Hyperspectral Unmixing Via Sparse Regression Optimization Problems and Algorithms

José M. Bioucas-Dias

Instituto de Telecomunicações
Instituto Superior Técnico
Technical University of Lisbon

Portugal

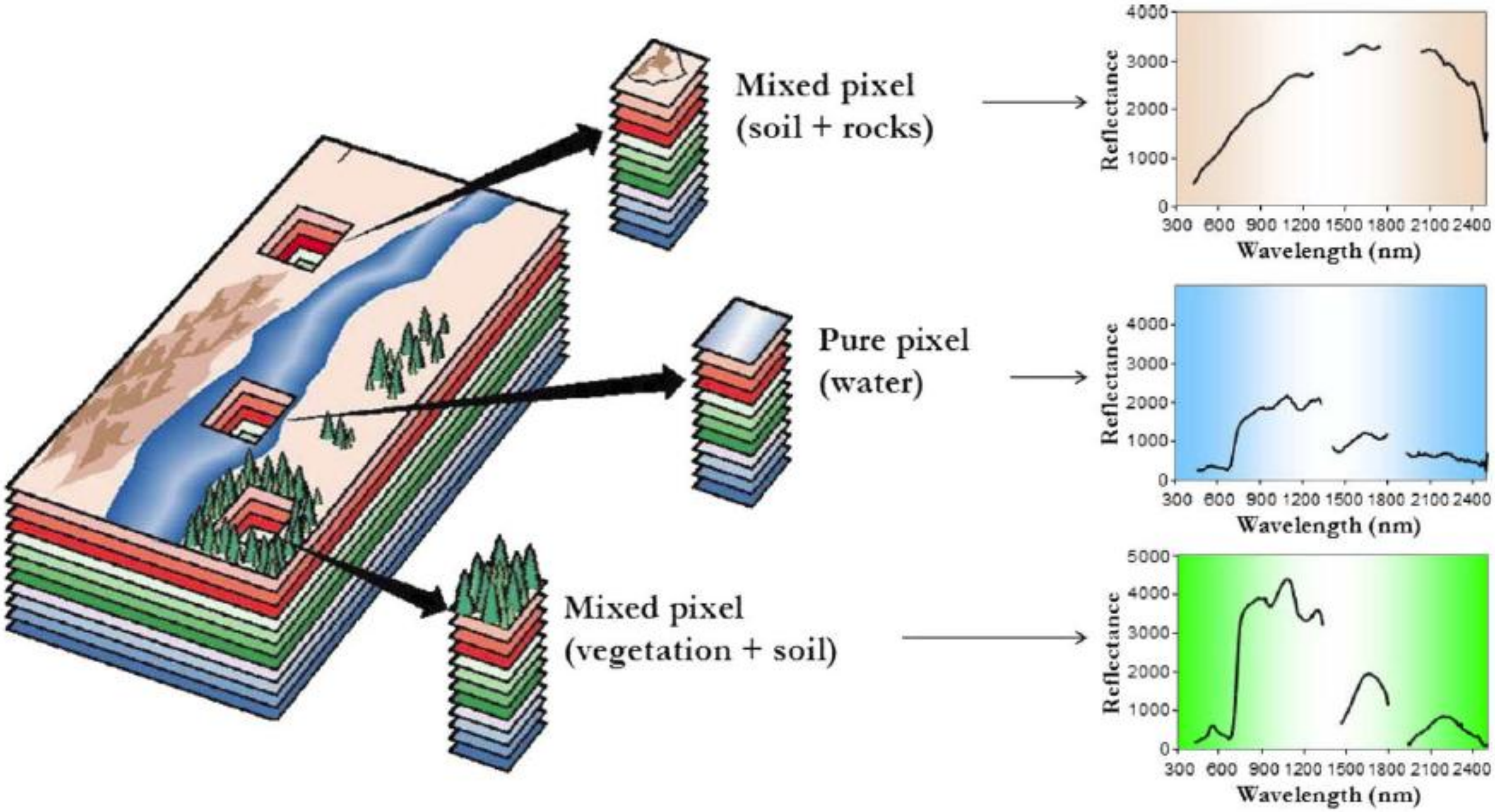
Joint work with *Mário A. T. Figueiredo*

INSTITUTO SUPERIOR TÉCNICO

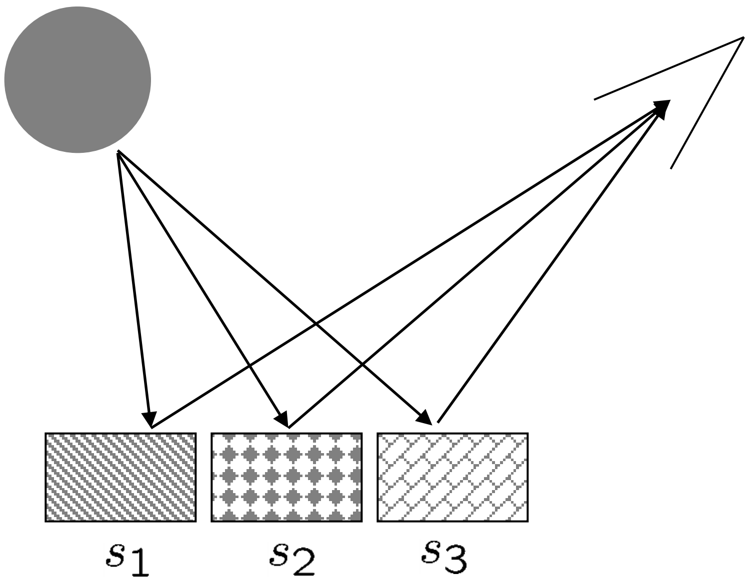
Outline

- Introduction to hyperspectral unmixing
- Hyperspectral unmixing via sparse regression.
- Recovery guarantees/convex and nonconvex solvers
- Improving hyperspectral sparse regression
  - structured sparsity
  - dictionary pruning
- Solving convex sparse hyperspectral unmixing and related convex inverse problems with ADMM
- Concluding remarks

Hyperspectral imaging (and mixing)



Linear mixing model (LMM)



$$\mathbf{y} = \sum_{i=1}^{p} s_i \boldsymbol{\rho}_i \qquad \boldsymbol{\rho}_i = \begin{bmatrix} \rho_{1i} \\ \rho_{2i} \\ \vdots \\ \rho_{mi} \end{bmatrix}$$

$$\mathbf{y} = \mathbf{M}\mathbf{s}$$

$$\mathbf{s} \equiv \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}$$

$$\mathbf{M} \equiv [\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \boldsymbol{\rho}_3]$$

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

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

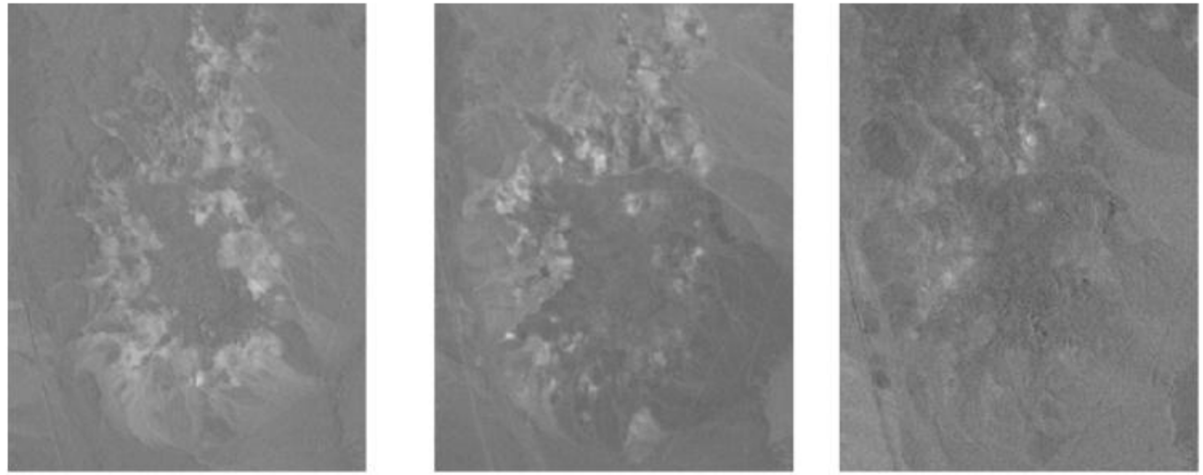
Hyperspectral unmixing

AVIRIS of Cuprite, Nevada, USA

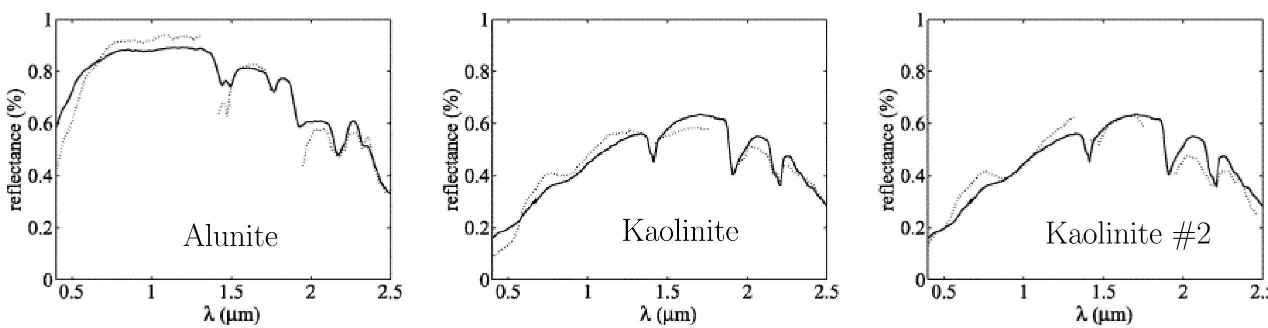
R – ch. 183 (2.10 μm)

G – ch. 193 (2.20 μm)

B – ch. 207 (2.34 μm)



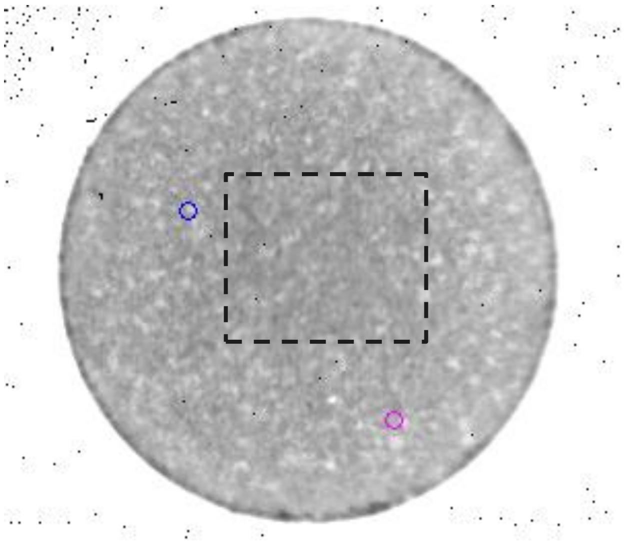
(a) (b) (c)

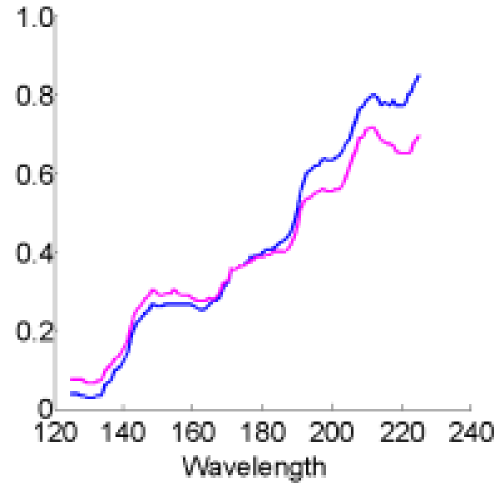


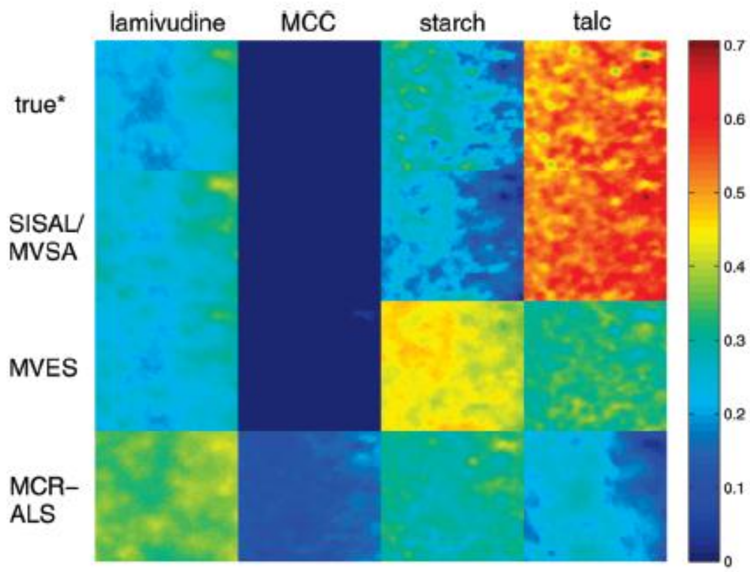
(a) (b) (c)

