Compressive Source Separation: Algorithms and Applications

Mohammad Golbabaee

Simon Arberet

Pierre Vandergheynst

Ecole Polytechnique Fédérale de Lausanne (EPFL)

UNLocX

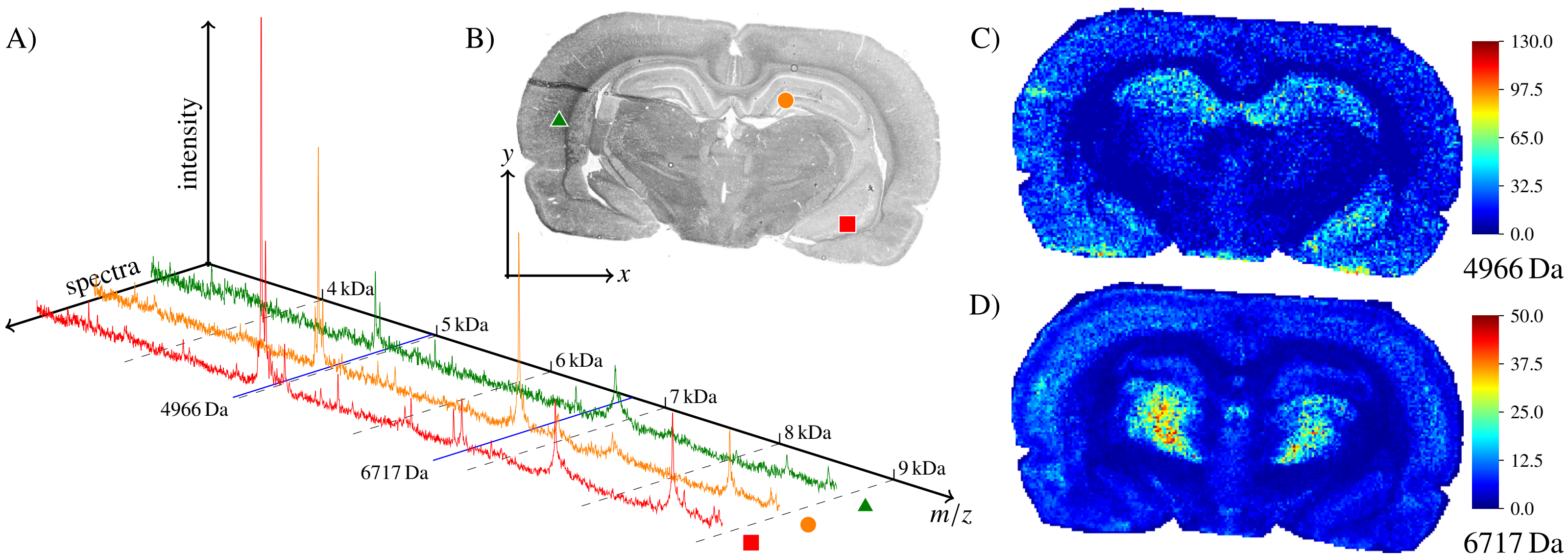
CIMI Workshop 2013, Toulouse

Motivation: MALDI Imaging

Molecular Mass Spectroscopy

Matrix-assisted laser desorption/ionization

Very high-dimensional: 3D X spectra !

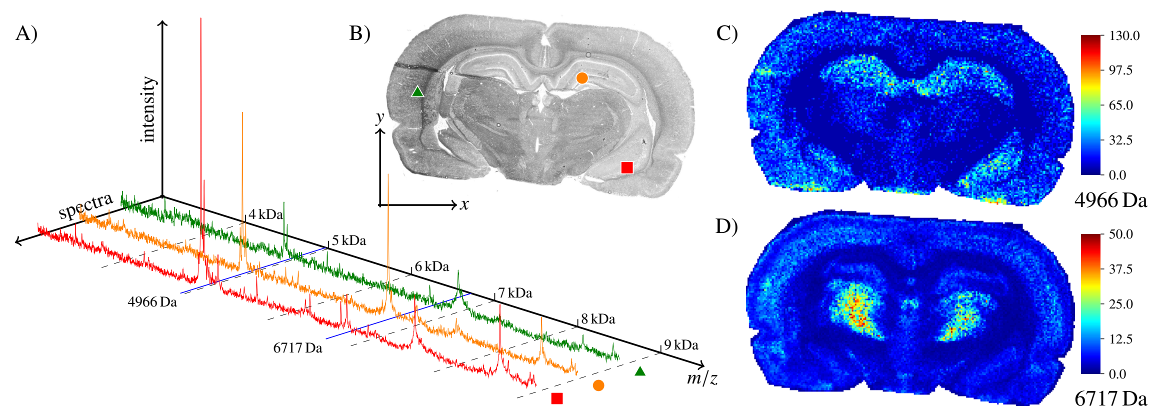


High-Dimensional Multichannel Data

$$X \in \mathbb{R}^{n_1 \times n_2}$$

Often very structured: spatial, spectral

$$X = \Phi\Theta$$



Very special structure: (linear) mixture model

Very special structure: (linear) mixture model

$$X = \mathbf{S}\mathbf{H}^T$$

Very special structure: (linear) mixture model

$$X = \mathbf{S}\mathbf{H}^T$$

See for instance [Bioucas-Dias et al., 2012]

Very special structure: (linear) mixture model

$$X = \mathbf{S}\mathbf{H}^T$$

See for instance [Bioucas-Dias et al., 2012]

Very special structure: (linear) mixture model

$$X = \mathbf{S}\mathbf{H}^T$$

See for instance [Bioucas-Dias et al., 2012]

Very special structure: (linear) mixture model

See for instance [Bioucas-Dias et al., 2012]

Very special structure: (linear) mixture model

$$X = \mathbf{\Psi\Theta H}^T$$

See for instance [Bioucas-Dias et al., 2012]

Maps to low-dimensional projections

Maps to low-dimensional projections

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

Maps to low-dimensional projections

Maps to low-dimensional projections

$$Y = \mathbf{A}(\mathbf{S}\mathbf{H}^T)$$

Maps to low-dimensional projections

Maps to low-dimensional projections

$$Y = \mathbf{A}(\mathbf{\Psi}\mathbf{\Theta}\mathbf{H}^T)$$

Maps to low-dimensional projections

$$Y = \mathbf{A}(\mathbf{\Psi\Theta H}^T)$$

Questions:

- How many projections ?
- Design of $\mathbf{A}$?

