Guaranteed recovery of a low-rank and joint-sparse matrix from incomplete and noisy measurements

Mohammad Golbabaee

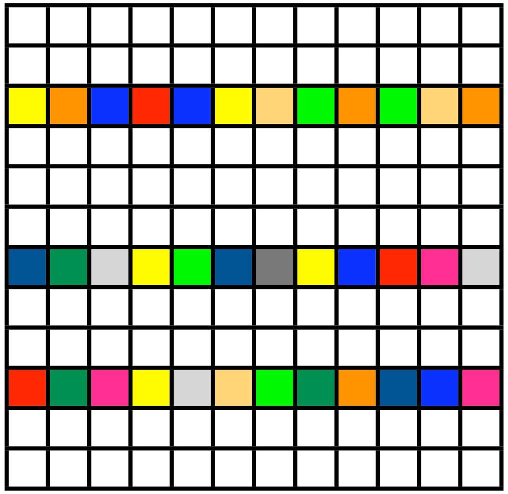
Pierre Vandergheynst

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Low-rank Joint Sparse Data Model

- Given a data matrix $X \in \mathbb{R}^{n_1 \times n_2}$ which is,
  - Joint sparse: only $k \ll n_1$ rows with nonzero elements
  - Low rank: $\mathbf{Rank}(X) = r \ll \min(k, n_2)$

$X:$



$$\mathbf{Card}(\mathrm{supp}(X)) \leq k$$

- If one knows position of the nonzero rows, the corresponding sub-matrix contains only $r(k + n_2 - r)$ degrees of freedom.

Compressed Sampling Mechanism

- Collecting $m \ll n_1 n_2$ linear measurements $y \in \mathbb{R}^m$:

$$y = \mathcal{A}(X) + z$$

  - $z \in \mathbb{R}^m$ noise vector
  - $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ sampling operator (linear mapping)

- Explicit matrix formulation: $\mathcal{A} \rightarrow \mathbf{A} \in \mathbb{R}^{m \times n_1 n_2}$

$$y = \mathbf{A} X_{\mathbf{vec}} + z$$

Compressed Sampling Mechanism

- Collecting $m \ll n_1 n_2$ linear measurements $y \in \mathbb{R}^m$:

$$y = \mathcal{A}(X) + z$$

  - $z \in \mathbb{R}^m$ noise vector
  - $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ sampling operator (linear mapping)

- Explicit matrix formulation: $\mathcal{A} \to \mathbf{A} \in \mathbb{R}^{m \times n_1 n_2}$

$$y = \mathbf{A} X_{\mathbf{vec}} + z$$

  - "Gaussian operator" $\mathcal{A}(.) \to$ $\mathbf{A}$ is i.i.d. Gaussian $\sim \mathcal{N}(0, 1/m)$

Compressed Sampling Mechanism

- Collecting $m \ll n_1 n_2$ linear measurements $y \in \mathbb{R}^m$:

$$y = \mathcal{A}(X) + z$$

  - $z \in \mathbb{R}^m$ noise vector
  - $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ sampling operator (linear mapping)

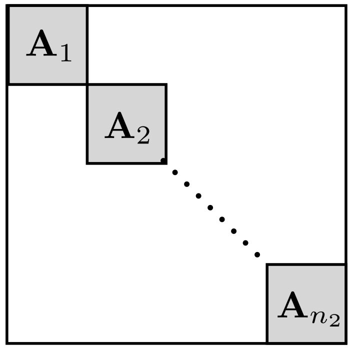
- Explicit matrix formulation: $\mathcal{A} \to \mathbf{A} \in \mathbb{R}^{m \times n_1 n_2}$

$$y = \mathbf{A} X_{\mathbf{vec}} + z$$

  - "Gaussian operator " $\mathcal{A}(.) \to$ $\mathbf{A}$ is i.i.d. Gaussian $\sim \mathcal{N}(0, 1/m)$
  - "i.i.d. Block-Diagonal" $\mathcal{A}$: Random block-diagonal $\mathbf{A}$ with i.i.d blocks (for *Distributed CS*)

    $\mathbf{A}_j \in \mathbb{R}^{\widehat{m} \times n_1} \quad \forall j \in \{1, ..., n_2\}$

