

# Hyperspectral Unmixing from A Convex Analysis and Optimization Perspective

Tsung-Han Chan<sup>†</sup>, Wing-Kin Ma<sup>\*</sup>, Chong-Yung Chi<sup>†</sup>, and A.ArulMurugan<sup>†</sup>

<sup>†</sup>Inst. of Commn. Eng., National Tsing Hua University, Taiwan  
Tel: +886-3-5731156, Fax: +886-3-5751787  
E-mail: cychi@ee.nthu.edu.tw

<sup>\*</sup>Dept. of Electronic Eng., Chinese Univ.of Hong Kong, Hong Kong  
Tel.: +852-31634350, Fax: +852-26035558  
E-mail: wkma@ee.cuhk.edu.hk

IEEE WHISPERS 2009, 26-28 AUGUST, GRENOBLE, FRANCE.



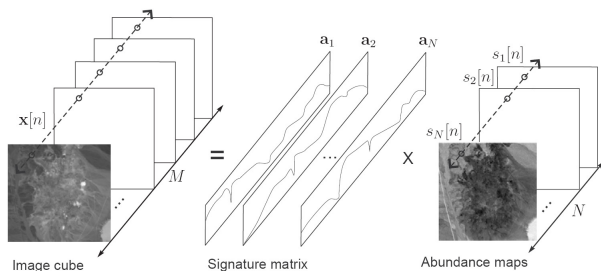
NATIONAL TSING HUA UNIVERSITY  
*Institute of Communications Engineering*



### **The Theme: Use a convex analysis perspective to view hyperspectral linear unmixing.**

- provide formulations & new interpretations for
  - dimension reduction
  - Craig's belief [Craig'94]
  - Winter's belief [Winter'99]
- **Theory:** prove that both Craig's & Winter's beliefs are optimal in the pure-pixel case.
- **Algorithms:** develop convex optimization based approximations for Craig's & Winter's beliefs.

# Problem Statement for Hyperspectral Unmixing

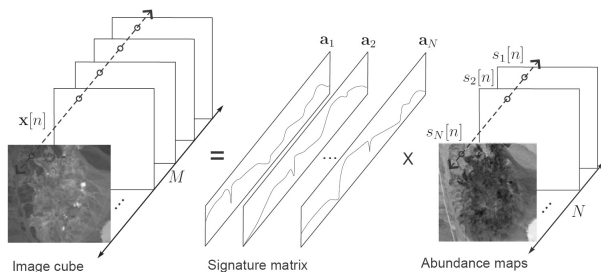


**Observed pixel vector:** (linear mixing model)

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n] = \sum_{i=1}^N s_i[n]\mathbf{a}_i, \quad n = 1, \dots, L \quad (1)$$

- $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_N] \in \mathbb{R}^{M \times N}$ ,  $\mathbf{a}_i$  is the  $i$ th **endmember signature**,
- $\mathbf{s}[n] = [s_1[n], \dots, s_N[n]]^T$  is the **abundance vector** of pixel  $n$ ,
- $M$  = no. of spectral bands,  $N$  = no. of endmember signatures, &  $L$  = no. of pixels.

# Problem Statement for Hyperspectral Unmixing



**Observed pixel vector:** (linear mixing model)

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n] = \sum_{i=1}^N s_i[n]\mathbf{a}_i, \quad n = 1, \dots, L \quad (2)$$

Some general assumptions:

- (A1) **(Non-negativity)**  $s_i[n] \geq 0$  for all  $i$  and  $n$ .
- (A2) **(Full-additivity)**  $\sum_{i=1}^N s_i[n] = 1$  for all  $n$ .
- (A3)  $\min\{L, M\} \geq N$  and  $\mathbf{a}_1, \dots, \mathbf{a}_N$  are linearly independent.

