilde{\mathbf{C}}, \tilde{\mathbf{d}})} \qquad (*)$$

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

Proposition. *Problem $(*)$ has a closed-form solution*

$$\mathbf{d} = \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{x}_i, \qquad \mathbf{C} = [\ \boldsymbol{q}_1(\mathbf{U}\mathbf{U}^T), \boldsymbol{q}_2(\mathbf{U}\mathbf{U}^T), \ldots, \boldsymbol{q}_{N-1}(\mathbf{U}\mathbf{U}^T)\ ]$$

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 $\mathbf{R}$.

Be reminded that $\boldsymbol{s}_i \in \mathbb{R}_+^L$. Hence, it is true that

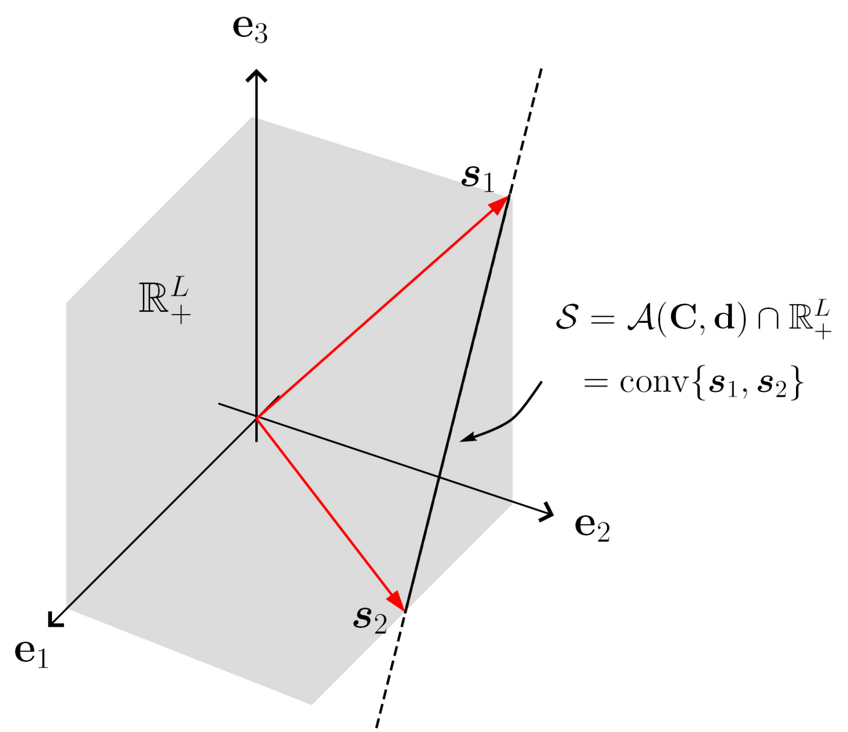
$$\boldsymbol{s}_i \in \text{aff}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} \cap \mathbb{R}_+^L = \mathcal{A}(\mathbf{C}, \mathbf{d}) \cap \mathbb{R}_+^L \triangleq \mathcal{S}$$

The following lemma arises from local dominance (A2):

Lemma. *Under* (A1) *and* (A2),

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

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



Summarizing the above results, a new nBSS criterion is as follows:

Theorem 1. **(CAMNS criterion)** *Under* (A1)-(A4), *the polyhedral set*

$$\mathcal{S} = \left\{ \boldsymbol{x} \in \mathbb{R}^L \;\middle|\; \boldsymbol{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \succeq \mathbf{0},\ \boldsymbol{\alpha} \in \mathbb{R}^{N-1} \right\}$$

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

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): $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N$ are linear independent.

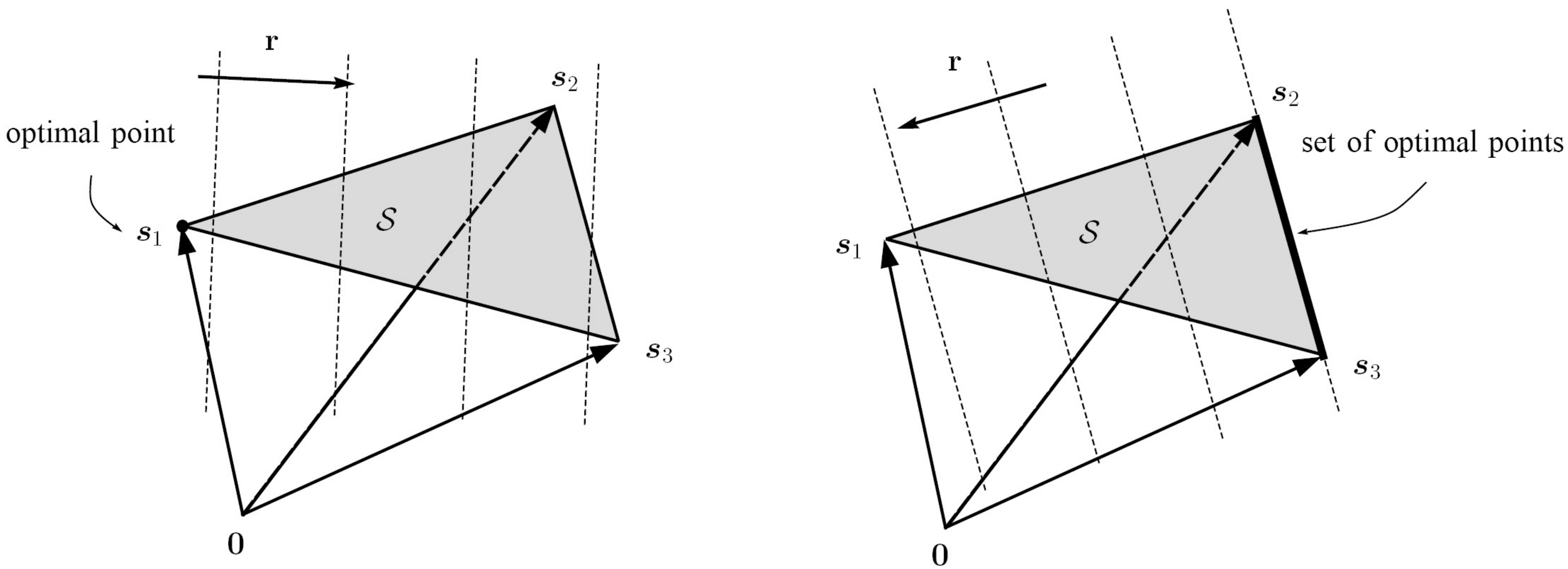
- By exploiting this property, we can locate all the extreme points by solving a sequence of LPs ($\approx 2N$ LPs at worst).