Outline

- 2 Problems
  - Sparse regression = dictionary of spectra known
    - Is it interesting in some applications ?
    - Can we use this information? Obtain theoretical guarantees ?
  - Sparse coding = blind: learn spectra and abundances
- CS: observe projections to low dimension
  - can we directly recover model parameters ?
  - can we use knowledge of spectra

CS of Multichannel Signals

Baseline: no structure $X \in \mathbb{R}^{n_1 \times n_2}$ $X_{vec} \in \mathbb{R}^{n_1 n_2}$

$$AX_{vec} := \mathcal{A}(X) \qquad y = AX_{vec} + z \qquad A \in \mathbb{R}^{m \times n_1 n_2}$$

$$\arg\min_{X_{vec}} \|X_{vec}\|_1 \quad s.t. \quad \|y - AX_{vec}\|_2 \leq \varepsilon$$

Recovery for $K(\ll n_1 n_2)$-sparse signals when $m = \mathcal{O}(K \log(n_1 n_2 / K))$

[Donoho, Candès-Romberg, Tao]

CS of Multichannel Signals

Typically impose spatial structure on sparsity model

CS of Multichannel Signals

Typically impose spatial structure on sparsity model

Via the dictionary: $\Psi = \mathbb{I}_{n_2} \otimes \Psi_{2D}$

$$\arg\min_{\Theta_{vec}} \|\Theta_{vec}\|_1 \quad s.t. \quad \|y - A\Psi\Theta_{vec}\|_2 \leq \varepsilon,$$

CS of Multichannel Signals

Typically impose spatial structure on sparsity model

Via the dictionary: $\Psi = \mathbb{I}_{n_2} \otimes \Psi_{2D}$

$$\arg\min_{\Theta_{vec}} \|\Theta_{vec}\|_1 \quad s.t. \quad \|y - A\Psi\Theta_{vec}\|_2 \leq \varepsilon,$$

k sparse per channel and RIP holds $\quad m \geq c\, n_2 k \log(n_1/k)$

CS of Multichannel Signals

Typically impose spatial structure on sparsity model

Via the dictionary: $\Psi = \mathbb{I}_{n_2} \otimes \Psi_{2D}$

$$\arg\min_{\Theta_{vec}} \|\Theta_{vec}\|_1 \quad s.t. \quad \|y - A\Psi\Theta_{vec}\|_2 \leq \varepsilon,$$

k sparse per channel and RIP holds $\quad m \geq c\, n_2 k \log(n_1/k)$

CS of Multichannel Signals

Typically impose spatial structure on sparsity model

Via the dictionary: $\Psi = \mathbb{I}_{n_2} \otimes \Psi_{2D}$

$$\arg\min_{\Theta_{vec}} \|\Theta_{vec}\|_1 \quad s.t. \quad \|y - A\Psi\Theta_{vec}\|_2 \leq \varepsilon,$$

k sparse per channel and RIP holds $\quad m \geq c\, n_2 k \log(n_1/k)$

Not exploiting spectral redundancies

Via other (block structured) sparsity penalties ?

[MMV, Davies-Eldar]

More structure ?

The Linear Mixture Case

$X = \mathbf{S}\mathbf{H}^T$

$\mathbf{S} \in \mathbb{R}^{n_1 \times \rho}$ Spatial abundance maps

$\mathbf{H} \in \mathbb{R}^{n_2 \times \rho}$ Spectra or endmembers

Each channel is a mixture $$X_j = \sum_{i=1}^{\rho} [\mathbf{H}]_{j,i}\, \mathbf{S}_i$$

Typically the number of endmembers is very small compared to the spatial and spectral dimensions

$$\rho << n_1,\ \rho << n_2$$

Spectra as Side Information ?

Assume we have a dictionary of spectra/endmembers

$\Phi = \mathbf{H} \otimes \mathrm{Id}_{n_1}$ $\qquad y = A\Phi\mathbf{S}_{vec} + z$

Spectra as Side Information ?

Assume we have a dictionary of spectra/endmembers

$$\Phi = \mathbf{H} \otimes \mathrm{Id}_{n_1}$$

$$y = \boxed{A\Phi}\mathbf{S}_{vec} + z$$

New sensing operator

Spectra as Side Information ?

Assume we have a dictionary of spectra/endmembers

$$\Phi = \mathbf{H} \otimes \mathrm{Id}_{n_1} \qquad y = \underbrace{A\Phi}_{\text{New sensing operator}}\mathbf{S}_{vec} + z$$

$$\mathbf{S}_{vec} = \mathbf{\Psi}\mathbf{\Theta}_{vec}$$

Spectra as Side Information ?

Assume we have a dictionary of spectra/endmembers

$$\Phi = \mathbf{H} \otimes \mathrm{Id}_{n_1} \qquad y = \boxed{A\Phi}\mathbf{S}_{vec} + z$$

New sensing operator

$$\mathbf{S}_{vec} = \mathbf{\Psi\Theta}_{vec} \longrightarrow$$ directly recover abundances

Spectra as Side Information ?

