Hyperspectral Unmixing in Remote Sensing: What Do Signal Processing People Learn from There?

Wing-Kin (Ken) Ma

July 3, 2015

Department of Electronic Engineering
The Chinese University of Hong Kong

A Talk Presented at KTH Royal Institute of Technology, Sweden
Hyperspectral Imaging

Courtesy to [Manolakis-Golowich-DiPietro2014].

- **Hyperspectral sensors** record EM scattering patterns over > 200 spectral bands, from visible to near-infrared wavelength at a resolution of 10nm.
• The high spectral resolution of hyperspectral imaging enables many possibilities.
  – we can distinguish different materials, as revealed by their spectral signatures.
Applications of Hyperspectral Imaging

- Hyperspectral imaging has found numerous applications in remote sensing, such as
  - mineral identification,
  - agriculture,
  - environment monitoring,
  - terrain classification, land-cover mapping,
  - object detection, change detection, ...

- It also finds applications in
  - planetary exploration (like Mars) and astrophysics, and
  - non-remote sensing scenarios such as food inspection, forensics, medical imaging and chemometrics.

- A key topic in hyperspectral imaging is **hyperspectral unmixing (HU)**.
  - the problem is to retrieve the materials’ spectral signatures and their compositions, both being unknown, from the observed hyperspectral image.
AVIRIS Cuprite image. Courtesy to USGS.
Hyperspectral Unmixing: Signal Model

- Modeling hyperspectral signals is a complex problem.
- Consider the simple but representative linear mixing model, wherein the incident light is assumed to interact only with one material before reflecting off.
Hyperspectral Unmixing: Signal Model

- Model:

\[ y[n] = \sum_{i=1}^{N} a_i s_i[n] + \nu[n] = A s[n] + \nu[n], \quad n = 1, \ldots, L, \]

where
- \( y[n] \in \mathbb{R}^M \) is the measured hyperspectral vector at pixel \( n \);
- \( A = [a_1, \ldots, a_N] \), \( a_i \in \mathbb{R}^M \) is an endmember signature vector;
- \( s[n] \in \mathbb{R}^N \) is the abundance vector at pixel \( n \); \( \nu[n] \) is noise.
Hyperspectral Unmixing: Signal Model

- **Assumptions** (standard):
  - \( s_i[n] \geq 0 \) for all \( i, n \) (non-negativity), \( \sum_{i=1}^{N} s_i[n] = 1 \) for all \( n \) (sum-to-one).
  - \( A \) has full column rank.
  - \( N \), the model order, is known.

\[
\begin{align*}
  x &= a_1 s_1[n] + a_2 s_2[n] + a_3 s_3[n] \\
  y[n] &= M x
\end{align*}
\]
Hyperspectral Unmixing: Problem Statement

• **Problem:** recover $\{s[n]\}_{n=1}^L$ from $\{y[n]\}_{n=1}^L$ without knowing $A$.
  
  – can also be stated as identifying $A$ from $\{y[n]\}_{n=1}^L$ without knowing $\{s[n]\}_{n=1}^L$.
  
  – essentially a **blind source separation (BSS)** problem from a signal processing (SP) viewpoint.
**Problem:** separate multiple speakers’ voices using an array of microphones.

**Challenge:** the location and propagation characteristics of each speaker are not known. Thus, the problem is blind.
Why not Apply BSS Algorithms from SP to Solve HU Directly?

- While BSS in SP is powerful, it’s not a magic that works for any scenario.

- All BSS methods make specific assumptions on the characteristics of \( \{s[n]\}_n \) and/or \( A \), and utilize them to achieve BSS.

**Example:** Independent component analysis (ICA), a well-known BSS framework, assumes that each \( s_i[n] \) is non-Gaussian & statistically independent of one other.

- The statistical independence assumption is reasonable for speech and audio signals.
- It is however violated in HU, due to the sum-to-one abundance constraint.
How about Tools in Machine Learning, like Non-Negative Matrix Factorization (NMF)?

- Problem: Given $Y = [y[1], \ldots, y[L]]$, solve

$$
\min_{A \geq 0, S \geq 0} \|Y - AS\|_F^2,
$$

where $A \in \mathbb{R}^{M \times N}$, $S \in \mathbb{R}^{N \times L}$.

- idea (intuitive): “learn the parts of objects” [Lee-Seung1999].

- widely used in machine learning and other fields.

- NP-hard in general [Vavasis2009].

- identifiability problem: the NMF solution may not be unique; i.e., there may exist two NMF solutions, denoted $(\hat{A}, \hat{S})$, $(\tilde{A}, \tilde{S})$, such that $\hat{A}$ does not equal $\tilde{A}$ even up to permutations and scaling of the columns.

- works in HU in practice, but often with modifications.
**BSS Arising from Hyperspectral Remote Sensing**

- Classical BSS algorithms developed in SP, e.g., ICA, do not fall into any of the mainstream HU approaches.