VCA [Nascimento, B-D, 2005]

NIR tablet imaging







[Lopes et al., 2010]

Spectral linear unmixing (SLU)

Given *N* spectral vectors of dimension *m:* $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$

Subject to the LMM: $\mathbf{Y} = \mathbf{M}\mathbf{S} + \mathbf{N}$, $\boldsymbol{S} \geq 0$, $\mathbf{1}_p^T \mathbf{S} = \mathbf{1}_N$

ANC: abundance nonnegative constraint

ASC: abundance sum-to-one constraint

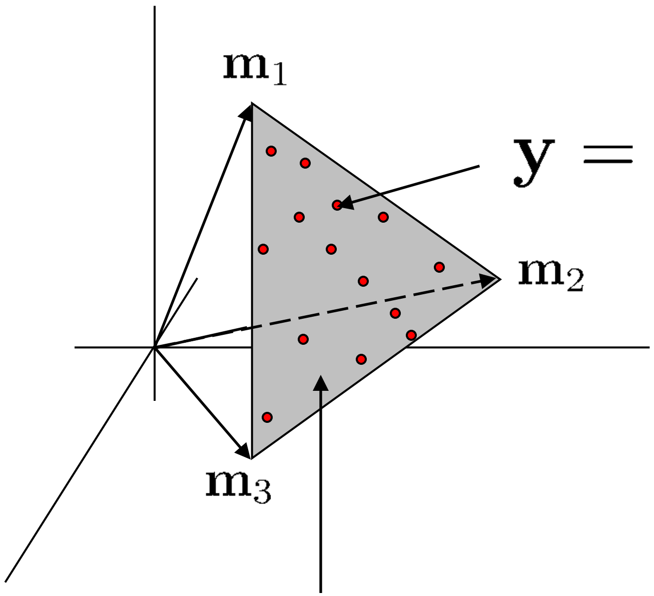
Determine:

- ❑ The mixing matrix $\mathbf{M}$ (*endmember spectra*)
- ❑ The *fractional abundance* vectors $\boldsymbol{S}$,

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

Geometrical view of SLU

$$\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_p]$$



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

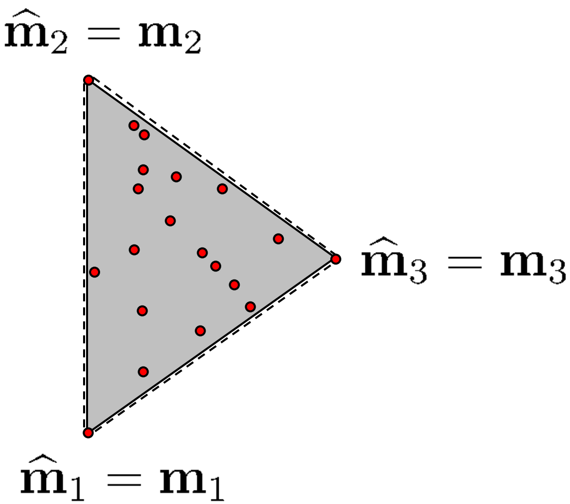
probability simplex (S_I)

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

Inferring **M** $\Leftrightarrow$ inferring the vertices of the simplex S_M

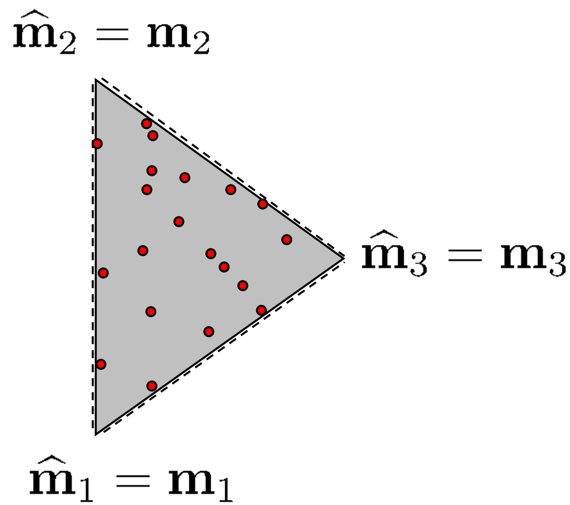
Minimum volume simplex (MVS)

Pure pixels



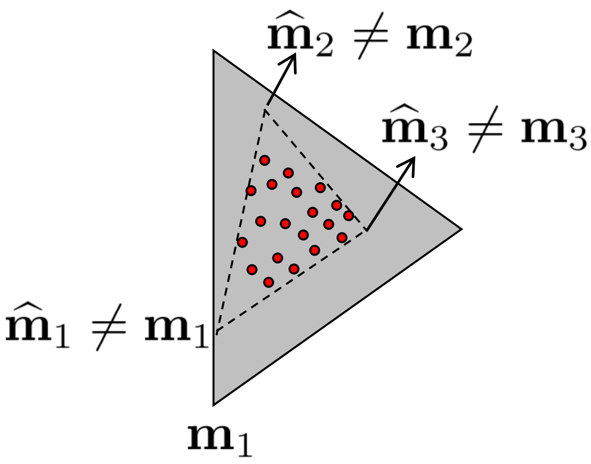
MVS works

No pure pixels



MVS works

No pure pixels



MVS does not work

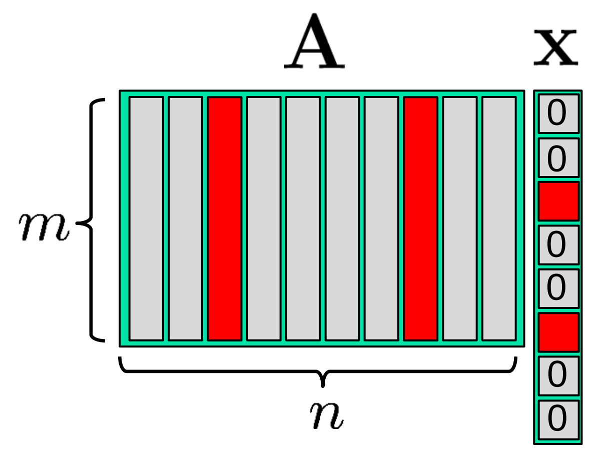
[B-D et al., IEEE JSTATRS, 2012]

Sparse regression-based SLU

- Key observation: Spectral vectors can be expressed as linear combinations of a few pure spectral signatures obtained from a (potentially very large) spectral library

[Iordache, B-D, Plaza, 11, 12]

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



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

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

- Advantage: sidesteps endmember estimation

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

Sparse reconstruction/compressive sensing

Key result: a sparse signal is exactly recoverable from an underdetermined linear system of equations in a computationally efficient manner via convex/nonconvex programming

[Candes, Romberg, Tao, 06] [Candes,Tao, 06] [Donoho, Tao, 06] [Blumensath, Davies, 09]

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$, $m < n$, and $\mathbf{x}^*$, such that $\mathbf{A}\mathbf{x}^* = \mathbf{y}$

$\mathbf{x}^*$ is the unique solution of $\mathbf{A}\mathbf{x} = \mathbf{y}$ if $2\|\mathbf{x}^*\|_0 < \text{spark}(\mathbf{A})$

$\mathbf{x}^*$ is the solution of the optimization problem

$$(P_0) \quad \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{y}$$

NP-hard

Sparse reconstruction/compressive sensing

Optimization strategies to cope with P_0 NP-hardness

Convex relaxation

(BP – Basis Pursuit) [Chen et al., 2001]

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ st: } \mathbf{A}\mathbf{x} = \mathbf{y}$$

(BPDN – BP denoising)

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ st: } \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \delta$$

(LASSO) [Tibshirani, 1996]

$$\min_{\mathbf{x}} (1/2)\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

Approximation algorithms

Bayesian CS [Ji et al., 2008]

CoSaMP – Compressive Sampling Matching Pursuit [Needell, Tropp, 2009]

IHT – Iterative Hard Thresholding [Blumensath, Davies, 09]

GDS - Gradient Descent Sparsification [Garg Khandekar, 2009]

HTP – Hard Thresholding Pursuit [Foucart, 10]

MP - Message Passing [Villa Schniter, 2012]

Exact recovery of sparse vectors

Recovery guarantees: linked with the restricted isometric property (RIP)

Restricted isometric constant: $\delta_p(\mathbf{A}), \quad \mathbf{A} \in \mathbb{R}^{m \times n}$:

$$(1-\delta)\|\mathbf{x}\|_2 \leq \|\mathbf{A}\mathbf{x}\| \leq (1+\delta)\|\mathbf{x}\|_2, \quad \|\mathbf{x}\|_0 \leq p$$

Many SR algorithms ensure exact recovery provided that:

$$\delta_t(\mathbf{A}) \leq \delta_*, \text{ for some t and } \delta_*$$

This condition is satisfied for random matrices provided that

$$m \simeq c\frac{t}{\delta_*^2}\log(n/t)$$

Algorithm	BP	HTP	CoSaMP	GDP	IHT
$\delta_t < \delta_*$	$\delta_{2s} < 0.465$	$\delta_{3s} < 0.577$	$\delta_{4s} < 0.384$	$\delta_{2s} < 0.333$	$\delta_{3s} < 0.555$
Ratio t/δ_*^2	9.243	**9**	27.08	18	12

(from [Foucart, 10])

Example: Gaussian matrices; signed signals

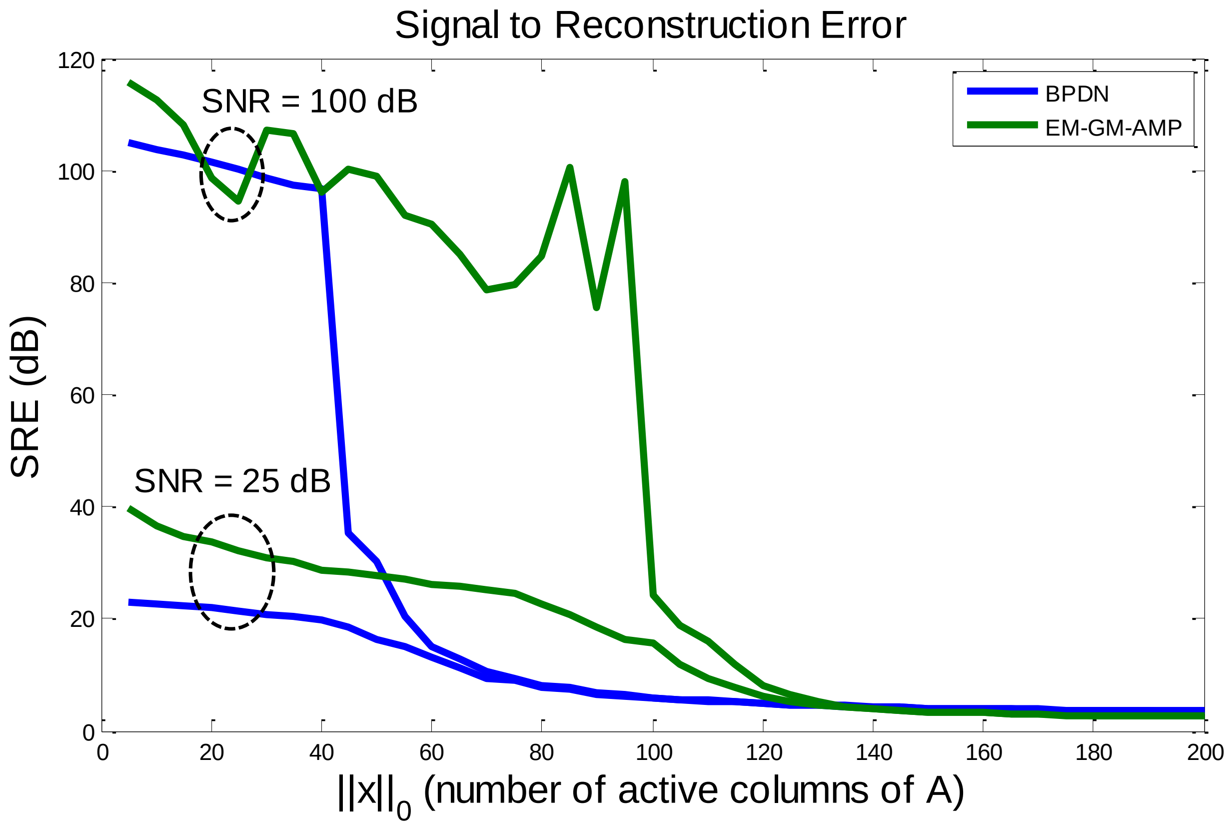
$$\mathbf{A} \in \mathbb{R}^{m\times n}; \quad (m = 200, n = 400); \quad \mathcal{N}(0,1); \text{ iid}; \quad x_i \sim \mathcal{N}(0,1)$$

