24th IEEE Signal Processing and Communications Applications Conference
16-19 May 2016
ZONGULDAK

# A Signal Processing Perspective on Hyperspectral Unmixing


*José M. Bioucas-Dias*

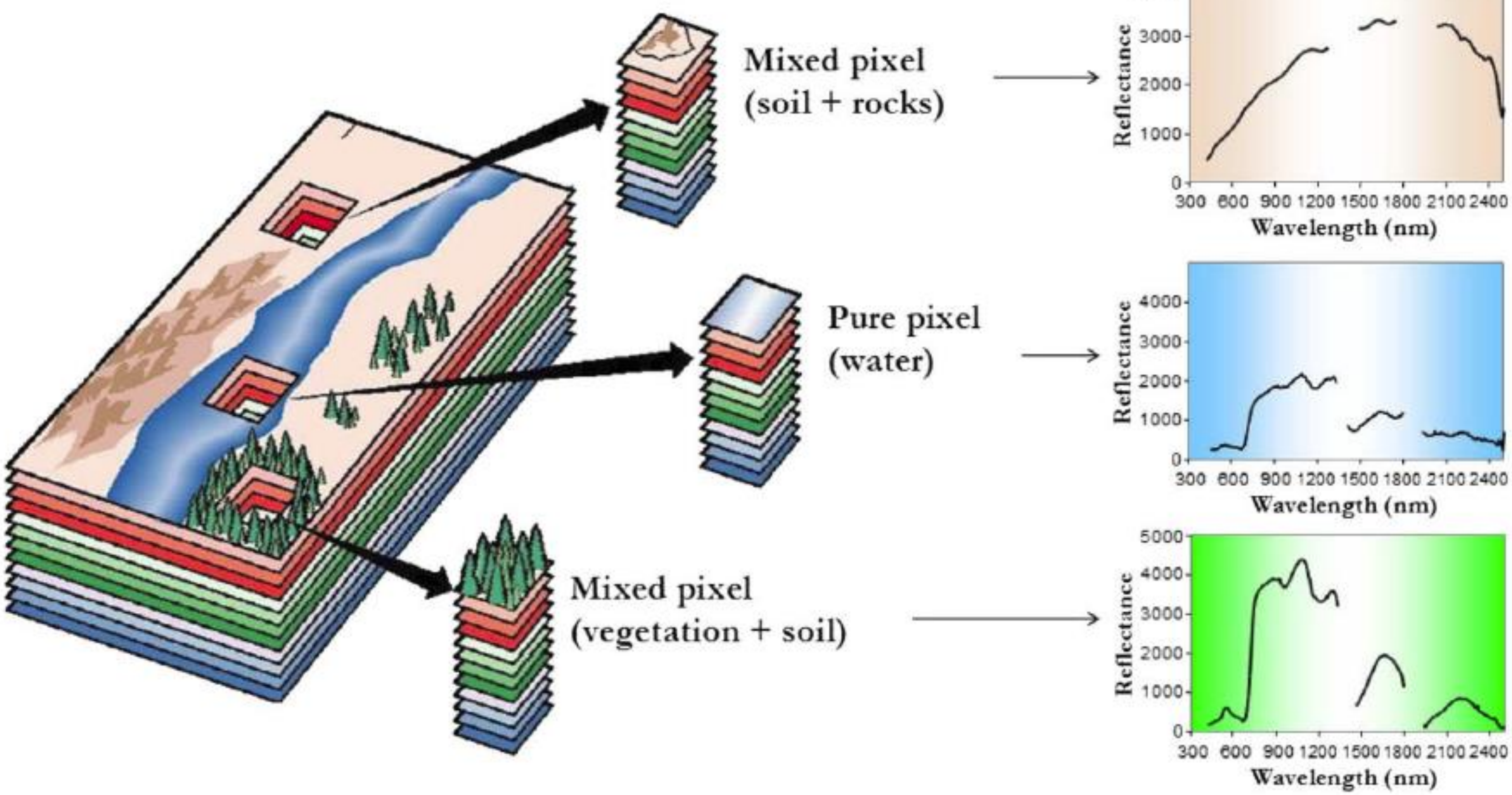
Instituto de Telecomunicações,
Instituto Superior Técnico
Universidade de Lisboa, Portugal


Workshop on Signal and Image Processing for Remote Sensing


FCT, project UID/EEA/50008/2013


Zonguldak, Turkey, May, 2016

# Hyperspectral imaging (and mixing)



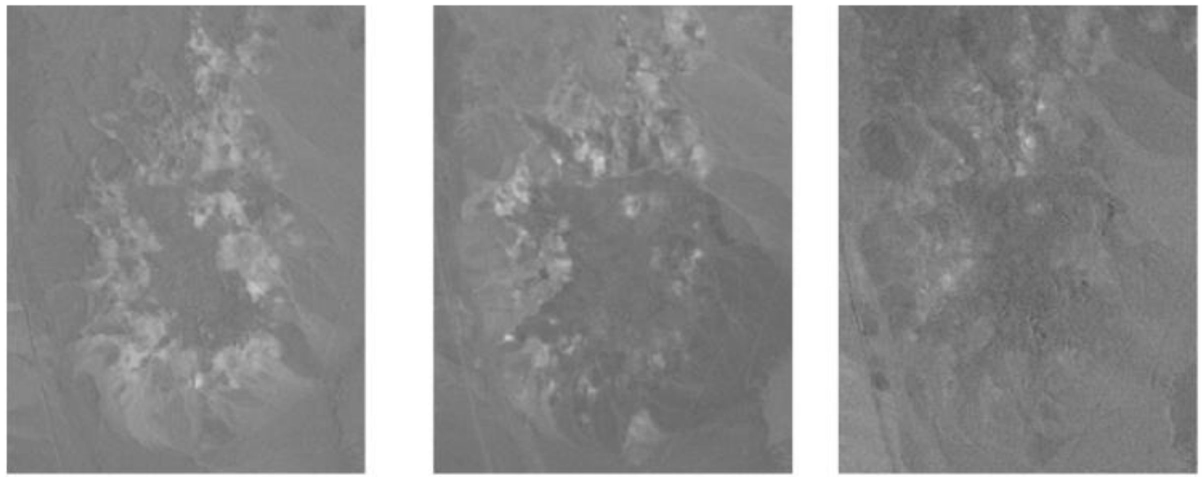
# Hyperspectral unmixing

AVIRIS of Cuprite
Nevada, USA

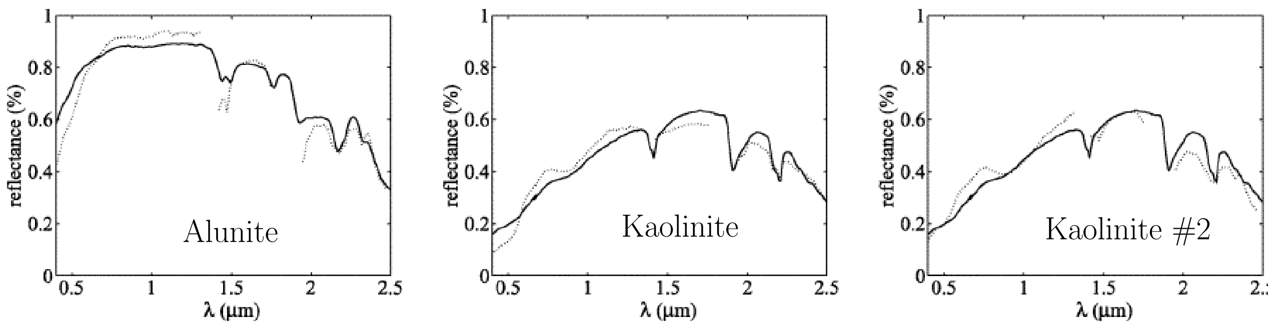
R – ch. 183 (2.10 $\mu\mathrm{m}$)

G – ch. 193 (2.20 $\mu\mathrm{m}$)

B – ch. 207 (2.34 $\mu\mathrm{m}$)



(a) (b) (c)




(a) (b) (c)

VCA [Nascimento, Bioucas, 05]

# Outline

- Mixing models
  - Linear
  - Nonlinear

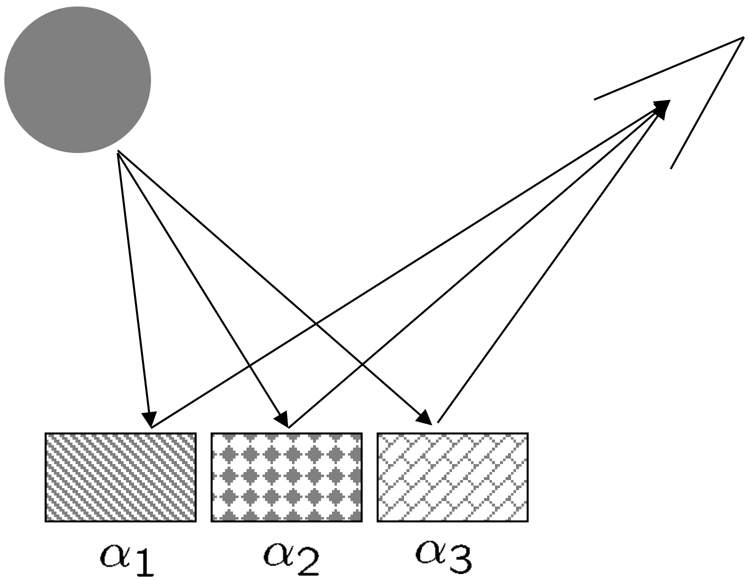
- Signal subspace identification

- Unmixing
  - Geometrical-based
  - Sparse regression-based

J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354-379, 2012.