    $\widehat{m} = m/n_2$: measurements per channel



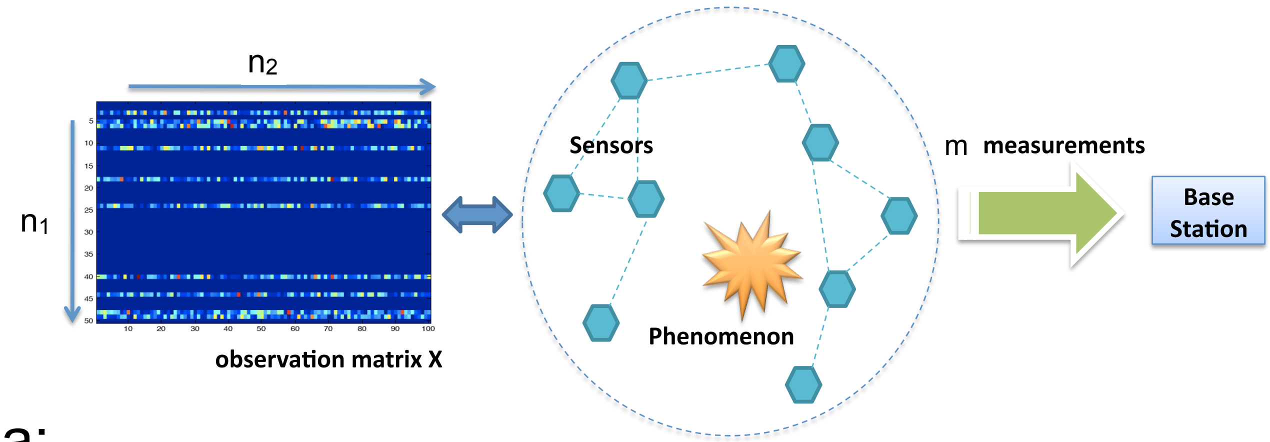
Multi-Array Signal Applications

Sensor networks

Monitoring a region which is affected by common phenomena

- Limited sources/causes & many correlated observations
  - Observations has joint-sparse representation in a basis.
  - Nonzero coefficients are linearly dependent.

[Baron et al., 2005]



<u>CS idea:</u>

- Distributed/Collaborative compressed sampling & Joint recovery
- Tradeoffs: Number of sensors v.s. complexity of each sensor

Structure-Aware Recovery (Prior Arts)

- l_2/l_1 norm minimization for joint-sparse data:

$$\arg\min_X \|X\|_{2,1} \quad \text{s.t.} \quad \|y - \mathcal{A}(X)\|_2 \leq \epsilon$$

  - Stable recovery guaranty by "Block-RIP", for Gaussian $\mathcal{A}$:

[Eldar Mishali, 2009]

$$m \gtrsim \mathcal{O}\Big(k \log(k/n_1) + kn_2\Big)$$

- Nuclear norm minimization for low-rank data:

$$\arg\min_X \|X\|_* \quad \text{s.t.} \quad \|y - \mathcal{A}(X)\|_2 \leq \epsilon$$

  - Stable recovery guaranty by "Rank-RIP", for Gaussian $\mathcal{A}$:

[Candes Plan, 2009]

$$m \gtrsim \mathcal{O}\Big(r(n_1 + n_2)\Big)$$

Structure-Aware Recovery (Prior Arts)

- l_2/l_1 norm minimization for joint-sparse data:

$$\arg\min_X \|X\|_{2,1} \quad \text{s.t.} \quad \|y - \mathcal{A}(X)\|_2 \leq \epsilon$$

  - Stable recovery guaranty by "Block-RIP", for Gaussian $\mathcal{A}$:

[Eldar Mishali, 2009]

$$m \gtrsim \mathcal{O}\Big(k\log(k/n_1) + kn_2\Big)$$

  - Increasing # channels ($n_2 \gg k\log(k/n_1)$) support recovery improves, however, for decoding the sparse coefficients, it requires $\widehat{m} \gtrsim \mathcal{O}(k)$ (log factor improvement w.r.t. l_1). Inter channels corrections neglected!