Algorithms:

(BPDN – SUnSAL) [B-D, Figueiredo, 2010]

$$\min_{\mathbf{x}} \ \|\mathbf{x}\|_1 \ \text{ st: } \ \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \delta$$

EM-BG-AMP
[Villa Schniter, 2012]



Example: Gaussian matrices; non-negative signals

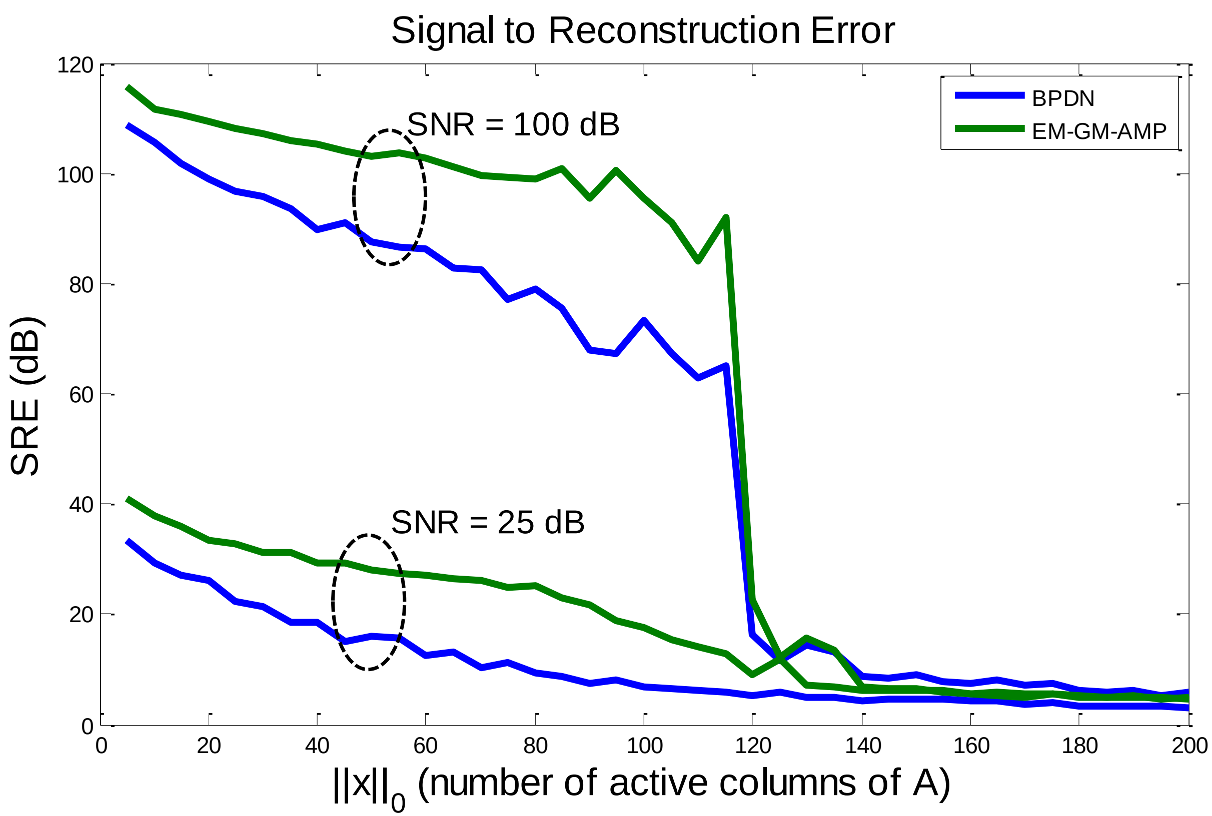
$\mathbf{A} \in \mathbb{R}^{m \times n}$; $\mathcal{N}(0,1)$ iid ; $m = 200, n = 400$; $x_i \sim$ UD in the simplex

Algorithms:

(BPDN – SUnSAL)

$$\min_{\mathbf{x}} \ \|\mathbf{x}\|_1 \ \text{ st: } \ \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \le \delta, \ \mathbf{x} \ge \mathbf{0}$$

EM-BG-AMP



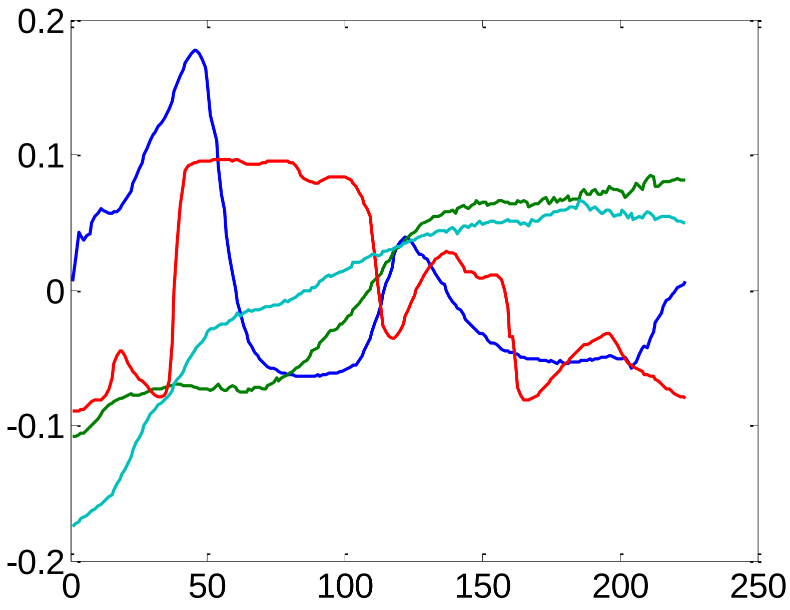
Hyperspectral libraries

Hyperspectral libraries exhibit poor RI constants

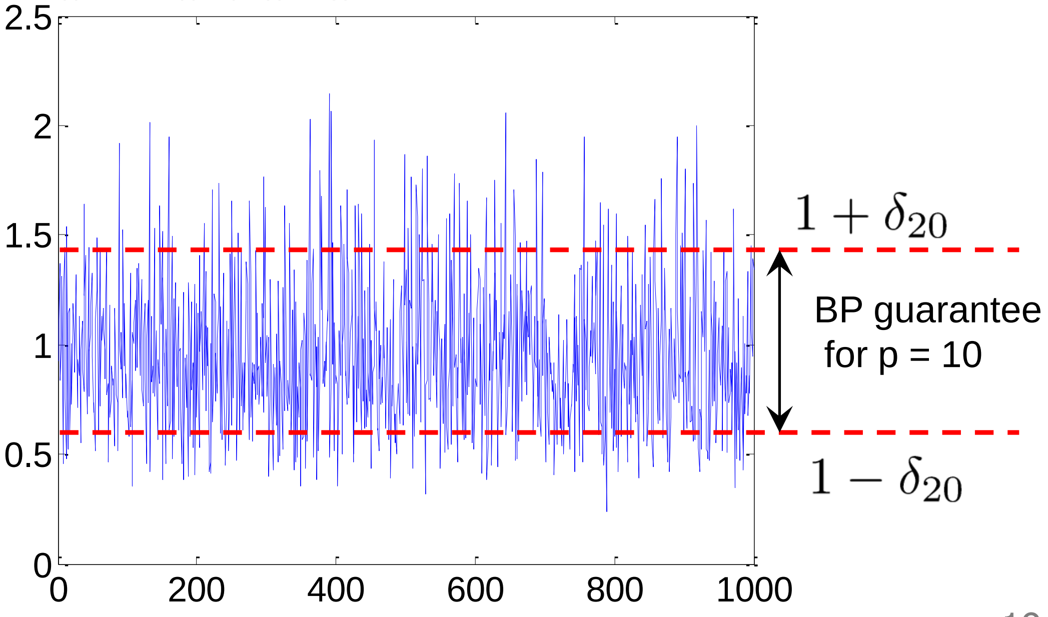
(Mutal coherence close to 1 [Iordache, B-D, Plaza, 11, 12])

Illustration: $\mathbf{A} \in \mathbb{R}^{m \times n}$; subset from USGS; $m = 200, n = 400$

Normalized and mean removed



$\|\mathbf{A}\mathbf{x}\|_2/\|\mathbf{x}\|_2$ $\|\mathbf{x}\|_0 = 10$



Example: Hyperspectral library; non-negative signals

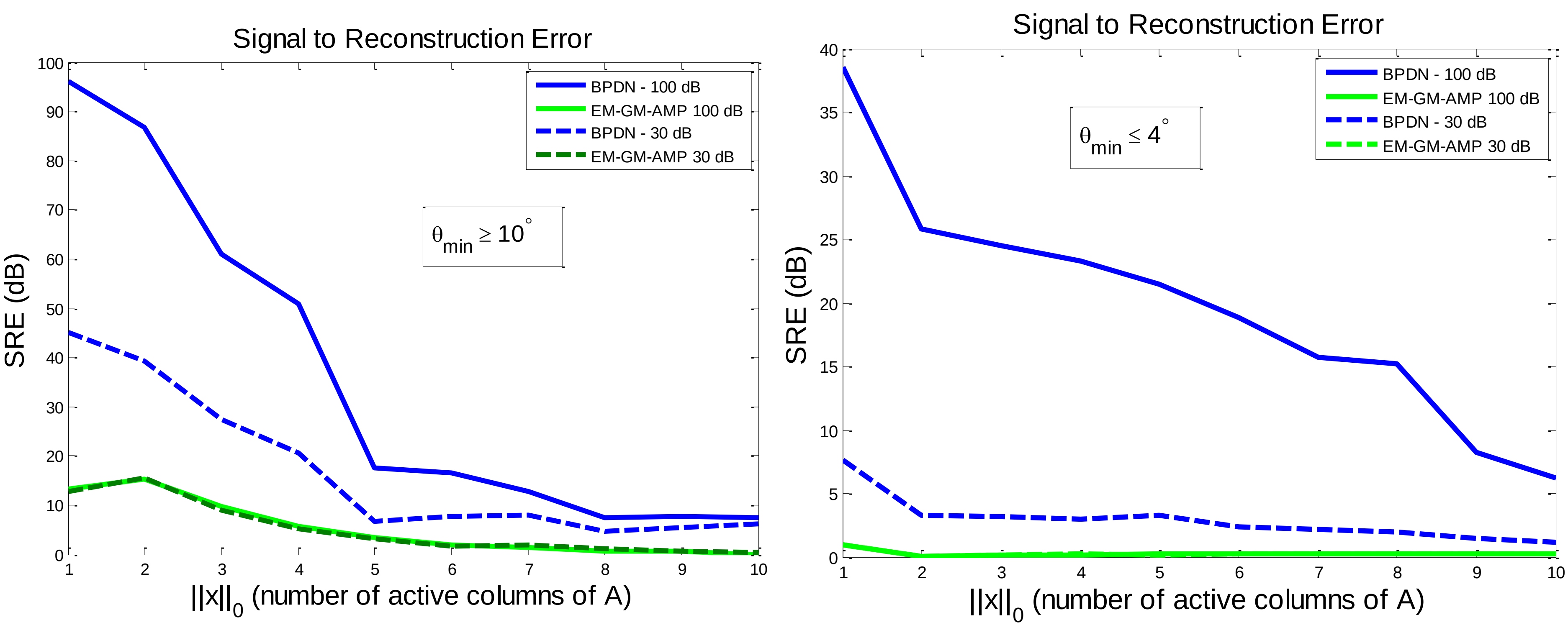
Illustration: $\mathbf{A} \in \mathbb{R}^{m \times n}$; subset from USGS; $m = 200, n = 300$