W.-K. Ma, J. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. Plaza, A. Ambikapathi and C.-Y. Chi, "A signal processing perspective on hyperspectral unmixing", *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 3 67-81, 2014

# Linear mixing model (LMM)




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

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

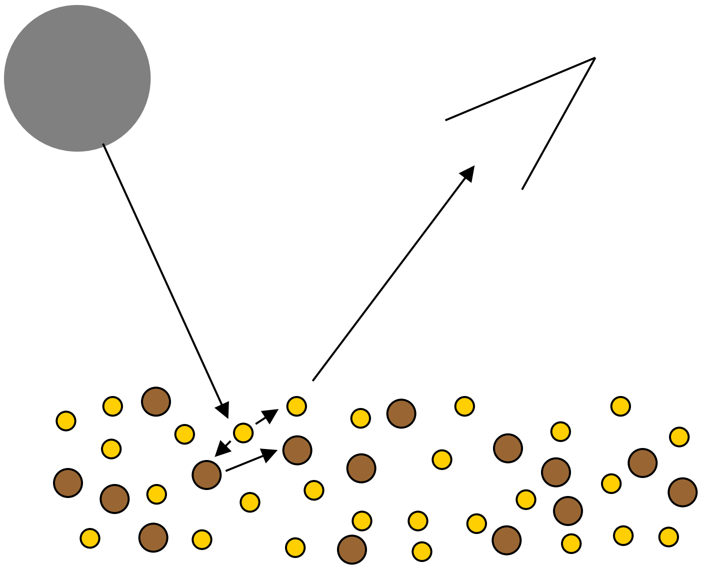
$$\mathbf{r} = \mathbf{M}\boldsymbol{\alpha}$$

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

Hyperspectral linear unmixing → Estimate $\mathbf{M}, \ \boldsymbol{\alpha}$

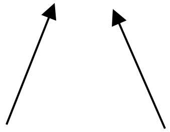
# Nonlinear mixing model

Intimate mixture (particulate media)



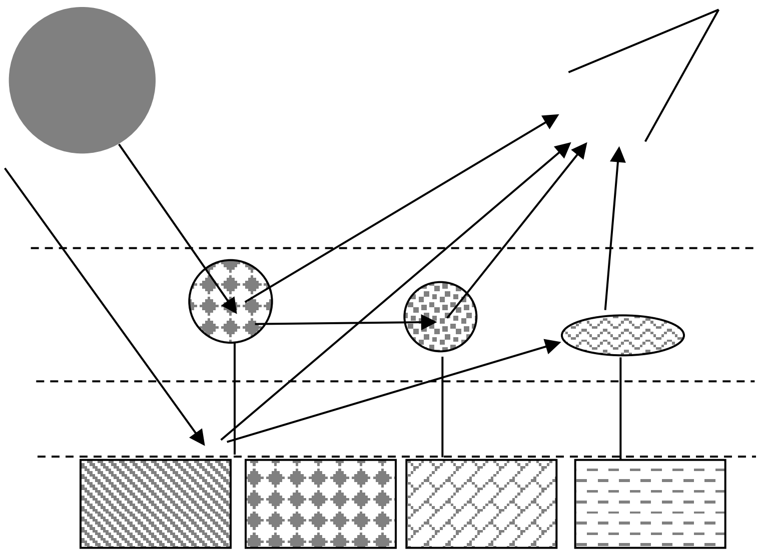
Radiative transfer theory

$$\mathbf{r} = f(\alpha, \theta)$$



material fractions media parameters

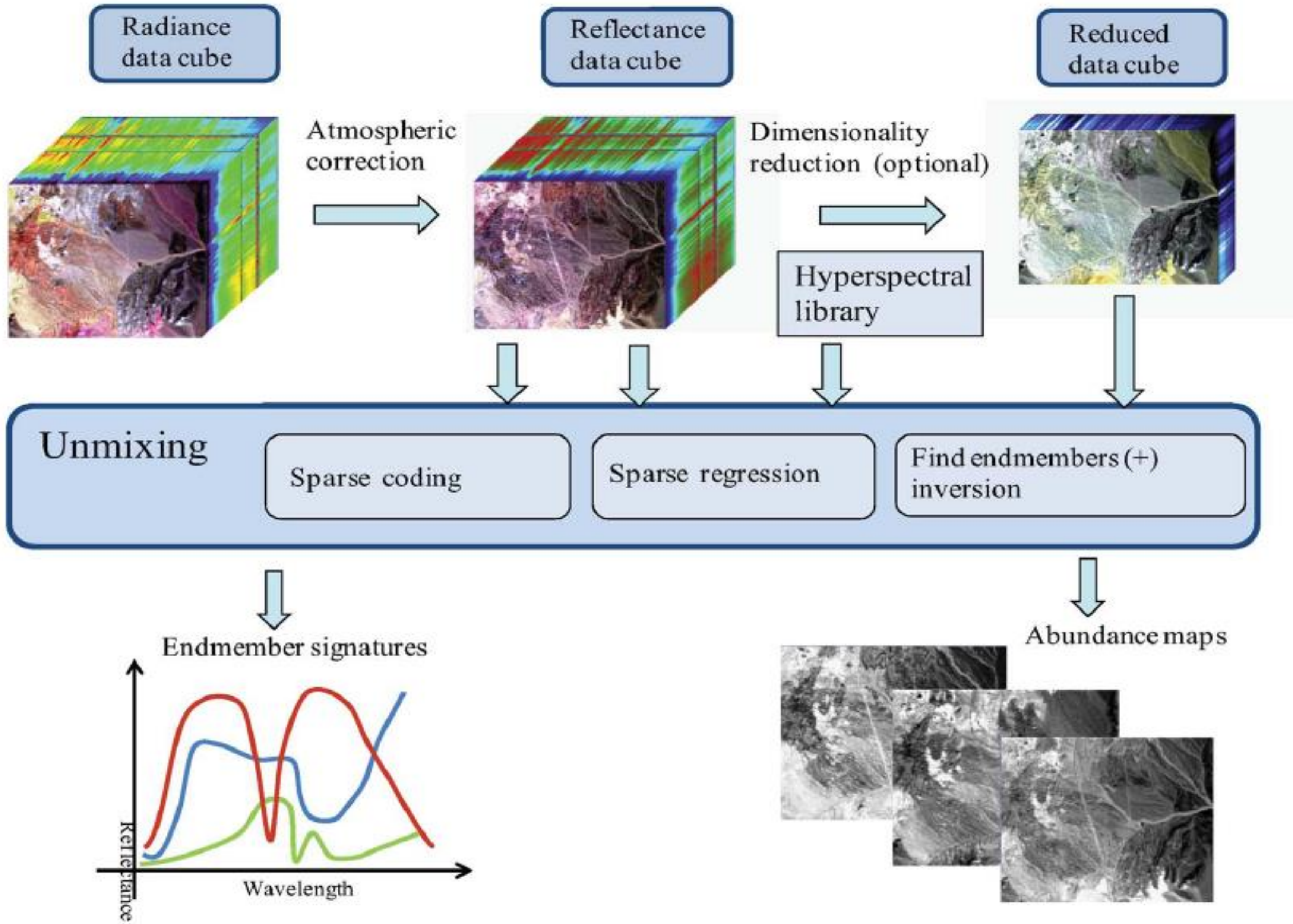
Two-layers: canopies+ground



$$\mathbf{r} = \underbrace{\sum_{i=1}^{p} \alpha_i \mathbf{m}_i}_{\text{single scattering}} + \underbrace{\sum_{\substack{i,j=1 \\ i \neq j}}^{p} \alpha_{i,j} \mathbf{m}_i \odot \mathbf{m}_j}_{\text{double scattering}}$$

single scattering double scattering

# Schematic view of the unmixing process



# Hyperspectral (linear) unmixing (HU)

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

$$\mathbf{Y} = \left[\mathbf{y}_i \in \mathbb{R}^L,\ i = 1, \ldots, N\right]$$

Subject to the LMM: $\mathbf{y}_i = \mathbf{M}\boldsymbol{\alpha}_i + \mathbf{n}_i,\quad \boldsymbol{\alpha}_i \geq 0,\quad \mathbf{1}^T\boldsymbol{\alpha}_i = 1$

ANC: abundance nonnegative constraint

ASC: abundance sum-to-one constraint

Determine:

- The mixing matrix $\mathbf{M}$ (*endmember spectra*)
- The *fractional abundance* vectors $\boldsymbol{\alpha}_i,\ i = 1, \ldots, N$

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

# Signal subspace identification

$$\mathbf{Y} = \mathbf{MS} + \mathbf{N} \quad \mathbf{S} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N] \quad \mathbf{N} = [\mathbf{n}_1, \ldots, \mathbf{n}_N]$$

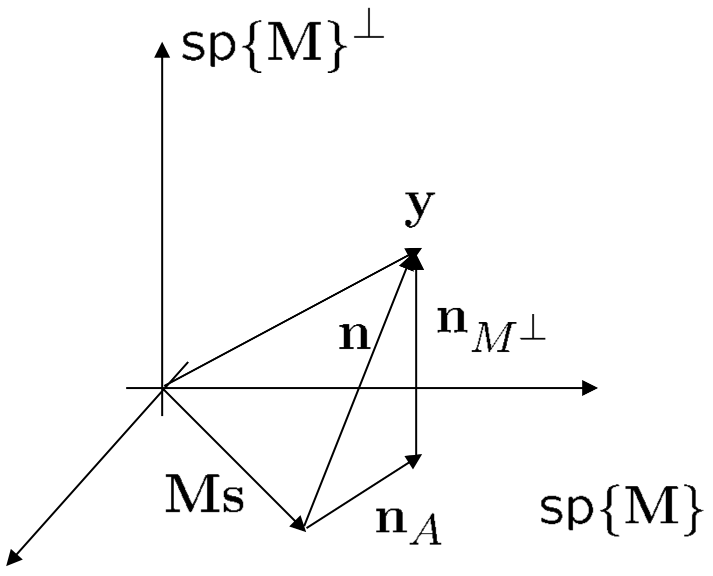
$$\dim(\mathbf{M}) = [L \times p] \quad L \gg p$$

Problem: Identify $\mathrm{sp}\{\mathbf{M}\}$ 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 $p/L$

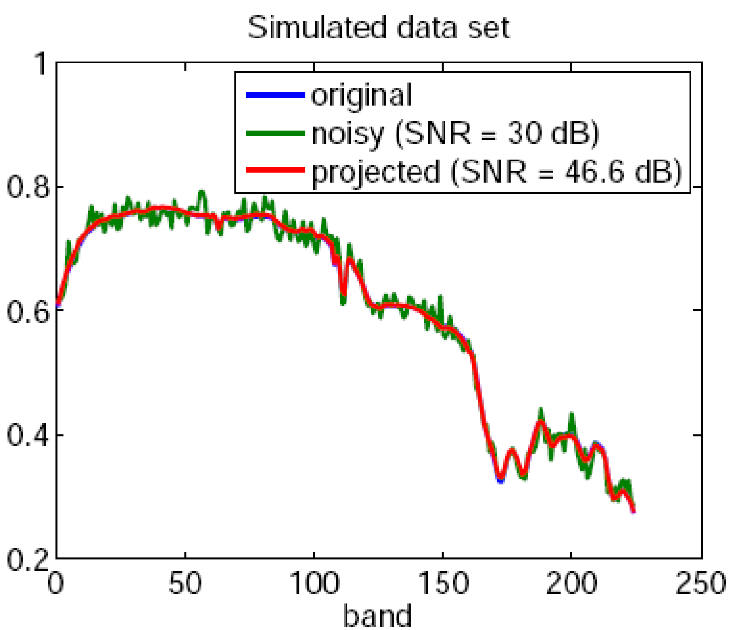


# Subspace identification algorithms

Known $p$, i.i.d. Gaussian noise: Exact ML solution [Scharf, 91]