Assume we have a dictionary of spectra/endmembers

$$\Phi = \mathbf{H} \otimes \mathrm{Id}_{n_1} \qquad\qquad y = A\Phi\mathbf{S}_{vec} + z$$

New sensing operator

$$\mathbf{S}_{vec} = \boldsymbol{\Psi}\boldsymbol{\Theta}_{vec} \longrightarrow$$ directly recover abundances

$$\arg\min_{\boldsymbol{\Theta}_{vec}} \|\boldsymbol{\Theta}_{vec}\|_1 \quad \text{s.t.} \quad \|y - A\Phi\boldsymbol{\Psi}\boldsymbol{\Theta}_{vec}\|_2 \leq \varepsilon$$

How do we choose A ? Influence of **H** ?

How can we use the knowledge of **H** ?

Fundamental Limits - 1

Our problem is of the form:

$$\arg\min_{\theta} \ \|\theta\|_1 \quad s.t. \quad \|y - A\mathbf{D}\theta\|_2 \leq \varepsilon$$

Compressive Sensing with a coherent dictionary D

But ...

Fundamental Limits - 1

Our problem is of the form:

$$\arg\min_{\theta} \|\theta\|_1 \quad s.t. \quad \|y - A\mathbf{D}\theta\|_2 \leq \varepsilon$$

Compressive Sensing with a coherent dictionary *D*

But ...

RIP on $A\mathbf{D}$ will impose very strong restrictions on the underlying $\mathbf{H}$: (recall $\mathbf{D} = (\mathbf{H} \otimes \mathrm{Id}_{n_1})\mathbf{\Psi}$)

Fundamental Limits - 1

Our problem is of the form:

$$\arg\min_{\theta} \; \|\theta\|_1 \quad s.t. \quad \|y - A\mathbf{D}\theta\|_2 \leq \varepsilon$$

Compressive Sensing with a coherent dictionary D

But ...

RIP on $A\mathbf{D}$ will impose very strong restrictions on the underlying $\mathbf{H}$: (recall $\mathbf{D} = (\mathbf{H} \otimes \mathrm{Id}_{n_1})\mathbf{\Psi}$)

$$\delta_k(A\mathbf{D}) < \sqrt{2} - 1 \Longrightarrow \xi(\mathbf{H}) < \sqrt{\sqrt{2} + 1}$$

$$\xi(\mathbf{H}) = \frac{\sigma_{\max}(\mathbf{H})}{\sigma_{\min}(\mathbf{H})}$$

Fundamental Limits - 2

Instead analyze with D-RIP for the sensing matrix

Fundamental Limits - 2

Instead analyze with D-RIP for the sensing matrix

$$(1-\delta_k^*)\|\mathbf{D}x\|_2^2 \leq \|A\mathbf{D}x\|_2^2 \leq (1+\delta_k^*)\|\mathbf{D}x\|_2^2$$

[Candès, Eldar, Needel, Randall]

Fundamental Limits - 2

Instead analyze with D-RIP for the sensing matrix

$$(1-\delta_k^*)\|\mathbf{D}x\|_2^2 \leq \|A\mathbf{D}x\|_2^2 \leq (1+\delta_k^*)\|\mathbf{D}x\|_2^2$$

[Candès, Eldar, Needel, Randall]

Control on $\mathbf{D}$: $\mathcal{L}_k(\mathbf{D})\|x\|_2 \leq \|\mathbf{D}x\|_2 \leq \mathcal{U}_k(\mathbf{D})\|x\|_2$

Fundamental Limits - 2

Instead analyze with D-RIP for the sensing matrix

$$(1-\delta_k^*)\|\mathbf{D}x\|_2^2 \le \|A\mathbf{D}x\|_2^2 \le (1+\delta_k^*)\|\mathbf{D}x\|_2^2$$

[Candès, Eldar, Needel, Randall]

Control on $\mathbf{D}$: $\mathcal{L}_k(\mathbf{D})\|x\|_2 \le \|\mathbf{D}x\|_2 \le \mathcal{U}_k(\mathbf{D})\|x\|_2$

If A satisfies the D-RIP for $\mathbf{H}\otimes \mathrm{Id}_{n_1}$ with constant:

$$\delta^*_{\gamma' k} < 1/3 \quad \text{where } \gamma' = 1 + 2\xi^2(\mathbf{H})$$

Then $\|\mathbf{\Theta}_{vec} - \widehat{\mathbf{\Theta}}_{vec}\|_2 \le c_0' \, k^{-1/2}\|\mathbf{\Theta}_{vec} - (\mathbf{\Theta}_{vec})_k\|_1 + c_1'\varepsilon$

Fundamental Limits - 2

Instead analyze with D-RIP for the sensing matrix

$$(1-\delta_k^*)\|\mathbf{D}x\|_2^2 \leq \|A\mathbf{D}x\|_2^2 \leq (1+\delta_k^*)\|\mathbf{D}x\|_2^2$$

[Candès, Eldar, Needel, Randall]

Control on $\mathbf{D}$: $\mathcal{L}_k(\mathbf{D})\|x\|_2 \leq \|\mathbf{D}x\|_2 \leq \mathcal{U}_k(\mathbf{D})\|x\|_2$

If A satisfies the D-RIP for $\mathbf{H} \otimes \mathrm{Id}_{n_1}$ with constant:

$$\delta^*_{\gamma' k} < 1/3 \quad \text{where } \gamma' = 1 + 2\xi^2(\mathbf{H})$$

Then $\|\boldsymbol{\Theta}_{vec} - \widehat{\boldsymbol{\Theta}}_{vec}\|_2 \leq c_0' \, k^{-1/2} \|\boldsymbol{\Theta}_{vec} - (\boldsymbol{\Theta}_{vec})_k\|_1 + c_1' \varepsilon$

Note that: $\gamma' k < n_1 n_2 \Rightarrow \ \xi(\mathbf{H}) \leq \sqrt{\dfrac{n_1 n_2/k - 1}{2}}$