Consider the following LP

$$
\begin{aligned}
p^\star = \ & \min_{\boldsymbol{s}} \ \mathbf{r}^T \boldsymbol{s} \\
& \text{s.t.} \ \boldsymbol{s} \in \mathcal{S}
\end{aligned}
\qquad (\dagger)
$$

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

- one of the extreme points of $\mathcal{S}$ (that is, one of the $\boldsymbol{s}_i$), or
- any point on a face of $\mathcal{S}$ (look rather unlikely, intuitively).


r
optimal point
s_1
s_2
s_3
$\mathcal{S}$
$\mathbf{0}$
r
set of optimal points
s_1
s_2
s_3
$\mathcal{S}$
$\mathbf{0}$

We can prove that getting a non-extreme-pt. solution is very unlikely:

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

$$\begin{aligned} p^\star = \ & \min_{\boldsymbol{s}} \ \mathbf{r}^T \boldsymbol{s} \\ & \text{s.t.} \ \boldsymbol{s} \in \mathcal{S} \end{aligned}$$

is uniquely given by $\boldsymbol{s}_i$ *for some* $i \in \{1, ..., N\}$.

- Suppose that we have found l extreme point, say, $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_l\}$.

- We can find the other extreme points, by using the linear independence of $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ to 'annihilate' the old extreme points.

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

$$\mathbf{B}[\boldsymbol{s}_1, \ldots, \boldsymbol{s}_l] = \mathbf{0} \qquad \mathbf{B}^T\mathbf{B} = \mathbf{I}_{L-l}$$

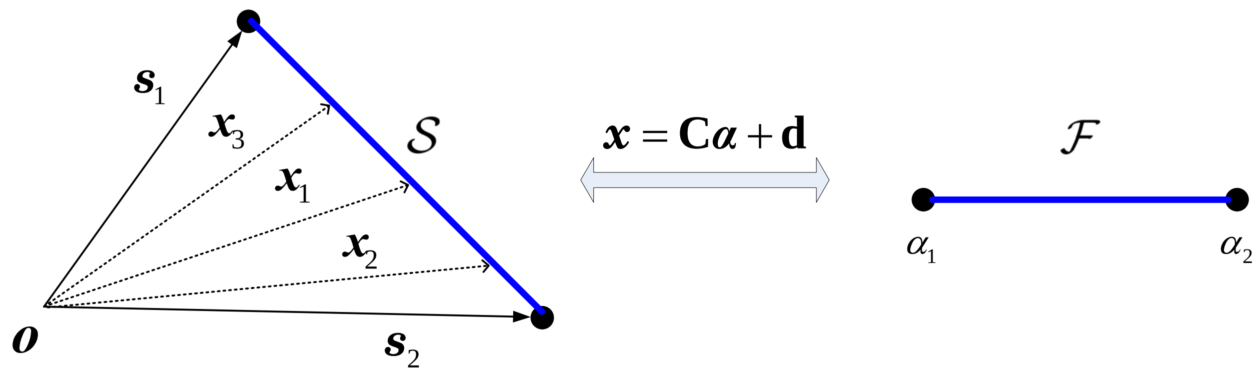
Then, with probability 1, at least one of the LPs

$$p^\star = \min_{\boldsymbol{s} \in \mathcal{S}} \mathbf{r}^T\boldsymbol{s} \qquad q^\star = \max_{\boldsymbol{s} \in \mathcal{S}} \mathbf{r}^T\boldsymbol{s}$$