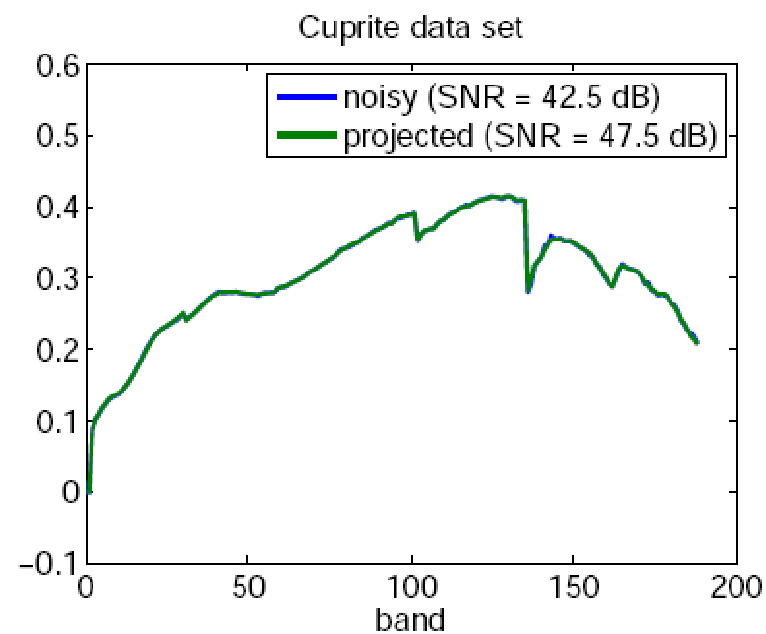
Unknown $p$, non-i.i.d. noise: Model order selection problem

Example with **HySime** [Bioucas, Nascimento, 08]




i.i.d. Gaussian noise

$L = 224,\ p = 5,\ N/p = 16.5\,\mathrm{dB}$

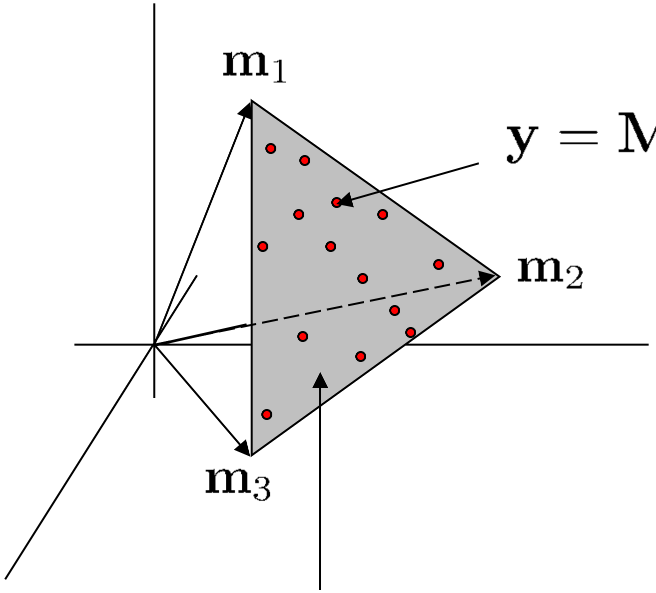



non-i.i.d. Gaussian noise

$L = 188,\ p = 15,\ N/p = 11.7\,\mathrm{dB}$

# Geometrical view of HU

$\mathbf{y} \in \mathbb{R}^L \qquad \mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_p]$



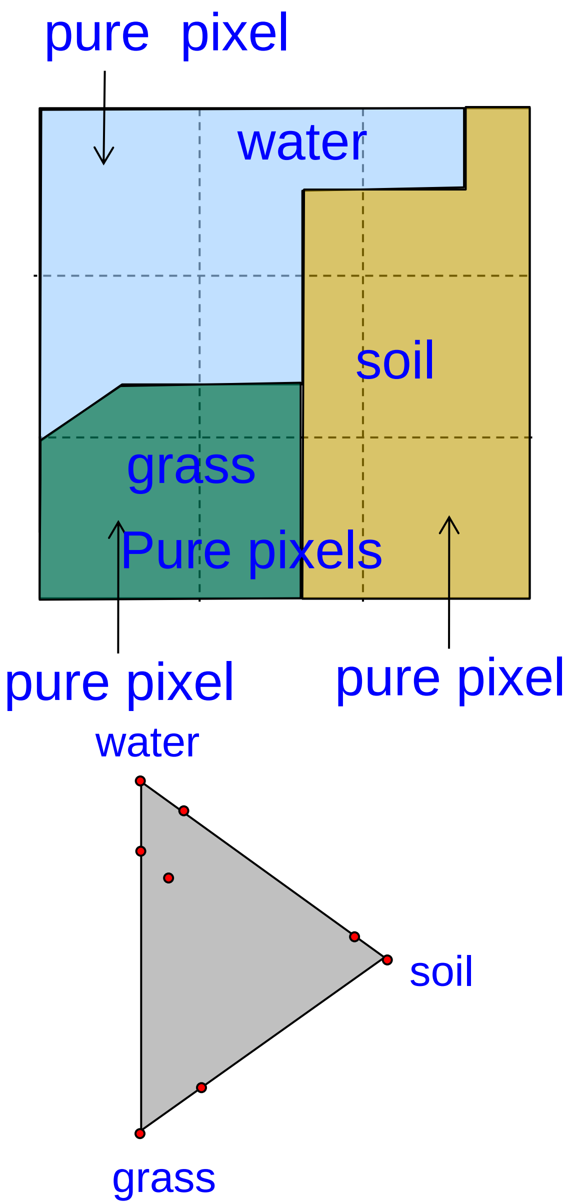

$$\sum_{j=1}^{p} \alpha_j = 1, \ \alpha_j \geq 0$$

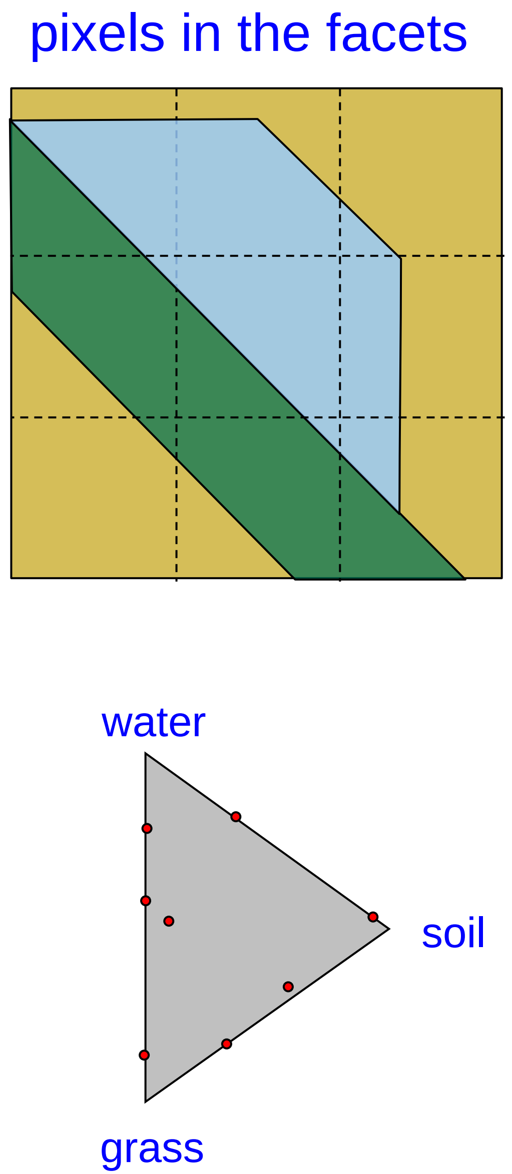
probability simplex $(S_I)$

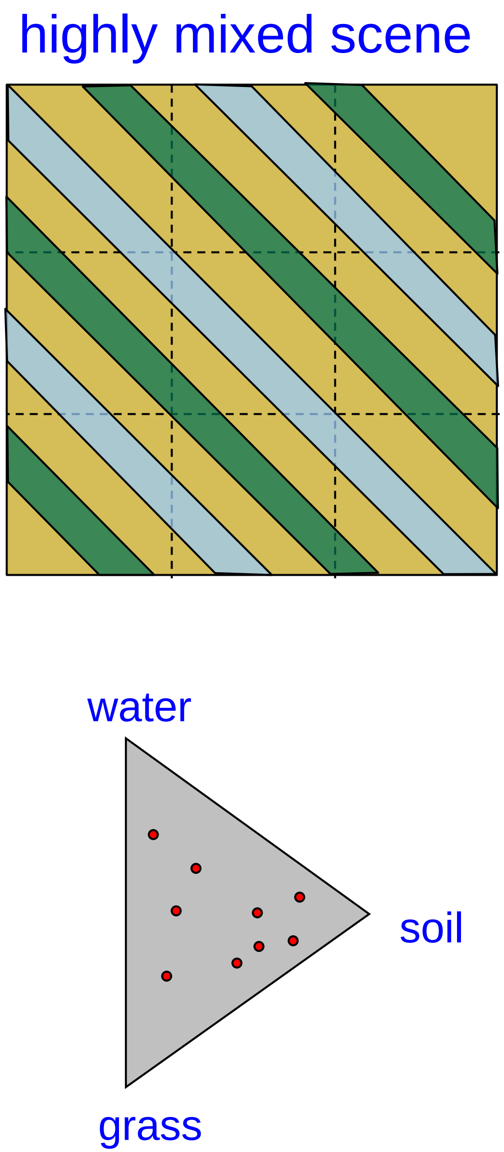
$S_M = \{\mathbf{x} \in \mathbb{R}^p \ : \ \mathbf{x} = \mathbf{M}\boldsymbol{\alpha}, \ \boldsymbol{\alpha} \in S_I\}$ ➡ $(p-1)$-simplex

Inferring $\mathbf{M} \Leftrightarrow$ inferring the vertices of the simplex $S_M$

# Classes of HU problems

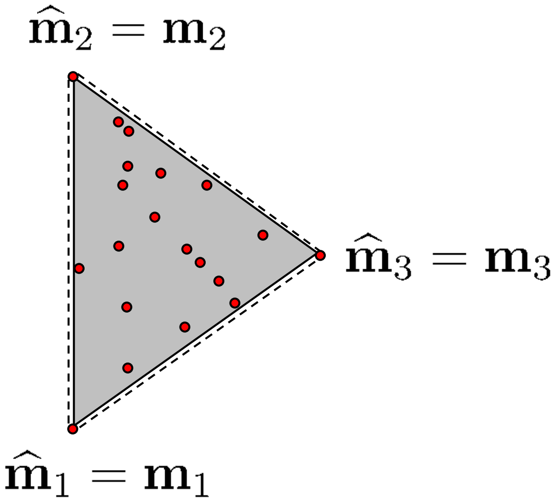


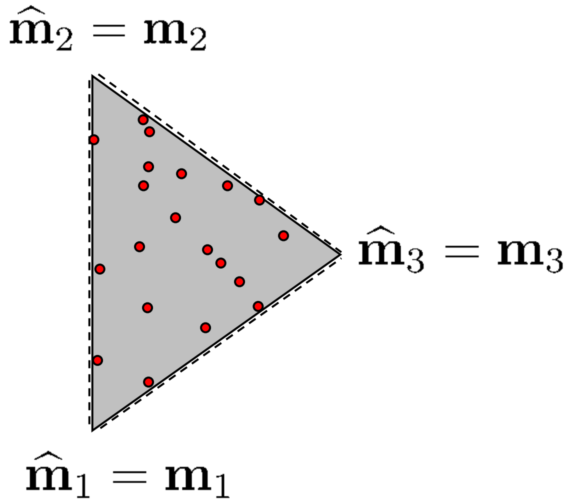
# Classes of HU problems

**Pure pixels**



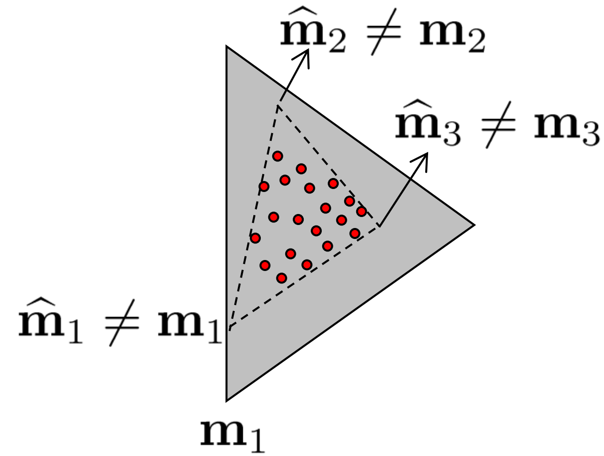

Well posed

**Pixels in the facets**




Well posed

**Highly mixed**




Ill-posed

## Algorithms

| Pure pixels | Pixels in the facets | Highly mixed |
| --- | --- | --- |
| Vertex pursuit | Facet estimation (minimum volume) | Statistical inference<br>Sparse regression |