Trivial Sampling Operators

"Decorrelation" sampling: $A = \mathbf{H}^\dagger \otimes \widetilde{A}$

$\widehat{m} \times n_1$ with $\widehat{m} \ll n_1$

$$
\begin{aligned}
y &= A\Phi \mathbf{S}_{vec} + z \\
&= \underbrace{(\mathbf{H}^\dagger \otimes \widetilde{A})}_{A}\underbrace{(\mathbf{H} \otimes \mathrm{Id}_{n_1})}_{\Phi} \mathbf{S}_{vec} + z, \\
&= \underbrace{(\mathrm{Id}_\rho \otimes \widetilde{A})}_{\triangleq \widetilde{A}_\rho} \mathbf{S}_{vec} + z.
\end{aligned}
$$

The analysis is then standard since $\mathbf{H}$ has disappeared

Constants will not depend on H, but effect on noise.

Fundamental Limits - 3

firsr case: n2 n1 matrix with sparsity k on spatial dimension

second case: mixture model: reduces sparsity by taking into account all channels and dense sampling uses it. But attention to constant

third case: each channel separately and each channel is k sparse, so roughly kn2 sparse

fourth case: mixture model reduces sparsity, decorrelation step removes effect of H.

With "decorrelation" sampling, the effect o

CS Acquisition Scheme	Dense	Dense	Uniform	Decorrelating
CS Recovery Approach	BPDN	SS-ℓ_1	SS-ℓ_1	SS-ℓ_1
CS measurements $m \gtrsim$	$\mathcal{O}\big(n_2 k \log(n_1/k)\big)$	$\mathcal{O}\big(k \log(\rho n_1/k)\big)$	$\mathcal{O}\big(n_2 k \log(n_1/k)\big)$	$\mathcal{O}\big(k \log(\rho n_1/k)\big)$
Constant depends on $\mathbf{H}$	-	Yes	Yes	No

no mixing structure

effect of $\mathbf{H}$

Applications

Full problem incorporates more constrains:

Applications

Full problem incorporates more constrains:

$$\sum_{j=1}^{\rho} [\mathbf{S}]_{i,j} = 1 \qquad \forall i \in \{1, \ldots, n_1\} \qquad\qquad [\mathbf{S}]_{i,j} \geq 0$$

Applications

Full problem incorporates more constrains:

$$\sum_{j=1}^{\rho} [\mathbf{S}]_{i,j} = 1 \qquad \forall i \in \{1, \ldots, n_1\} \qquad [\mathbf{S}]_{i,j} \geq 0$$

$$\begin{aligned} \arg\min_{\boldsymbol{\Theta}} \quad & \|\boldsymbol{\Theta}_{vec}\|_1 \\ \text{subject to} \quad & \|y - A\Phi\boldsymbol{\Psi}\boldsymbol{\Theta}_{vec}\|_2 \leq \varepsilon \\ & \Psi_{2\mathrm{D}}\, \boldsymbol{\Theta}\, \mathbb{I}_\rho = \mathbb{I}_{n_1} \\ & \boldsymbol{\Psi}\boldsymbol{\Theta}_{vec} \geq 0. \end{aligned}$$

Applications

Full problem incorporates more constrains:

$$\sum_{j=1}^{\rho} [\mathbf{S}]_{i,j} = 1 \qquad \forall i \in \{1, \ldots, n_1\} \qquad\qquad [\mathbf{S}]_{i,j} \geq 0$$

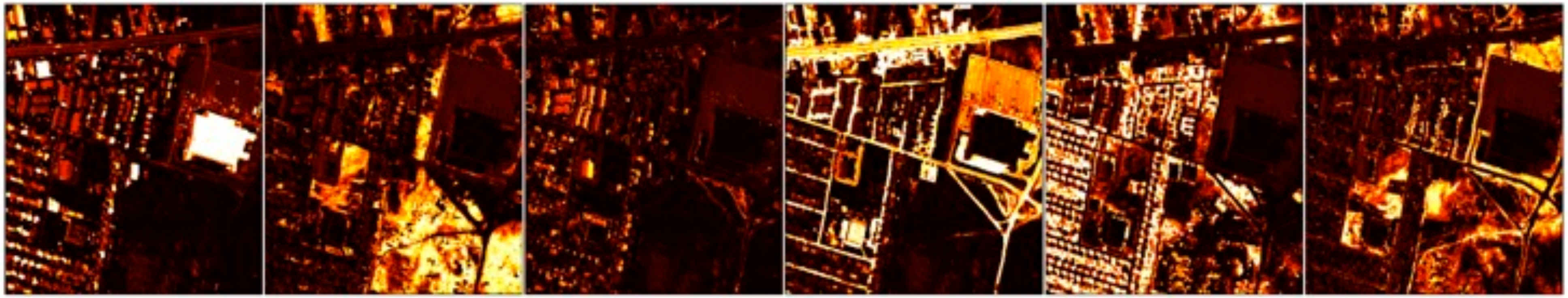
$$\begin{aligned} \arg\min_{\boldsymbol{\Theta}} \quad & \|\boldsymbol{\Theta}_{vec}\|_1 \\ \text{subject to} \quad & \|y - A\Phi\boldsymbol{\Psi}\boldsymbol{\Theta}_{vec}\|_2 \leq \varepsilon \\ & \Psi_{2\mathrm{D}}\,\boldsymbol{\Theta}\,\mathbb{I}_{\rho} = \mathbb{I}_{n_1} \\ & \boldsymbol{\Psi}\boldsymbol{\Theta}_{vec} \geq 0. \end{aligned}$$

$$\arg\min_{\mathbf{S}} \; f_1(\mathbf{S}) + f_2(\mathbf{S}) + f_3(\mathbf{S})$$

$$f_1(\mathbf{S}) = \mathcal{P}(\mathbf{S}),\; f_2(\mathbf{S}) = i_{\mathcal{B}_2}(\mathbf{S}),\; f_3(\mathbf{S}) = i_{\mathcal{B}_{\Delta+}}(\mathbf{S})$$

