Hyperspectral Unmixing
Geometrical, Statistical, and Sparse Regression-Based Approaches

José M. Bioucas-Dias
Instituto de Telecomunicações,
Instituto Superior Técnico
Universidade de Lisboa, Portugal

Société Française de Photogrammétrie et Télédétection
Grenoble, France, May, 2016

FCT, project UID/EEA/50008/2013
Hyperspectral imaging (and mixing)
Hyperspectral unmixing

AVIRIS of Cuprite, Nevada, USA

R – ch. 183 (2.10 μm)
G – ch. 193 (2.20 μm)
B – ch. 207 (2.34 μm)

VCA [Nascimento, B-D, 2005]
Outline

- Mixing models
  - Linear
  - Nonlinear

- Signal subspace identification

- Unmixing
  - Geometrical-based
  - Statistical-based
  - Sparse regression-based


Linear mixing model (LMM)

Incident radiation interacts only with one component (checkerboard type scenes)

\[
\mathbf{r} = \sum_{i=1}^{p} \alpha_i \mathbf{m}_i \\
\mathbf{m}_i = \begin{bmatrix}
\rho_{i1} \\
\rho_{i2} \\
\vdots \\
\rho_{iL}
\end{bmatrix}
\]

\[
\mathbf{r} = \mathbf{M} \mathbf{\alpha}
\]

\[
\mathbf{M} \equiv [\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3] \\
\mathbf{\alpha} = \begin{bmatrix}
\alpha_1 \\
\alpha_2 \\
\alpha_2
\end{bmatrix}
\]

Hyperspectral linear unmixing

\[\text{Estimate } \mathbf{M}, \mathbf{\alpha}\]
Nonlinear mixing model

Intimate mixture (particulate media) Two-layers: canopies+ground

Radiative transfer theory

\[ r = f(\alpha, \theta) \]

- material fractions
- media parameters

\[ r = \sum_{i=1}^{p} \alpha_i m_i + \sum_{i,j=1 \atop i \neq j}^{p} \alpha_{i,j} m_i \odot m_j \]

- single scattering
- double scattering
Schematic view of the unmixing process

1. Radiance data cube
2. Atmospheric correction
3. Reflectance data cube
4. Dimensionality reduction (optional)
5. Hyperspectral library
6. Reduced data cube

Unmixing
- Sparse coding
- Sparse regression
- Find endmembers (+) inversion

Endmember signatures

Abundance maps
Spectral linear unmixing (SLU)

Given $N$ spectral vectors of dimension $L$:

$$Y = \left\{ y_i \in \mathbb{R}^L, \ i = 1, \ldots, N \right\}$$

Subject to the LMM: $y = M\alpha + n$, \(\alpha \geq 0, \ 1^T\alpha = 1\)

Determine:
- The mixing matrix $M$ (endmember spectra)
- The fractional abundance vectors $\alpha$

$\Rightarrow$ SLU is a blind source separation problem (BSS)
Subspace identification

\[ y = \mathbf{M} \alpha + \mathbf{n} \quad \text{dim}(\mathbf{M}) = [L \times p] \quad L \gg p \]

Problem: Identify \( \text{span}\{\mathbf{M}\} \)
the subspace generated by the columns of \( \mathbf{M} \)

Reasoning underlying DR

1. Lightens the computational complexity
2. Attenuates the noise power by a factor of \( p/L \)
Subspace identification algorithms

**Exact ML solution** [Scharf, 91] (known $p$, i.i.d. Gaussian noise)

**PCA** - Principal component analysis (unknown $p$, i.i.d. noise)

**NAPC** - Noise adjusted principal components [Lee et al., 90]

**MNF** - Maximum noise fraction [Green et al., 88]

**HFC** - Harsanyi-Farrand-Chang [Harsanyi et al., 93]

**NWHFC** - [Chang, Du, 94]

**HySime** - Hyperspectral signal identification by minimum error [B-D, Nascimento, 08]

**GENE** - geometry-based estimation of number of endmembers [ArulMurugan, 13]

**RMT** - [Kritchman, Nadler, 2009], [Cawse et al., 11], [Halimi et al., 16]

---

Example (HySime)

$$L = 224 \quad p = 3$$

$$n \sim \mathcal{N}(0, \sigma^2 I)$$

$$\sigma = 0.1$$
Geometrical view of SLU

\[ M = [m_1, \ldots, m_p] \]

\[ y = M\alpha \]

\[ \sum_{j=1}^{p} \alpha_j = 1, \quad \alpha_j \geq 0 \]

probability simplex \( (S_I) \)

\[ S_M = \{ x \in \mathbb{R}^p : x = M\alpha, \alpha \in S_I \} \rightarrow (p-1) \text{- simplex} \]