Algorithms:

(BPDN – SUnSAL)

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

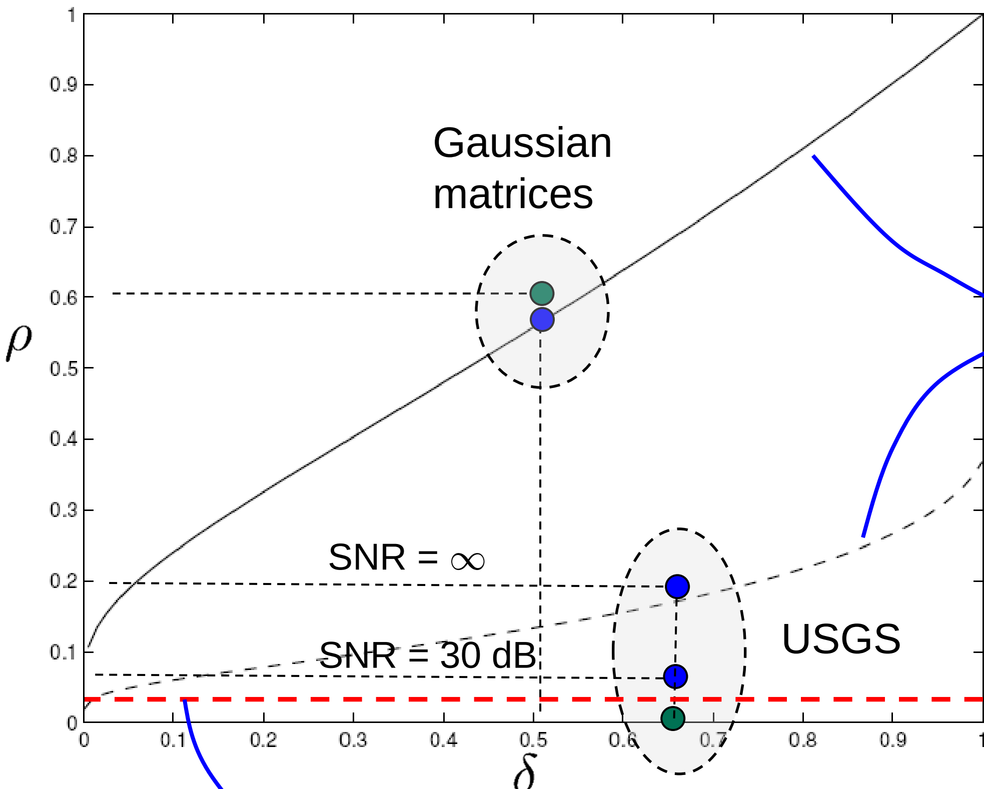
EM-BG-AMP



Phase transition curves

$\ell_1, -/\ell_0$ equivalence for non-negative signals $(\mathbf{A} \in \mathbb{R}^{m \times n})$

$\delta = m/n$ undersamplig factor $\rho = \|\mathbf{x}\|_0/m$ fractional sparsity



EM-BG-AMP

BPDN

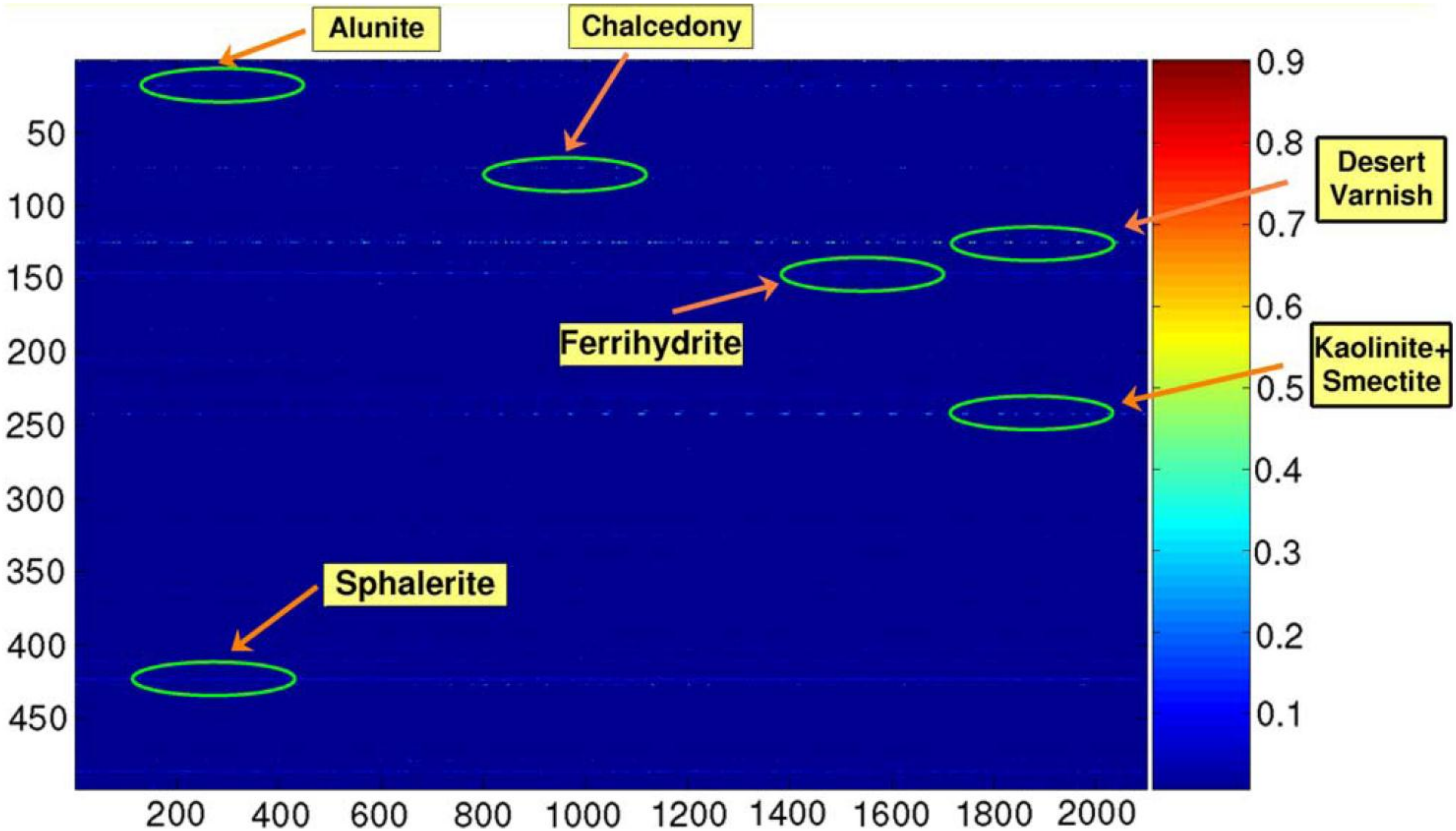
$\mathbf{A}$ - orthogonal projector

Results linked with the k-neighborliness property of polytopes

[Donoho, Tanner, 2005]

$\rho = 0.025$ (enough for many hyperspectral applications)

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 $p \leq 5$

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

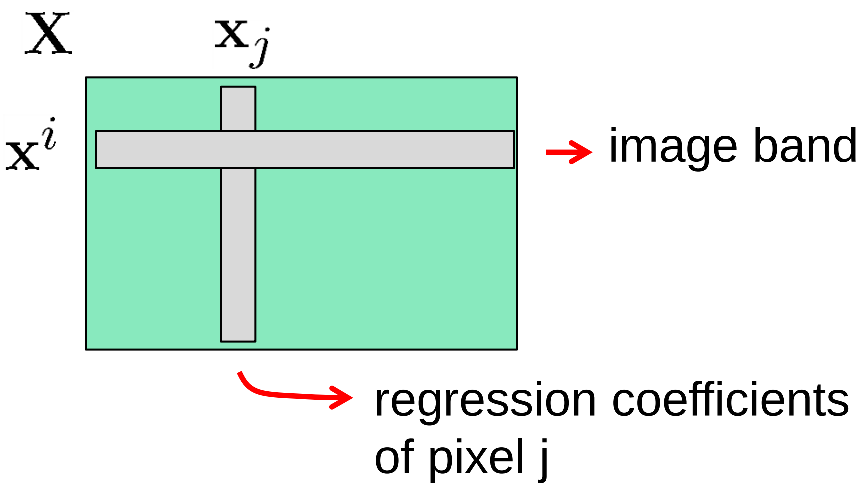
- Structured sparsity + subspace structure
  (pixels in a give data set share the same support)
- Spatial contextual information
  (pixels belong to an image)

Beyond l_1 pixelwise regularization

Rationale: introduce new sparsity-inducing regularizers to counter the sparse regression limits imposed by the high coherence of the hyperspectral libraries.

Let's rewrite the LMM as $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N}$

$$\begin{cases} \mathbf{Y} \in \mathbb{R}^{m\times N} \\ \mathbf{A} \in \mathbb{R}^{m\times n} \\ \mathbf{X} \in \mathbb{R}^{n\times N} \end{cases}$$



Constrained total variation sparse regression (CTVSR)

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

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

[Iordache, B-D, Plaza, 11]

total Variation of i-th band

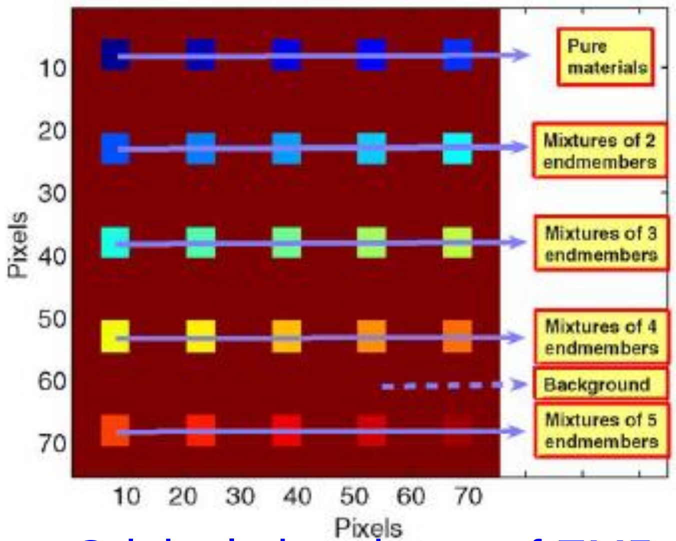
$$\phi_{TV}(\mathbf{X}) := \sum_{i=i}^{n} \|\mathbf{L}\mathbf{x}^i\|_1 = \sum_{i=1}^{n}\sum_{k=1}^{N}\sqrt{([\mathbf{D}_h\mathbf{x}^i]_k)^2 + ([\mathbf{D}_v\mathbf{x}^i]_k)^2}$$

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

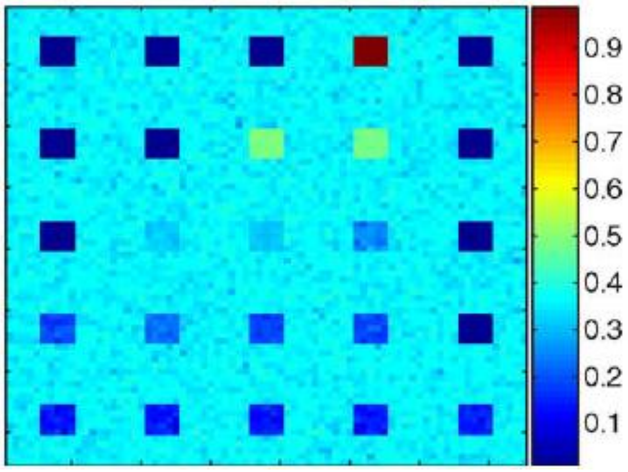
Ilustrative examples with simulated data : SUnSAL-TV