- It turns out that the developments of HU in remote sensing have led to a new branch of BSS approaches not seen in classical BSS studies.
  - HU also has strong link to other topics, such as text mining in machine learning.

- **Aim of this talk:** review beautiful ideas of HU from an SP people’s eyes

- **Scope:**
  - pure pixel search
  - convex geometry
  - sparse regression, if time allows
  - non-negative matrix factorization, also if time allows

Other approaches like Bayesian solutions will not be covered, but note that they are also important.
A Recent Overview


Acknowledgment: Xiao Fu for helping prepare the early version of this slides.
Pure Pixel Search
Pure Pixels

(a) TERRAIN hyperspectral image set. (b) Spectral signatures of hand picked pixels.

- While a captured scene is generally a mixture of endmembers, there are cases where some pixels contain only one endmember.
**Definition:** Endmember $i$ is said to have a pure pixel if, for some index $\ell_i$, $s[\ell_i] = e_i$, where $e_i$ is a unit vector, with $[e_i]_j = 0$ for all $j \neq i$ and $[e_i]_i = 1$.

- Note that in the noiseless case, $y[\ell_i] = A e_i = a_i$.

![Diagram](image-url)
**Pure Pixels**

**Definition:** Endmember $i$ is said to have a pure pixel if, for some index $\ell_i$,

$$s[\ell_i] = e_i,$$

where $e_i$ is a unit vector, with $[e_i]_j = 0$ for all $j \neq i$ and $[e_i]_i = 1$.

- Note that in the noiseless case, $y[\ell_i] = Ae_i = a_i$.

- **Implication:**
  - Suppose that every endmember has a pure pixel, and there is no noise.
  - If we know $\ell_1, \ldots, \ell_N$, then $[y[\ell_1], \ldots, y[\ell_N]] = A$ — and the problem is solved!
**Pure Pixels**

**Definition:** Endmember $i$ is said to have a pure pixel if, for some index $\ell_i$,

$$s[\ell_i] = e_i,$$

where $e_i$ is a unit vector, with $[e_i]_j = 0$ for all $j \neq i$ and $[e_i]_i = 1$.

- Note that in the noiseless case, $y[\ell_i] = Ae_i = a_i$.

- **Implication:**
  - Suppose that every endmember has a pure pixel, and there is no noise.
  - If we know $\ell_1, \ldots, \ell_N$, then $[y[\ell_1], \ldots, y[\ell_N]] = A$ — and the problem is solved!

- **Problem:** find the pure pixel indices.
Successive Projection Algorithm (SPA)

- a simple and efficient algorithm for pure pixel search.

- prerequisite required to understand SPA: basic linear algebra.
  - In comparison, understanding the principle of ICA requires a good background on statistics and probability theory.
SPA: How it Works

- **Assumptions:** no noise, and the pure pixel assumption.

- The index
  
  $$\hat{\ell}_1 = \arg \max_{n=1,\ldots,L} \|y[n]\|_2^2,$$

  where $\| \cdot \|_2$ is the 2-norm, is a pure pixel index.

- Proof:
  
  $$\|y[n]\|_2 = \left\| \sum_{i=1}^{N} s_i[n] a_i \right\|_2 \leq \sum_{i=1}^{N} \|s_i[n] a_i\|_2$$

  (triangle inequality)

  $$= \sum_{i=1}^{N} s_i[n] \|a_i\|_2$$

  ($s_i[n] \geq 0$)

  $$\leq \max_{i=1,\ldots,N} \|a_i\|_2.$$  

  ($\sum_{i=1}^{N} s_i[n] = 1$)

  Equality above hold if and only if $s[n] = e_j$, where $j = \arg \max_{i=1,\ldots,N} \|a_i\|_2$. 
SPA: How it Works

- Let $P_{\perp X} = I - X(X^TX)^{-1}X^T$ be the orthogonal complement projector of $X$.

- Suppose we previously obtained $k - 1$ (distinct) pure pixels

$$\hat{A}_{1:k-1} \triangleq [ y[\hat{\ell}_1], \ldots, y[\hat{\ell}_{k-1}] ] = A_{1:k-1}.$$  

The index

$$\hat{\ell}_k = \arg\max_{n=1,\ldots,L} \|P_{\perp \hat{A}_{1:k-1}} y[n]\|_2^2,$$

is a pure pixel index of a new endmember.

- Proof: Since $P_{\perp \hat{A}_{1:k-1}} a_i = 0$ for $i = 1, \ldots, k - 1$, and $P_{\perp \hat{A}_{1:k-1}} a_i \neq 0$ otherwise,

$$P_{\perp \hat{A}_{1:k-1}} y[n] = \sum_{i=k}^{N} s_i[n] P_{\perp \hat{A}_{1:k-1}} a_i.$$

By the same trick as the previous, we obtain the desired result.
The SPA pseudocode:

**Algorithm:** SPA

**input** \( \{y[n]\}_{n=1}^{L}, N \).

\( \hat{l}_1 = \arg \max_{n=1,...,L} \|y[n]\|_2^2 \).