Inferring \( M \) ⇔ inferring the vertices of the simplex \( S_M \)
Classes of SLU problems

Pure pixels

\[ \hat{m}_2 = m_2 \]

\[ \hat{m}_3 = m_3 \]

\[ \hat{m}_1 = m_1 \]

Well posed

Pixels in the facets

\[ \hat{m}_2 = m_2 \]

\[ \hat{m}_3 = m_3 \]

\[ \hat{m}_1 = m_1 \]

Well posed

Highly mixed

\[ \hat{m}_2 \neq m_2 \]

\[ \hat{m}_3 \neq m_3 \]

\[ \hat{m}_1 \neq m_1 \]

Ill-posed

Algorithms

<table>
<thead>
<tr>
<th>Vertex pursuit</th>
<th>Facet estimation</th>
<th>Statistical inference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Sparse regression</td>
</tr>
</tbody>
</table>
Unmixing frameworks

- Geometrical (blind)
  Exploits parallelisms between the linear mixing model and properties of convex sets
  **Application scenarios:** pure pixels, pixels in facets

- Statistical (blind, semi-blind)
  Approaches linear unmixing as a statistical inference problem
  **Application scenarios:** all

- Sparse regression (semi-blind)
  Approaches linear unmixing as a sparse regression problem
  **Application scenarios:** all
Simplex vertex pursuit

Hard assumption
The data set contains at least one pure pixel of each material

- Search endmembers in the data set
- Computationally light

PPI - [Boardman, 93]; N-FINDR - [Winter, 99]; IEA - [Neville et al., 99];
AMEE – [Plaza et al, 02]; SMACC – [Gruninger et al., 04]
VCA - [Nascimento, B-D, 03, 05]; SGA - [Chang et al., 06]
AVMAX, SVMAX - [Chan, et al., 11]; RNMF- [Gillis & Vavasis, 12,14];
SD-SOMP, SD-ReOMP - [Fu et al.13, 15];
Simplex vertex pursuit

N-FINDR

Iteratively increase $|\det(m_1, \ldots, m_p)|$

$m_1, \ldots, m_p \in \{y_1, \ldots, y_N\}$

AVMAX

$$\max_{m_1, \ldots, m_p} |\det(m_1, \ldots, m_p)|$$

$s.t.: m_1, \ldots, m_p \in \text{conv}\{y_1, \ldots, y_N\}$

VCA

For $k = 1 : p$

$m := \arg \max_{y_i} \| (P_M^\perp \xi)^T y_i \|$ 

$M := [M, m]$ 

SVMAX

For $k = 1 : p$

$m := \arg \max_{y_i} \| (P_M^\perp y_i) \|$ 

$M := [M, m]$
Unmixing example

HYDICE sensor $\mathbf{M} \in \mathbb{R}^{210 \times 6}$  $N = 500 \times 307$  resolution = 0.75m
Simplex facet estimation

- Minimum-volume constrained nonnegative matrix factorization (MVC-NMF) (inspired by NMF [Lee, Seung, 01])

\[
(\hat{M}, \hat{S}) = \arg\min_{M \in \mathbb{R}^{L \times P}, S \in \mathbb{R}^{P \times N}} \frac{1}{2} \| Y - MS \|_F^2 + \tau V(M) + \lambda \phi(S)
\]

s.t.: \( M \succeq 0, \ S \succeq 0, \ 1^T S = 1^T_N \),

volume regularizer

DPFT - [Craig, 90]; CCA - [Perczel et al., 89] (seminal works on MVC)
ICE - [Breman et al., 2004] \( V(M) \equiv \text{quadratic}, \ \phi = 0 \);
MVC-NMF - [Miao, Qi, 07] \( V(M) = |\det(MM^T)|, \ \phi = 0 \);
SPICE - [Zare, Gader, 2007] \( V(M) \equiv \text{quadratic}, \ \phi \equiv \text{weighted } \ell_1 \)
L1/2 - NMF - [Qian, Jia, Zhou, Robles-Kelly, 11] \( V = 0, \ \phi(S) = \sum_{ij} |a_{ij}|^{1/2} \)
CoNMF - [Li, B-D, Plaza, 12] \( V(M) \equiv \text{quadratic}, \ \phi(S) = \|S\|_{2,1} \)
Minimum volume simplex algorithms