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

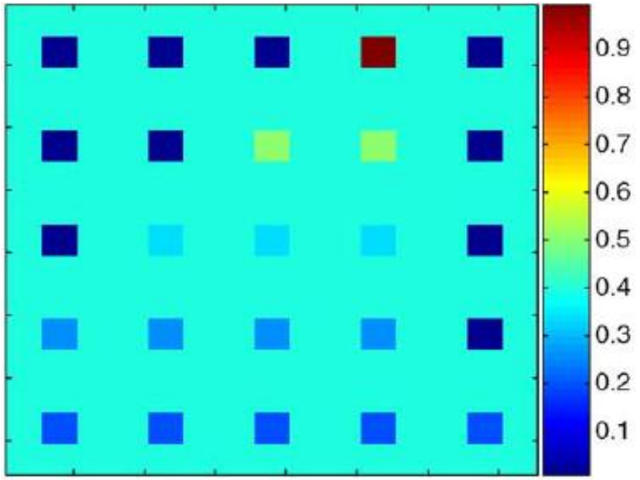
Original data cube



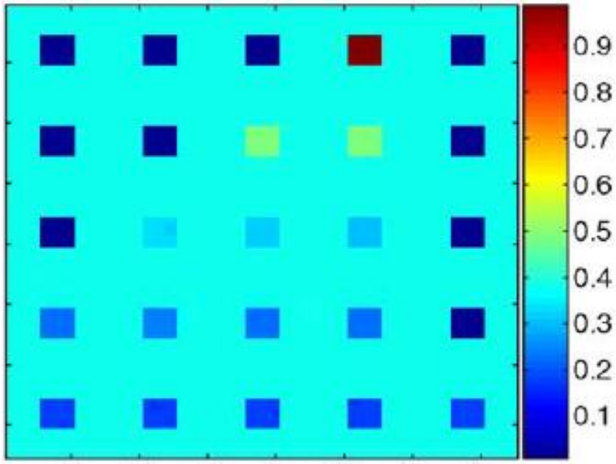
SUnSAL estimate



Original abundance of EM5



SUnSAL-TV estimate



Constrained colaborative sparse regression (CCSR)

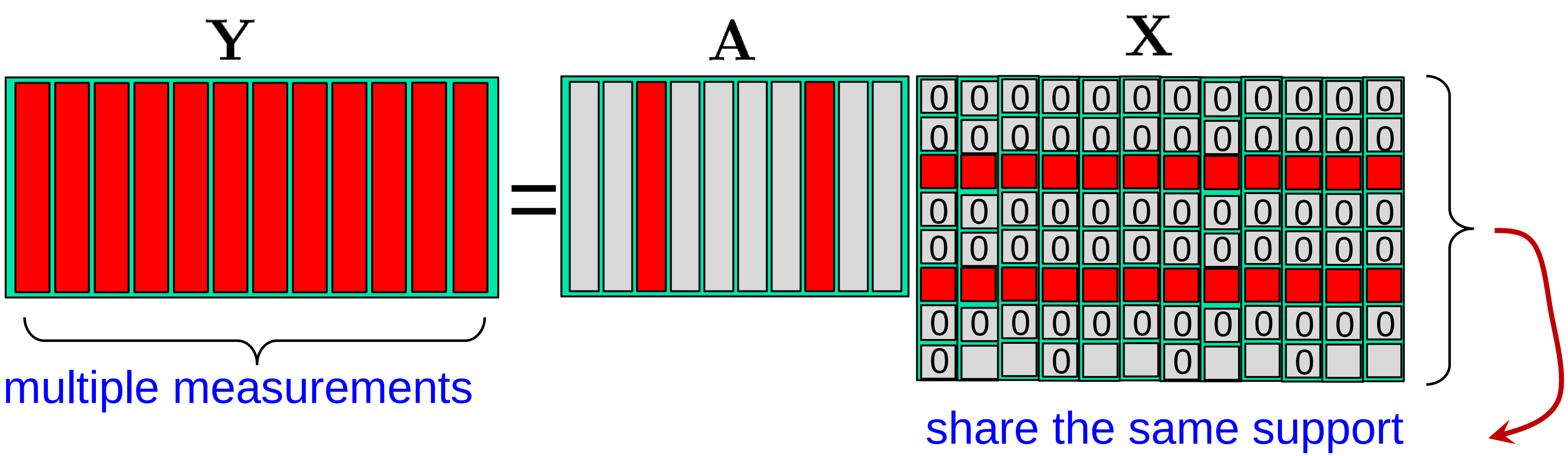
$$\min_{\mathbf{X}} \ (1/2)\|\mathbf{A}\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}\|_{2,1}$$

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

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

[Iordache, B-D, Plaza, 11, 12]

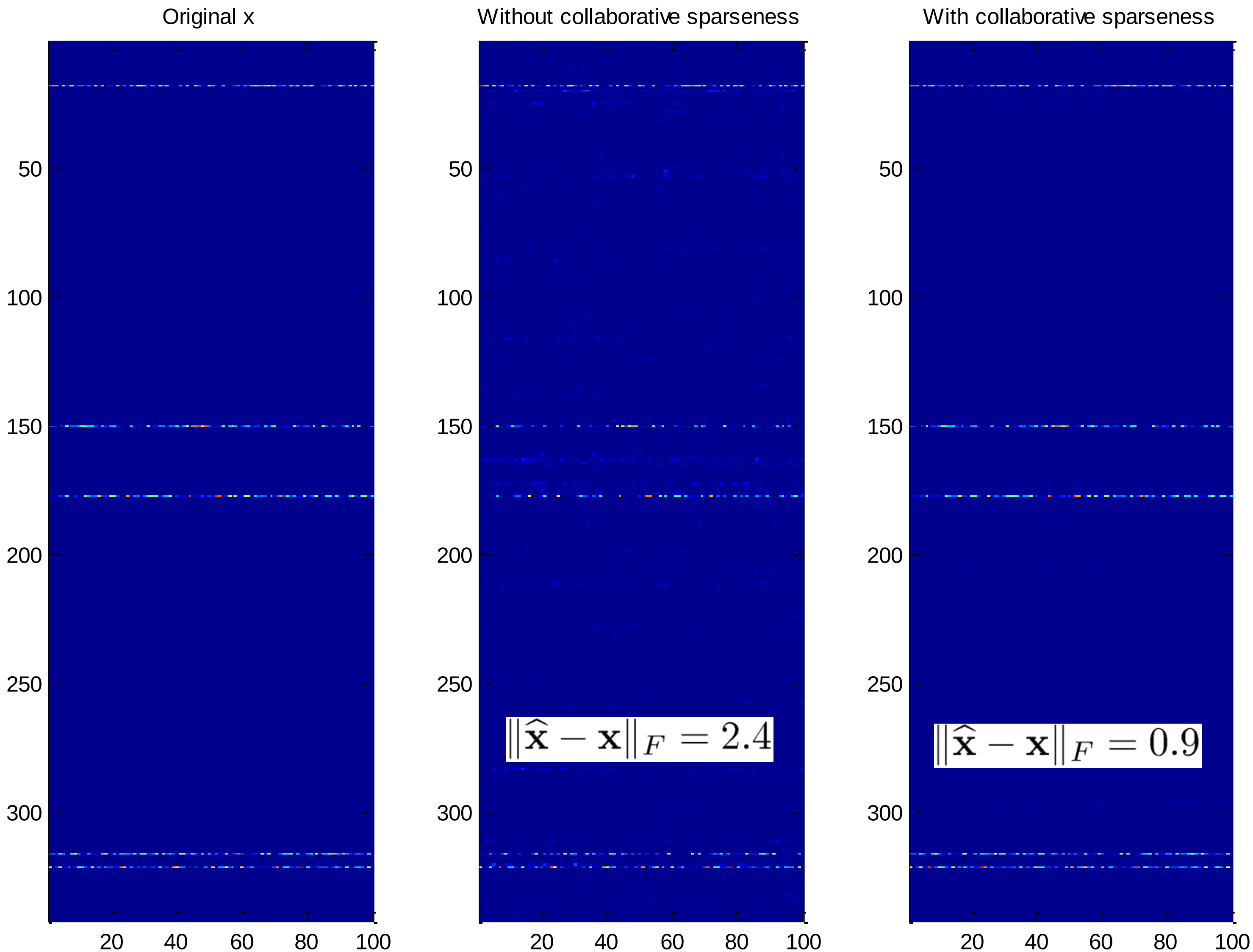
[Turlach, Venables, Wright, 2004]



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

Ilustrative examples with simulated data : CSUnSAL

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



SNR = 35dB

time = 10 sec

Multiple measurements

The multiple measurement vector (MMV) problem

$$\text{minimize } \|\mathbf{X}\|_0$$

$$\text{subject to: } \mathbf{Y} = \mathbf{A}\mathbf{X}$$

$\|\mathbf{X}\|_0$ - number of non-null rows of $\mathbf{X}$

MMV has a unique solution iff

[Feng, 1997], [Chen, Huo, 2006], [Davies, Eldar, 2012]

$$\|\mathbf{X}\|_0 < \frac{\text{spark}(\mathbf{A}) + \text{rank}(\mathbf{Y}) - 1}{2}$$

MMV gain

If $\text{rank}(\mathbf{Y}) = \|\mathbf{X}\|_0$, the above bound is achieved using the multiple sinal classification (MUSIC) algorithm

Endemember identification with MUSIC

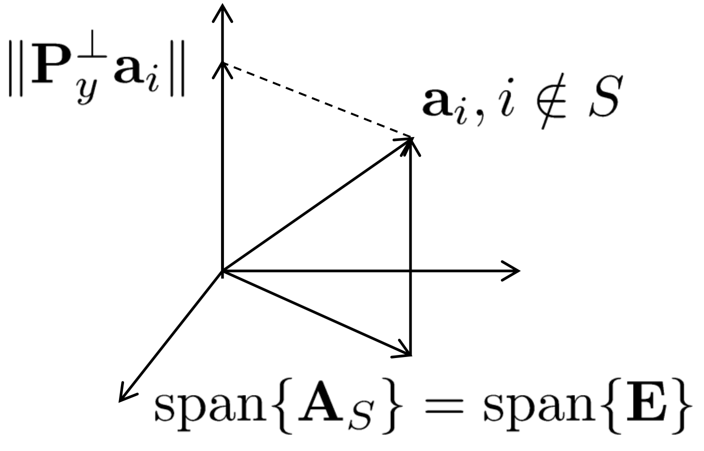
$\mathbf{Y} = \mathbf{A}\mathbf{X}$ (noiseless measurements)

MUSIC algorithm

1) Compute $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_p]$, the first *p* eigenvalues of $\mathbf{R}_y = \mathbf{Y}\mathbf{Y}^T/n$

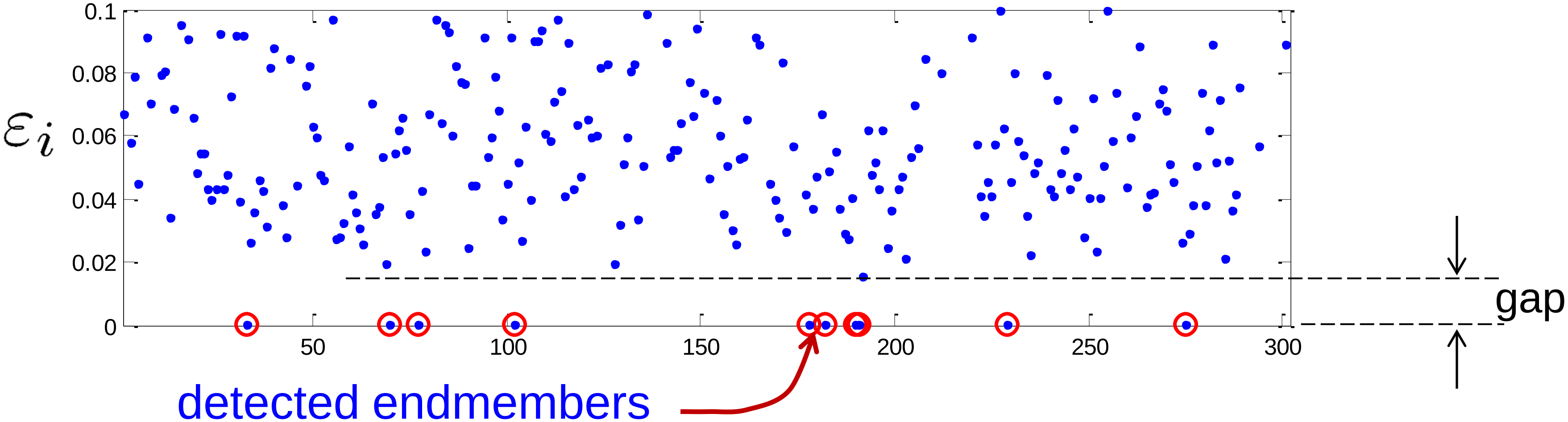
2) Compute $\varepsilon_i = \dfrac{\|\mathbf{P}_y^{\perp}\mathbf{a}_i\|}{\|\mathbf{a}_i\|}$, for $i = 1, \ldots, m$ and set