\( \hat{A} = y[\hat{l}_1] \).

for \( k = 2, \ldots, N \),

\( \hat{l}_k = \arg \max_{n=1,...,L} \|P_{\hat{A}}^\perp y[n]\|_2^2 \).

\( \hat{A} := [\hat{A}, y[\hat{l}_k]] \).

**output** \( \hat{A} \).

**Fact** [Chan-Ma-Ambikapathi-Chi2011]: In the noiseless case and under the pure pixel assumption, SPA exactly identifies all the endmember signatures \( a_1, \ldots, a_N \).

**Question:** what about the noisy case?
SPA in the Noisy Case

- SPA is shown to be robust to noise.

**Theorem [Gillis-Vavasis2014]:** Let $\epsilon = \max_{n=1,\ldots,L} \|\mathbf{v}[n]\|_2$ be the noise level. Under the pure pixel assumption and assuming that the noise level satisfies

$$
\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(A)}{\sqrt{N} \kappa^2(A)}\right),
$$

where $\sigma_{\min}(A)$ and $\kappa(A)$ denote the smallest singular value and condition number of $A$, resp., SPA identifies all the endmember signatures up to error $\mathcal{O}(\epsilon \cdot \kappa^2(A))$; more precisely,

$$
\max_{i=1,\ldots,N} \min_{j=1,\ldots,N} \|a_i - \hat{a}_j\|_2 \leq \mathcal{O}\left(\epsilon \cdot \kappa^2(A)\right).
$$

- **Implication:** SPA’s noise robustness depends on the conditioning of $A$. In particular, more similar endmembers $\implies$ higher noise sensitivity.
Historical Note on SPA

• SPA has been repeatedly rediscovered.
  – To our best knowledge, SPA first appears in chemometrics [Araújo et al. 2001].
  – It is very similar to the automatic target generation process (ATGP) [Ren-Chang2003], developed in remote sensing.
  – The same algorithm can be derived by at least two different ways, namely simplex volume maximization [Chan-Ma-Ambikapathi-Chi2011] and self-dictionary sparse regression [Fu-Ma-Chan-Bioucas2015].
  – SPA is closely related to the modified Gram-Schmidt algorithm with column pivoting in numerical linear algebra.

• Vertex component analysis (VCA) [Nascimento-Bioucas2003] is a highly popularized HU algorithm. VCA and SPA use essentially the same principles, esp., successive nulling.
Convex Geometry
Convex Geometry (CG): Some History

• We have seen how HU can be easily handled under the pure pixel assumption.

• Actually, the pure pixel concepts came from the study of convex geometry (CG) of hyperspectral signals [Boardman-Kruse-Green1995].

• In hyperspectral remote sensing, Craig’s work in the 1990s is considered most seminal in introducing CG [Craig1990], [Craig1994].
  – The subsequent works by Boardman and Winter, which popularized pure pixel search, should also be mentioned [Boardman-Kruse-Green1995], [Winter1999].

• Intriguingly, CG has been discovered or rediscovered several times in different fields.
  – geology [Imbrie1964], also [Full-Ehrlich-Klovan1981];
  – chemometrics [Perczel et al. 1989];
  – nuclear magnetic resonance spectroscopy [Naanaa-Nuzillard2005];
  – signal processing theory and methods [Chan-Ma-Chi-Wang2008].
• **convex hull**: \( \text{conv}\{a_1, \ldots, a_N\} = \{y = \sum_{i=1}^{N} a_i \theta_i \mid \theta \geq 0, 1^T \theta = 1\} \).

• **simplex**: \( \text{conv}\{a_1, \ldots, a_N\} \) is a simplex if \( a_1, \ldots, a_N \) are affinely independent.
  
  – the set of all vertices of a simplex is \( \{a_1, \ldots, a_N\} \).
Convex Geometry Observation

- Recall the noiseless signal model

\[ y[n] = \sum_{i=1}^{N} s_i[n] a_i, \]

and that \( s_i[n] \geq 0, \sum_{i=1}^{N} s_i[n] = 1. \)

Apparently, we have

\[ y[n] \in \text{conv}\{a_1, \ldots, a_N\}. \]

- **Observation**: each hyperspectral pixel \( y[n] \) lies in the simplex of the ground-truth endmembers \( \{a_1, \ldots, a_N\} \).
Convex Geometry Observation

- Recall the noiseless signal model

\[ y[n] = \sum_{i=1}^{N} s_i[n] a_i, \]

and that \( s_i[n] \geq 0, \sum_{i=1}^{N} s_i[n] = 1. \) Apparently, we have

\[ y[n] \in \text{conv}\{a_1, \ldots, a_N\}. \]

- **Observation**: each hyperspectral pixel \( y[n] \) lies in the simplex of the ground-truth endmembers \( \{a_1, \ldots, a_N\} \).