The **affine hull** of  $\{\mathbf{a}_1, \dots, \mathbf{a}_N\} \subset \mathbb{R}^M$  is defined as:

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

An affine hull can always be represented by

$$\mathcal{A}(\mathbf{C}, \mathbf{d}) \triangleq \{ \mathbf{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \mid \boldsymbol{\alpha} \in \mathbb{R}^P \}$$

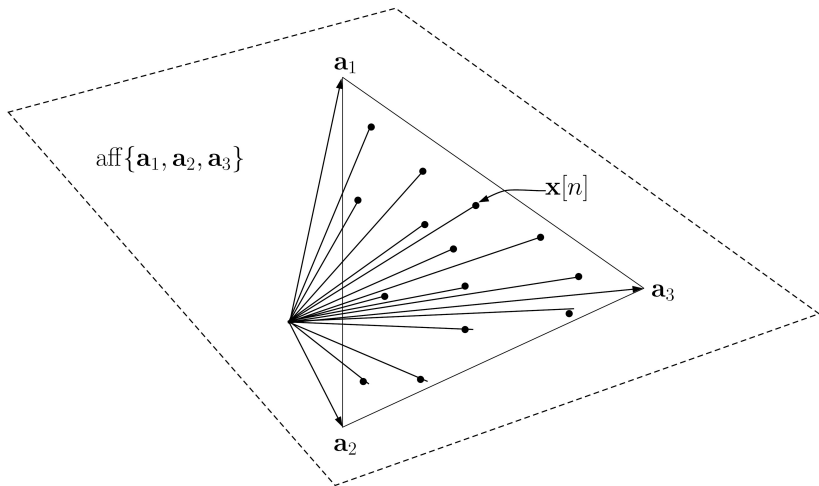
for some  $\mathbf{C} \in \mathbb{R}^{N \times P}$ ,  $\mathbf{d} \in \mathbb{R}^N$ , &  $P \leq N - 1$ .

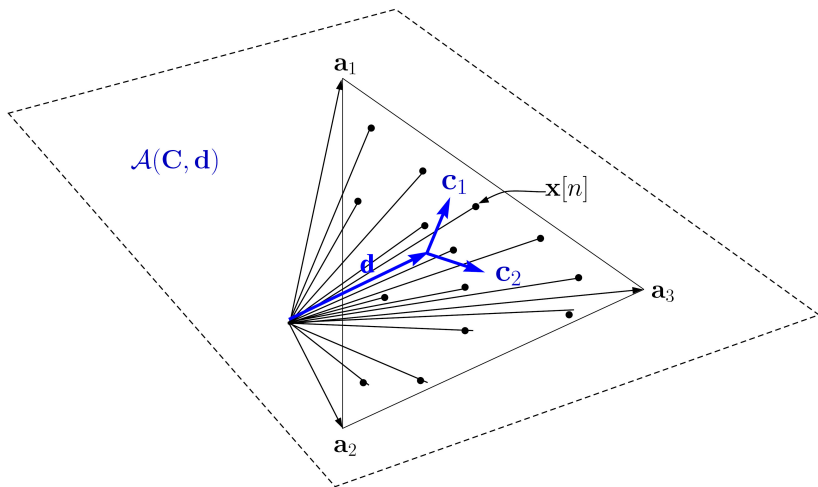
Recall  $\mathbf{x}[n] = \sum_{i=1}^N s_i[n] \mathbf{a}_i$ . Under (A2) and (A3), we have

$$\mathbf{x}[n] \in \text{aff}\{\mathbf{a}_1, \dots, \mathbf{a}_N\} = \mathcal{A}(\mathbf{C}, \mathbf{d}), \quad \forall n = 1, \dots, L,$$

with  $P = N - 1$ .