- Nuclear norm minimization for low-rank data:

$$\arg\min_X \|X\|_* \quad \text{s.t.} \quad \|y - \mathcal{A}(X)\|_2 \leq \epsilon$$

  - Stable recovery guaranty by "Rank-RIP", for Gaussian $\mathcal{A}$:

[Candes Plan, 2009]

$$m \gtrsim \mathcal{O}\Big(r(n_1 + n_2)\Big)$$

Structure-Aware Recovery (Prior Arts)

- l_2/l_1 norm minimization for joint-sparse data:

$$\arg\min_X \|X\|_{2,1} \quad \text{s.t.} \quad \|y - \mathcal{A}(X)\|_2 \leq \epsilon$$

  - Stable recovery guaranty by "Block-RIP", for Gaussian $\mathcal{A}$:

[Eldar Mishali, 2009]

$$m \gtrsim \mathcal{O}\Big(k \log(k/n_1) + kn_2\Big)$$

  - Increasing # channels ($n_2 \gg k\log(k/n_1)$) support recovery improves, however, for decoding the sparse coefficients, it requires $\widehat{m} \gtrsim \mathcal{O}(k)$ (log factor improvement w.r.t. *l₁*). Inter channels corrections neglected!

- Nuclear norm minimization for low-rank data:

$$\arg\min_X \|X\|_* \quad \text{s.t.} \quad \|y - \mathcal{A}(X)\|_2 \leq \epsilon$$

  - Stable recovery guaranty by "Rank-RIP", for Gaussian $\mathcal{A}$:

[Candes Plan, 2009]

$$m \gtrsim \mathcal{O}\Big(r(n_1 + n_2)\Big)$$

  - Sparsity of data is neglected (performance degrades as $n_1 \gg n_2$) !

LTS EPFL LTS2 ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Our Approach

Convex Optimizations for LR-JS Recovery

- "Low-rank and joint-sparse" matrix recovery by one of the following three convex minimizations:

P1:
$$\begin{aligned} \arg\min_X \quad & \|X\|_* \\ \text{subject to} \quad & \|y - \mathcal{A}(X)\|_2 \leq \epsilon, \\ & \|X\|_{2,1} \leq \gamma. \end{aligned}$$

P2:
$$\begin{aligned} \arg\min_X \quad & \|X\|_{2,1} \\ \text{subject to} \quad & \|y - \mathcal{A}(X)\|_2 \leq \epsilon, \\ & \|X\|_* \leq \tau. \end{aligned}$$

P3:
$$\begin{aligned} \arg\min_X \quad & \|X\|_{2,1} + \lambda\|X\|_* \\ \text{subject to} \quad & \|y - \mathcal{A}(X)\|_2 \leq \epsilon. \end{aligned}$$

- Solutions of P1-3 coincides for proper regularizations.

Convex Optimizations for LR-JS Recovery

- "Low-rank and joint-sparse" matrix recovery by one of the following three convex minimizations:

P1:

$$\begin{aligned} \arg\min_{X} \quad & \|X\|_* \\ \text{subject to} \quad & \|y - \mathcal{A}(X)\|_2 \leq \epsilon, \\ & \|X\|_{2,1} \leq \gamma. \end{aligned}$$

Here, we focus on P1.

P2:

$$\begin{aligned} \arg\min_{X} \quad & \|X\|_{2,1} \\ \text{subject to} \quad & \|y - \mathcal{A}(X)\|_2 \leq \epsilon, \\ & \|X\|_* \leq \tau. \end{aligned}$$

P3:

$$\begin{aligned} \arg\min_{X} \quad & \|X\|_{2,1} + \lambda \|X\|_* \\ \text{subject to} \quad & \|y - \mathcal{A}(X)\|_2 \leq \epsilon. \end{aligned}$$

- Solutions of P1-3 coincides for proper regularizations.

Theoretical Bounds

LR-JS Restricted Isometry Property

- **Definition:** *For integers k=1, 2, ... and r = 1, 2, ...,* $\mathcal{A}$ *satisfies the "restricted isometry property", if for all k-joint sparse and rank r matrices X we have,*

$$(1-\delta_{r,k})\|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq (1+\delta_{r,k})\|X\|_F^2.$$

The RIP constant $\delta_{r,k}$ *is the smallest constant for which the property above holds.*

RIP Random Sampling Operators

- **Theorem:** *Let $\mathcal{A}$ be a random mapping obeying the following concentration bound for any $X \in \mathbb{R}^{n_1 \times n_2}$ and 0< t <1,*

$$\mathcal{P}\left(\left| \|\mathcal{A}(X)\|_2^2 - \|X\|_F^2 \right| > t\|X\|_F^2 \right) \leq C \exp\left(-c\, m\right),$$

where C and c are fixed constants given t. Then, $\mathcal{A}$ satisfies RIP with constant $\delta_{r,k}$, with probability greater than $1 - Ce^{-\kappa_0 m}$, if number of measurements are greater than

$$m \geq \kappa_1 \big(k \log(n_1/k) + kr + n_2 r\big),$$

κ_0 and κ_1 are fixed constant for a given $\delta_{r,k}$.