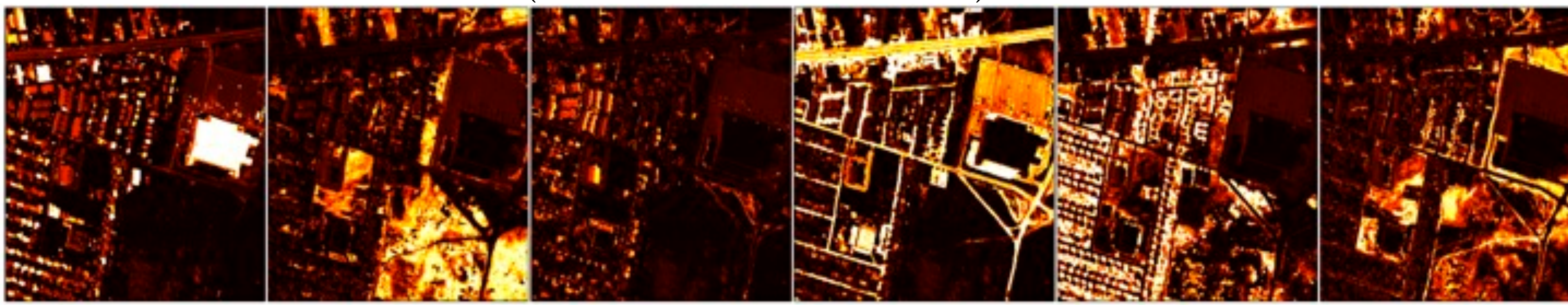
Hyper Spectral Imaging

$\mathbf{S}$: Sources (element abundancies) $\mathbf{S} \in \mathbb{R}^{n_1 \times \rho}$