# An Geometry Illustration for $N = 3$





## Lemma 1 (Affine set fitting) [Chan'08]

Under (A2) and (A3), we can show that

$$\mathcal{A}(\mathbf{C}, \mathbf{d}) = \text{aff}\{\mathbf{x}[1], \dots, \mathbf{x}[L]\}.$$

Moreover,  $(\mathbf{C}, \mathbf{d})$  can be obtained from  $\mathbf{x}[1], \dots, \mathbf{x}[L]$  by

$$\mathbf{d} = \frac{1}{L} \sum_{n=1}^L \mathbf{x}[n], \quad \mathbf{C} = [\mathbf{q}_1(\mathbf{U}\mathbf{U}^T), \mathbf{q}_2(\mathbf{U}\mathbf{U}^T), \dots, \mathbf{q}_{N-1}(\mathbf{U}\mathbf{U}^T)],$$

where  $\mathbf{U} = [\mathbf{x}[1] - \mathbf{d}, \dots, \mathbf{x}[L] - \mathbf{d}] \in \mathbb{R}^{M \times L}$ , and  $\mathbf{q}_i(\mathbf{R})$  denotes the eigenvector associated with the  $i$ th principal eigenvalue of  $\mathbf{R}$ .

- In the presence of noise in the model, Lemma 1 is still optimal in yielding the least squares approximation error in the fitting.



## Lemma 1 (Affine set fitting) [Chan'08]

Under (A2) and (A3), we can show that

$$\mathcal{A}(\mathbf{C}, \mathbf{d}) = \text{aff}\{\mathbf{x}[1], \dots, \mathbf{x}[L]\}.$$

Moreover,  $(\mathbf{C}, \mathbf{d})$  can be obtained from  $\mathbf{x}[1], \dots, \mathbf{x}[L]$  by

$$\mathbf{d} = \frac{1}{L} \sum_{n=1}^L \mathbf{x}[n], \quad \mathbf{C} = [\mathbf{q}_1(\mathbf{U}\mathbf{U}^T), \mathbf{q}_2(\mathbf{U}\mathbf{U}^T), \dots, \mathbf{q}_{N-1}(\mathbf{U}\mathbf{U}^T)],$$

where  $\mathbf{U} = [\mathbf{x}[1] - \mathbf{d}, \dots, \mathbf{x}[L] - \mathbf{d}] \in \mathbb{R}^{M \times L}$ , and  $\mathbf{q}_i(\mathbf{R})$  denotes the eigenvector associated with the  $i$ th principal eigenvalue of  $\mathbf{R}$ .

## Relationship to principal component analysis (PCA) [Jolliffe'86]

- The operations of affine set fitting are exactly the same as PCA.
- But affine set fitting has no statistical assumption, it is an outcome of (deterministic) convex geometry.

# Dimension Reduction

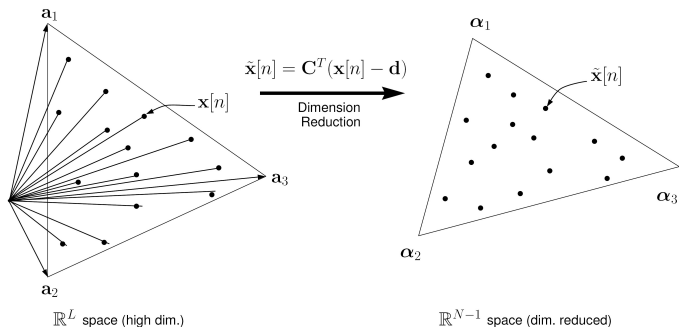
Since  $\mathbf{x}[n] \in \mathcal{A}(\mathbf{C}, \mathbf{d})$ , its affine representation is

$$\mathbf{x}[n] = \mathbf{C}\tilde{\mathbf{x}}[n] + \mathbf{d} \in \mathbb{R}^M.$$

Then the **dimension-reduced pixel**  $\tilde{\mathbf{x}}[n]$  is given by

$$\tilde{\mathbf{x}}[n] = \mathbf{C}^T(\mathbf{x}[n] - \mathbf{d}) = \sum_{i=1}^N s_i[n]\alpha_i \in \mathbb{R}^{N-1},$$

where  $\alpha_i = \mathbf{C}^T(\mathbf{a}_i - \mathbf{d})$  is the  $i$ th **dimension-reduced endmember**.