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

On alternatives of implementing CAMNS

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

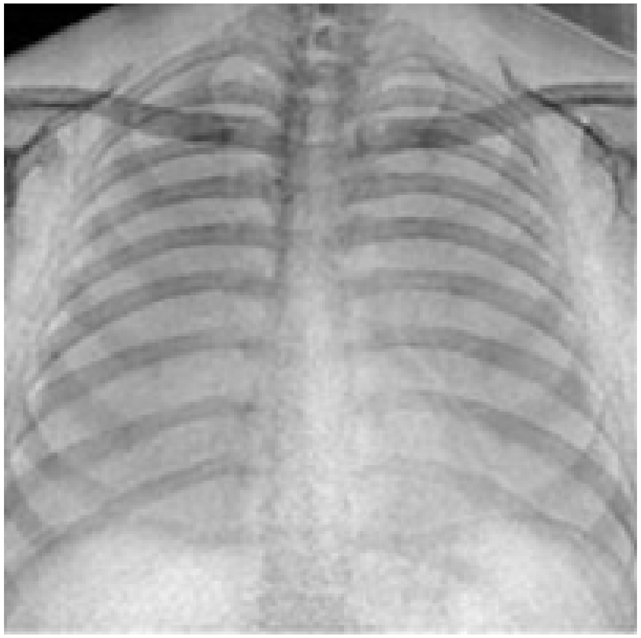


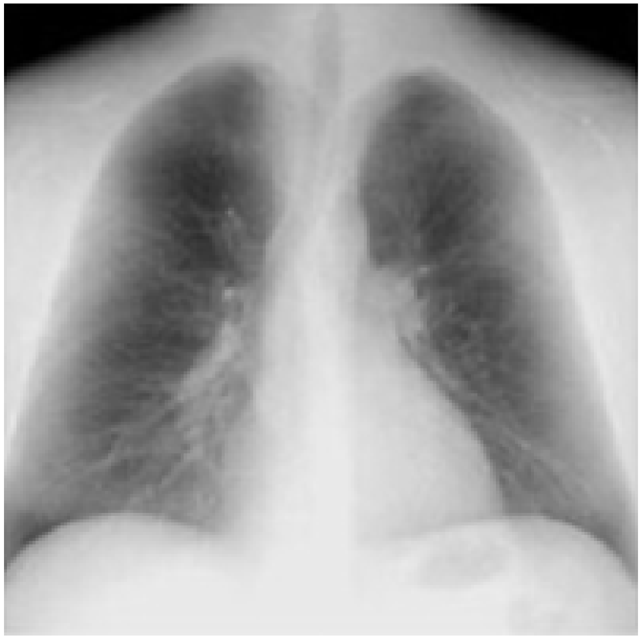
- The set $\mathcal{F}$ has a smaller vector dim. (note that $L \gg N$). Also it is a simplex with extreme pts related to those of $\mathcal{S}$ in a one-to-one manner.

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

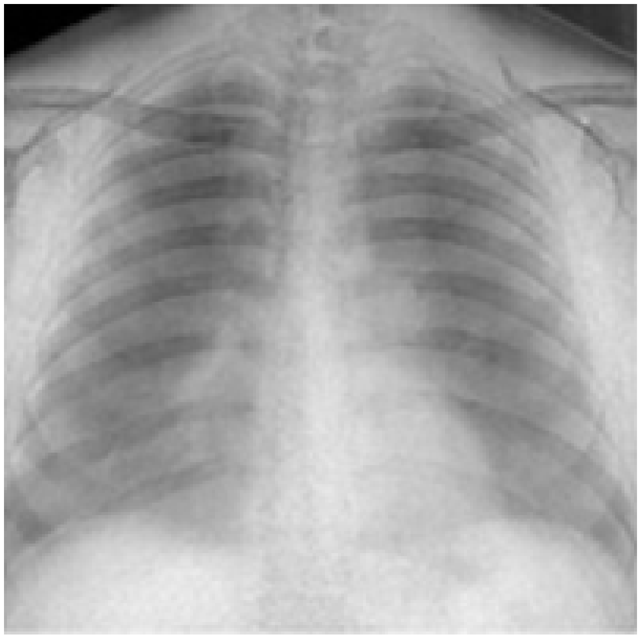
- For $N = 3$, $\mathcal{F}$ is a triangle on $\mathbb{R}^2$ and there is also a simple way for locating its extreme points.