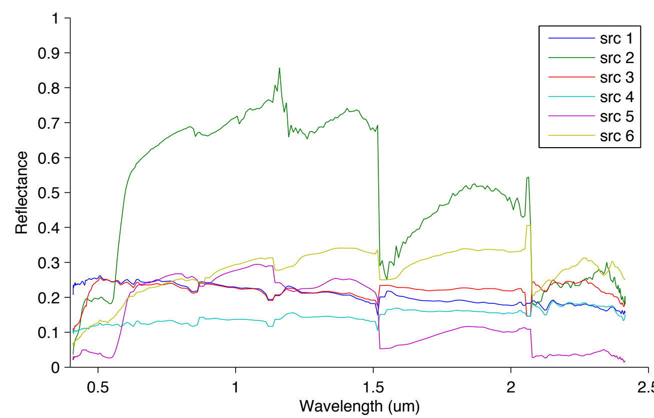


Hyper Spectral Imaging

S: Sources (element abundancies) $\mathbf{S} \in \mathbb{R}^{n_1 \times \rho}$

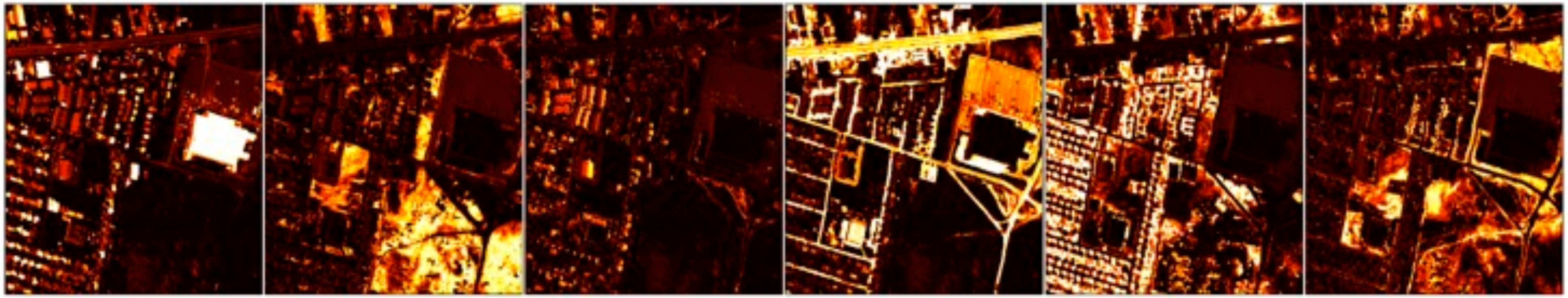


A: Spectra (depending on modality) $\mathbf{A} \in \mathbb{R}^{n_2 \times \rho}$

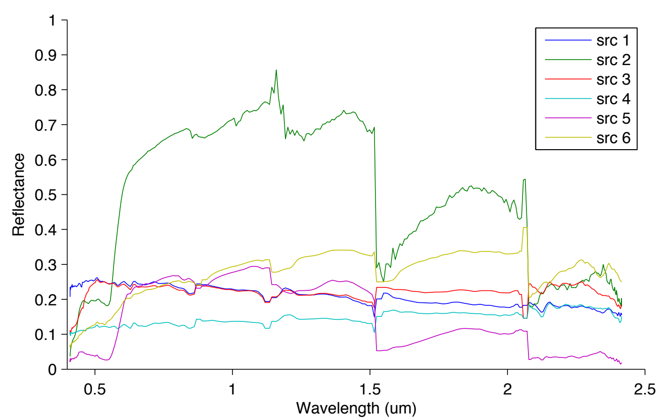


Hyper Spectral Imaging

S: Sources (element abundancies) $\mathbf{S} \in \mathbb{R}^{n_1 \times \rho}$



A: Spectra (depending on modality) $\mathbf{A} \in \mathbb{R}^{n_2 \times \rho}$

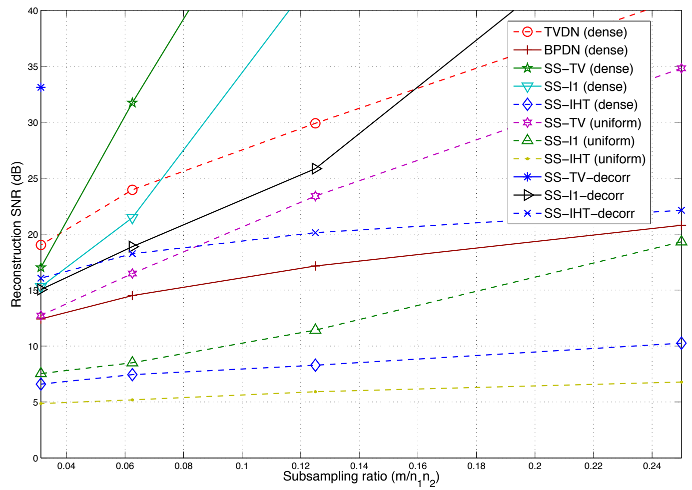


Each pixel is a weighted combination of source spectra: $y = \mathbf{S}\mathbf{A}^T$

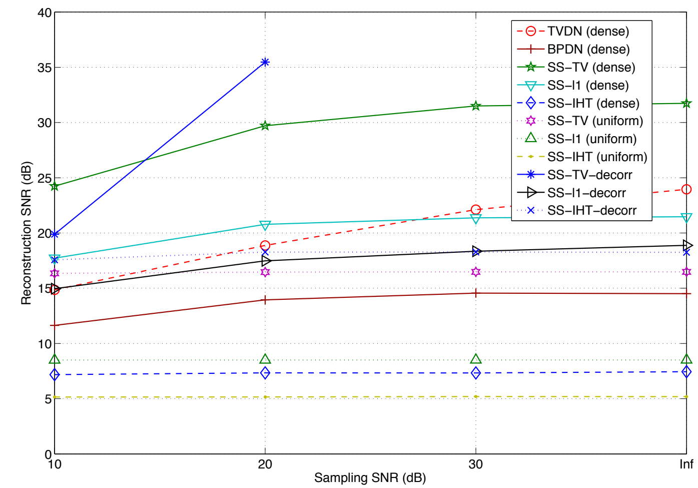
Some experiments

- We have implemented various problems
  - with/without linear mixture model
  - simple sparse wavelet model, TV
- We compared different algorithms
  - PPXA, a variant with IHT, ...
- We used several sensing matrices
  - dense, uniform, decorr, varied the core matrix (random conv, ...)
- We compared on various datasets
  - synthetic, real, CASSI ...

Results



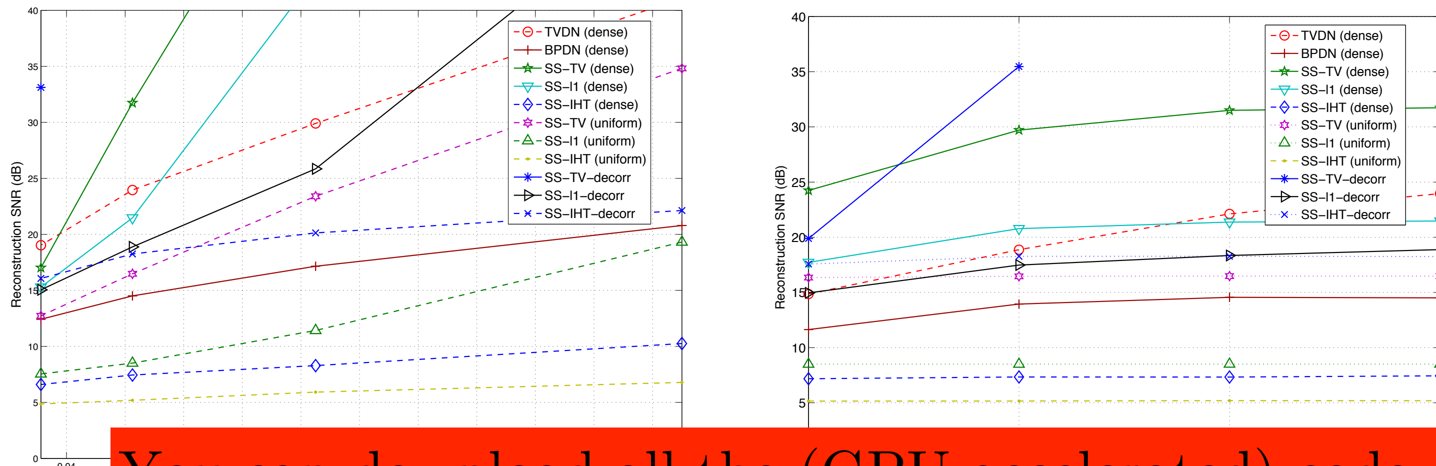
(a) Reconstruction SNR vs. subsampling ratio (noiseless sampling)



(b) Reconstruction SNR vs. sampling SNR (subsampling ratio:1/16)