# 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

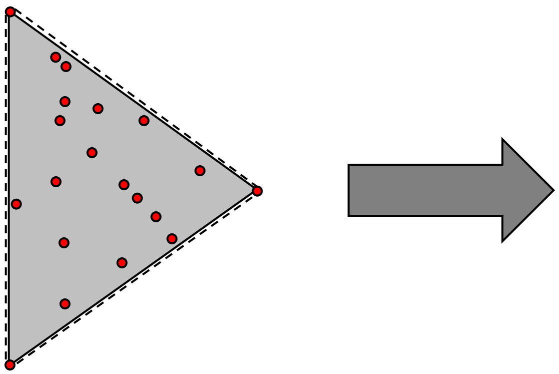
  **Application scenarios:** all

- Sparse regression (semi-blind)
  Approaches linear unmixing as a sparse regression problem

  **Application scenarios:** all

# Simplex vertex pursuit

Assumption: the data set contains at least one pure pixel of each material



- Search endmembers in the data set
- Computationally light

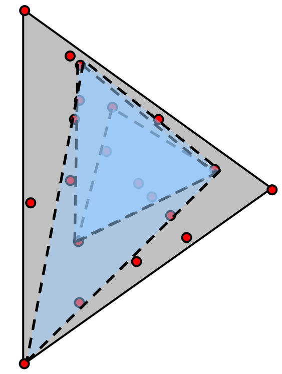
**PPI** - [Boardman, 93]; **N-FINDR** - [Winter, 99]; **IEA** - [Neville *et al*., 99];
**AMEE** – [Plaza *et al*, 02]; **SMACC** – [Gruninger *et al.*, 04]
**VCA** - [Nascimento,Bioucas, 03, 05]; **SGA** - [Chang *et al*., 06]
**AVMAX**, **SVMAX** - [Chan, et al., 11]; **RNMF**- [Gillis,Vavasis, 12,14];
**SD-SOMP**, **SD-ReOMP** - [Fu,Ma,Chan,Bioucas15]

# Simplex vertex pursuit

**N-FINDR**

$$\text{iteratively increase } |\det(\mathbf{m}_1, \ldots, \mathbf{m}_p)|$$

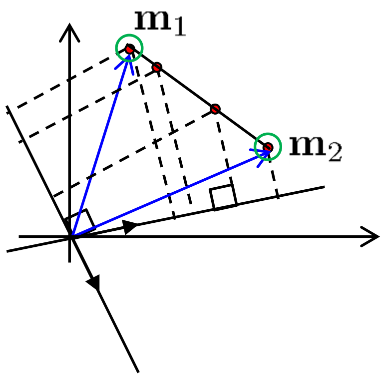
$$\mathbf{m}_1, \ldots, \mathbf{m}_p \in \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$$



**AVMAX**

$$\max_{\mathbf{m}_1, \ldots, \mathbf{m}_p} \quad |\det(\mathbf{m}_1, \ldots, \mathbf{m}_p)|$$

$$\text{st:} \quad \mathbf{m}_1, \ldots, \mathbf{m}_p \in \text{conv}\{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$$




**VCA**

for $k = 1 : p$

$$\mathbf{m} := \arg\max_{\mathbf{y}_i} \|(P_{\mathbf{M}}^{\perp}\xi)^T \mathbf{y}_i\|$$

$$\mathbf{M} := [\mathbf{M}, \ \mathbf{m}]$$

**SPA**
**SVMAX**

for $k = 1 : p$

$$\mathbf{m} := \arg\max_{\mathbf{y}_i} \|(P_{\mathbf{M}}^{\perp}\mathbf{y}_i\|$$

$$\mathbf{M} := [\mathbf{M}, \ \mathbf{m}]$$

# Recovery guarantees and stability

- Noiseless case: under the pure pixel assumption, AVMAX,VCA, SPA, and SVMAX exactly identifies all the endmember signatures [Nascimento,Bioucas,03,05], [Chan et al.,11], [Gillis,Vavasis,12,14]

- Noisy case: [Gillis,Vavasis,14]

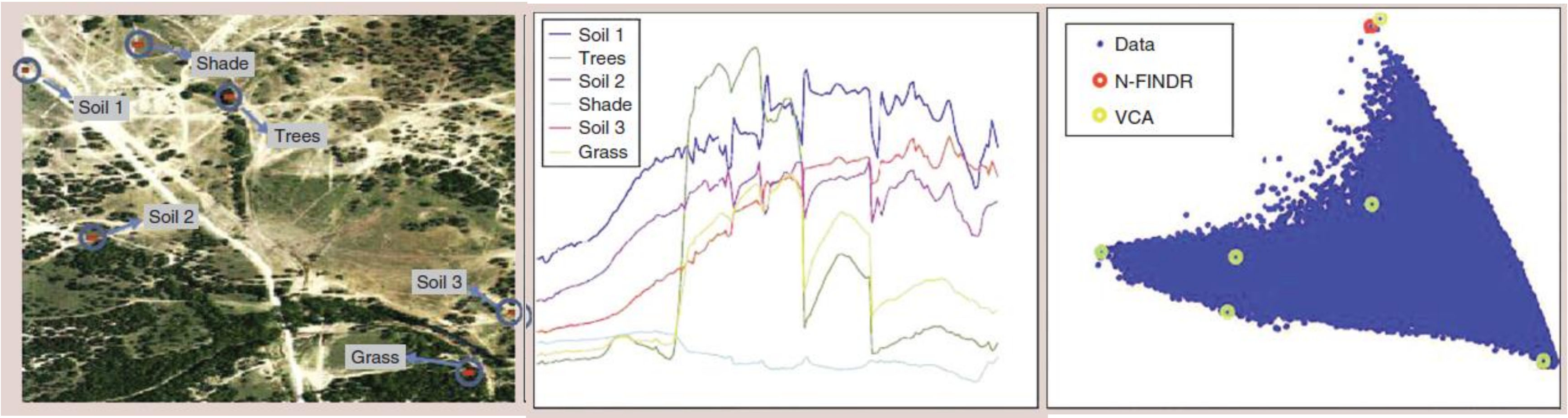
  - $\kappa = \sigma_{\min}(\mathbf{M}) / \sigma_{\min}(\mathbf{M})$
  - $\epsilon = \max_{1 \leq i \leq N} ||n_i||_2$ $\quad \epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}}{\sqrt{p}\kappa^2}\right)$
  - Pure pixel assumption

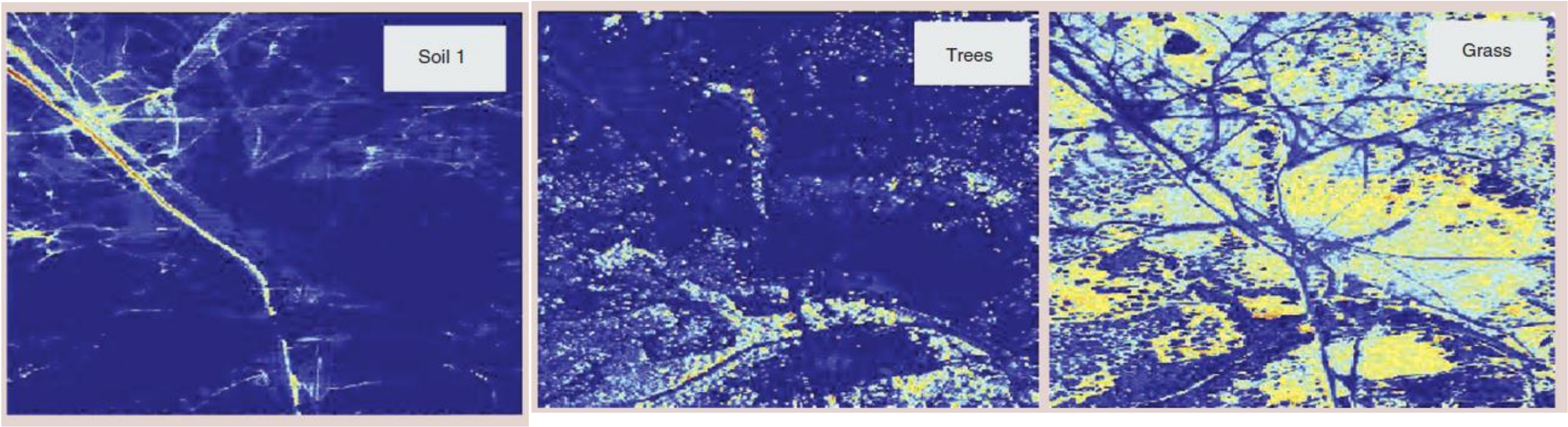
Then SPA identifies all the endmember signatures $\{\mathbf{m}_1, \ldots, \mathbf{m}_p\}$ up to error $\mathcal{O}\left(\epsilon\, \kappa^2\right)$ more precisely, we have

$$\max_{1 \leq i \leq p} \min_{1 \leq j \leq p} ||\mathbf{m}_i - \hat{\mathbf{m}}_j||_2 \leq \mathcal{O}\left(\epsilon\, \kappa^2\right)$$

# Unmixing example

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



# Simplex facet estimation

$$\mathbf{Y} = \mathbf{MS} + \mathbf{N}$$

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

$$(\widehat{\mathbf{M}}, \widehat{\mathbf{S}}) = \arg \min_{\mathbf{M}\in\mathbb{R}^{L\times p}, \mathbf{S}\in\mathbb{R}^{p\times N}} \frac{1}{2}\|\mathbf{Y} - \mathbf{MS}\|_F^2 + \tau \text{Vol}(\mathbf{M}) + \lambda\phi(\mathbf{S})$$

$$\text{s.t.:} \quad \mathbf{M} \geq 0, \quad \mathbf{S} \geq 0, \quad \mathbf{1}^T\mathbf{S} = \mathbf{1}_N^T,$$

volume regularizer

---

**DPFT** - [Craig,90]; **CCA** - [Perczel et al., 89] (seminal works on MVC)

**ICE** - [Breman *et al*., 04] ($V(\mathbf{M}) \equiv \text{quadratic}, \ \phi = 0$ );

**MVC-NMF** - [Miao,Qi, 07] ($V(\mathbf{M}) = |\det(\mathbf{M}^T\mathbf{M})|, \ \phi = 0$);

**SPICE** - [Zare,Gader, 07] ($V(\mathbf{M}) \equiv \text{quadratic}, \ \phi \equiv \text{weighted } \ell_1$)

**L1/2 - NMF** - [Qian,Jia,Zhou,Robles-Kelly, 11] ($V = 0$ , $\phi(\mathbf{S}) = \sum_{ij} |\alpha_{ij}|^{1/2}$ )

**CoNMF** - [Li,Bioucas,Plaza,12,16] ($V(\mathbf{M}) \equiv \text{quadratic}, \ \phi(\mathbf{S}) = \|\mathbf{S}\|_{2,1}$)