The **convex hull** of  $\{\alpha_1, \dots, \alpha_N\} \subset \mathbb{R}^M$  is defined as:

$$\text{conv}\{\alpha_1, \dots, \alpha_N\} = \left\{ \mathbf{x} = \sum_{i=1}^N \theta_i \alpha_i \mid \boldsymbol{\theta} \succeq \mathbf{0}, \sum_{i=1}^N \theta_i = 1 \right\}$$

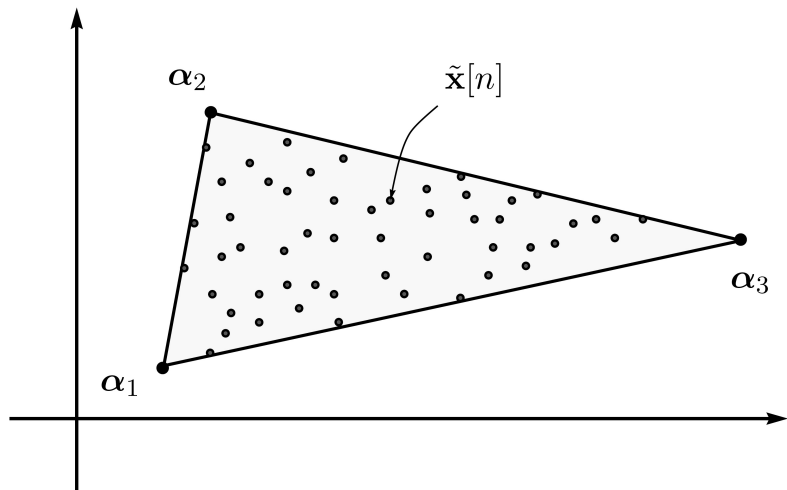
A convex hull  $\text{conv}\{\alpha_1, \dots, \alpha_N\} \in \mathbb{R}^M$  is called a **simplex** if  $M = N - 1$  &  $\alpha_1, \dots, \alpha_N$  are affinely independent.

Recall  $\tilde{\mathbf{x}}[n] = \sum_{i=1}^N s_i[n] \alpha_i$ ,  $s_i[n] \geq 0 \forall i, n$ ,  $\sum_{i=1}^N s_i[n] = 1$ .

### Lemma 2 (Simplex geometry) [Chan'09]

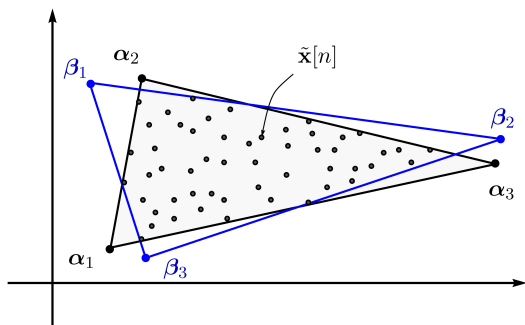
Under (A1), (A2), and (A3), all the  $\tilde{\mathbf{x}}[1], \dots, \tilde{\mathbf{x}}[L]$  are confined by a **simplex**  $\text{conv}\{\alpha_1, \dots, \alpha_N\}$ :

$$\tilde{\mathbf{x}}[n] \in \text{conv}\{\alpha_1, \dots, \alpha_N\} \subset \mathbb{R}^{N-1}, \forall n$$



**Question:** Could we estimate  $\alpha_1, \dots, \alpha_N$  from  $\tilde{\mathbf{x}}[1], \dots, \tilde{\mathbf{x}}[L]$ ?

# One Possible Approach— Craig's Belief



Formulation: Min. Volume Simplex Fitting [Chan'09] [Li-Bioucas'08]

$$\begin{aligned} \min_{\beta_1, \dots, \beta_N} \quad & V(\beta_1, \dots, \beta_N) \\ \text{s.t.} \quad & \tilde{\mathbf{x}}[n] \in \text{conv}\{\beta_1, \dots, \beta_N\}, \quad \forall n, \end{aligned} \quad (3)$$

where  $V(\beta_1, \dots, \beta_N)$  is the volume of  $\text{conv}\{\nu_1, \dots, \nu_N\}$ .