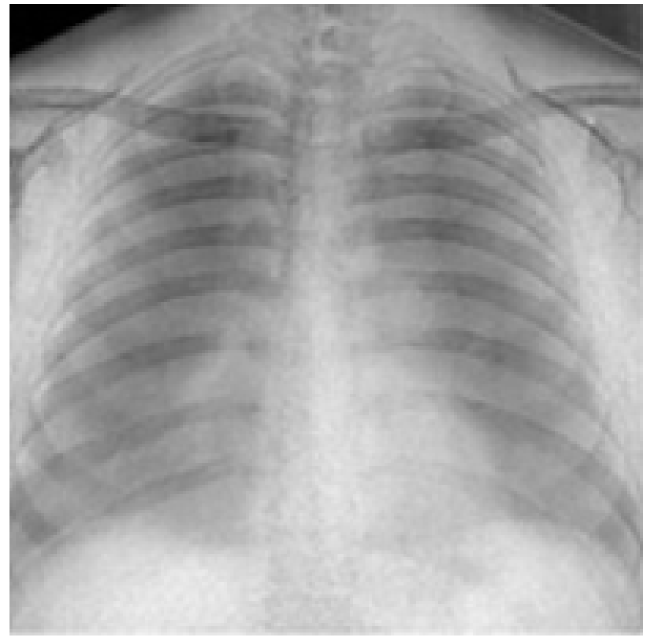
Simulation example 1: Dual energy X-Ray



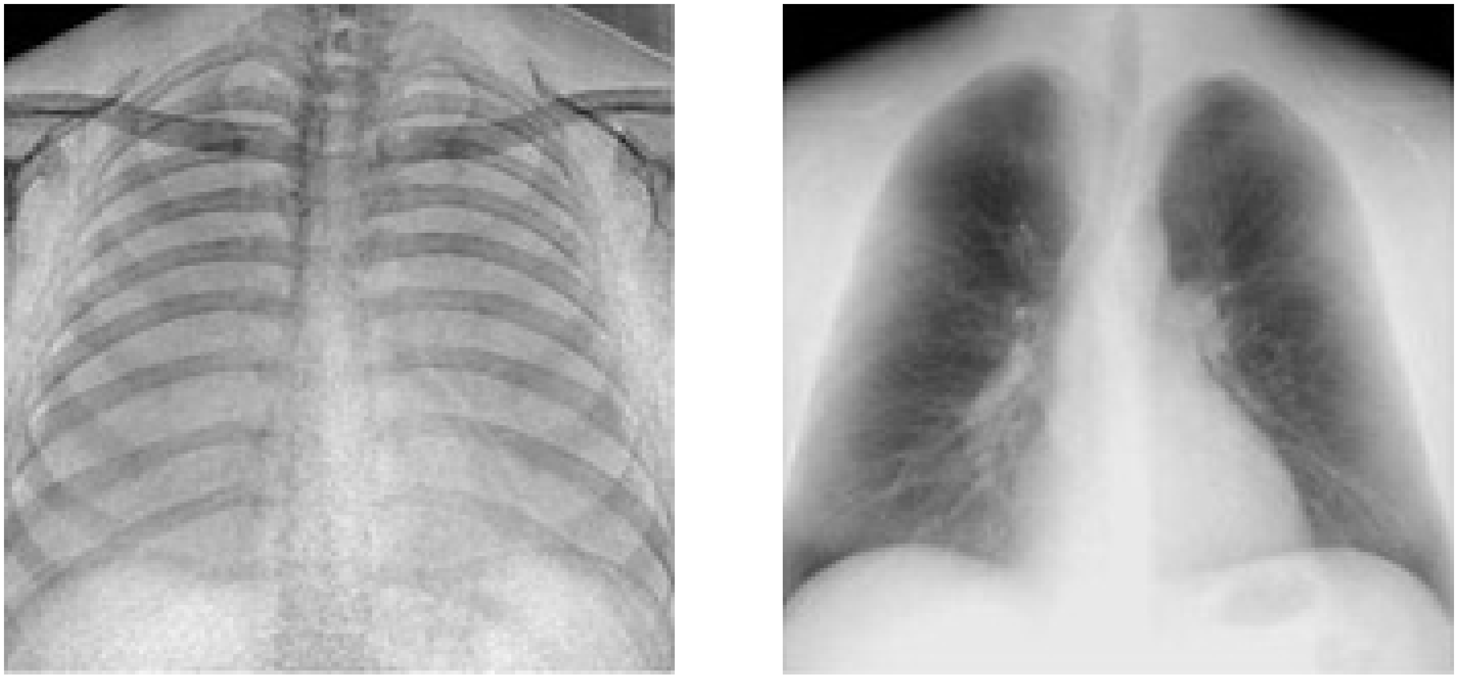


Original sources

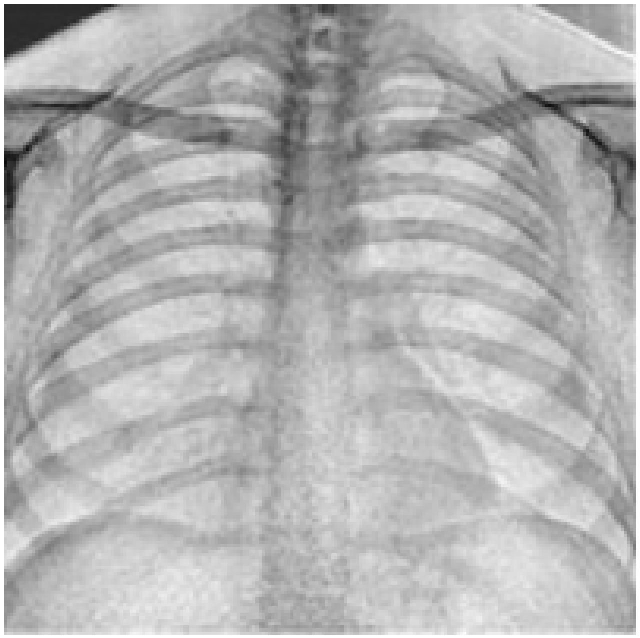


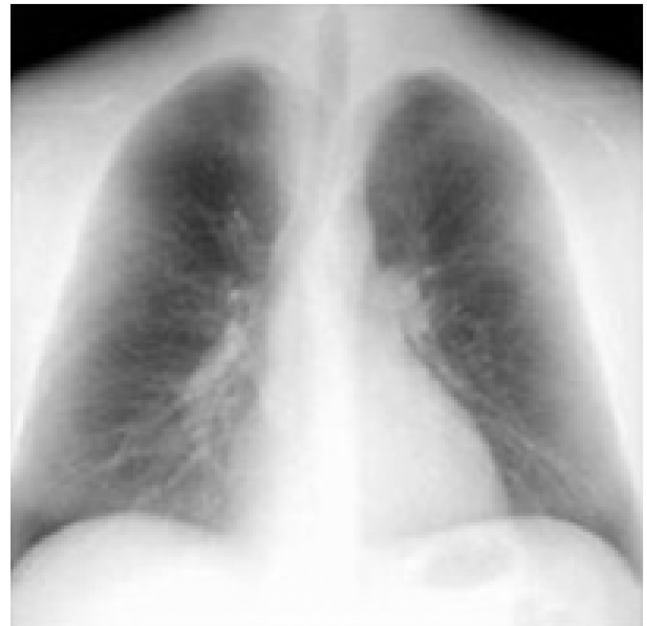