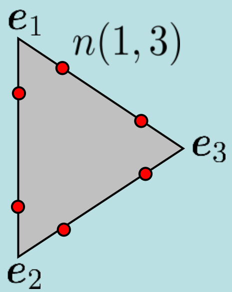
# Indentifiability and convergence guarantees

- Minimum volume criterium ($\mathbf{N}=\mathbf{0}$)

$$\min_{\mathbf{M}} \mathrm{Vol}(\mathbf{M}) \quad \text{s.t.}\ \ \mathbf{Y}=\mathbf{MS},\ \ \mathbf{S}\geq 0,\ \ \mathbf{1}^T\mathbf{S}=\mathbf{1}_N^T$$

Indentifiability [Lin,Ma,Li,Chi,Ambikapathi,15]: the above optimization yields the true matrix $\mathbf{M}$ subject to ordering permutations if for any $i,j \in \{1,\ldots,p\}$ there is a pixel $n=n(i,j)$ such that

$$\boldsymbol{\alpha}_n = a_n\boldsymbol{e}_i + (1-a_n)\boldsymbol{e}_j \qquad \begin{cases} 2/3 & p=3 \\ 1/2 & p\geq 4 \end{cases} < a_n \leq 1$$




- Proximal alternating optimization to solve MVC-NMF [Lina,Bioucas,16]
  - Converge to stationary points
  - Nonconvex problem. Initialization matters. No recovery guarantees

# Minimum volume simplex. Formulation on $\mathbf{Q} \equiv \mathbf{M}^{-1}$

**MVSA** – Minimum volume simplex analysis [Li,Bioucas,08]

Optimization variable $\mathbf{Q} \equiv \mathbf{M}^{-1} \quad \Longrightarrow \quad \mathbf{QY} = \mathbf{S}$

$$\widehat{\mathbf{Q}} = \arg\max_{\mathbf{Q}} \; \log(|\det(\mathbf{Q})|)$$

$$\text{s.t.:} \quad \mathbf{1}_p^T \mathbf{Q} = \mathbf{q}_m, \quad \mathbf{QY} \geq 0$$

ASC ANC

- Convex constraint set
- Nonconvex, but do not get trapped in poor local minima

**MVSA** solves a sequence of quadratic programs

**MVES** [Chan et al., 09]

- ❑ 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

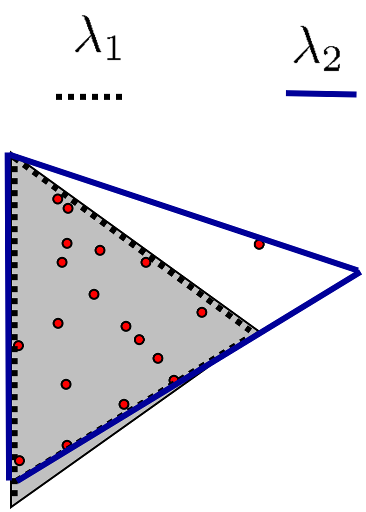
# Robust minimum volume simplex algorithms: outliers

**SISAL** – Simplex identification via split augmented Lagrangian  [Bioucas,09]

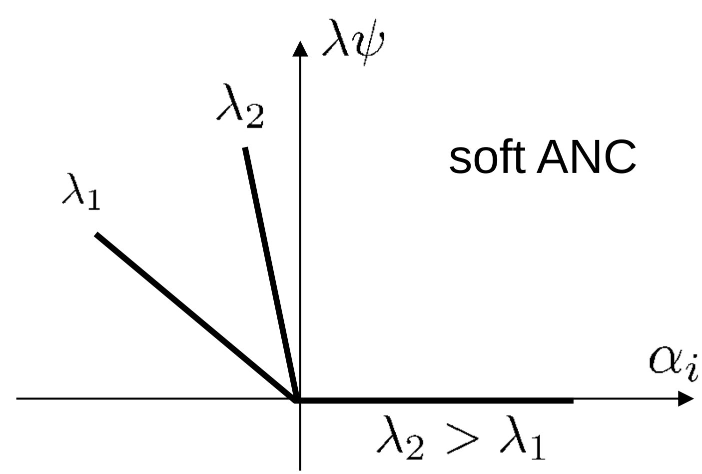
$$\widehat{\mathbf{Q}} = \arg\max_{\mathbf{Q}} \log(|\det(\mathbf{Q})|) - \lambda\phi(\mathbf{QY})$$

$$\text{s.t.:} \quad \mathbf{1}_p^T\mathbf{Q} = \mathbf{q}_m$$

ASC



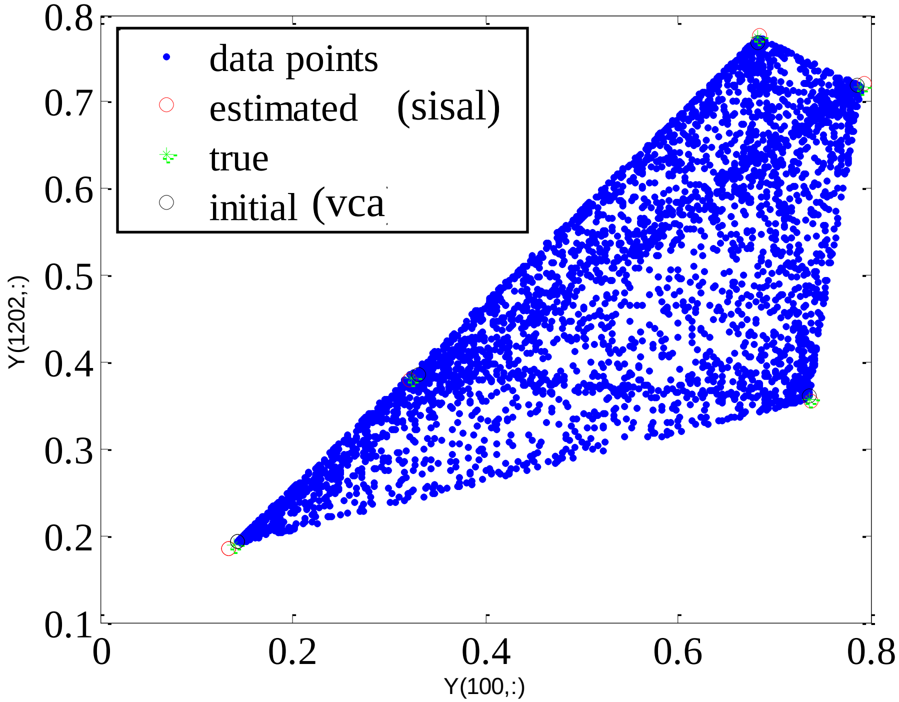

$\lambda_2 > \lambda_1$

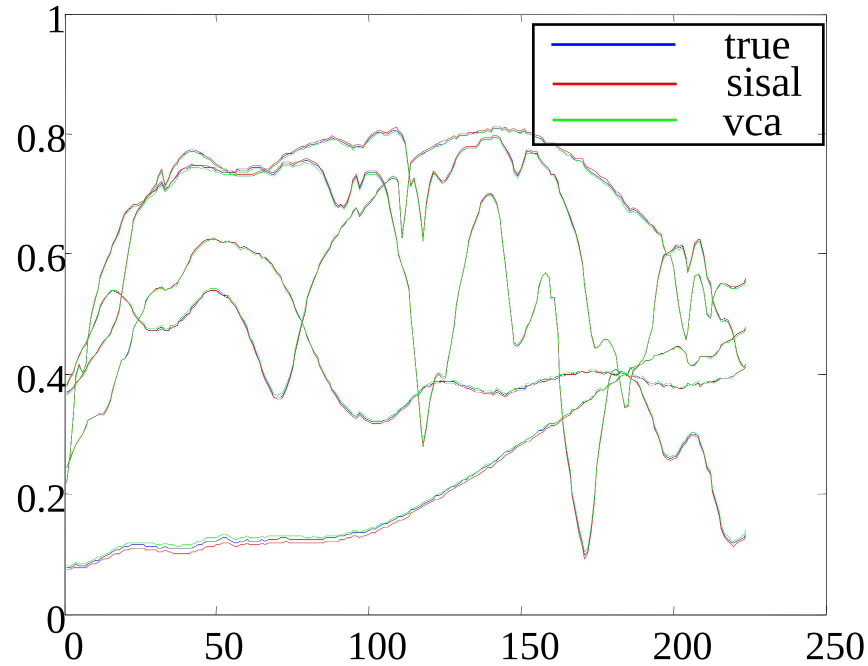



- **SISAL** solves a sequence of convex subproblems using ADMM
- $(\lambda = \infty) \Rightarrow \mathbf{MVES} \equiv (\mathbf{MVES},\ \mathbf{SISAL})$

# Example: data set contains pure pixels

$$N = 5000 \qquad p = 5 \qquad \max_{\alpha_i} = 1, \text{ for } i = 1, \ldots, p$$



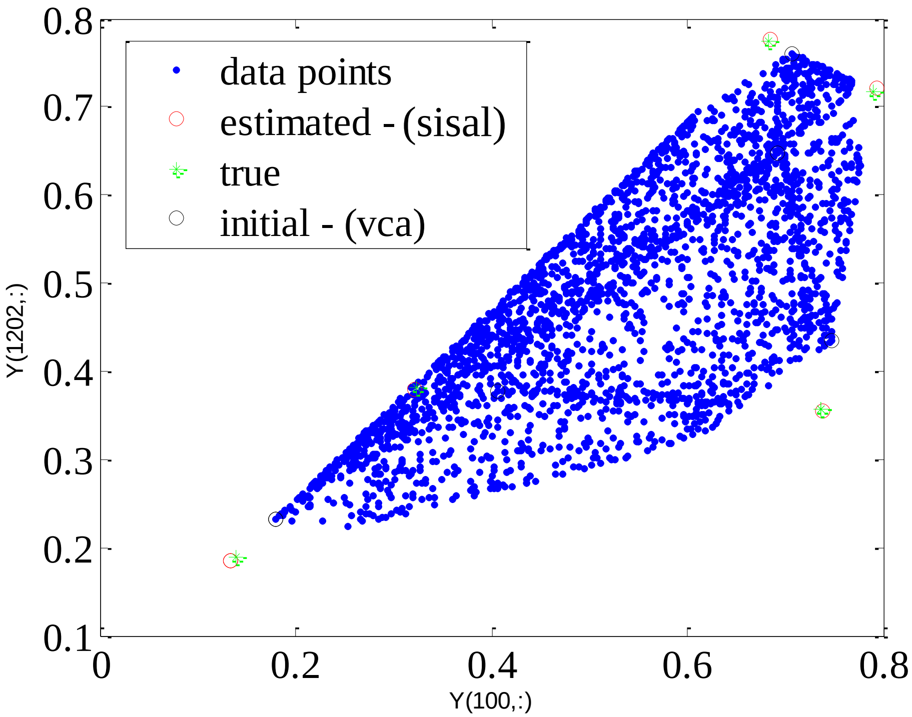

Time:

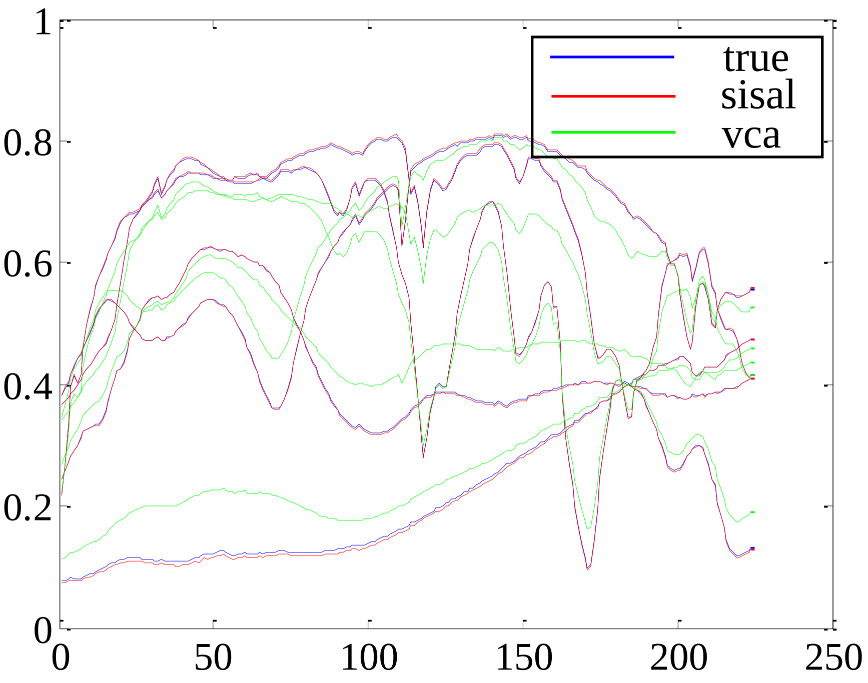
VCA → 0.5 sec

SISAL →2 sec

# Example: Data set does not contain pure pixels

$$N = 5000 \qquad p = 5 \qquad \max_{\alpha_i} = 0.8, \text{ for } i = 1, \ldots, p$$

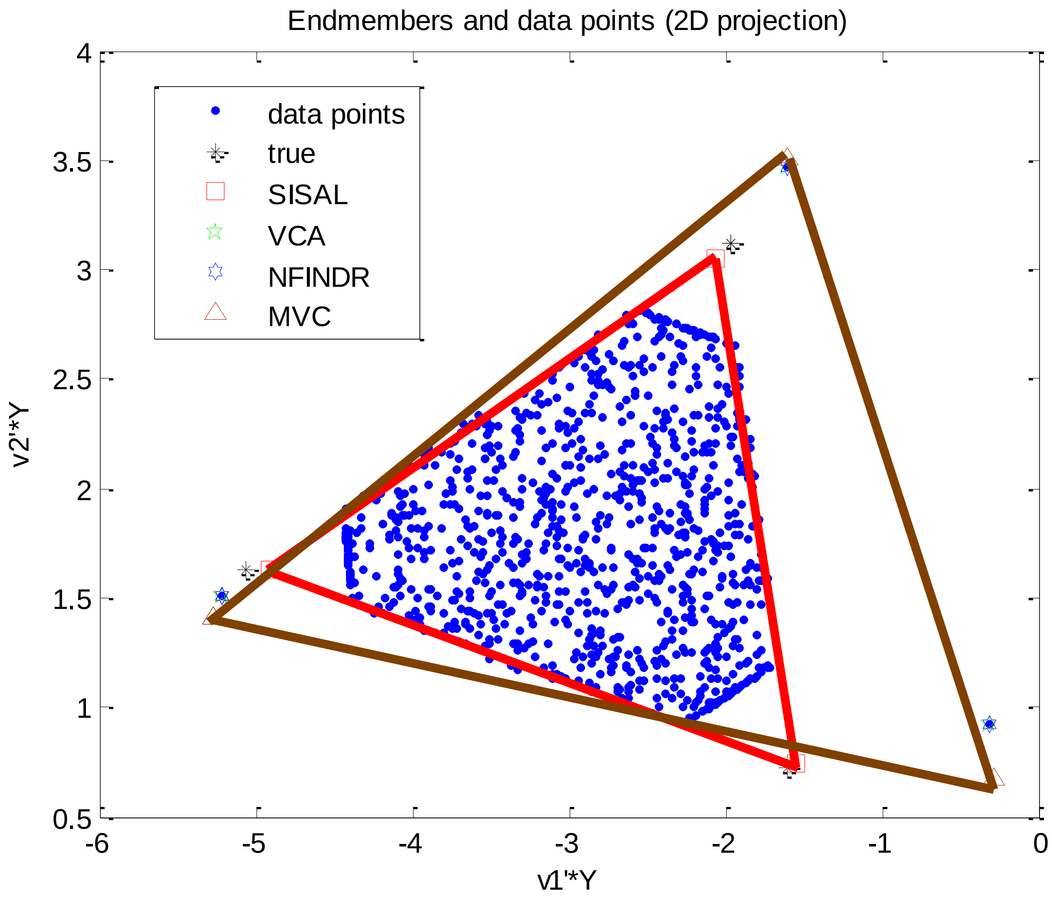


# No pure pixels and outliers

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

$$\text{no. outliers} = 3$$




ERROR(mse):

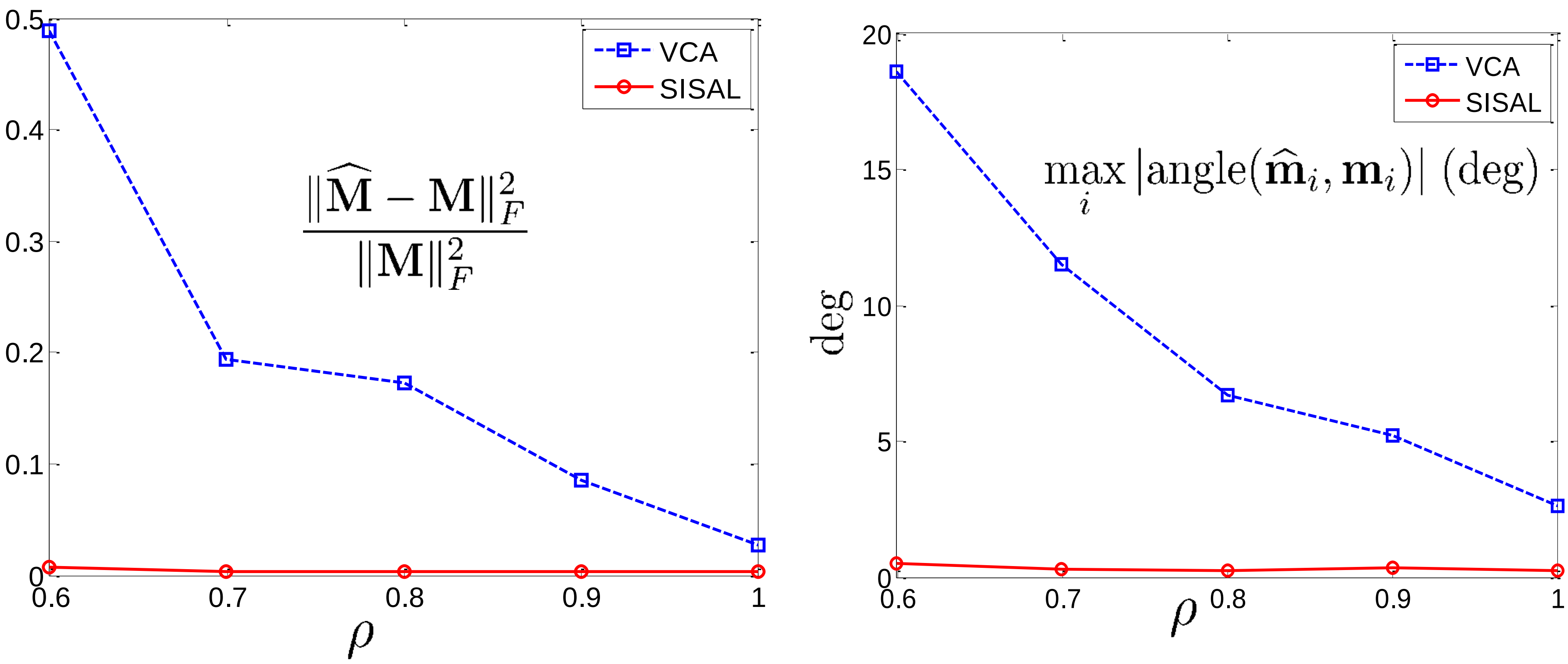
SISAL = 0.03
VCA = 0.88
NFINDR = 0.88
MVC-NMF = 0.90

TIMES (sec):

SISAL = 0.61,
VCA = 0.20
NFINDR = 0.25
MVC-NMF = 25

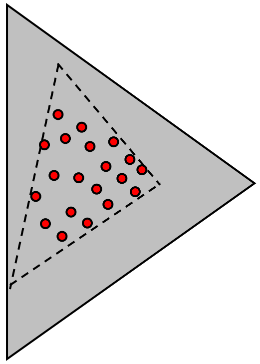
# Simulation results

Monte Carlo simulation: $p = 4, \quad \rho = \max \boldsymbol{\alpha}_i$ (purity) $\mathrm{SNR} = 35\,\mathrm{dB}$



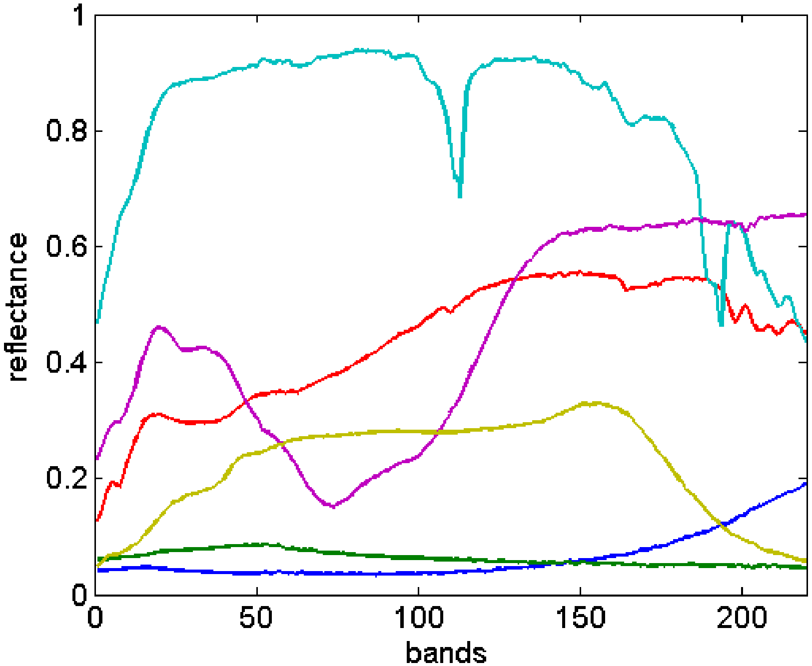
# Sparse regression-based HU

- Geometric-based methods do not work in highly mixed datasets



- Resort to sparse regression based on spectral libraries

- Example: USGS (500 spectral signatures of minerals)