$\mathbf{M} = \mathbf{A}_S$ with $S = \{i \,:\, \varepsilon_i = 0, i = 1, \ldots, m\}$



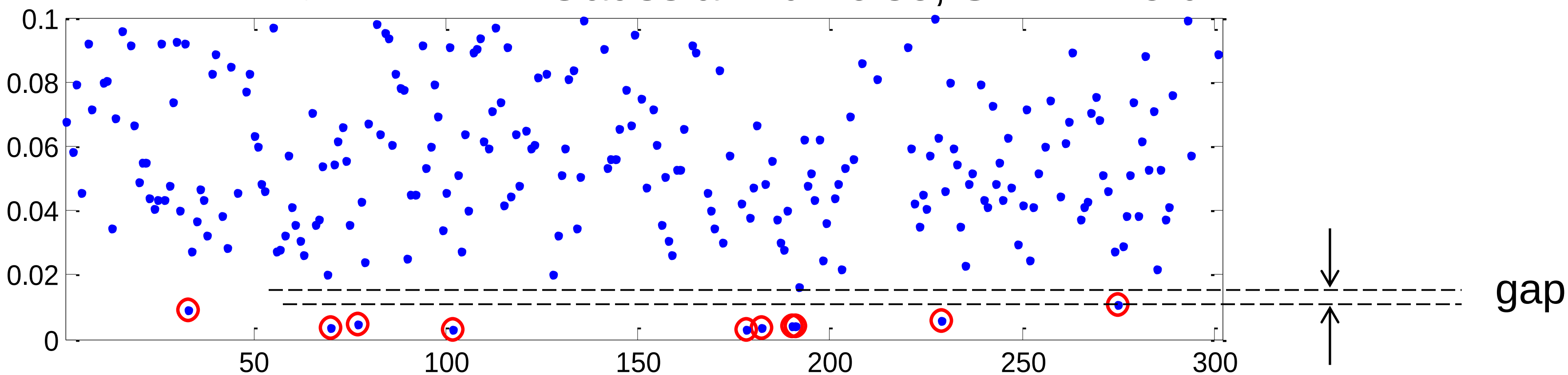
Examples (simulated data)

A – USGS ($\geq 3^\circ$) p = 10, N=5000, X - uniform over the simplex, SNR = ∞,



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

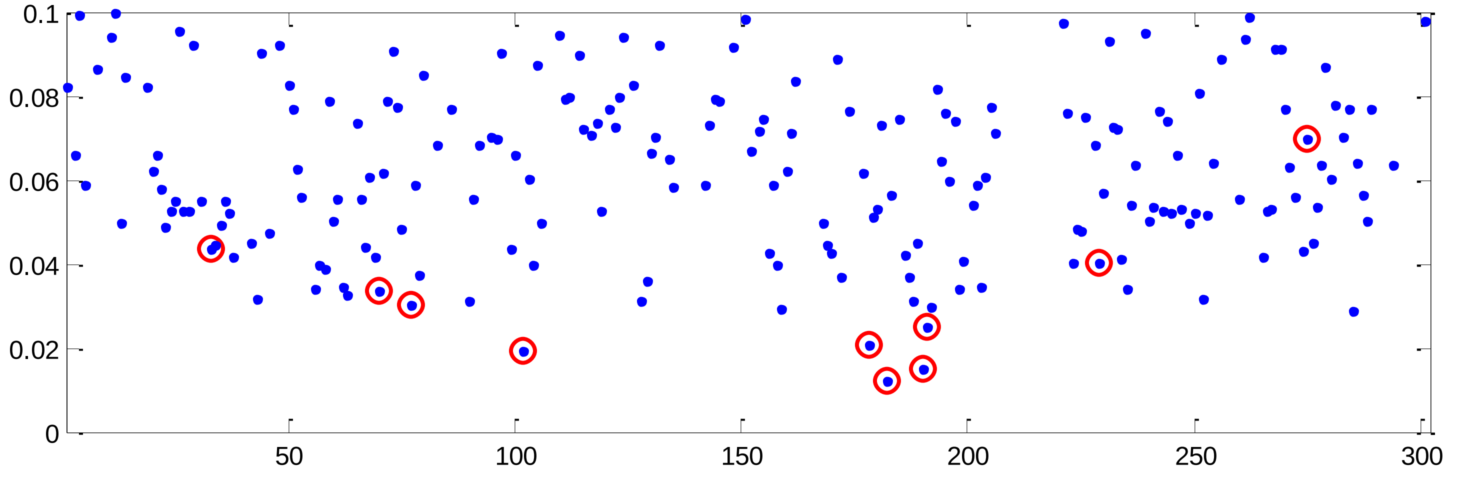
Gaussian iid noise, SNR = 25 dB



Examples (simulated data)

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N}$$

colored noise, SNR = 25 dB



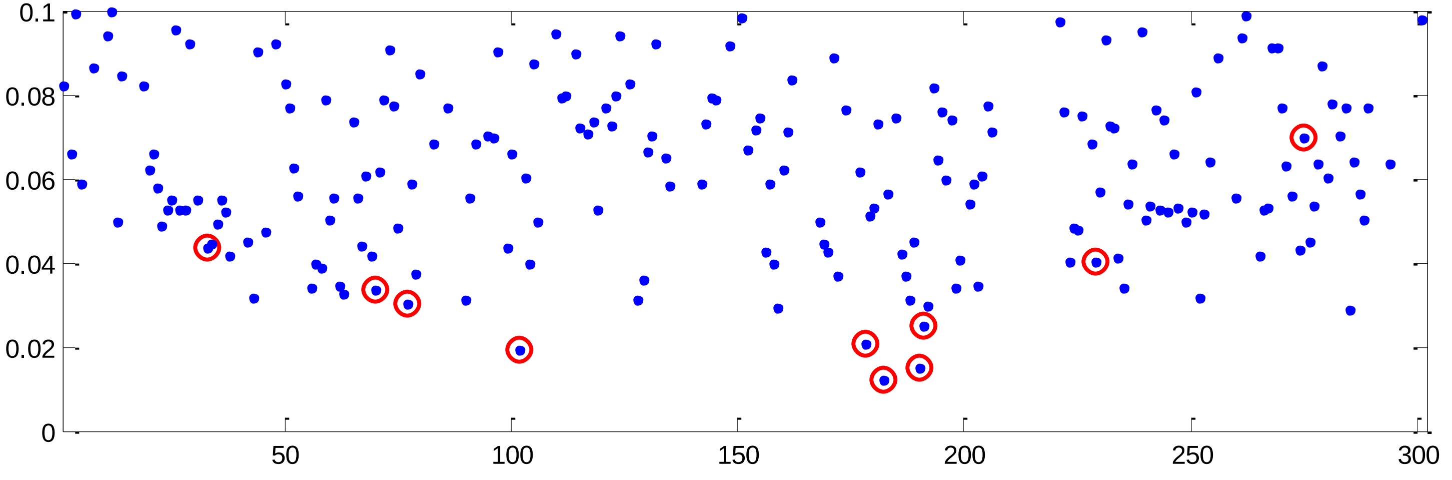
cause of the large projection errors: poor identification of the subspace signal

$$\mathbf{R}_y \simeq \mathbf{A}_S \mathbf{R}_x \mathbf{A}_S^T + \mathbf{R}_n \quad (\mathbf{R}_n \neq \sigma^2 \mathbf{I}) \quad \Rightarrow \text{span}\{\mathbf{E}\} \neq \text{span}\{\mathbf{A}_S\}$$

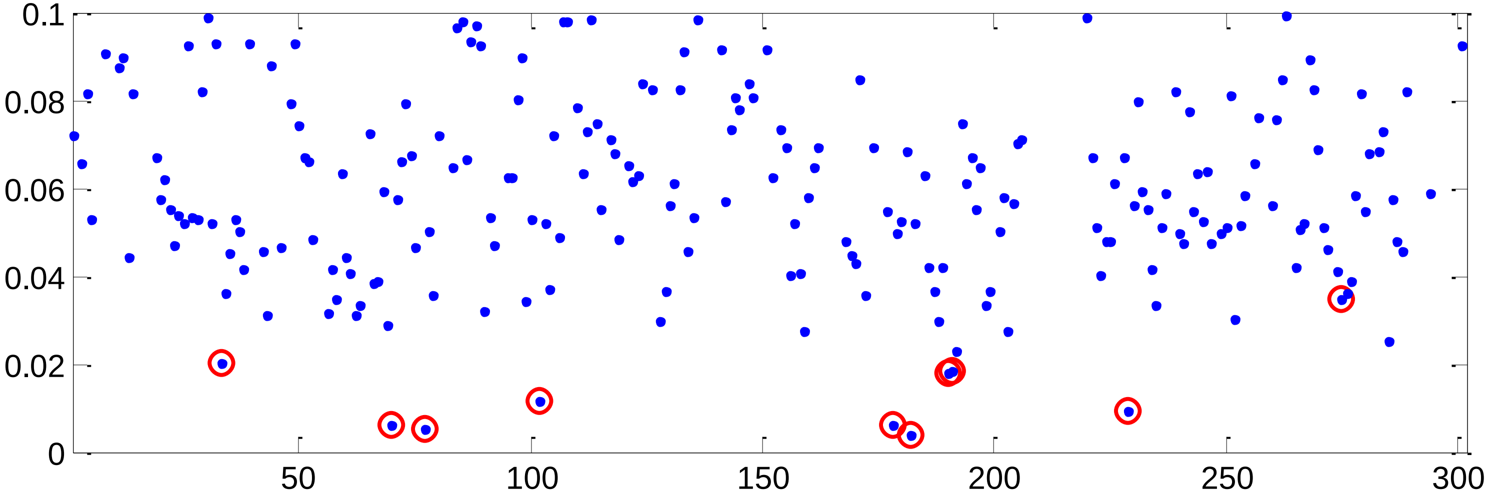
cure: identify the signal subspace

Signal subspace identification

colored noise, SNR = 25 dB (incorrect signal subspace)



colored noise, SNR = 25 dB (signal subspace identified with HySime, [BD, Nascimento, 2008])



Proposed MUSIC – Colaborative SR algorithm

MUSIC-CSR algorithm [B-D, 2012]

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

2) Compute $\varepsilon_i = \frac{\|\mathbf{P}_y^{\perp}\mathbf{a}_i\|}{\|\mathbf{a}_i\|}$, 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

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

[B-D, Figueiredo, 2012]

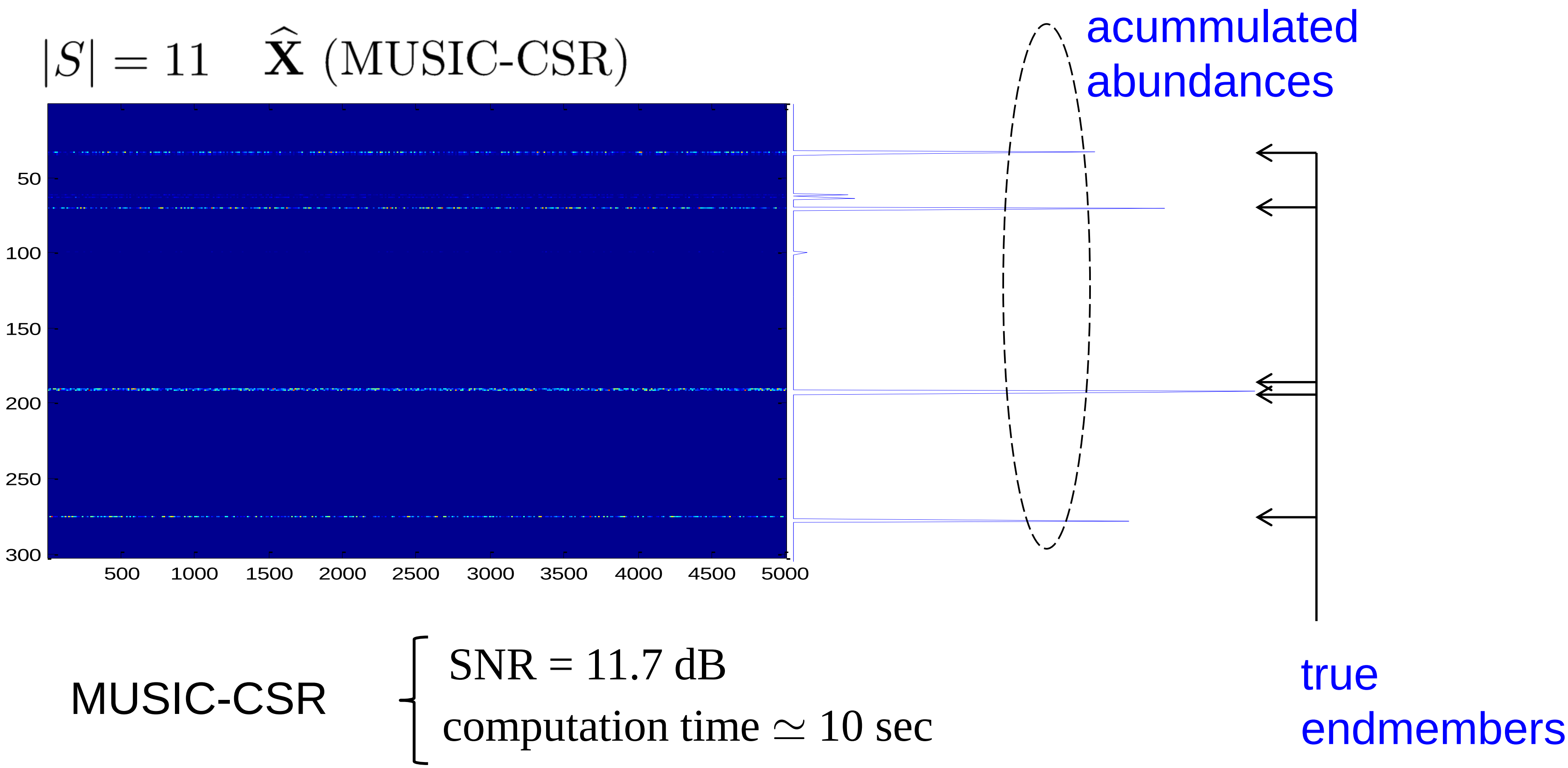
Related work: CS-MUSIC [Kim, Lee, Ye, 2012]

