Hyperspectral Unmixing from A Convex Analysis and Optimization Perspective

Tsung-Han Chan†, Wing-Kin Ma*, Chong-Yung Chi†, and A.ArulMurugan†

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

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

\[ x[n] = As[n] = \sum_{i=1}^{N} s_i[n]a_i, \quad n = 1, \ldots, L \]  

- \( A = [a_1, \ldots, a_N] \in \mathbb{R}^{M \times N} \), \( a_i \) is the \( i \)th endmember signature,
- \( s[n] = [s_1[n], \ldots, s_N[n]]^T \) 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)

\[ \mathbf{x}[n] = \mathbf{A}s[n] = \sum_{i=1}^{N} s_i[n] \mathbf{a}_i, \quad n = 1, \ldots, L \] (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 \( \mathbf{a}_1, \ldots, \mathbf{a}_N \) are linearly independent.
The affine hull of \( \{a_1, \ldots, a_N\} \subset \mathbb{R}^M \) is defined as:

\[
\text{aff}\{a_1, \ldots, a_N\} = \left\{ x = \sum_{i=1}^{N} \theta_i a_i \mid \theta \in \mathbb{R}^N, \sum_{i=1}^{N} \theta_i = 1 \right\}.
\]

An affine hull can always be represented by

\[
\mathcal{A}(C, d) \triangleq \left\{ x = C\alpha + d \mid \alpha \in \mathbb{R}^P \right\}
\]

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

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

\[
x[n] \in \text{aff}\{a_1, \ldots, a_N\} = \mathcal{A}(C, d), \quad \forall n = 1, \ldots, L,
\]

with \( P = N - 1. \)
An Geometry Illustration for $N = 3$

$\text{aff}\{a_1, a_2, a_3\}$

$x[n]$
An Geometry Illustration for $N = 3$
Lemma 1 (Affine set fitting) [Chan’08]

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

\[ A(C, d) = \text{aff}\{x[1], \ldots, x[L]\}. \]

Moreover, \((C, d)\) can be obtained from \(x[1], \ldots, x[L]\) by

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

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

- 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

\[ A(C, d) = \text{aff}\{x[1], \ldots, x[L]\}. \]

Moreover, \((C, d)\) can be obtained from \(x[1], \ldots, x[L]\) by

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

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

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

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

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

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

where $\alpha_i = C^T(a_i - d)$ is the $i$th dimension-reduced endmember.
The **convex hull** of \( \{ \alpha_1, \ldots, \alpha_N \} \subset \mathbb{R}^M \) is defined as:

\[
\text{conv}\{\alpha_1, \ldots, \alpha_N\} = \left\{ x = \sum_{i=1}^{N} \theta_i \alpha_i \left| \begin{array}{c} \theta \succeq 0, \\ \sum_{i=1}^{N} \theta_i = 1 \end{array} \right. \right\}
\]

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

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

**Lemma 2 (Simplex geometry) [Chan’09]**

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

\[
\tilde{x}[n] \in \text{conv}\{\alpha_1, \ldots, \alpha_N\} \subset \mathbb{R}^{N-1}, \ \forall n
\]
Question: Could we estimate $\alpha_1, \ldots, \alpha_N$ from $\tilde{x}[1], \ldots, \tilde{x}[L]$?
One Possible Approach—Craig’s Belief

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

\[
\min_{\beta_1,\ldots,\beta_N} V(\beta_1, \ldots, \beta_N)
\]

s.t. \( \tilde{x}[n] \in \text{conv}\{\beta_1, \ldots, \beta_N\}, \ \forall \ n, \)

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

Inspired by Craig’s belief: find a minimum-volume simplex enclosing all data points \( \tilde{x}[1], \ldots, \tilde{x}[L] \). [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 $i \in \{1, \ldots, N\}$, there exists at least one pixel index $\ell_i$ such that $x[\ell_i] = a_i$.

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

Under (A1)-(A4), the globally optimal solution of the min. simplex volume problem is exactly $\alpha_1, \ldots, \alpha_N$, corresponding to the true endmembers $a_i = C\alpha_i + d$. 
Another Possible Approach— Winter’s Belief

Formulation: Max. Volume Simplex Fitting

\[
\max_{\nu_1, \ldots, \nu_N \in \mathbb{R}^{N-1}} \quad V(\nu_1, \ldots, \nu_N)
\]
\[
\text{s.t.} \quad \nu_i \in \text{conv}\{\tilde{x}[1], \ldots, \tilde{x}[L]\}, \quad \forall \ i,
\]