Noise SNR	$+\infty$ dB				30 dB				10 dB			
Sampling rate	1/4	1/8	1/16	1/32	1/4	1/8	1/16	1/32	1/4	1/8	1/16	1/32
`SS-IHT`*(dense sampling)*	0.69	0.61	0.57	0.48	0.71	0.6	0.57	0.48	0.7	0.6	0.57	0.48
`SS-l1`*(dense sampling)*	**1.0**	**1.0**	0.95	0.81	**1.0**	**1.0**	0.95	0.8	**1.0**	0.98	0.91	0.73
`SS-TV`*(dense sampling)*	**1.0**	**1.0**	**1.0**	0.92	**1.0**	**1.0**	**1.0**	0.91	**1.0**	**1.0**	**0.98**	0.88
`SS-IHT`*(uniform sampling)*	0.43	0.38	0.31	0.25	0.43	0.37	0.31	0.26	0.43	0.37	0.3	0.26
`SS-l1`*(uniform sampling)*	0.97	0.73	0.45	0.31	0.95	0.73	0.48	0.3	0.96	0.75	0.42	0.3
`SS-TV`*(uniform sampling)*	**1.0**	0.98	0.9	0.76	**1.0**	0.97	0.89	0.74	**1.0**	0.97	0.88	0.74
`SS-IHT-decorr`	0.98	0.98	0.96	0.94	0.99	0.98	0.96	0.94	0.98	0.97	0.95	0.92
`SS-l1-decorr`	**1.0**	0.99	0.97	0.92	**1.0**	0.99	0.96	0.91	0.98	0.95	0.92	0.87
`SS-TV-decorr`	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	0.99	**0.98**	**0.96**

Results



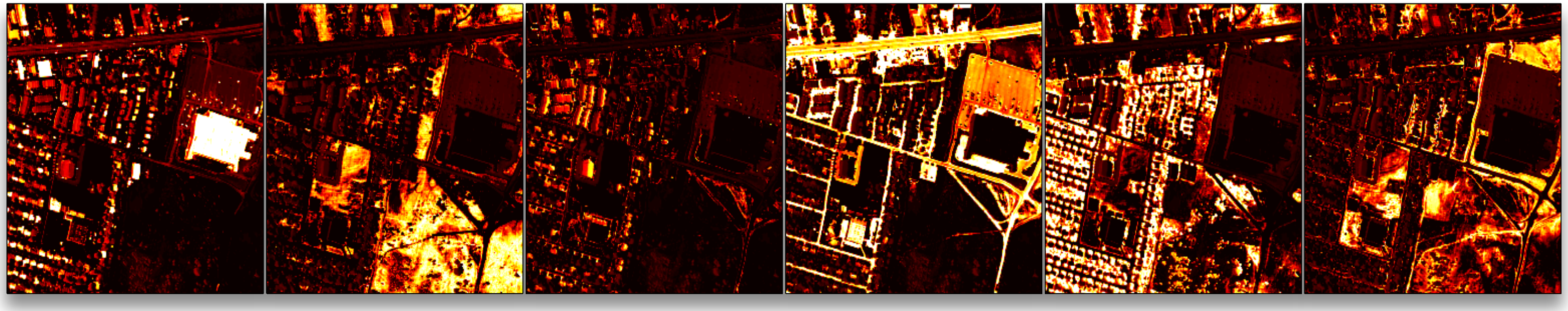
(a) Recons

sampling)

You can download all the (GPU accelerated) code, datasets and scripts for experiments on our web page.
Code uses UnlocBox: http://unlocbox.sourceforge.net
Have fun :)

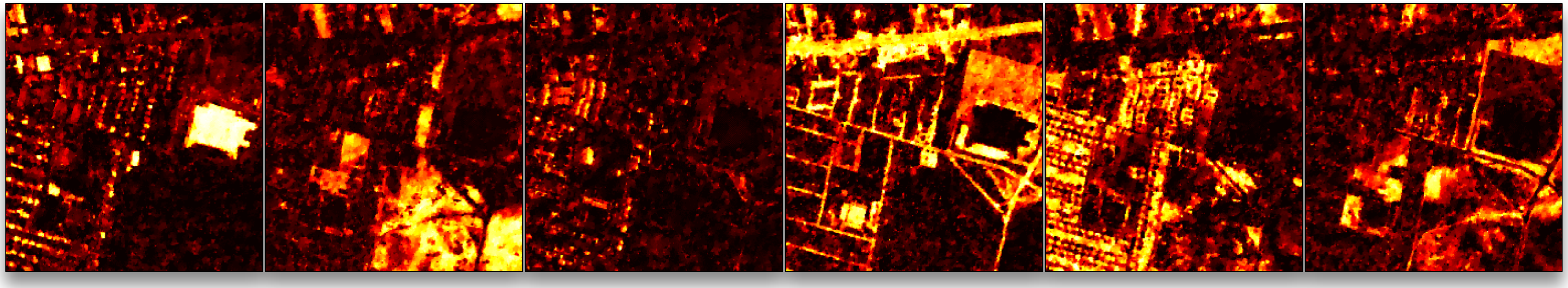
SS-IHT*(dense sampling)*	0.69	0.61	0.57	0.48	0.71	0.6	0.57	0.48	0.7	0.6	0.57	0.48
SS-l1*(dense sampling)*	**1.0**	**1.0**	0.95	0.81	**1.0**	**1.0**	0.95	0.8	**1.0**	0.98	0.91	0.73
SS-TV*(dense sampling)*	**1.0**	**1.0**	**1.0**	0.92	**1.0**	**1.0**	**1.0**	0.91	**1.0**	**1.0**	**0.98**	0.88
SS-IHT*(uniform sampling)*	0.43	0.38	0.31	0.25	0.43	0.37	0.31	0.26	0.43	0.37	0.3	0.26
SS-l1*(uniform sampling)*	0.97	0.73	0.45	0.31	0.95	0.73	0.48	0.3	0.96	0.75	0.42	0.3
SS-TV*(uniform sampling)*	**1.0**	0.98	0.9	0.76	**1.0**	0.97	0.89	0.74	**1.0**	0.97	0.88	0.74
SS-IHT-decorr	0.98	0.98	0.96	0.94	0.99	0.98	0.96	0.94	0.98	0.97	0.95	0.92
SS-l1-decorr	**1.0**	0.99	0.97	0.92	**1.0**	0.99	0.96	0.91	0.98	0.95	0.92	0.87
SS-TV-decorr	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	0.99	**0.98**	**0.96**

Hyper-Spectral Imaging



(a) Ground truth

From 3% of the original data:



(e) SS-TV-decorr, source reconstruction SNR: 8.64 dB

Blind Case

Very high dimensional data (thousands of channels)

Typical problem: Source Separation

Given the dictionary of spectra $\mathbf{A}$ and the data y

V Recover the source abundances, factorizing $y = \mathbf{S}\mathbf{A}^T$

Blind Case

Very high dimensional data (thousands of channels)

Typical problem: Source Separation

Given the dictionary of spectra $\mathbf{A}$ and the data y

V Recover the source abundances, factorizing $y = \mathbf{S}\mathbf{A}^T$

Can this be achieved even when?