(N < k and iid noise)

MUSIC – CSR results

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

$|S| = 11 \quad \widehat{\mathbf{X}}$ (MUSIC-CSR)

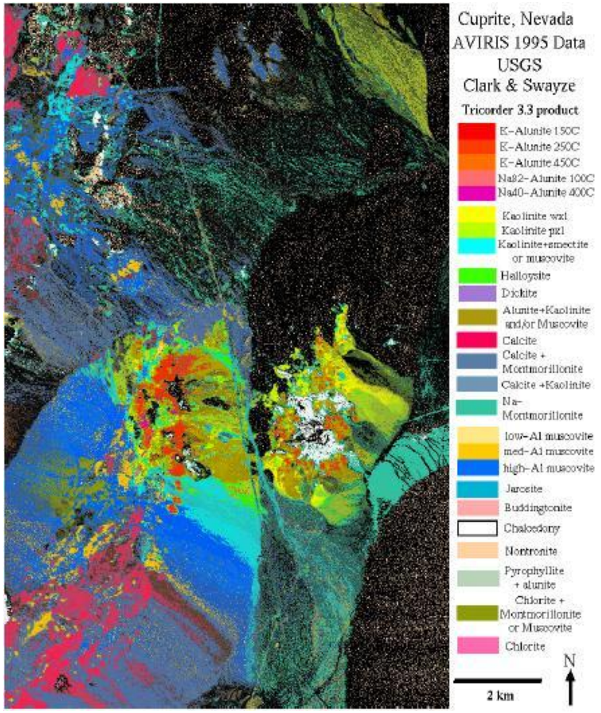


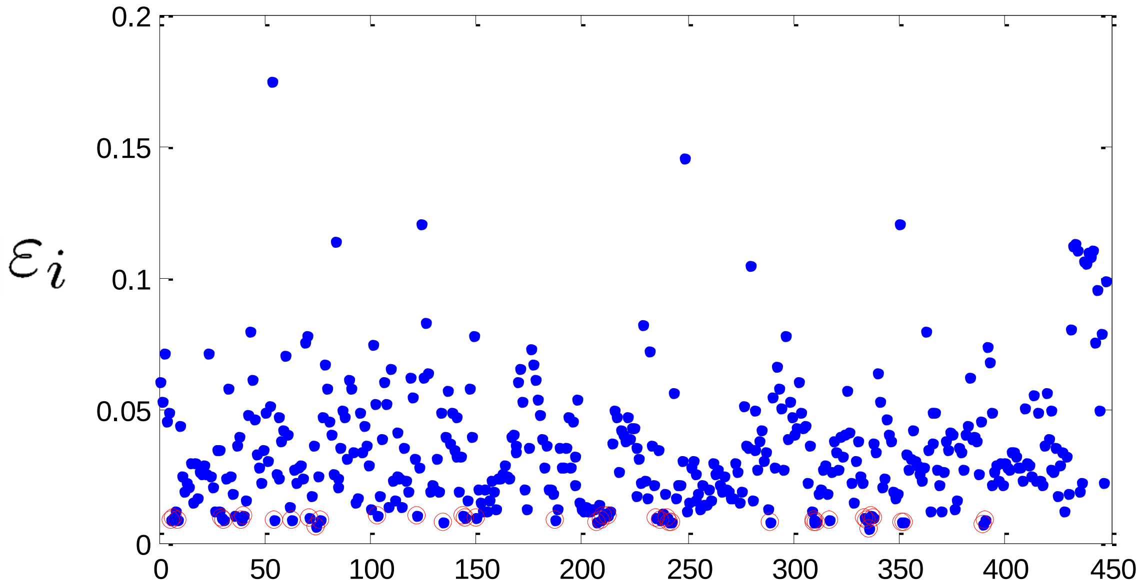
MUSIC-CSR $\begin{cases} \text{SNR} = 11.7 \text{ dB} \\ \text{computation time} \simeq 10 \text{ sec} \end{cases}$

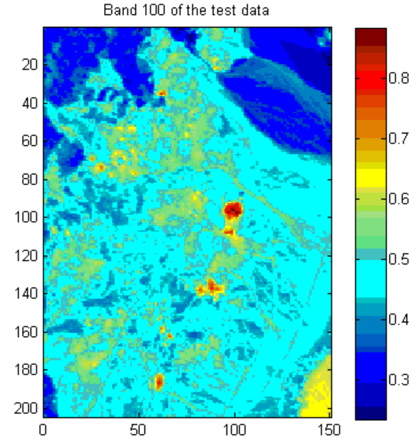
CSR $\begin{cases} \text{SNR} = 0 \text{ dB} \\ \text{computation time} \simeq 600 \text{ sec} \end{cases}$

Results with CUPRITE

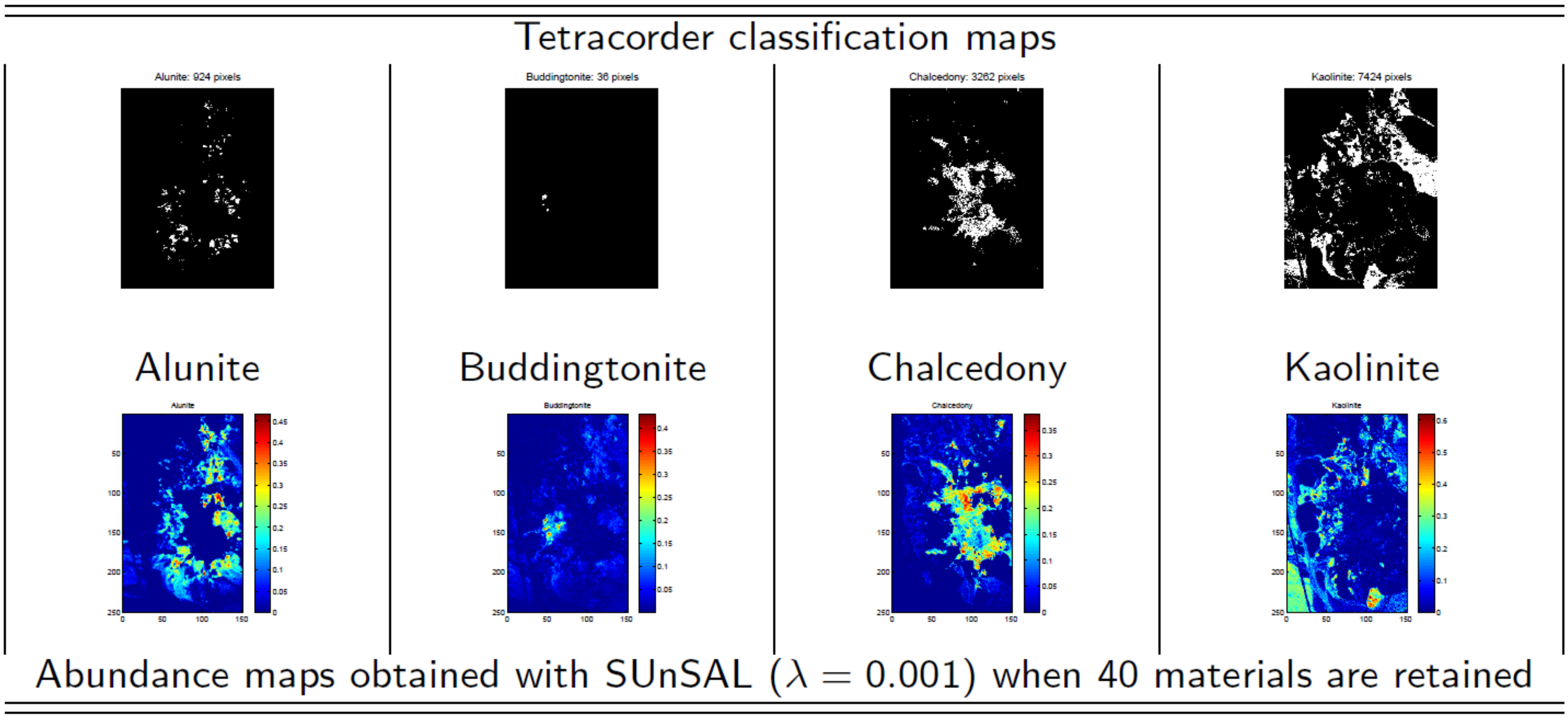
- size: 350x350 pixels
- spectral library: 302 materials (minerals) from the USGS library
- spectral bands: 188 out of 224
  (noisy bands were removed)
- spectral range: 0.4 – 2.5 um
- spectral resolution: 10 nm
- “validation” map: Tetracorder***







Results with real data



Note: Good spatial distribution of the endmembers

Processing times: 2.6 ms/pixel using the full library; 0.22ms/pixels using the pruned library with 40 members

Convex optimization problems in SLU

Constrained least squares (CLS)

$$\min_{\mathbf{X}} \ \|\mathbf{AX}-\mathbf{Y}\|_F^2$$
$$\text{subject to: } \ \mathbf{X} \geq \mathbf{0}$$

$$\|\mathbf{X}\|_F^2 := \operatorname{tr}\{\mathbf{X}\mathbf{X}^T\}$$

Fully constrained least squares (FCLS)

$$\min_{\mathbf{X}} \ \|\mathbf{AX}-\mathbf{Y}\|_F^2$$
$$\text{subject to: } \ \mathbf{X} \geq \mathbf{0}, \ \mathbf{1}_n^T\mathbf{X} = \mathbf{1}_N^T$$

$\mathbf{X} \geq \mathbf{0}$ → ANC

$\mathbf{1}_n^T\mathbf{X} = \mathbf{1}_N^T$ → ASC

Constrained sparse regression (CSR)

$$\min_{\mathbf{X}} \ (1/2)\|\mathbf{AX}-\mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}\|_1$$
$$\text{subject to: } \ \mathbf{X} \geq \mathbf{0}$$

$$\|\mathbf{X}\|_1 := \sum_{i=1}^N \|\mathbf{x}_i\|_1$$

Convex optimization problems in SLU

Constrained basis pursuit (CBP)

$$\min_{\mathbf{X}} \ \|\mathbf{X}\|_1$$
$$\text{subject to: } \ \mathbf{AX} = \mathbf{Y}, \ \ \mathbf{X} \geq \mathbf{0}$$

CBP denoising (CBPDN)

$$\min_{\mathbf{X}} \ \|\mathbf{X}\|_1$$
$$\text{subject to: } \ \|\mathbf{AX} - \mathbf{Y}\|_F \leq \delta, \ \ \mathbf{X} \geq \mathbf{0}$$

Constrained colaborative sparse regression (CCSR)

$$\min_{\mathbf{X}} \ (1/2)\|\mathbf{AX} - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}\|_{2,1}$$
$$\text{subject to: } \ \mathbf{X} \geq \mathbf{0}, \ \ \mathbf{1}_n^T\mathbf{X} = \mathbf{1}_N^T$$