**MVSA** – Minimum volume simplex analysis [Li, B-D, 08]

**Optimization variable** \( Q \equiv M^{-1} \implies QY = S \)

\[
\hat{Q} = \arg\max_Q \log(|\det(Q)|)
\]

s.t.: \( 1_p^T Q = q_m \), \( QY \geq 0 \)

**ANC**

**ASC**

**MVSA** solves a sequence of quadratic programs

**MVES** [Chan, Chi, Huang, Ma, 2009]

- Solves a sequence of linear programs by exploiting the cofactor expansion of \( \det(Q) \)
- The existence of pure pixels is a sufficient condition for exact identification of the true endmembers
Robust minimum volume simplex algorithms: outliers

SISAL – Simplex identification via split augmented Lagrangian [B-D,09]

\[
\hat{Q} = \arg \max_Q \log(|\det(Q)|) - \lambda \phi(QY)
\]

s.t.: \[1^T_p Q = q_m\]

\(\lambda_2 > \lambda_1\)

- SISAL solves a sequence of convex subproblems using ADMM
- \((\lambda = \infty) \Rightarrow \text{MVES} \equiv (\text{MVES, SISAL})\)
Example: data set contains pure pixels

\[ N = 5000 \quad p = 5 \quad \max \alpha_i = 1, \text{ for } i = 1, \ldots, p \]

Time:
- VCA \(\rightarrow\) 0.5 sec
- SISAL \(\rightarrow\) 2 sec

IWMSDF, SYSU - 2014
Example: Data set does not contain pure pixels

\[ N = 5000 \quad \rho = 5 \quad \max_{\alpha_i} = 0.8, \text{ for } i = 1, \ldots, \rho \]
No pure pixels and outliers

\[ N = 1000 \quad p = 3 \quad \max_{\alpha_i} = 0.8, \text{ for } i = 1, \ldots, p \]

no. outliers = 3

ERROR (mse):
- SISAL = 0.03
- VCA = 0.88
- NFINDR = 0.88
- MVC - NMF = 0.90

TIMES (sec):
- SISAL = 0.61
- VCA = 0.20
- NFINDR = 0.25
- MVC - NMF = 25
Real data: ArtImageDataA (converted into absorbance)

Tiles 1,2 (‘Prussian blue’ + oil)
Tiles 15,16 (‘Ultramarine blue’ + oil)

Estimate endmember signatures by SISAL
Geometrical Approaches: Limitations

- determined by a small number of pixels; some may be outliers
- MVS – Computationally heavy
- PPI, N-FINDR, VCA, SGA, AMEE, SVMAX, AVMAX depend on the existence of pure pixels in the data
- Do not work in highly mixed scenarios
Statistical approaches

- **observation model** \( Y = MS + N \)

- **prior (Bayesian framework)** \( p_S(S)p_M(M) \)

- **posterior density**

\[
p_{M,S|Y}(M, S|Y) = p_{Y|M,S}(Y|M, S)p_M(M)p_S(S)/p_Y(Y)
\]

- **inference**

\[
(M, S)_{\text{MAP}} \equiv \arg \max_{M, S} p_{M,S|Y}(M, S|Y)
\]
\[
= \arg \min_{M, S} - \log p_{Y|M,S}(Y|M, S) - \log p_M(M) - \log p_S(S)
\]

\[
\hat{M}_{\text{MMSE}} \equiv \mathbb{E}[M|Y] = \int M p_{M|Y}(M|Y) dM
\]

\[
\hat{S}_{\text{MMSE}} \equiv \mathbb{E}[S|Y] = \int S p_{S|Y}(S|Y) dS.
\]
Spectral linear unmixing and ICA/IFA

- Formally, SLU is a linear source separation problem
- Independent Component Analysis (ICA) come to mind

**ICA**
- *Fastica*, [Hyvarinen & Oja, 2000]
- *Jade*, [Cardoso, 1997]
- *Bell and Sejnowski*, [Bell and Sejnowski, 1995]

**IFA**
- *IFA*, [Moulines *et al.*, 1997], [Attias, 1999]
Statistical approaches: ICA

\[ y = M\alpha + n \]

Assumptions

1. Fractional abundances (sources) are independent

\[ p_\alpha(\alpha) = p_1(\alpha_1)p_2(\alpha_2) \ldots p_k(\alpha_p) \]

2. Non-Gaussian sources

Endmembers compete for the same area

\[ \sum_{j=1}^{k} \alpha_j = 1 \]

Sources are Dependent

ICA does not apply [Nascimento, B-D, 2005]
Representative Bayesian approaches

[Parra et al., 00]  \( (S \leftarrow \text{u.d. on the simplex}, M \leftarrow \text{AR model}, \text{MAP}) \)