$y = \mathcal{A}(\mathbf{S}\mathbf{A}^T)$ Indirect/degraded observations

The dictionary of spectra $\mathbf{A}$ is unknown

Blind Source Separation

HSI Compressive Blind Source Separation (CS-BSS)

$$\underset{\mathbf{S},\mathbf{A}}{\arg\min} \quad \|y - \mathcal{A}(\mathbf{S}\mathbf{A}^T)\|_{\ell_2}^2$$

subject to

$$\sum_j^{\rho} \|\mathbf{S}_j\|_{TV} \leq \tau$$
$$\sum_{i=1}^{\rho} [\mathbf{S}]_{i,j} = 1$$
$$[\mathbf{S}]_{i,j} \geq 0$$

Source image constraints

$$\|\Psi_{1D}^T \mathbf{A}\|_{\ell_1} \leq \gamma$$
$$[\mathbf{A}]_{i,j} \geq 0.$$

Spectral signature constraints

- Bi-convex minimization

Algorithm: alternating convex minimization

1- Initialize A at random

2- Source recovery given A

3- Mixture recovery given S

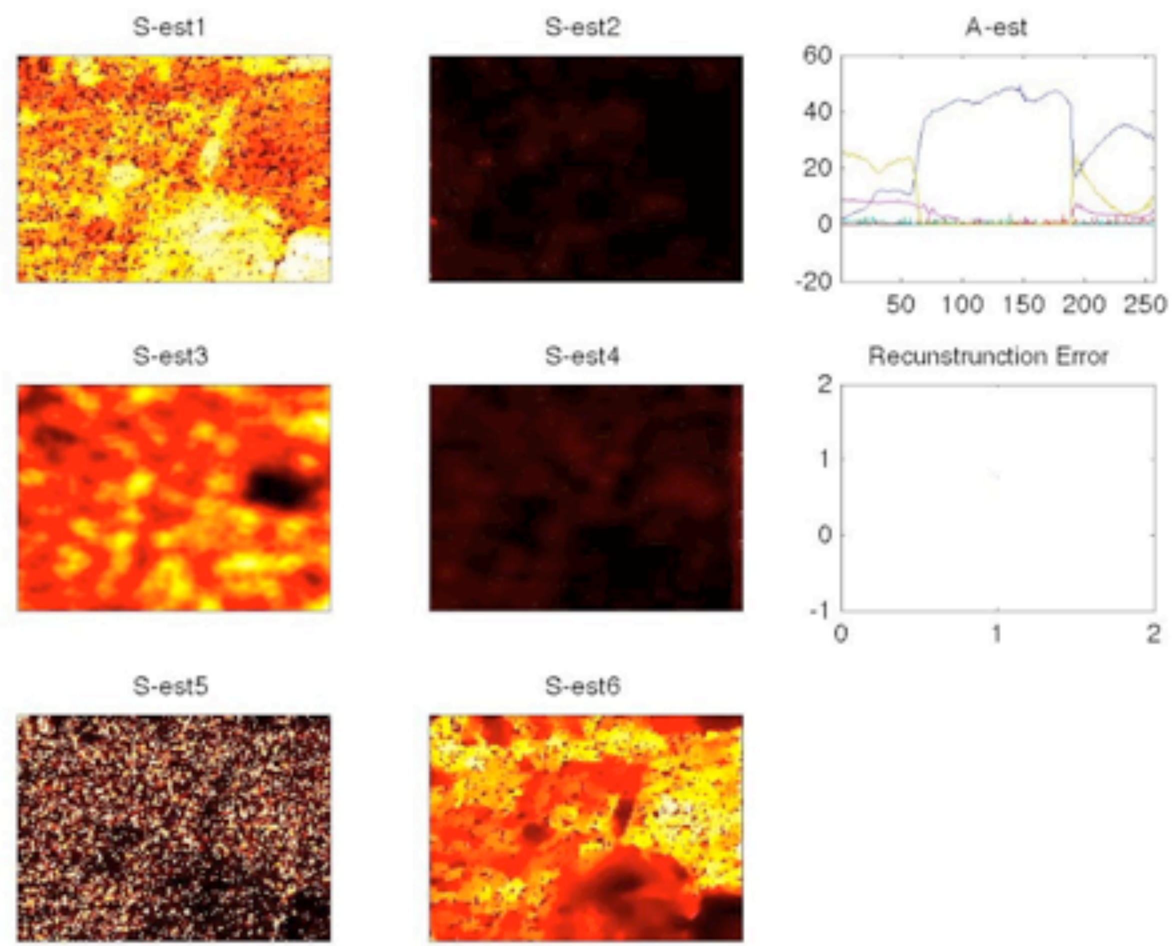
4- Repeat 2-3 until convergence

Goal: Compute source maps directly from compressed measurements

Separation or Segmentation in the compressed domain

Blind Separation

Blind Separation



Outlook

- Significant challenges ahead in signal processing
  - Big Data
  - Ubiquitous but Cheap Sensing (i.e dirty signals)
- Sometimes, no need to reconstruct
  - clustering
  - classification
- Methods that would allow principled and guaranteed task-based processing very appealing