- **Problem**: identify the vertices of \( \text{conv}\{a_1, \ldots, a_N\} \) from \( \{y[n]\}_{n=1}^{L} \).
Convex Geometry Observed in Craig’s Work [Craig 1994]

Minimum-Volume Transforms for Remotely Sensed Data

Maurice D. Craig

Abstract—Scatter diagrams for multispectral remote sensing data tend to be triangular, in the two-band case, pyramidal for three bands, and so on. They radiate away from the so-called darkpoint, which represents the scanner’s response to an unilluminated target. A minimum-volume transform may be described ( provisionally ) as a nonorthogonal linear transformation of the multivariate data to new axes passing through the dark point, with directions chosen such that they ( for two bands, or the new coordinate planes ( for three bands, etc. ) embrace the data cloud as tightly as possible.

The reason for the observed shapes of scatter diagrams is to be found in the theory of linear mixing at the subfootprint scale. Thus, suitably defined, minimum-volume transforms can often be used to unmix images into new spatial variables showing the proportions of the different cover types present, a type of enhancement that is not only intense, but physically meaningful. The present paper furnishes details for constructing computer programs to effect this operation. It will serve as a convenient technical source that may be referenced in subsequent, more profusely illustrated publications that address the intended application, the mapping of surface mineralogy.

I. INTRODUCTION

This paper describes processing algorithms for two closely related transformations, both applicable to radiance data from multispectral scanners. It supplies de-

Fig. 1. Two-band triangular scatter plot for a 512 × 512 subscene of a Landsat Thematic Mapper image (actually WRS 111-075, Nullagine, W.A., acquired August 18, 1986).

away from the so-called dark point, the scanner’s response to a target of nil reflectance in all bands (see Fig. 1).

This appearance of bivariate scatter diagrams now sug-
Dimension Reduction

- By exploiting the affine nature of $y[n]$, we can perform dimension reduction

$$x[n] = \sum_{i=1}^{N} s_i[n] b_i$$

where $b_1, \ldots, b_N \in \mathbb{R}^{N-1}$ are dimension-reduced endmembers.
**Simplex Volume Maximization: Intuition**

Winter’s belief [Winter1999]: the true endmembers may be located by finding a collection of pixels whose simplex volume is the largest.

![Diagram showing simplex volume maximization](image)
Winter’s belief [Winter1999]: the true endmembers may be located by finding a collection of pixels whose simplex volume is the largest.
Winter’s belief [Winter1999]: the true endmembers may be located by finding a collection of pixels whose simplex volume is the largest.
Winter’s belief [Winter1999]: the true endmembers may be located by finding a collection of pixels whose simplex volume is the largest.
Winter’s belief [Winter1999]: the true endmembers may be located by finding a collection of pixels whose simplex volume is the largest.

- Winter’s belief led to N-FINDR, an HU algorithm class that has been widely used and has numerous variations.
Simplex Volume Maximization: Formulation

- volume maximization (VolMax) formulation [Chan-Ma-Ambikapathi-Chi2011]:

\[
\begin{align*}
\max_{\mathbf{B}} \ & \ vol(\mathbf{B}) \\
\text{s.t.} \ & \ b_i \in \text{conv}\{\mathbf{x}[1], \ldots, \mathbf{x}[L]\}, \\
& \quad i = 1, \ldots, N,
\end{align*}
\]

where

\[
vol(\mathbf{B}) = \frac{1}{(N-1)!} \left| \det \left( \begin{bmatrix} b_1 & \cdots & b_N \end{bmatrix} \right) \right|
\]

is the simplex volume of \( \text{conv}\{\mathbf{b}_1, \ldots, \mathbf{b}_N\} \).
Simplex Volume Maximization: Optimization

- VolMax formulation:

\[
\max_B \text{vol}(B) \\
\text{s.t. } b_i \in \text{conv}\{x[1], \ldots, x[L]\}, \forall i.
\]

- VolMax is NP-hard in general [Çivril et al. 2009].
Simplex Volume Maximization: Optimization

• VolMax formulation:

\[
\max_B \text{vol}(B) \\
\text{s.t. } b_i \in \text{conv}\{x[1], \ldots, x[L]\}, \forall i.
\]

• Consider alternating optimization.

  – generate an iterate \(\hat{B}\) via cyclic updates of

  \[
  \hat{b}_k := \arg \max_{b_k} \text{vol}(\hat{B}_{-k}, b_k) \text{ s.t. } b_k \in \text{conv}\{x[1], \ldots, x[L]\}, \ k = 1, \ldots, N.
  \]

  – the update has a closed form: \(\hat{b}_k = x[\ell_k], \ \ell_k = \arg \max_{n=1, \ldots, L} \text{vol}(\hat{B}_{-k}, x_k)\).

  – the algorithm is identical to an existing N-FINDR variant, SC-N-FINDR [Wu-Chu-Chang2008].