- Inspired by **Craig's belief**: find a minimum-volume simplex enclosing all data points  $\tilde{\mathbf{x}}[1], \dots, \tilde{\mathbf{x}}[L]$ . [Craig'94].

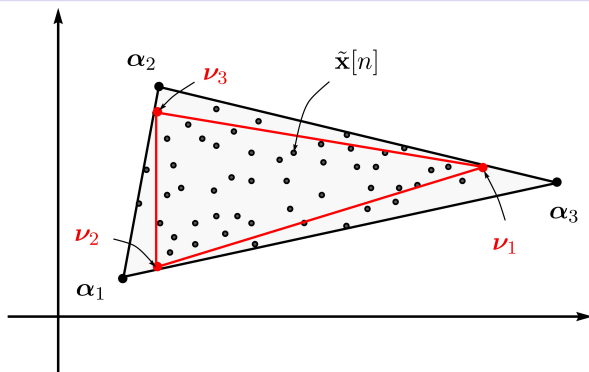
- Craig's belief is sound intuitively. But can we prove some theoretical guarantee of it?
- We prove a sufficient condition for the min. volume simplex problem as follows.

**Pure pixel assumption:**

(A4) For each  $i \in \{1, \dots, N\}$ , there exists at least one pixel index  $\ell_i$  such that  $\mathbf{x}[\ell_i] = \mathbf{a}_i$ .

**Theorem 1** (Endmember identifiability of Craig's belief)

Under (A1)-(A4), the globally optimal solution of the min. simplex volume problem is exactly  $\alpha_1, \dots, \alpha_N$ , corresponding to the true endmembers  $\mathbf{a}_i = \mathbf{C}\alpha_i + \mathbf{d}$ .



Formulation: Max. Volume Simplex Fitting

$$\begin{aligned} \max_{\nu_1, \dots, \nu_N \in \mathbb{R}^{N-1}} \quad & V(\nu_1, \dots, \nu_N) \\ \text{s.t.} \quad & \nu_i \in \text{conv}\{\tilde{\mathbf{x}}[1], \dots, \tilde{\mathbf{x}}[L]\}, \quad \forall i, \end{aligned} \quad (4)$$

- Inspired by **Winter's belief**: find a maximum-volume simplex enclosed by  $\text{conv}\{\tilde{\mathbf{x}}[1], \dots, \tilde{\mathbf{x}}[L]\}$  [Winter'99].

## Theorem 2 (Endmember identifiability of Winter's belief)

Under (A1)-(A4), the globally optimal solution of max. simplex volume problem is exactly  $\alpha_1, \dots, \alpha_N$ , corresponding to the true endmembers  $\mathbf{a}_i = \mathbf{C}\alpha_i + \mathbf{d}$ .

By Theorem 1 and Theorem 2, we can conclude that

## Relation between Craig's and Winter's beliefs

Both the min. & max. simplex volume problems can perfectly identify the endmembers in the pure pixel case.



## Formulation: Maximum Volume Simplex Fitting

$$\begin{aligned} \max_{\substack{\boldsymbol{\nu}_i \in \mathbb{R}^{N-1} \\ \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N \in \mathbb{R}^L}} V(\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N) \\ \text{s.t. } \boldsymbol{\nu}_i = \tilde{\mathbf{X}}\boldsymbol{\theta}_i, \quad \boldsymbol{\theta}_i \succeq \mathbf{0}, \quad \mathbf{1}_L^T \boldsymbol{\theta}_i = 1 \quad \forall i, \end{aligned}$$

where  $\tilde{\mathbf{X}} = [ \tilde{\mathbf{x}}[1], \dots, \tilde{\mathbf{x}}[L] ] \in \mathbb{R}^{(N-1) \times L}$ .

- The maximum simplex volume problem is a nonconvex optimization problem: The constraints are convex, but the objective

$$V(\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N) = \left| \det \left( \begin{bmatrix} \boldsymbol{\nu}_1 & \cdots & \boldsymbol{\nu}_N \\ 1 & \cdots & 1 \end{bmatrix} \right) \right| / (N-1)!$$

is **nonconcave**.