Inspired by Winter’s belief: find a maximum-volume simplex enclosed by \(\text{conv}\{\tilde{x}[1], \ldots, \tilde{x}[L]\}\) [Winter’99].
Theorem 2 (Endmember identifiability of Winter’s belief)

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

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.
Solving the Max. Simplex Volume Problem

Formulation: Maximum Volume Simplex Fitting

\[
\begin{align*}
\max_{\nu_1, \ldots, \nu_N \in \mathbb{R}^{N-1}, \theta_1, \ldots, \theta_N \in \mathbb{R}^L} & \quad V(\nu_1, \ldots, \nu_N) \\
\text{s.t.} & \quad \nu_i = \tilde{X}\theta_i, \quad \theta_i \succeq 0, \quad 1^T_L\theta_i = 1 \quad \forall \ i,
\end{align*}
\]

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

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

\[
V(\nu_1, \ldots, \nu_N) = \left| \det \left( \begin{bmatrix} \nu_1 & \cdots & \nu_N \\ 1 & \cdots & 1 \end{bmatrix} \right) \right| / (N - 1)!
\]

is nonconcave.

- Maximizing \( V(\nu_1, \ldots, \nu_N) \) w.r.t. each \( \nu_i \) is however easy, with convex optimization.
Formulation: Maximum Volume Simplex Fitting

\[
\max_{\nu_i \in \mathbb{R}^{N-1}, \theta_1, \ldots, \theta_N \in \mathbb{R}^L} V(\nu_1, \ldots, \nu_N)
\]

s.t. \(\nu_i = \tilde{X}\theta_i, \quad \theta_i \succeq 0, \quad 1^T L \theta_i = 1 \forall i,\)

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

- By cofactor expansion,
  \[V(\nu_1, \ldots, \nu_N) \propto |b_j^T \nu_j + (-1)^{N+j} \det(V_{Nj})|,\]

  where \(b_j \) & \(V_{ij}\) are variables dependent on \(\nu_1, \ldots, \nu_{j-1}, \nu_{j+1}, \ldots, \nu_N.\)

- \(V(\nu_1, \ldots, \nu_N)\) is absolute affine w.r.t. each \(\nu_j.\)

- Maximization w.r.t. \(\nu_j\) can be globally optimally solved by two linear programs (LPs).
Formulation: Maximum Volume Simplex Fitting

\[
\max_{\nu_i \in \mathbb{R}^{N-1}, \theta_1, \ldots, \theta_N \in \mathbb{R}^L} V(\nu_1, \ldots, \nu_N) \\
\text{s.t. } \quad \nu_i = \tilde{X}\theta_i, \quad \theta_i \succeq 0, \quad 1^T_L \theta_i = 1 \quad \forall \ i,
\]

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

Alternating Method

Repeat

solve the \( j \)th partial maximization problem

\[
(\hat{\nu}_j, \hat{\theta}_j) := \arg \max_{\nu_j, \theta_j} V(\nu_1, \ldots, \nu_N) \\
\text{s.t. } \quad \nu_j = \tilde{X}\theta_j, \quad \theta_j \succeq 0, \quad 1^T_L \theta_j = 1
\]

by two LPs

update \( j := (j \mod N) + 1 \).

Until some stopping rule is satisfied.
Formulation: Minimum Volume Simplex Fitting

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

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

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

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

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

We can use alternating linear programming again!
100 Monte Carlo runs were performed.

\( x[n] \): 1000 synthetic pixels \((L = 1000)\).

\( a_1, \ldots, a_N \): selected from USGS library \((M = 417)\) [Clark’93].

\( s[n] \): Dirichlet distribution [Nascimento’05].

**Performance index:** Root-mean-square spectral angle (error performance measure) is defined as

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

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

where \( \Pi_N \) is the set of all the permutations of \( \{1, 2, \ldots, N\} \).

\[\dagger\] \( s_i = [s_i[1], \ldots, s_i[L]]^T \) denotes the \( i \)th abundance map, and \( \hat{a}_i \) and \( \hat{s}_i \) denote the estimated \( a_i \) and \( s_i \), respectively.
Six endmembers ($N = 6$) from USGS library were selected.

We generated seven data sets with different purity levels $\rho = 0.7, 0.75, \ldots, 1$ for performance evaluation.

**Purity level**

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

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

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

⇒ 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 ($\phi_{en}$).

$^0$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 ($\phi_{ab}$).
Conclusions

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!)
<table>
<thead>
<tr>
<th>Reference</th>
<th>Title and Details</th>
</tr>
</thead>
</table>

*Thank You for Your Attention!*