Acmite NMNH133746

Cassiterite HS279.3B

Fassaite HS118.3B

Kaolinite KGa-2 (pxyl)

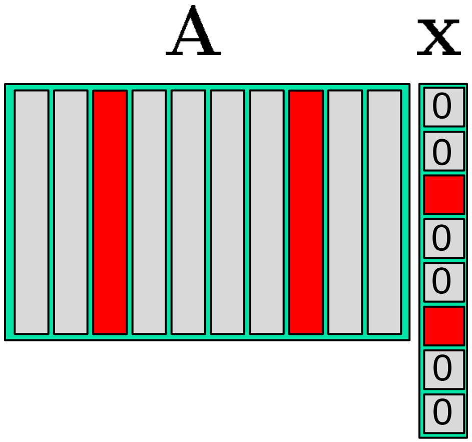
Olivine GDS70.c GSB 70um

Sphalerite S102-8

# Sparse regression-based HU

- Spectral vectors are expressed as linear combinations of a few pure spectral signatures obtained from a (potentially very large) spectral library [Iordache,Bioucas,Plaza,11,12]

$$\mathbf{y} = \sum_{i \in S} \mathbf{a}_i \mathbf{x}_i = \mathbf{A}\mathbf{x}$$




- Unmixing: given $\mathbf{y}$ and $\mathbf{A}$, find the sparsest solution of

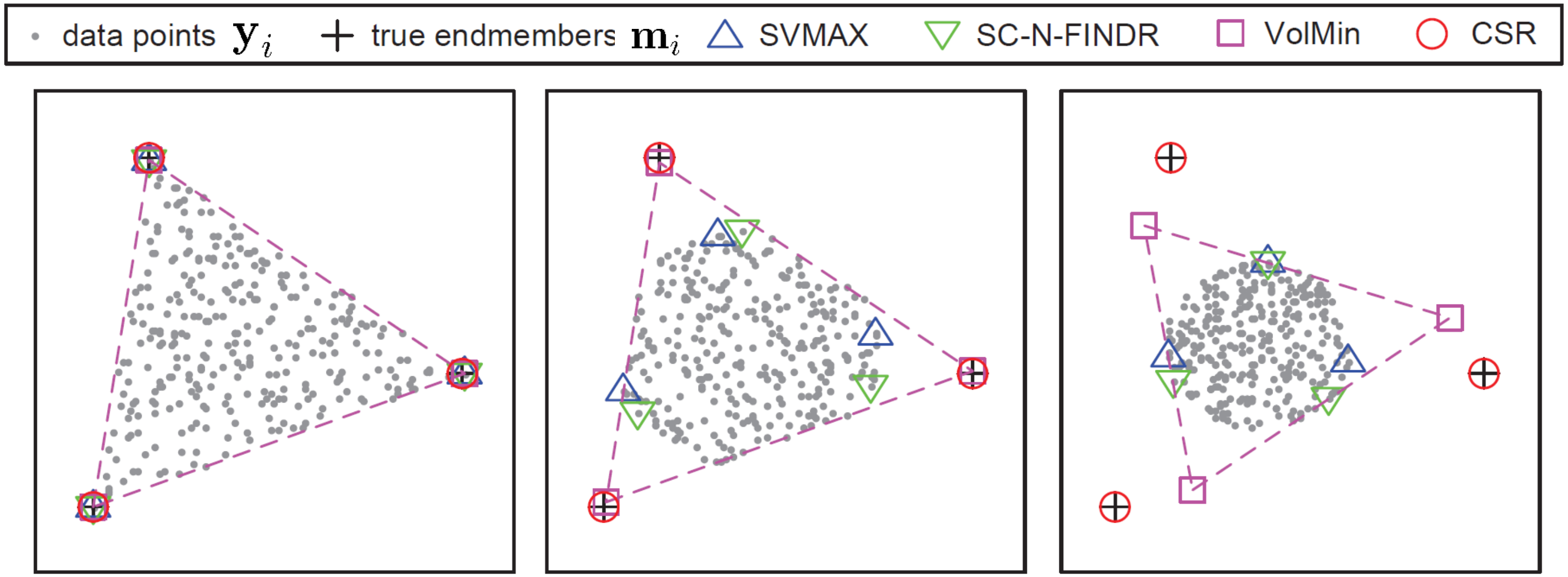
$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

- Advantage: sidesteps endmember estimation

- Disadvantage: **Combinatorial problem !!!**

# Simulation results

Highly mixed dataset: $p = 3$



# Convex approximations to P0

- CBPDN – Constrained basis pursuit denoising

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \ \text{subject to} \ \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \delta, \ \ \mathbf{x} \geq \mathbf{0},$$

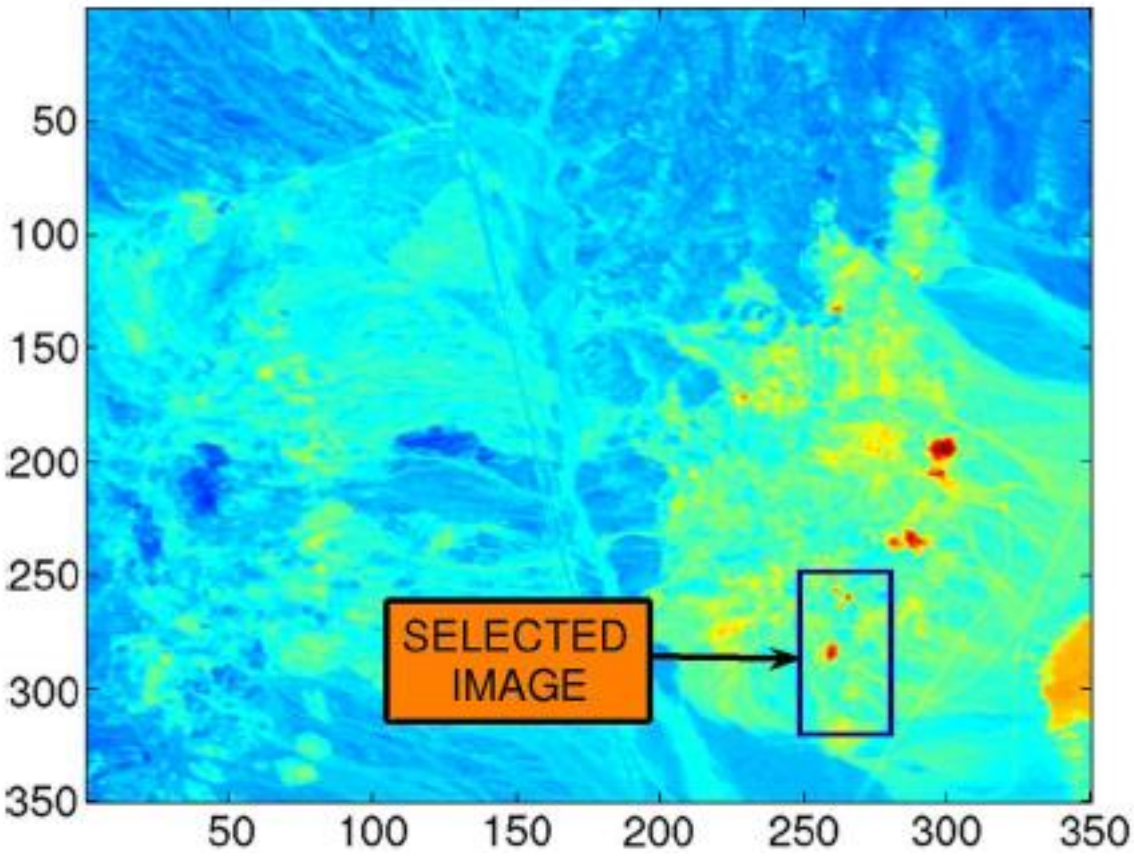
- Equivalent problem

$$\min_{\mathbf{x}} (1/2)\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \lambda\|\mathbf{x}\|_1, \quad \mathbf{x} \geq 0$$

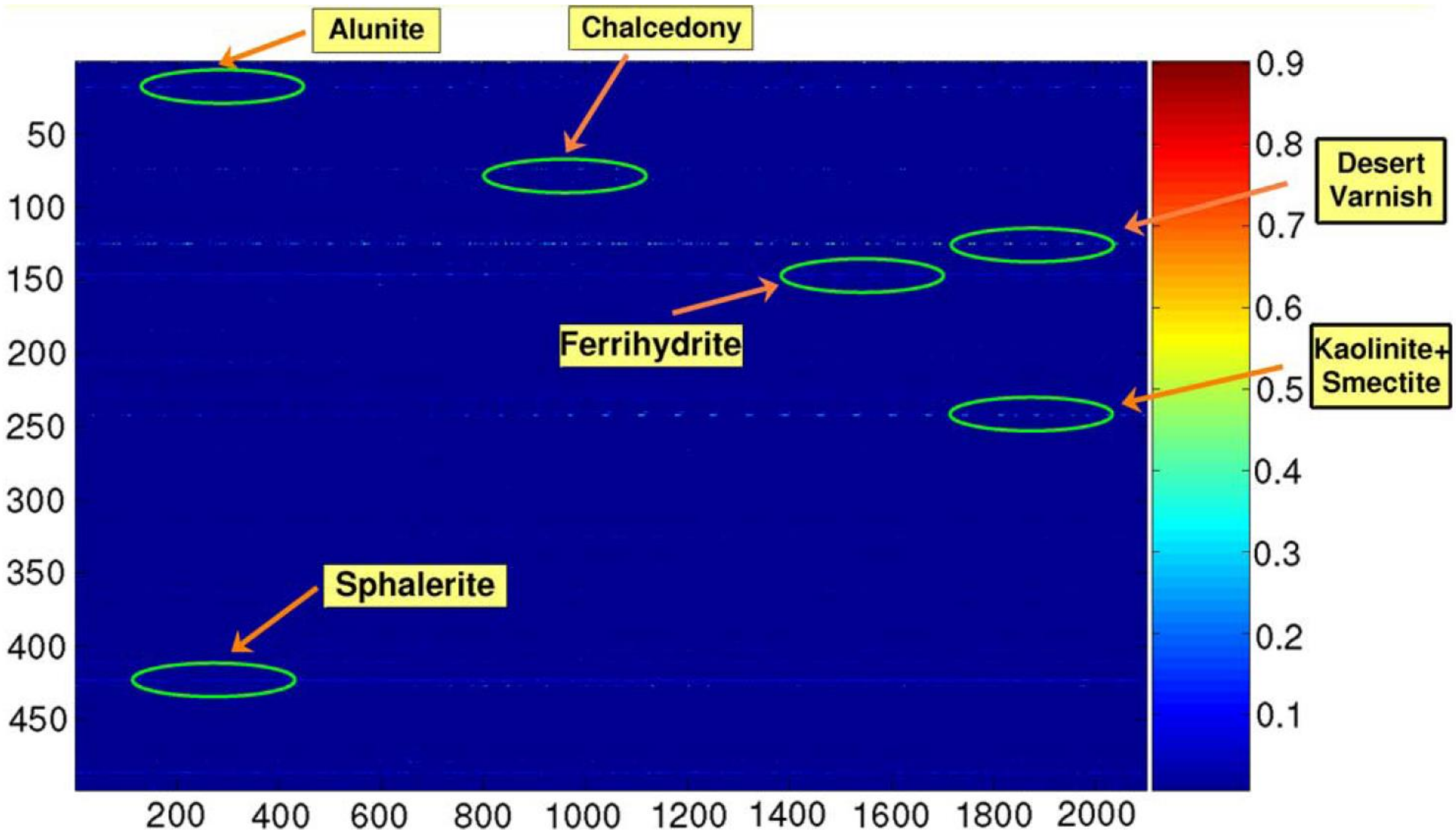
- Result: In given circumstances, related with the coherence among the columns of matrix **A**, BP(DN) yields the sparsest solution ([Donoho 06], [Candès et al. 06]).