  – a similar algorithm was also derived in text mining [Arora-Ge-Halpern-Mimno et al. 2013]. In the same ref., the algorithm is proved to be robust to noise.
Simplex Volume Maximization: Optimization

- VolMax formulation:

\[
\max_B \text{vol}(B) \\
\text{s.t. } b_i \in \text{conv}\{\mathbf{x}[1], \ldots, \mathbf{x}[L]\}, \forall i.
\]

- Consider a recursive heuristic.
  - Let \( f_i = [b_i^T, 1]^T, F = [f_1, \ldots, f_N] \). We have

\[
\text{vol}(B) \propto |\det(F)| = \|f_1\|_2 \cdot \|P_{F_{1:2}}^\perp f_2\|_2 \cdot \|P_{F_{1:3}}^\perp f_3\|_2 \cdots \|P_{F_{1:(N-1)}}^\perp f_N\|_2
\]

  - SVMAX [Chan-Ma-Ambikapathi-Chi2011]: for \( k = 1, \ldots, N \), do

\[
\hat{b}_k = \arg\max_{b_k} \|P_{\hat{F}_{1:(k-1)}}^\perp f_k\|_2^2 \text{ s.t. } b_k \in \text{conv}\{\mathbf{x}[1], \ldots, \mathbf{x}[L]\},
\]

where \( \hat{F}_{1:(k-1)} = [\hat{f}_1, \ldots, \hat{f}_{k-1}] \), \( \hat{f}_i = [\hat{b}_i^T, 1]^T \).

  - the resulting algorithm is almost the same as SPA.
Simplex Volume Maximization Without Pure Pixels

- Assume the case of no pure pixels.

\[ b_1, \ldots, b_N \]

[Chan-Ma-Ambikapathi-Chi2011]
Simplex Volume Maximization Without Pure Pixels

- Assume the case of no pure pixels.
Simplex Volume Maximization Without Pure Pixels

- Assume the case of no pure pixels.
Simplex Volume Maximization Without Pure Pixels

- Assume the case of no pure pixels.
Simplex Volume Maximization Without Pure Pixels

- Assume the case of no pure pixels.

- Assume the noiseless case. VolMax can perfectly identify the true endmember signatures $b_1, \ldots b_N$ if and only if the pure pixel assumption holds [Chan-Ma-Ambikapathi-Chi2011].
**Simplex Volume Minimization: Intuition**

**Craig’s belief [Craig1994]:** the true endmembers may be located by finding a data enclosing simplex whose volume is the smallest.
Craig’s belief [Craig1994]: the true endmembers may be located by finding a data enclosing simplex whose volume is the smallest.
**Simplex Volume Minimization: Intuition**

Craig’s belief [Craig1994]: the true endmembers may be located by finding a data enclosing simplex whose volume is the smallest.
Simplex Volume Minimization: Intuition

Craig’s belief [Craig1994]: the true endmembers may be located by finding a data enclosing simplex whose volume is the smallest.

- It seems that volume minimization (VolMin) can still identify the true endmembers even if the pure pixel assumption does not exactly hold.
Simplex Volume Minimization: Formulation

- VolMin formulation:

\[
\min_B \text{vol}(B) \\
\text{s.t. } x[n] \in \text{conv}\{b_1, \ldots, b_N\}, \\
\quad n = 1, \ldots, L.
\]

- Challenge:
  - both the objective function and constraints of VolMin are nonconvex.
  - NP-hard in general [Packer2008]
Simplex Volume Minimization: Optimization

- By transformation of a simplex to a polyhedron [Li-Bioucas2008], [Chan-Chi-Huang-Ma2009], we can recast VolMin as

\[
\begin{align*}
\min \limits_B & \quad \text{vol}(B) \\
\text{s.t.} & \quad x[n] \in \text{conv}\{b_1, \ldots, b_N\}, \forall n
\end{align*}
\]

\[\iff \max \limits_{H,g} \quad |\text{det}(H)| \]

\[\text{s.t. } Hx[n] - g \geq 0, \quad (Hx[n] - g)^T 1 \leq 1, \forall n\]

where \(H = [b_1 - b_N, \ldots, b_{N-1} - b_N]^{-1}, g = Hb_N.\)
Simplex Volume Minimization: Optimization

- By transformation of a simplex to a polyhedron [Li-Bioucas2008], [Chan-Chi-Huang-Ma2009], we can recast VolMin as

\[
\min_B \text{vol}(B) \quad \text{s.t.} \quad x[n] \in \text{conv}\{b_1, \ldots, b_N\}, \forall n
\]

\[
\max_{H,g} |\det(H)| \quad \iff \quad \text{s.t.} \quad Hx[n] - g \geq 0, \\
(Hx[n] - g)^T 1 \leq 1, \forall n
\]

- The problem at the RHS has convex constraints, although the objective function is still nonconvex.

- **MVSA** [Li-Bioucas2008], **SISAL** [Bioucas2009]: apply iterative linear approximation.