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

Constrained total variation (CTV)

$$\min_{\mathbf{X}} \ (1/2)\|\mathbf{AX} - \mathbf{Y}\|_F^2 + \lambda_1\|\mathbf{X}\|_1 + \lambda_2\phi_{TV}(\mathbf{X})$$
$$\text{subject to: } \ \mathbf{X} \geq \mathbf{0}$$

Convex optimization problems in SLU

Structure of the optimization problems: $\min_{\mathbf{X} \in C} \sum_{i=1}^{J} g_i(\mathbf{X})$

convex functions (g_i)

convex set (C)

$$\Leftrightarrow \min_{\mathbf{X}} \sum_{i=1}^{J} g_i(\mathbf{X}) + \iota_C(\mathbf{X})$$

Source of difficulties: large scale ($n \times N \gtrsim 10^7$); nonsmoothness

Line of attack: alternating direction method of multiplies (ADMM)

[Glowinski, Marrocco, 75], [Gabay, Mercier, 76]

Alternating Direction Method of Multipliers (ADMM)

Unconstrained (convex) optimization problem: $\min_{\mathbf{z}\in\mathbb{R}^d} \; f_1(\mathbf{z}) + f_2(\mathbf{G}\,\mathbf{z})$

ADMM [Glowinski, Marrocco, 75], [Gabay, Mercier, 76]

$$\mathbf{z}_{k+1} = \arg\min_{\mathbf{z}} f_1(\mathbf{z}) + \frac{\mu}{2}\|\mathbf{G}\,\mathbf{z} - \mathbf{u}_k - \mathbf{d}_k\|^2$$

$$\mathbf{u}_{k+1} = \arg\min_{\mathbf{u}} f_2(\mathbf{u}) + \frac{\mu}{2}\|\mathbf{G}\,\mathbf{z}_{k+1} - \mathbf{u} - \mathbf{d}_k\|^2$$

$$\mathbf{d}_{k+1} = \mathbf{d}_k - (\mathbf{G}\,\mathbf{z}_{k+1} - \mathbf{u}_{k+1})$$

Interpretations: variable splitting + augmented Lagrangian + NLBGS;

Douglas-Rachford splitting on the dual [Eckstein, Bertsekas, 92];

split-Bregman approach [Goldstein, Osher, 08]

A Cornerstone Result on ADMM [Eckstein, Bertsekas, 1992]

Consider the problem

$$\min_{\mathbf{z}\in\mathbb{R}^d} \; f_1(\mathbf{z}) + f_2(\mathbf{G}\,\mathbf{z})$$

Let f_1 and f_2 be closed, proper, and convex and $\mathbf{G}$ have full column rank.

Let $(\mathbf{z}_k,\; k=0,1,2,...)$ be the sequence produced by ADMM, with $\mu > 0$, then, if the problem has a solution, say $\bar{\mathbf{z}}$, then

$$\lim_{k\rightarrow\infty} \mathbf{z}_k = \bar{\mathbf{z}}$$

The theorem also allows for inexact minimizations, as long as the errors are absolutely summable.

Convergence rate: $O(1/\varepsilon)$ [He, Yuan, 2011]

ADMM for two or more functions [Figueiredo, B-D, 09]

Consider a more general problem $$\min_{\mathbf{z}\in\mathbb{R}^d} \sum_{j=1}^{J} g_j(\mathbf{H}^{(j)}\mathbf{z}) \qquad (P)$$

$g_j : \mathbb{R}^{p_j} \to \bar{\mathbb{R}}$ — Proper, closed, convex functions

$\mathbf{H}^{(j)} \in \mathbb{R}^{p_j \times d}$ — Arbitrary matrices

There are many ways to write (P) as $\min_{\mathbf{z}\in\mathbb{R}^d} f_1(\mathbf{z}) + f_2(\mathbf{G}\,\mathbf{z})$

We adopt:

$$f_1(\mathbf{z}) = 0, \qquad \mathbf{G} = \begin{bmatrix} \mathbf{H}^{(1)} \\ \vdots \\ \mathbf{H}^{(J)} \end{bmatrix}, \qquad f_2\left(\begin{bmatrix} \mathbf{u}^{(1)} \\ \vdots \\ \mathbf{u}^{(J)} \end{bmatrix}\right) = \sum_{j=1}^{J} g_j(\mathbf{u}^{(j)})$$

Another approach in [Goldfarb, Ma, 09, 11]

Applying ADMM to more than two functions

$$\mathbf{z}_{k+1} = \left[\sum_{j=1}^{J} (\mathbf{H}^{(j)})^* \mathbf{H}^{(j)}\right]^{-1} \left(\sum_{j=1}^{J} (\mathbf{H}^{(j)})^* \left(\mathbf{u}_k^{(j)} + \mathbf{d}_k^{(j)}\right)\right)$$

$$\mathbf{u}_{k+1}^{(1)} = \arg\min_{\mathbf{u}} g_1(\mathbf{u}) + \frac{\mu}{2}\|\mathbf{u} - \mathbf{H}^{(1)}\mathbf{z}_{k+1} + \mathbf{d}_k^{(1)}\|^2 = \text{prox}_{g_1/\mu}(\mathbf{H}^{(1)}\mathbf{z}_{k+1} - \mathbf{d}_k^{(j)})$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$\mathbf{u}_{k+1}^{(J)} = \arg\min_{\mathbf{u}} g_J(\mathbf{u}) + \frac{\mu}{2}\|\mathbf{u} - \mathbf{H}^{(J)}\mathbf{z}_{k+1} + \mathbf{d}_k^{(J)}\|^2 = \text{prox}_{g_1/\mu}(\mathbf{H}^{(1)}\mathbf{z}_{k+1} - \mathbf{d}_k^{(j)})$$

$$\mathbf{d}_{k+1}^{(1)} = \mathbf{d}_k^{(1)} - (\mathbf{H}^{(1)}\mathbf{z}_{k+1} - \mathbf{u}_{k+1}^{(1)})$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\mathbf{d}_{k+1}^{(J)} = \mathbf{d}_k^{(J)} - (\mathbf{H}^{(J)}\mathbf{z}_{k+1} - \mathbf{u}_{k+1}^{(J)})$$

Conditions for easy applicability: inexpensive proximity operators

inexpensive matrix inversion

Constrained sparse regression (CSR)

Problem $\min_{\mathbf{X}} (1/2)\|\mathbf{AX}-\mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}\|_1, \quad \text{subject to: } \mathbf{X} \geq \mathbf{0}$

Equivalent formulation $\min_{\mathbf{X}} (1/2)\|\mathbf{AX}-\mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}\|_1 + \iota_{\mathbb{R}_+}(\mathbf{X})$

indicator of the first orthant

Template: $\min_{\mathbf{X}} \sum_{j=1}^{J=3} g_j(\mathbf{H}^{(j)}\mathbf{X})$

Mapping:

$$\mathbf{H}^{(1)} = \mathbf{A}, \quad g_1(\mathbf{Z}) = (1/2)\|\mathbf{Z}-\mathbf{Y}\|_F^2$$

$$\mathbf{H}^{(2)} = \mathbf{I}, \quad g_2(\mathbf{Z}) = \lambda\|\mathbf{Z}\|_1$$

$$\mathbf{H}^{(3)} = \mathbf{I}, \quad g_3(\mathbf{Z}) = \iota_{\mathbb{R}_+}(\mathbf{Z})$$

Proximity operators:

$$\text{prox}_{g_1/\mu}(\mathbf{W}) = \frac{\mathbf{Y}+\mu\mathbf{W}}{1+\mu}$$

$$\text{prox}_{g_2/\mu}(\mathbf{W}) = \text{soft}(\mathbf{W}, \lambda/\mu)$$

$$\text{prox}_{g_3/\mu}(\mathbf{W}) = \max\{\mathbf{0}, \mathbf{W}\}$$

Matrix inversion can be precomputed

(typical sizes 200~300 x 500~1000)

$$\left[\sum_{j=1}^{J}(\mathbf{H}^{(j)})^*\,\mathbf{H}^{(j)}\right]^{-1} = \left[\mathbf{A}^T\mathbf{A}+2\mathbf{I}\right]^{-1}$$

Spectral unmixing by split augmented Lagrangian (SUnSAL) [B-D, Figueiredo, 2010]

Related algorithm (split-Bregman view) in [Szlam, Guo, Osher, 2010]

Concluding remarks

- Sparse regression framework, used with care, yields effective hyperspectral unmixing results.

- Critical factors
  - High mutual coherence of the hyperspectral libraries
  - Non-linear mixing and noise
  - Acquisition and calibration of hyperspectral libraries

- Favorable factors
  - Hyperspectral mixtures are highly sparse

- ADMM is a very flexible and efficient tool solving the hyperspectral sparse regression optimization problems

References

M. V. Afonso, J. Bioucas-Dias, and M. Figueiredo, “An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems”, IEEE Transactions on Image Processing, vol. 20, no. 3, pp. 681-695, 2011.

J. Bioucas-Dias and M. Figueiredo, “Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing”, in 2nd IEEE GRSS Workshop on Hyperspectral Image and Signal Processing-WHISPERS'2010, Raykjavik, Iceland, 2010.

J. 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

S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” SIAM Rev., vol. 43, no. 1, pp. 129–159, 2001.

D. Gabay and B. Mercier, “A dual algorithm for the solution of nonlinear variational problems via finite-element approximations,” Comput. Math. Appl., vol. 2, pp. 17–40, 1976.

References

R. Glowinski and A. Marroco, “Sur l’approximation, par elements finis d’ordre un, et la resolution, par penalisation-dualité d’une classe de problemes de dirichlet non lineares,” Rev. Française d’Automat. Inf. Recherche Opérationelle, vol. 9, pp. 41–76, 1975.

T. Goldstein and S. Osher, “The split Bregman algorithm for regularized problems,” Comput. Appl. Math., Univ. California, Los Angeles, Tech. Rep. 08–29, 2008.

B. He and X. Yuan, X, “On convergence rate of the Douglas-Rachford operator splitting method”, 2011.

J. Eckstein, D. Bertsekas, “On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators”, Math. Progr., vol. 5, pp. 293-318, 1992

E. Esser, “Applications of Lagrangian-based alternating direction methods and connections to split-Bregman,” Comput. Appl. Math., Univ. California, Los Angeles, Tech. Rep. 09–31, 2009.

M-D, Iordache, J. Bioucas-Dias and A. Plaza, "Sparse unmixing of hyperspectral data", IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no. 6, pp. 2014-2039, 2011

References

M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Hyperspectral unmixing with sparse group lasso", in IEEE International Geoscience and Remote Sensing Symposium (IGARSS'11), Vancouver, Canada, 2011.

M.-D. Iordache, J. Bioucas-Dias and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing", IEEE Transactions on Geoscience and Remote Sensing, vol. PP, no. 99, pp. 1-19, 2012.

C. Li, T. Sun, K. Kelly, and Y. Zhang, "A Compressive Sensing and Unmixing Scheme for Hyperspectral Data Processing", IEEE Transactions on Image Processing, 2011

G. Martin, J.Bioucas-Dias, A. Plaza, "Hyperspectral coded aperture (hyca): a new technique for hyperspectral compressive sensing", IEEE International Geoscience and Remote Sensing Symposium (IGARSS'12), Munich, Germany, 2012

J. Nascimento and J. Bioucas-Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data", IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 4, pp. 898-910, 2005.

References

P. Sprechmann, I. Ramirez, G. Sapiro, G., and, Y. Eldar, “Collaborative hierarchical sparse modeling”, IEEE 4th Annual Conference on Information Sciences and Systems (CISS), 2010, pp. 1-6, 2010

R. Tibshirani, “Regression shrinkage and selection via the lasso”, “Journal of the Royal Statistical Society. Series B (Methodological”, pp. 267--288, 2006.

A Turlach, W. Venables, and S. Wright, “Atomic decomposition by basis pursuit,” SIAM Rev., vol. 43, no. 1, pp. 129–159, 2001.

Q. Zhang, R. Plemmons, D. Kittle, D. Brady, S. Prasad, “Joint segmentation and reconstruction of hyperspectral data with compressed measurements”, Applied Optics, 2011.

X. Zhao, F. Wang, T. Huang, M. Ng and R. Plemmons, “Deblurring and Sparse Unmixing For Hyperspectral Images”, Preprint, May 2012.