- Maximizing  $V(\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N)$  w.r.t. each  $\boldsymbol{\nu}_i$  is however easy, with convex optimization.

## Formulation: Maximum Volume Simplex Fitting

$$\begin{aligned} \max_{\substack{\boldsymbol{\nu}_i \in \mathbb{R}^{N-1} \\ \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N \in \mathbb{R}^L}} V(\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N) \\ \text{s.t. } \boldsymbol{\nu}_i = \tilde{\mathbf{X}}\boldsymbol{\theta}_i, \quad \boldsymbol{\theta}_i \succeq \mathbf{0}, \quad \mathbf{1}_L^T \boldsymbol{\theta}_i = 1 \quad \forall i, \end{aligned}$$

where  $\tilde{\mathbf{X}} = [ \tilde{\mathbf{x}}[1], \dots, \tilde{\mathbf{x}}[L] ] \in \mathbb{R}^{(N-1) \times L}$ .

- By cofactor expansion,

$$V(\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N) \propto \left| \mathbf{b}_j^T \boldsymbol{\nu}_j + (-1)^{N+j} \det(\boldsymbol{\nu}_{Nj}) \right|,$$

where  $\mathbf{b}_j$  &  $\boldsymbol{\nu}_{ij}$  are variables dependent on  $\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_{j-1}, \boldsymbol{\nu}_{j+1}, \dots, \boldsymbol{\nu}_N$ .

- $V(\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N)$  is absolute affine w.r.t. each  $\boldsymbol{\nu}_j$ .
- Maximization w.r.t.  $\boldsymbol{\nu}_j$  can be globally optimally solved by two linear programs (LPs).

## Formulation: Maximum Volume Simplex Fitting

$$\begin{aligned} \max_{\substack{\boldsymbol{\nu}_i \in \mathbb{R}^{N-1} \\ \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N \in \mathbb{R}^L}} V(\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N) \\ \text{s.t. } \boldsymbol{\nu}_i = \tilde{\mathbf{X}}\boldsymbol{\theta}_i, \quad \boldsymbol{\theta}_i \succeq \mathbf{0}, \quad \mathbf{1}_L^T \boldsymbol{\theta}_i = 1 \quad \forall i, \end{aligned}$$

where  $\tilde{\mathbf{X}} = [ \tilde{\mathbf{x}}[1], \dots, \tilde{\mathbf{x}}[L] ] \in \mathbb{R}^{(N-1) \times L}$ .

**Alternating Method****Repeat**

solve the  $j$ th partial maximization problem

$$\begin{aligned} (\hat{\boldsymbol{\nu}}_j, \hat{\boldsymbol{\theta}}_j) := \arg \max_{\boldsymbol{\nu}_j, \boldsymbol{\theta}_j} V(\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N) \\ \text{s.t. } \boldsymbol{\nu}_j = \tilde{\mathbf{X}}\boldsymbol{\theta}_j, \quad \boldsymbol{\theta}_j \succeq \mathbf{0}, \quad \mathbf{1}_L^T \boldsymbol{\theta}_j = 1 \end{aligned}$$

by two LPs

update  $j := (j \text{ modulo } N) + 1$ .

**Until** some stopping rule is satisfied.

## Formulation: Minimum Volume Simplex Fitting

$$\begin{aligned} \min_{\mathbf{B}, \beta_N, s'[1], \dots, s'[L]} \quad & |\det(\mathbf{B})| \\ \text{s.t.} \quad & s'[n] \succeq \mathbf{0}, \quad \mathbf{1}_{N-1}^T s'[n] \leq 1, \\ & \tilde{\mathbf{x}}[n] = \beta_N + \mathbf{B}s'[n], \quad \forall n = 1, \dots, L. \end{aligned}$$

Let  $\mathbf{H} = \mathbf{B}^{-1} \in \mathbb{R}^{(N-1) \times (N-1)}$  and  $\mathbf{g} = \mathbf{B}^{-1}\beta_N \in \mathbb{R}^{N-1}$ .