- **MVES** [Chan-Chi-Huang-Ma2009]: apply row-by-row alternating optimization (hint: by cofactor expansion, \(\det(H)\) is linear w.r.t. each row of \(H\)).
Further Discussion with Simplex Volume Minimization

- In the noiseless case, VolMin can be shown to be equivalent to a stochastic maximum-likelihood estimator [Nascimento-Bioucas2012].

- Numerical evidence has suggested that VolMin may work without pure pixels.
Numerical Demo.: Three Endmembers, Pure Pixel Case

- data points
- true endmembers
- SVMAX
- SC–N–FINDR
- VolMin
- CSR
Numerical Demo.: Three Endmembers, No-Pure Pixel Case
Simulation Results: Mean Squared Error (MSE) Comparison

A Monte Carlo result. $N = 8$. "Purity $\rho$" describes the pixel purity: $\rho = 1$ corresponds to the pure pixel case, and $\rho = 1/\sqrt{N}$ the most heavily mixed case.
**Unique Identifiability of VolMin**

- **Question:** can VolMin uniquely identify the true endmembers—provably?

- An answer is recently provided in [Lin-Ma-Li-Chi-Ambikapathi2015].
No Pure-Pixel Case

**Assumption:** For every $i, j \in \{1, \ldots, N\}$, $i \neq j$, there exists an index $n(i, j)$ such that

$$s[n(i, j)] = \alpha_{ij} e_i + (1 - \alpha_{ij}) e_j,$$

for some $\frac{1}{2} < \alpha_{ij} \leq 1$. 

![Diagram](image-url)
No Pure-Pixel Case

Assumption: For every $i, j \in \{1, \ldots, N\}$, $i \neq j$, there exists an index $n(i, j)$ such that

$$s[n(i, j)] = \alpha_{ij}e_i + (1 - \alpha_{ij})e_j,$$

for some $\frac{1}{2} < \alpha_{ij} \leq 1$.

**Theorem [Lin-Ma-Li-Chi-Ambikapathi2015]:** In the noiseless case and under the above assumption, the VolMin solution is uniquely and exactly given by $a_1, \ldots, a_N$ subject to ordering permutations if

1. $N = 3$ and $\alpha_{ij} > \frac{2}{3}$ for all $i, j$, or if

2. $N \geq 4$. 

\[ [523x533] No Pure-Pixel Case [523x533] \]

\[ Assumption: \] For every $i, j \in \{1, \ldots, N\}$, $i \neq j$, there exists an index $n(i, j)$ such that

\[ s[n(i, j)] = \alpha_{ij} e_i + (1 - \alpha_{ij}) e_j, \]

for some $\frac{1}{2} < \alpha_{ij} \leq 1$.

**Theorem [Lin-Ma-Li-Chi-Ambikapathi2015]:** In the noiseless case and under the above assumption, the VolMin solution is uniquely and exactly given by $a_1, \ldots, a_N$ subject to ordering permutations if

1. $N = 3$ and $\alpha_{ij} > \frac{2}{3}$ for all $i, j$, or if

2. $N \geq 4$. 

\[ [523x533] No Pure-Pixel Case [523x533] \]
VolMin with Outliers

- VolMin is sensitive to outliers.

![Diagram showing VolMin with Outliers](image-url)
VolMin with Outliers

- VolMin is sensitive to outliers.

![Diagram showing VolMin with Outliers]
VolMin with Outliers

- VolMin is sensitive to outliers.

![Diagram showing VolMin with Outliers](image_url)
VolMin with Outliers

- VolMin is sensitive to outliers.

- The outlier problem may be handled by employing an outliers-removing pre-processing stage [Chan-Ambikapathi-Ma-Chi2013], or by using robust VolMin formulations (e.g., SISAL [Bioucas2009]).
Sparse Regression
Motivation

- In geoscience and remote sensing, a tremendous amount of effort has been spent on recording materials’ spectral samples.
  - the U.S. Geological Survey (USGS) library contains > 1,300 spectral samples covering materials such as minerals, rocks, liquids, artificial materials, vegetations, microorganisms . . .

- **Idea:** instead of assuming unknown $A$, why not use the library to do HU?
A Row-Sparse Model

- Redefine $A = [a_1, \ldots, a_K]$ as a dictionary of $K$ spectral signature samples.

- **Model:** Let $Y = [y[1], \ldots, y[L]]$.

$$Y = AC + V,$$

where $C \in \mathbb{R}^{K \times L}$ is row-sparse.

- **Problem:** find a row-sparse $C$. 
Collaborative Sparse Regression

- Let $\|C\|_{\text{row-0}}$ be the row-0 quasinorm, or the number of nonzero rows of $C$.

- Assume the noiseless case (for tutorial purposes).