Observations

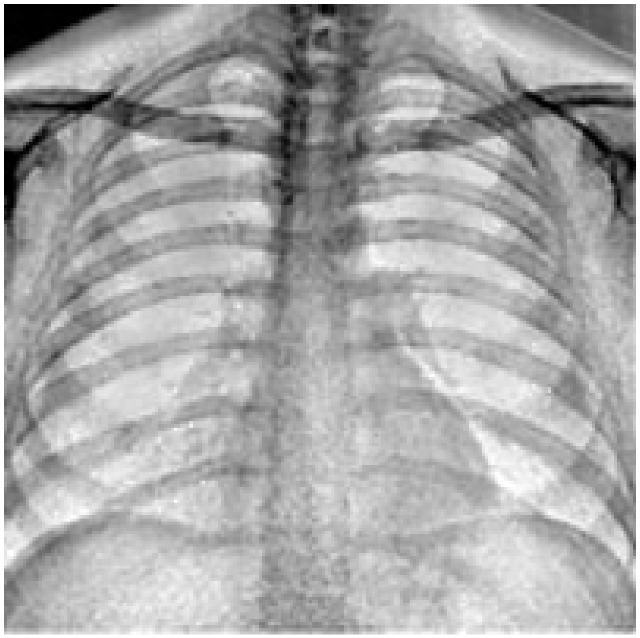


Separated sources by **CAMNS**



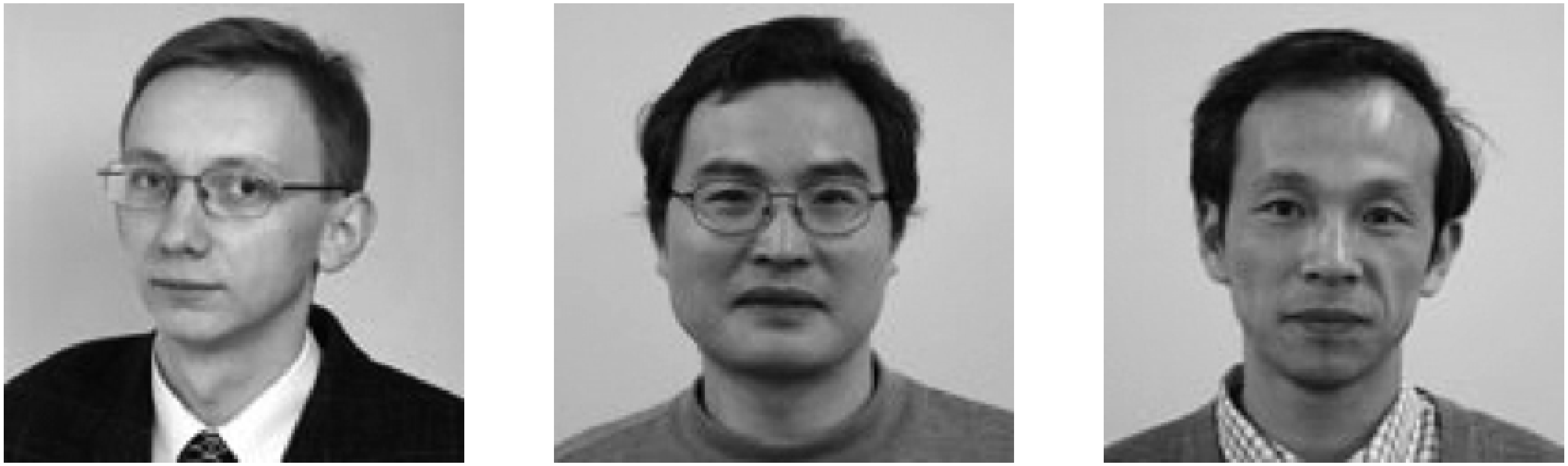


Separated sources by **nICA** (a benchmarked nBSS method)