RIP Random Sampling Operators

- **Theorem:** *Let* $\mathcal{A}$ *be a random mapping obeying the following concentration bound for any* $X \in \mathbb{R}^{n_1 \times n_2}$ *and 0< t <1,*

$$\mathcal{P}\left(\left| \|\mathcal{A}(X)\|_2^2 - \|X\|_F^2 \right| > t\|X\|_F^2 \right) \leq C \exp\left(-c\, m\right),$$

where C and c are fixed constants given t. Then, $\mathcal{A}$ *satisfies RIP with constant* $\delta_{r,k}$*, with probability greater than* $1 - Ce^{-\kappa_0 m}$*, if number of measurements are greater than*

$$m \geq \kappa_1 \big(k \log(n_1/k) + kr + n_2 r\big),$$

κ_0 *and* κ_1 *are fixed constant for a given* $\delta_{r,k}$*.*

- **Corollary:** Gaussian, Bernoulli or sub-Gaussian random $\mathcal{A}$, satisfy RIP whenever the number of the measurements scales as in above.

Reconstruction Performance

- **Theorem.** *For $\mathcal{A}$ satisfying RIP ($\delta_{6r,2k} \leq \delta^*$) and $\|z\|_2 \leq \epsilon$, the solution $\widehat{X}$ to P1 obeys the following bound:*

$$\|X - \widehat{X}\|_F \leq c\Big(\frac{\|X - X^{\#}_{r,k}\|_{2,1}}{\sqrt{k}} + \frac{\|X - X^{\#}_{r,k}\|_*}{\sqrt{r}}\Big) + c'\epsilon$$

$X^{\#}_{r,k}$ is the best rank r and k-joint sparse matrix which minimizes the error term. δ^*, c and c‘ are fixed constants.

Reconstruction Performance

- **Theorem.** *For $\mathcal{A}$ satisfying RIP ($\delta_{6r,2k} \leq \delta^*$) and $\|z\|_2 \leq \epsilon$, the solution $\widehat{X}$ to P1 obeys the following bound:*

$$\|X - \widehat{X}\|_F \leq c\left(\frac{\|X - X^{\#}_{r,k}\|_{2,1}}{\sqrt{k}} + \frac{\|X - X^{\#}_{r,k}\|_*}{\sqrt{r}}\right) + c'\epsilon$$

  $X^{\#}_{r,k}$ *is the best rank r and k-joint sparse matrix which minimizes the error term. δ^*, c and c‘ are fixed constants.*

  For sub-Gaussian measurement ensembles, if

$$m \geq \mathcal{O}\big(k \log(n_1/k) + kr + n_2 r\big)$$

  - Exact recovery for noiseless and exact LR-JS matrices ($X = X^{\#}_{r,k}$).
  - Stability against noise and “non-exact” LR-JS data.

- It outperforms previous methods for setups with $r \ll k \ll n_1$

LTS EPFL

LTS2

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Implementation

P1:
$$\boxed{\begin{aligned} \arg\min_{X} \quad & \|X\|_* \\ \text{subject to} \quad & \|y - \mathcal{A}(X)\|_2 \leq \epsilon, \\ & \|X\|_{2,1} \leq \gamma. \end{aligned}}$$

Implementation

P1: $$\boxed{\begin{aligned} \arg\min_X \quad & \|X\|_* \\ \text{subject to} \quad & \|y - \mathcal{A}(X)\|_2 \leq \epsilon, \\ & \|X\|_{2,1} \leq \gamma. \end{aligned}}$$

$f_1(X)$

$\mathcal{C}_{\epsilon,\tau}(y, \mathcal{A})$

$$P_1 = \arg\min_X \; f_1(X) + i_{\mathcal{C}}(X)$$ where, $$i_{\mathcal{C}}(X) = \begin{cases} 0 & \text{if } X \in \mathcal{C} \\ +\infty & \text{otherwise} \end{cases}$$