- Collaborative sparse regression (CSR) [Iordache-Bioucas-Plaza2013]:

\[
\min_C \| C \|_{\text{row-0}} \\
\text{s.t. } Y = AC, \ C \geq 0.
\]

- processed by convex relaxations, e.g., replacing $\|C\|_{\text{row-0}}$ by

\[
\|C\|_{2,1} = \sum_{k=1}^{K} \|c^k\|_2.
\]

- in practice, a regularized formulation is employed

\[
\min_{C \geq 0} \| Y - AC \|_F^2 + \lambda \| C \|_{2,1}
\]

for some $\lambda > 0$, and a fast custom-derived solver (e.g., ADMM) is required.
Numerical Demo.: Three Endmembers, Very Heavy Mixtures

- data points
- true endmembers
- SVMAX
- SC–N–FINDR
- VolMin
- CSR
Limitations of Sparse Regression

- **Challenge 1:** The number of dictionary samples $K$ is usually large (say, of order of hundreds), and some spectral samples are very similar.
  - high mutual coherence with $A \implies$ performance degradation with CSR.
  - a larger $K \implies$ a higher computational requirement for CSR
    * For example, $L = 10^4$ pixels and $K = 500$ dictionary members result in $5 \times 10^6$ optimization variables for CSR.
  - remedy: prune the dictionary beforehand, via some cheaper algorithms [Iordache-Bioucas-Plaza2014].

- **Challenge 2:** There may be mismatches between the spectral samples and the true endmembers’ spectral signatures.
Self-Dictionary Sparse Regression

• Suppose that we use the observed data $Y$ itself as the dictionary (what?):

$$\min_C \|C\|_{\text{row}-0} \quad \text{s.t. } Y = YC, \ C \geq 0, \ 1^T C = 1^T. \quad (\dagger)$$

– the problem turns out to be equivalent to pure pixel search—without knowing $N$ [Esser-Moller-Osher-Sapiro-Xin2012], [Elhamifar-Sapiro-Vidal2012].


• Apart from convex relaxations, one can also employ greedy pursuit to handle the problem.

– consider applying a popular greedy method, simultaneous orthogonal matching pursuit (SOMP) [Chen-Huo2006], [Tropp-Gilbert-Strauss2006].

– a special case of the algorithm is identical to SPA! [Fu-Ma-Chan-Bioucas2015]
Non-Negative Matrix Factorization
NMF

- **Problem:** Given $Y \in \mathbb{R}^{M \times L}$ and $N < \min\{M, L\}$, solve

  $$\min_{A \geq 0, S \geq 0} \|Y - AS\|_F^2,$$

  with $A \in \mathbb{R}^{M \times N}$, $S \in \mathbb{R}^{N \times L}$.

  - a dimensionality reduction problem.
  - popularized by [Lee-Seung1999].
  - widely used in machine learning and other fields, has numerous applications.
  - in HU, NMF is used as a tool to extract the true endmember matrix $A$ and true abundance matrix $S = [s[1], \ldots, s[N]]$. 
NMF Problem Natures

**Problem:** Given \( Y \in \mathbb{R}^{M \times L} \) and \( N < \min\{M, L\} \), solve

\[
\min_{A \geq 0, S \geq 0} \|Y - AS\|_F^2,
\]

with \( A \in \mathbb{R}^{M \times N}, S \in \mathbb{R}^{N \times L} \).

- NP-hard in general [Vavasis2009].
- The NMF solution may not be unique (subject to scaling and permutations).
  - not desirable in HU or BSS, where we always want to uniquely identify \((A, S)\).
  - the NMF solution is unique if some specific conditions with \((A, S)\) are satisfied; e.g., the separability assumption [Donoho-Stodden2003], which is identical to the pure pixel assumption.
NMF in HU: Formulations

• A generic NMF formulation for HU:

\[
\min_{A \geq 0, S \in S^L} \|Y - AS\|_F^2 + \lambda \cdot g(A) + \mu \cdot h(S),
\]

where \( S^L = \{ S \mid S \geq 0, 1^T S = 1^T \} \), \( g, h \) are regularizers, \( \lambda, \mu > 0 \).

• Idea: make the problem more well posed by appropriate regularization.

• very pragmatic developments; less to say in theory.
NMF in HU: Formulations

- A generic NMF formulation for HU:

$$\min_{A \geq 0, S \in S^L} \|Y - AS\|_F^2 + \lambda \cdot g(A) + \mu \cdot h(S),$$

where $S^L = \{S \mid S \geq 0, 1^T S = 1^T\}$, $g, h$ are regularizers, $\lambda, \mu > 0$.