[Moussaoui et al., 06,a,b], [Dobigeon et al., 09,a,b], [Dobigeon et al., 09,b], [Arngreen, 11]

- data term associated with the LMM
- \( S \) – Uniform on the simplex
- conjugate prior distributions for some unknown parameters
- Infer MMSE estimates by Markov chain Monte Carlo algorithms

DECA - [Nascimento, B-D 09, 14]

- data term associated with a noiseless LMM
- \( S \) – Dirichlet mixture model
- MDL based inference of the number of Dirichlet modes
- MAP inference (GEM - algorithm)
DECA – Dependent component analysis

\[ p_{Y|M}(Y|M, \theta) = \left( \prod_{i=1}^{N} p_{\alpha}(M^{-1}y_i, \theta) \right)^{|\text{det}(M^{-1})|^N} \]

parameters of the Dirichlet mixture

minimum volume term
DECA – Results on Cuprite

- Data term associated with the LMM
- $S$ – Dirichlet mixture model
- Latent label process enforcing adjacent pixels to have the same label
- Spatial prior: tree-structured sticky hierarchical Dirichlet process (SHDP)
- MMSE inference by MCMC
- Model order inference (number of endmembers)

[Mittelman, Dobigeon, Hero, 12]
Sparse regression-based SLU

- Spectral vectors can be expressed as **linear combinations of a few pure spectral signatures** obtained from a (potentially very large) spectral library
  
  \[ y = \sum_{i \in S} a_i x_i = Ax \]

- **Unmixing**: given \( y \) and \( A \), find the sparsest solution of

  \[ y = Ax \]

- **Advantage**: sidesteps endmember estimation

- **Disadvantage**: **Combinatorial problem !!!**
Convex approximations to P0

CBPDN – Constrained basis pursuit denoising

$$\min_x \|x\|_1 \text{ subject to } \|y - Ax\|_2 \leq \delta, \ x \geq 0,$$

Equivalent problem

$$\min_x (1/2)\|y - Ax\|^2 + \lambda\|x\|_1, \ x \geq 0$$

Striking result: In given circumstances, related with the coherence among the columns of matrix $A$, BP(DN) yields the sparsest solution ([Donoho 06], [Candès et al. 06]).

Efficient solvers for CBPDN: SUNSAL, CSUNSAL
[B-D, Figueiredo, 10]
Real data – AVIRIS Cuprite

[Iordache, B-D, Plaza, 11, 12]
Real data – AVIRIS Cuprite

Iordache, B-D, Plaza, 11, 12
Sparse reconstruction of hyperspectral data: Summary

Bad news: Hyperspectral libraries have poor RI constants

Good news: Hyperspectral mixtures are highly sparse, very often $p \cdot 5$

Surprising fact: Convex programs (BP, BPDN, LASSO, …) yield much better empirical performance than non-convex state-of-the-art competitors

Directions to improve hyperspectral sparse reconstruction

- Structured sparsity + subspace structure (pixels in a given data set share the same support)
- Spatial contextual information (pixels belong to an image)
Constrained total variation sparse regression (CTVSR)

$$\min_{X} \frac{1}{2} \|AX - Y\|_F^2 + \lambda_1 \|X\|_1 + \lambda_2 \phi_{TV}(X)$$

subject to: \(X \geq 0\)

$$\phi_{TV}(X) := \sum_{i=1}^{N} \|Lx^i\|_1 = \sum_{i=1}^{n} \sum_{k=1}^{N} \sqrt{([D_h x^i_k]^2 + ([D_v x^i_k]^2)}$$

Related work \[Zhao, Wang, Huang, Ng, Plemmons, 12\]

Other Regularizers:

- vector total variation (VTV)! promotes piecewise smooth vectors \[Bresson, Chan, 02\], \[Goldluecke et al., 12\], \[Yuan, Zhang, Shen, 12\]

- convex generalizations of Total Variation based on the Structure Tensor \[Lefkimmiatis et al., 13\]

- sparse representation (2D, 3D) in the wavelet domain
Ilustrative examples with simulated data: SUnSAL-TV

$\mathbf{A} \in \mathbb{R}^{224 \times 240}$ (from USGS library)  \hspace{1cm} (m = 224, N = 75 \times 75, k = 5)

Original data cube

SUnSAL estimate

Original abundance of EM5

SUnSAL-TV estimate
Constrained collaborative sparse regression (CCSR)

\[
\min_{\mathbf{X}} \frac{1}{2} \| \mathbf{AX} - \mathbf{Y} \|_F^2 + \lambda \| \mathbf{X} \|_{2,1}
\]

subject to: \( \mathbf{X} \geq 0, \quad \mathbf{1}_n^T \mathbf{X} = \mathbf{1}_N^T \)

[1ordache, B-D, Plaza, 11, 12]

\[
\| \mathbf{X} \|_{2,1} := \sum_{i=1}^n \| \mathbf{x}^i \|_2
\]

[Turlach, Venables, Wright, 2004]

Theoretical guarantees (superiority of multichannel): the probability of recovery failure decays exponentially in the number of channels.  