Then,  $s'[n] = \mathbf{B}^{-1}(\tilde{\mathbf{x}}[n] - \beta_N) = \mathbf{H}\tilde{\mathbf{x}}[n] - \mathbf{g}$ .

Then the problem can be transformed as [Li-Bioucas'08], [Chan'09]

$$\begin{aligned} \max_{\mathbf{H}, \mathbf{g}} \quad & |\det(\mathbf{H})| \\ \text{s.t.} \quad & \mathbf{H}\tilde{\mathbf{x}}[n] - \mathbf{g} \succeq \mathbf{0}, \\ & \mathbf{1}_{N-1}^T (\mathbf{H}\tilde{\mathbf{x}}[n] - \mathbf{g}) \leq 1, \quad \forall n = 1, \dots, L. \end{aligned} \tag{5}$$

We can use alternating linear programming again!

- 100 Monte Carlo runs were performed.
- $\mathbf{x}[n]$ : 1000 synthetic pixels ( $L = 1000$ ).
- $\mathbf{a}_1, \dots, \mathbf{a}_N$ : selected from USGS library ( $M = 417$ ) [Clark'93].
- $\mathbf{s}[n]$ : Dirichlet distribution [Nascimento'05].
- **Performance index:** Root-mean-square spectral angle (error performance measure) is defined as

$$\phi_{en} = \min_{\pi \in \Pi_N} \sqrt{\frac{1}{N} \sum_{i=1}^N \left[ \arccos \left( \frac{\mathbf{a}_i^T \hat{\mathbf{a}}_{\pi_i}}{\|\mathbf{a}_i\| \|\hat{\mathbf{a}}_{\pi_i}\|} \right) \right]^2}$$

$$\phi_{ab} = \min_{\pi \in \Pi_N} \sqrt{\frac{1}{N} \sum_{i=1}^N \left[ \arccos \left( \frac{\mathbf{s}_i^T \hat{\mathbf{s}}_{\pi_i}}{\|\mathbf{s}_i\| \|\hat{\mathbf{s}}_{\pi_i}\|} \right) \right]^2}$$

where  $\Pi_N$  is the set of all the permutations of  $\{1, 2, \dots, N\}$ .<sup>†</sup>

---

<sup>†</sup>  $\mathbf{s}_i = [s_i[1], \dots, s_i[L]]^T$  denotes the  $i$ th abundance map, and  $\hat{\mathbf{a}}_i$  and  $\hat{\mathbf{s}}_i$  denote the estimated  $\mathbf{a}_i$  and  $\mathbf{s}_i$ , respectively.

- Six endmembers ( $N = 6$ ) from USGS library were selected.
- We generated seven data sets with different purity levels  $\rho = 0.7, 0.75, \dots, 1$  for performance evaluation.

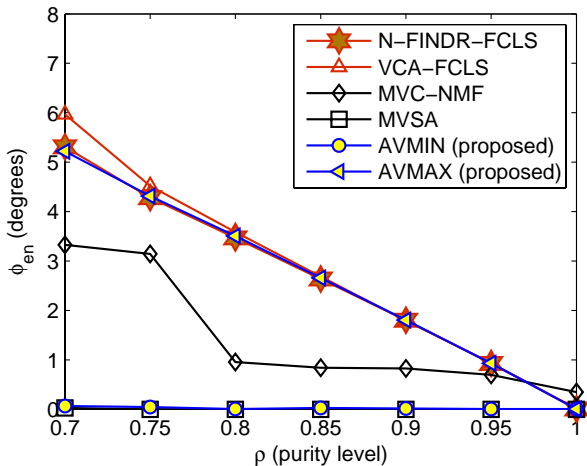
### Purity level

A data set with *purity level*  $\rho$  denotes a set of  $L$  observed pixels with all the purities  $\rho_1, \dots, \rho_L$  in the range  $[\rho - 0.1, \rho]$ , where

$$\frac{1}{\sqrt{N}} \leq \rho_n = \|\mathbf{s}[n]\| \leq 1$$

is a purity measure for an observed pixel  $\mathbf{x}[n] (= \sum_{i=1}^N s_i[n] \mathbf{a}_i)$ . The closer to unity the value of  $\rho_n$ , the more a single endmember  $\mathbf{a}_i$  dominates in  $\mathbf{x}[n]$ .