- Examples:
  - MVC-NMF [Miao-Qi2007]: $g(A) = \text{simplex volume}$, $h(S) = 0$.
  - ICE [Berman et al. 2004]: $g(A) = \sum_{i \neq j} \|a_i - a_j\|_2^2$ (approx. vol.), $h(S) = 0$.
  - dictionary learning [Charles-Olshausen-Rozell2011]: $g(A) = 0$, $h(S) = \|S\|_{1,1}$.
  - SPICE [Zare-Gader2007]: same $g$ as in ICE, $h(S) = \sum_i \gamma_i \|s_i\|_1$.

  and many, many more.
**NMF in HU: Algorithms**

\[
\min_{A \geq 0, S \in S^L} \|Y - AS\|_F^2 + \lambda \cdot g(A) + \mu \cdot h(S),
\]

- Most algorithms follow a two-block alternating optimization strategy: generate a sequence of iterates \(\{(A^{(k)}, S^{(k)})\}_k\) via

  \[
  A^{(k)} = \arg \min_{A \geq 0} \|Y - AS^{(k-1)}\|_F^2 + \lambda \cdot g(A),
  \]

  \[
  S^{(k)} = \arg \min_{S \in S^L} \|Y - A^{(k)}S\|_F^2 + \mu \cdot h(S).
  \]

- for practical reasons, the updates above are often inexact.
- convergence to a stationary point (for inexact updates) has still to be thoroughly analyzed.
- initialization: it’s common to use VCA (pure pixel search) to initialize.
Related Topics and Applications
Pure Pixels and Non-Negative Matrix Factorization (NMF)

- NMF has received much attention in machine learning and text mining.
- Recently, there has been much interest in an NMF problem subclass called near-separable NMF [Arora-Ge-Kannan-Moitra2012]—which assumes pure pixels.

\[ Y = \ldots \approx \ldots = AS \]

weights to reconstruct each article

topics found in different articles
• **Dynamic fluorescent imaging (DFI)** exploits highly specific and bio-compatible fluorescent contrast agents to interrogate small animals for drug development and disease research.

• DFI images are linear mixtures of the anatomical maps of different organs.
• Separated anatomical maps using a convex geometry-based method [Chan-Ma-Chi-Wang2008]. Biomedical images have many dark points, and that makes pure pixel a good assumption.
Video Summarization in Computer Vision

The video summarization result shown in [Elhamifar-Sapiro-Vidal2012]. Courtesy to the above reference.

**Problem:** find a (small) subset of video frames that can best represent the whole set of video frames.

**In the video summarization context, pure pixels in HU become those most representative video frames.**
Pure Pixels, PARAFAC, and Classical BSS

• Consider the PARAFAC problem.
  
  – **Model** (symmetric PARAFAC):
    
    \[ R_m = AD_m A^T, \quad D_m = \text{Diag}(d_m), \quad m = 1, \ldots, M. \]
    
  – **Problem:** retrieve \( A \) from \( R_1, \ldots, R_M \).

• **Suppose** that \( d_m \geq 0, \quad d_m^T 1 = 1 \), and the pure pixel assumption holds for \( \{d_m\} \).

• **PARAFAC by SPA:** Since

\[ y_m \triangleq \text{vec}(R_m) = (A \odot A)d_m, \]

where \( A \odot A = [a_1 \otimes a_1, \ldots, a_N \otimes a_N] \), we may

1. use SPA to obtain \( A \odot A \) from \( y_1, \ldots, y_M \);
2. extract \( A \) from \( A \odot A \) (hint: \( \text{vec}^{-1}(a_i \otimes a_i) = a_i a_i^T \)).
Pure Pixels, PARAFAC, and Classical BSS

• **Application:** the classical speech BSS [Fu-Ma-Huang-Sidiropoulos2015].
  
  – each $R_m$ is a local covariance for a short-time interval.
  – The locally sparse natures of speech make pure pixel a reasonable assumption.
Performance comparison of various BSS algorithms. ‘ProSPA’ is a modified version of SPA, custom-designed for the blind speech separation application.
Conclusion and Discussion

• We have reviewed HU using a perspective of SP theory and methods.

• HU started out with strong intuitions from remote sensing researchers, most notably, Craig, Boardman and Winter, rather than rigorous SP or math.

• Now, HU has emerged as a research topic that has tight connections to other fields such as SP, machine learning, and optimization.

• Ideas originated from remote sensing, particularly, pure pixels and CG, are elegant and powerful, and they have been rediscovered or adopted in other topics like BSS and NMF.

• On the other hand, HU is rapidly absorbing ideas from other fields, e.g., optimization and sparsity. Its fundamental theory and methods have been greatly improved recently, and it’s no longer easy to draw a line between fields.
Conclusion and Discussion

• We have reviewed HU using a perspective of SP theory and methods.

• HU started out with strong intuitions from remote sensing researchers, most notably, Craig, Boardman and Winter, rather than rigorous SP or math.

• Now, HU has emerged as a research topic that has tight connections to other fields such as SP, machine learning, and optimization.

• Ideas originated from remote sensing, particularly, pure pixels and CG, are elegant and powerful, and they have been rediscovered or adopted in other topics like BSS and NMF.

• On the other hand, HU is rapidly absorbing ideas from other fields, e.g., optimization and sparsity. Its fundamental theory and methods have been greatly improved recently, and it’s no longer easy to draw a line between fields.

Thank you!
References