Implementation

P1:
$$\boxed{\begin{aligned} &\arg\min_X \ \|X\|_* \\ &\text{subject to } \|y - \mathcal{A}(X)\|_2 \le \epsilon, \\ &\qquad\qquad \|X\|_{2,1} \le \gamma. \end{aligned}}$$

$f_1(X)$

$\mathcal{C}_{\epsilon,\tau}(y, \mathcal{A})$

$$P_1 = \arg\min_X \ f_1(X) + i_{\mathcal{C}}(X) \quad \text{where,} \quad i_{\mathcal{C}}(X) = \begin{cases} 0 & \text{if } X \in \mathcal{C} \\ +\infty & \text{otherwise} \end{cases}$$

- Sum of two "lower semi-continuous" convex functions *[Combettes,Pesquet'10]*

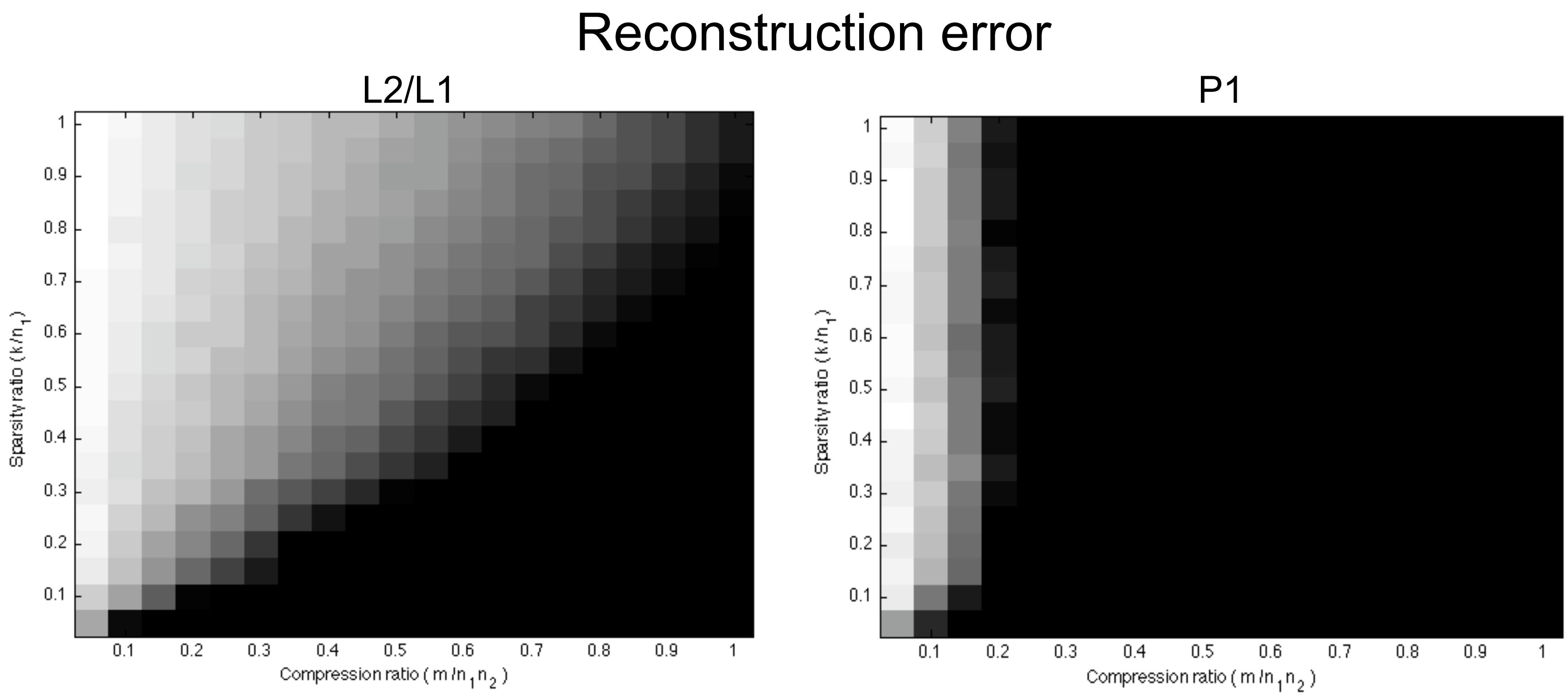
  ⇒ Douglas-Rachford algorithm *[Douglas, Rachford,1956]*