[ Eldar, Rauhut, 11 ]
Illustrative examples with simulated data: CSUnSAL

\[ \mathbf{A} \in \mathbb{R}^{224 \times 350} \text{ (from USGS library)} \quad \mathbf{x} \in \mathbb{R}^{350 \times 100} \text{ (sparsity k = 5)} \]

\[ \| \hat{\mathbf{x}} - \mathbf{x} \|_F = 2.4 \quad \| \hat{\mathbf{x}} - \mathbf{x} \|_F = 0.9 \]

SNR = 35dB
time = 10 sec
MUSIC – Colaborative SR algorithm

### MUSIC-CSR algorithm [Iordahe, B-D, Plaza, 2013]

1) Estimate the signal subspace $\text{span}\{A_S\}$ using, e.g. the HySime algorithm.

2) Compute $\varepsilon_i = \frac{\|P_y a_i\|}{\|a_i\|} , \text{ for } i = 1, \ldots, m$ and define the index set $S = [i : \varepsilon_i \leq \delta, i = 1, \ldots, m]$

3) Solve the collaborative sparse regression optimization

$$\min_x \left( \frac{1}{2} \|Y - A_S X\|^2 + \lambda \|X\|_{2,1}, \ X \geq 0 \right)$$

[B-D, Figueiredo, 2012]

Related work: CS-MUSIC [Kim, Lee, Ye, 2012]

(N < k and iid noise)
MUSIC – CSR results

A – USGS ($\geq 3^\circ$), Gaussian shaped noise, SNR = 25 dB, $k = 5$, $m = 300$

$|S| = 11 \quad \hat{X} \quad (\text{MUSIC-CSR})$

\begin{align*}
\text{MUSIC-CSR} & \quad \begin{cases}
\text{SNR} = 11.7 \text{ dB} \\
\text{computation time '10 sec}
\end{cases} \\
\text{CSR} & \quad \begin{cases}
\text{SNR} = 0 \text{ dB} \\
\text{computation time '600 sec}
\end{cases}
\end{align*}
Brief Concluding remarks

- HU is a hard inverse problem (noise, bad-conditioned direct operators, nonlinear mixing phenomena)

- HU calls for sophisticated math tools and framework (statistical inference, optimization, machine learning)

- The research efforts devoted to non-linear mixing models are increasing

Linear mixing

- Apply geometrical approaches when there are data vectors near or over the simplex facets

- Apply statistical methods in highly mixed data sets

- Apply sparse regression methods, if there exits a spectral library for the problem in hand
Spectral nonlinear unmixing (SNLU). Just a few topics

- **Detecting nonlinear mixtures in polynomial post-nonlinear mixing model**, [Altmann, Dobigeon, Tourneret, 11,13]
  \[ y = M\alpha + b(M\alpha) \odot (M\alpha) + n \]
  hypothesis test

- **Bilinear unmixing model**, [Fan, Hu, Miller, Li, 09], [Nascimento, B-D, 09], [Halimi, Altmann, Dobigeon, Tourneret, 11,11]
  \[ y = \sum_{i=1}^{p} \alpha_i m_i + \sum_{i,j=1}^{p} \alpha_{i,j} m_i \odot m_j + n \]

- **Kernel-based unmixing algorithms to specifically account for intimate mixtures** [Broadwater, Chellappa, Burlina, 07], [Broadwater, Banerjee, 09,10, 11], [Chen, Richard, Ferrari, Honeine, 13]

---


Sparse regression-based SLU

\[ A \in \mathbb{R}^{L \times m} \]  
(library, \( p \ll L < m \), undetermined system)

Problem – P0

\[
\min_{x} \|x\|_0 \quad \text{subject to} \quad \|y - Ax\|_2 \leq \delta, \quad x \geq 0,
\]

Very difficult (NP-hard)

Approximations to P0:
OMP – orthogonal matching pursuit [Pati et al., 2003]
BP – basis pursuit [Chen et al., 2003]
BPDN – basis pursuit denoising
IHT (see [Blumensath, Davis, 11], [Kyrillidis, Cevher, 12])