- Efficient solvers for CBPDN: **SUNSAL, CSUNSAL** [Bioucas,Figueiredo, 10]

# Real data – AVIRIS Cuprite



# Real data – AVIRIS Cuprite




[Iordache, Bioucas,Plaza,11]

# Sparse reconstruction of hyperspectral data: Summary

Bad news: Hyperspectral libraries have poor RI constants

Good news: Hyperspectral mixtures are highly sparse, very often $k \leq 5$

Surprising fact: Convex programs (BP, BPDN, LASSO, …) yield useful better empirical in regimes for which we don't have recovery guarantees

Directions to improve hyperspectral sparse reconstruction

- Structured sparsity + subspace structure
  (pixels in a give data set share the same support)
- Spatial contextual information (pixels belong to an image)

# Constrained colaborative sparse regression (CCSR)

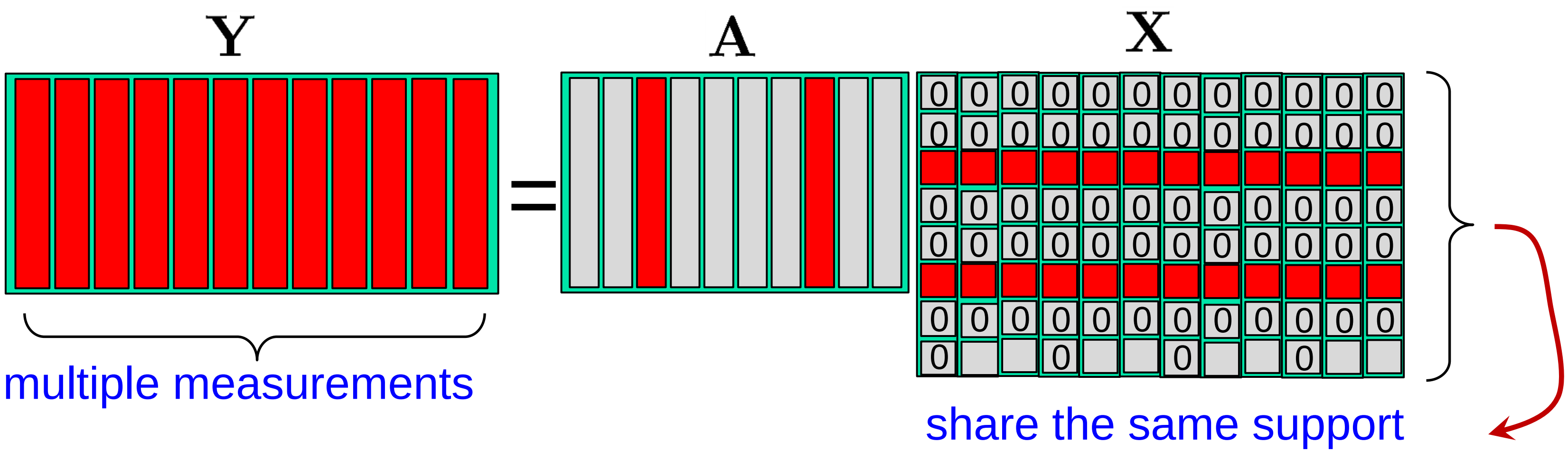
$$\min_{\mathbf{X}} (1/2)\|\mathbf{AX}-\mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}\|_{2,1}$$

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

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

[Iordache, Bioucas, Plaza, 11, 12]

[Turlach, Venables, Wright, 04]




Theoretical guaranties (superiority of multichanel) : the probability of recovery failure decays exponentially in the number of channels. [Eldar, Rauhut, 11]

# Constrained total variation sparse regression (CTVSR)

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

[Iordache, B-D, Plaza, 11]

$$\text{subject to: } \mathbf{X} \geq \mathbf{0}$$

Total Variation of i-th band

$$\phi_{TV}(\mathbf{X}) := \sum_{i=i}^{n} \|\mathbf{L}\mathbf{x}^i\|_1 = \sum_{i=1}^{n} \sum_{k=1}^{N} \sqrt{([\mathbf{D}_h \mathbf{x}^i]_k)^2 + ([\mathbf{D}_v \mathbf{x}^i]_k)^2}$$

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

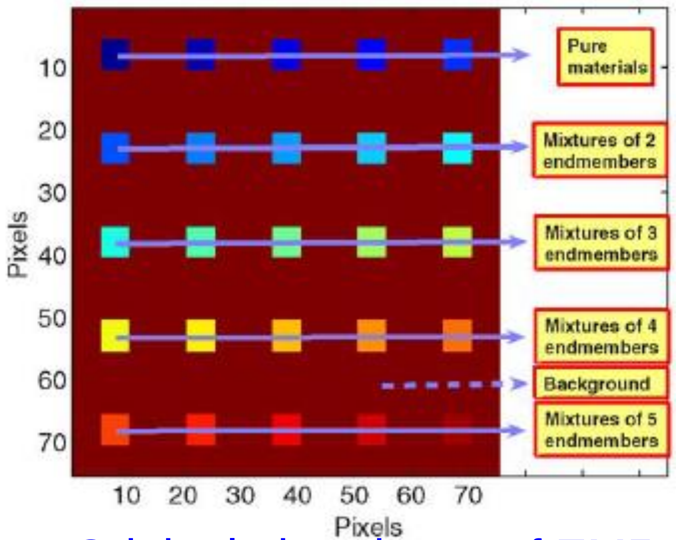
Other Regularizers:

- vector total variation (VTV)$\rightarrow$ promotes piecewise smoo 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

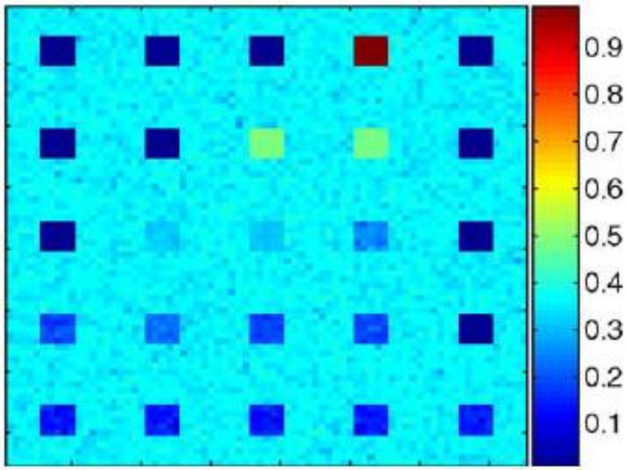
# Ilustrative examples with simulated data : SUnSAL-TV

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

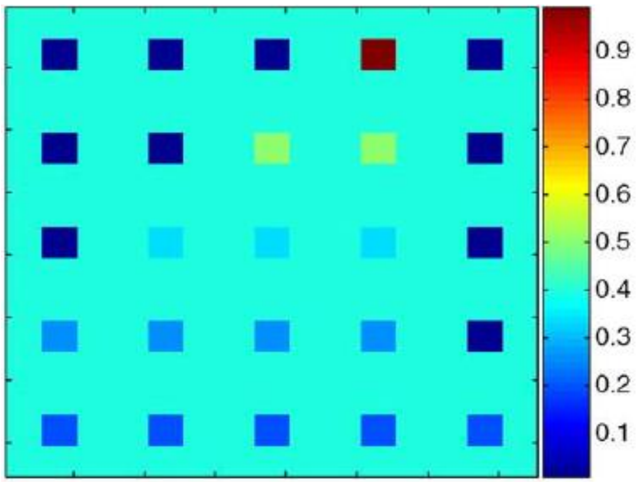
Original data cube



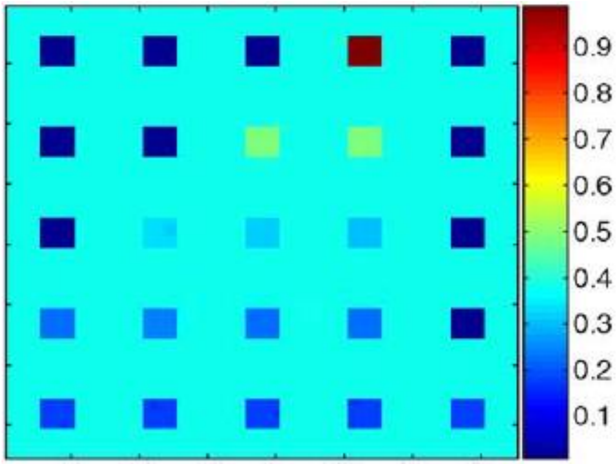

SUnSAL estimate




Original abundance of EM5




SUnSAL-TV estimate



# MUSIC – Colaborative SR algorithm

MUSIC-CSR algorithm [Iordache, Bioucas, Plaza, 13]

1) Estimate the signal subspace $\mathrm{span}\{\mathbf{A}_S\}$using, e.g. the HySime algorithm.

2) Compute $\varepsilon_i = \dfrac{\|\mathbf{P}_y^{\perp}\mathbf{a}_i\|}{\|\mathbf{a}_i\|}$, for $i = 1, \ldots, m$ and define

   the index set $S = [i : \varepsilon_i \leq \delta, i = 1, \ldots, m]$

3) Solve the colaborative sparse regression optimization

$$\min_{\mathbf{X}} (1/2)\|\mathbf{Y} - \mathbf{A_S}\mathbf{X}\|^2 + \lambda\|\mathbf{X}\|_{2,1}, \quad \mathbf{X} \geq 0$$

[Bioucas, Figueiredo,12]

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

(N < k and iid noise)

# Brief Concluding remarks

- HU is a hard inverse problem (noise, bad-conditioned direct operators, nonlinear mixing phenomena)
- HU calls for sophisticated math tools and frameworks (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