⇒ The generated data for  $\rho = 1$  includes some highly pure pixels.



**Figure:** Simulation results of the endmember estimates obtained by the various algorithms under test for different purity levels ( $\phi_{en}$ ).

<sup>0</sup>VCA: Vertex component analysis [Nascimento'05]

MVC-NMF: Minimum volume constrained nonnegative matrix factorization [Miao'07]

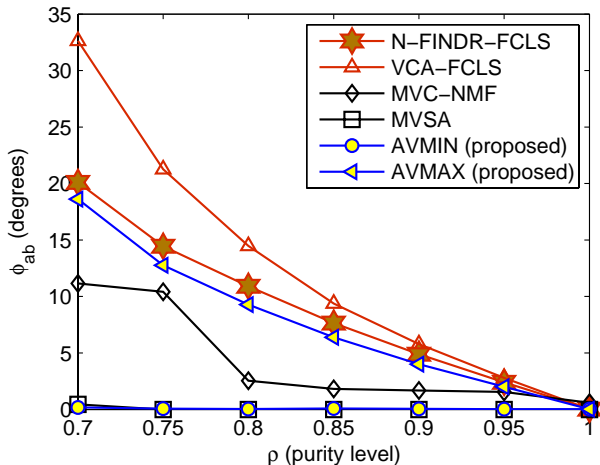


Figure: Simulation results of the abundance estimates obtained by the various algorithms under test for different purity levels ( $\phi_{ab}$ ).

<sup>0</sup>MVSA: Minimum volume simplex analysis [Li-Bioucas'08]



- We have provided a **convex analysis and optimization perspective** to hyperspectral unmixing, from dimension reduction, criteria, to algorithms.
- Open questions arising:
  - theoretical endmember identifiability conditions without pure pixels (positive by simulations, but a tricky analysis problem...)
  - other possible formulations (using determinant as the objective is not the only way out!)

- Chan'08 T.-H. Chan, W.-K. Ma, C.-Y. Chi, and Y. Wang, "A convex analysis framework for blind separation of non-negative sources," *IEEE Trans. Signal Processing*, vol. 56, no. 10, pp. 5120-5134, Oct. 2008.
- Chan'09 T.-H. Chan, C.-Y. Chi, Y.-M. Huang and W.-K. Ma, "A convex analysis based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," in *International Conference on Acoustics, Speech and Signal Processing*, Taipei, Taiwan, April 19-24, 2009, pp. 1089-1092.
- Clark'93 R. N. Clark, G. A. Swayze, A. Gallagher, T. V. King, and W. M. Calvin, "The U.S. geological survey digital spectral library: version 1: 0.2 to 3.0  $\mu\text{m}$ ," in *U.S. Geol. Surv., Denver, CO.*, 1993, pp. 93-592.
- Craig'94 M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542-552, May 1994.
- Jolliffe'86 I. T. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 1986.
- Miao'07 L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765-777, Mar. 2007.
- Li-Bioucas'08 J. Li and J. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE International Geoscience and Remote Sensing Symposium*, vol. 4, Boston, MA, Aug. 8-12, 2008, pp. 2369-2371.
- Nascimento'05 J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898-910, Apr. 2005.
- Winter'99 M. E. Winter, "N-findr: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proc. SPIE Conf. Imaging Spectrometry*, Pasadena, CA, Oct. 1999, pp. 266-275.
- Strang'06 G. Strang, *Linear Algebra and Its Application*, 4th ed. CA: Thomson, 2006.

*Thank You for Your Attention!*