Separated sources by **NMF** (yet another benchmarked nBSS method)

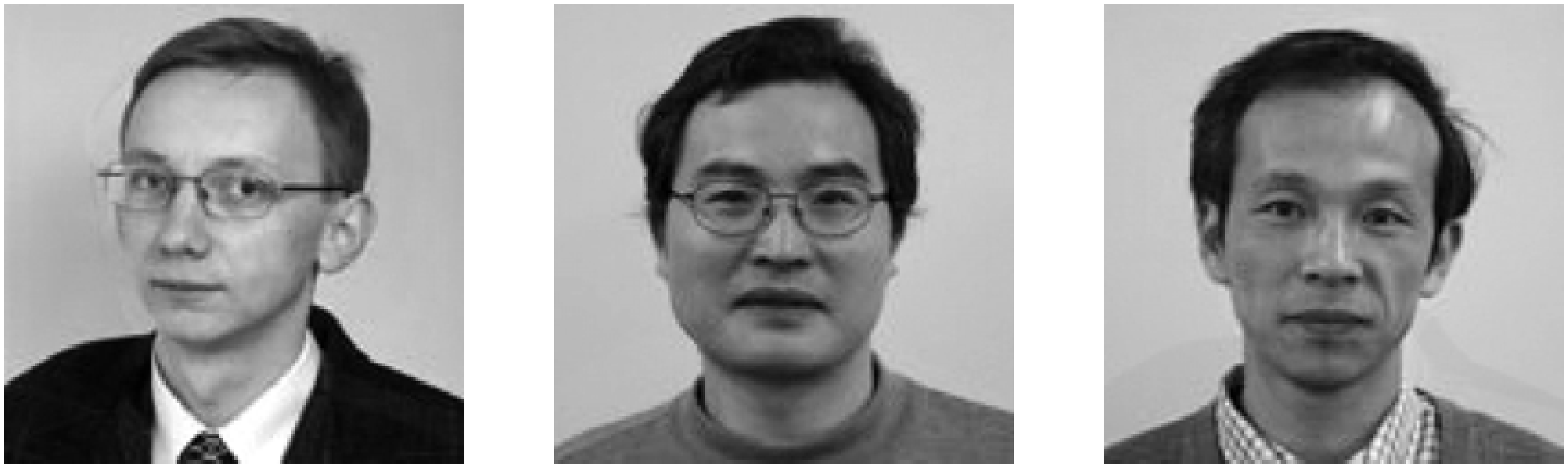
Simulation example 2: Human faces



Original sources

Observations



Separated sources by **CAMNS**

Separated sources by **nICA**

Separated sources by **NMF**

Simulation example 3: Ghosting

Original sources

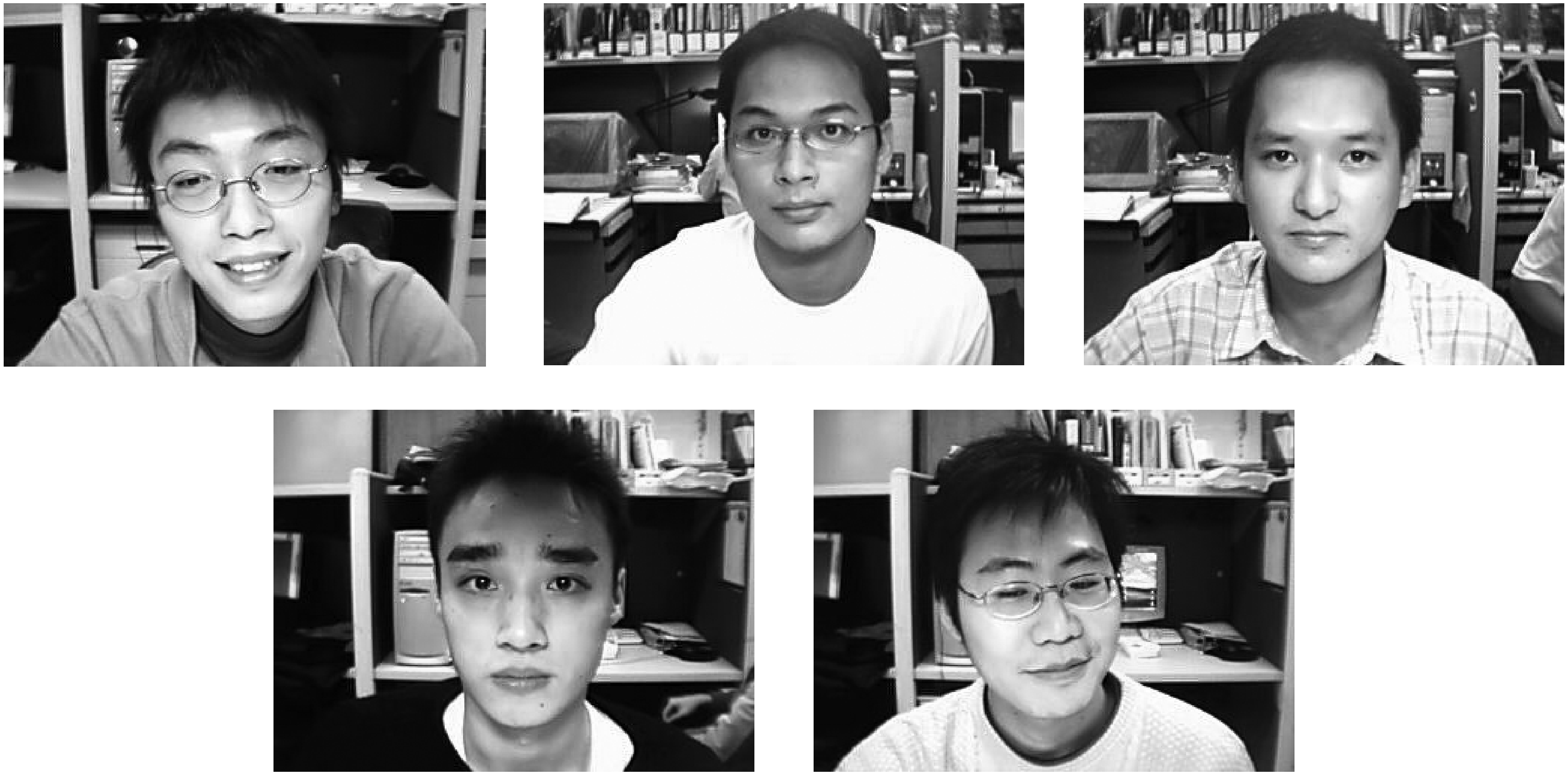
Observations

Separated sources by **CAMNS**

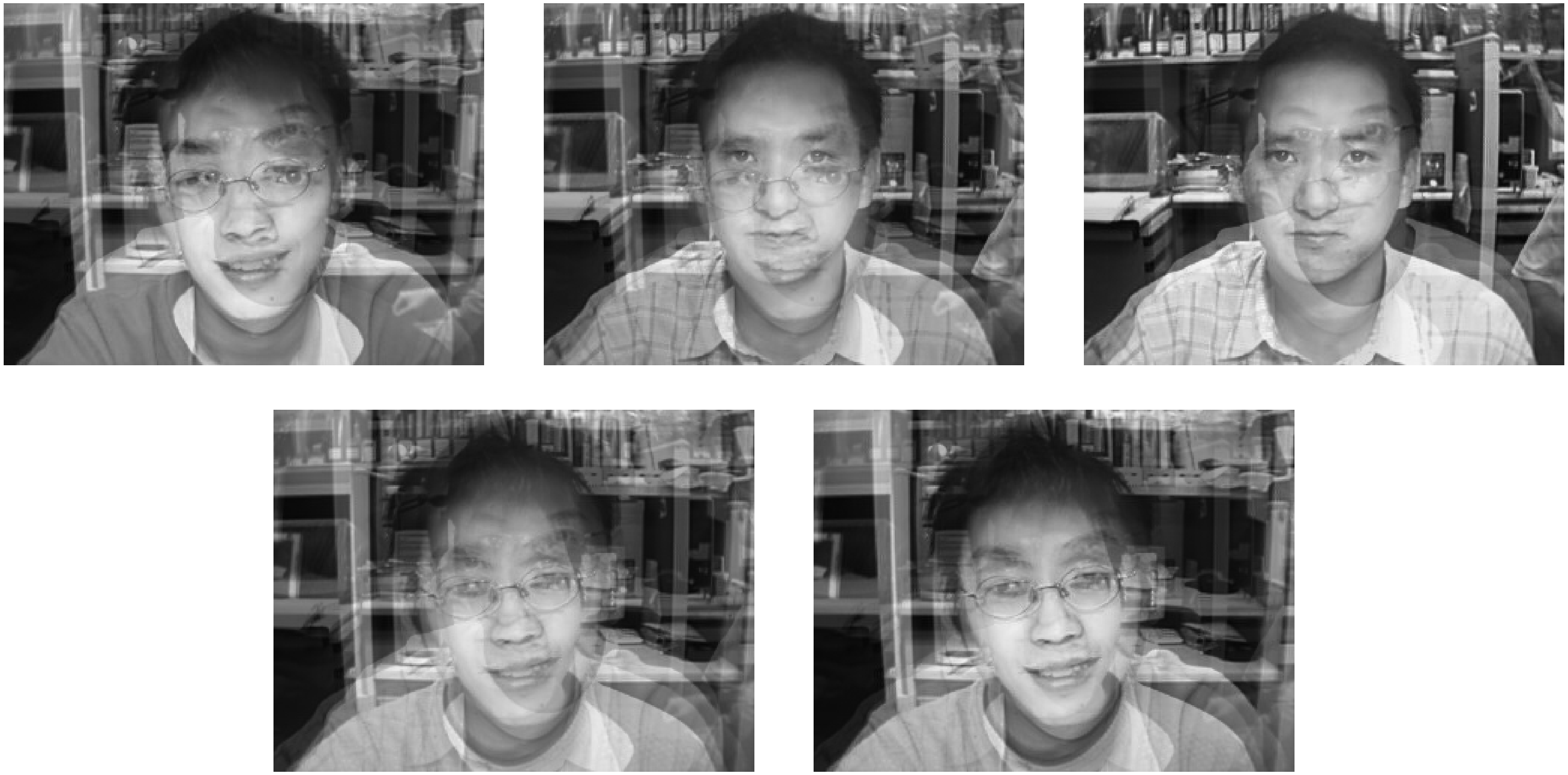
Separated sources by **nICA**

Separated sources by **NMF**

Simulation example 4: Five of my students



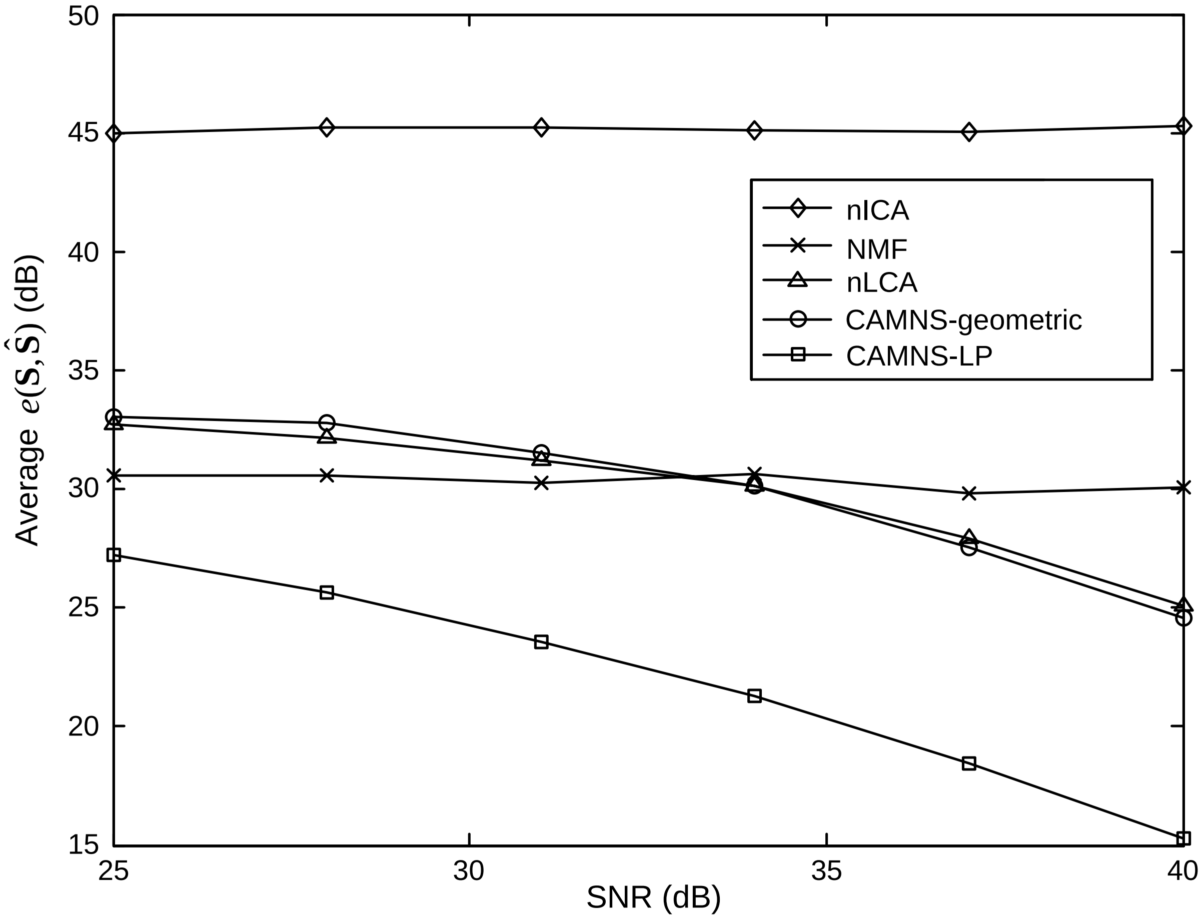
Original sources



Observations

Separated sources by **CAMNS**

Simulation example 5: Monte Carlo performance for $N = 3$



Average sum squared errors of the sources with respect to SNRs.

Conclusion

- A convex analysis framework, called CAMNS has been developed for nBSS.
- 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 realize CAMNS. Its complexity is polynomial (specifically, $\mathcal{O}(L^{1.5}(N-1)^2)$).
- A number of simulation results indicate that CAMNS performs very well even in the presence of dependent sources.
- The source codes is available at `http://www.ee.cuhk.edu.hk/~wkma`

References

[Wang *et al.* 2003] Y. Wang, J. Xuan, R. Srikanchana, & P. L. Choyke, "Modeling and reconstruction of mixed functional and molecular patterns," *Int'l Journal Biomedical Imaging*, pp. 1–9, 2005.

[Plumbley 2003] M. D. Plumbley, "Algorithms for nonnegative independent component analysis," *IEEE Trans. Neural Networks*, vol. 14, no. 3, pp. 534–543, May 2003.

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

[Chan-Ma-Chi-Wang 2008] T.-H. Chan, W.-K. Ma, C.-Y. Chi, & Y. Wang, "A convex analysis framework for blind separation of non-negative sources," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5120-5134, Oct. 2008.