  1. $\mathrm{prox}_{\lambda f_1}(X) = \mathcal{S}_\lambda\big(\Sigma(X)\big)$, soft thresholding of the singular values.
  2. Dijkstra method for projection into intersection $\mathcal{C}_{\epsilon,\tau}(y, \mathcal{A})$. *[Boyle Dijkstra,1986]*

  \- Complexity dominated by (1), and (2) if ***A*** is not a tight frame.

Numerical Experiments

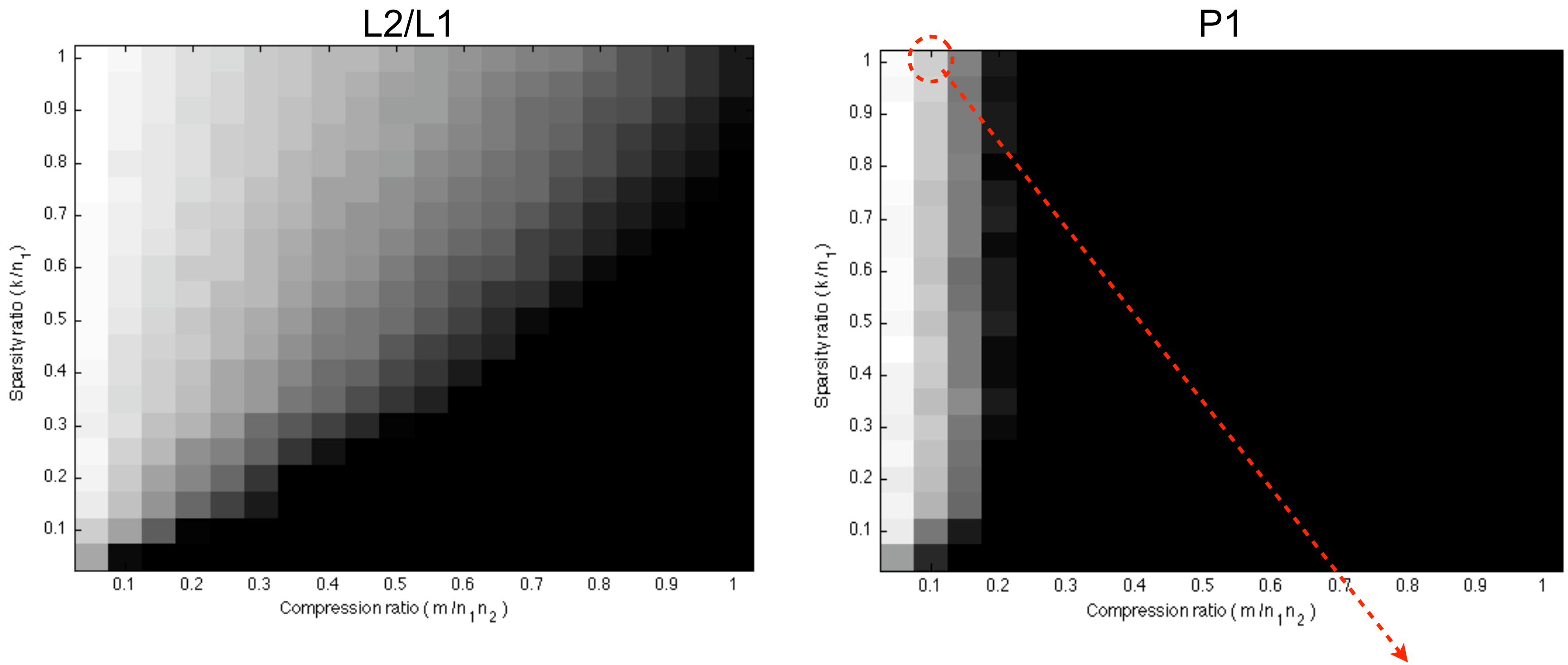
P1 v.s. L2/L1



- 40x40 random data matrix, **Rank**(X) = 2, and Gaussian $\mathcal{A}$
- P1: recovery for compression rates below sparsity ratio!

P1 v.s. L2/L1

Reconstruction error

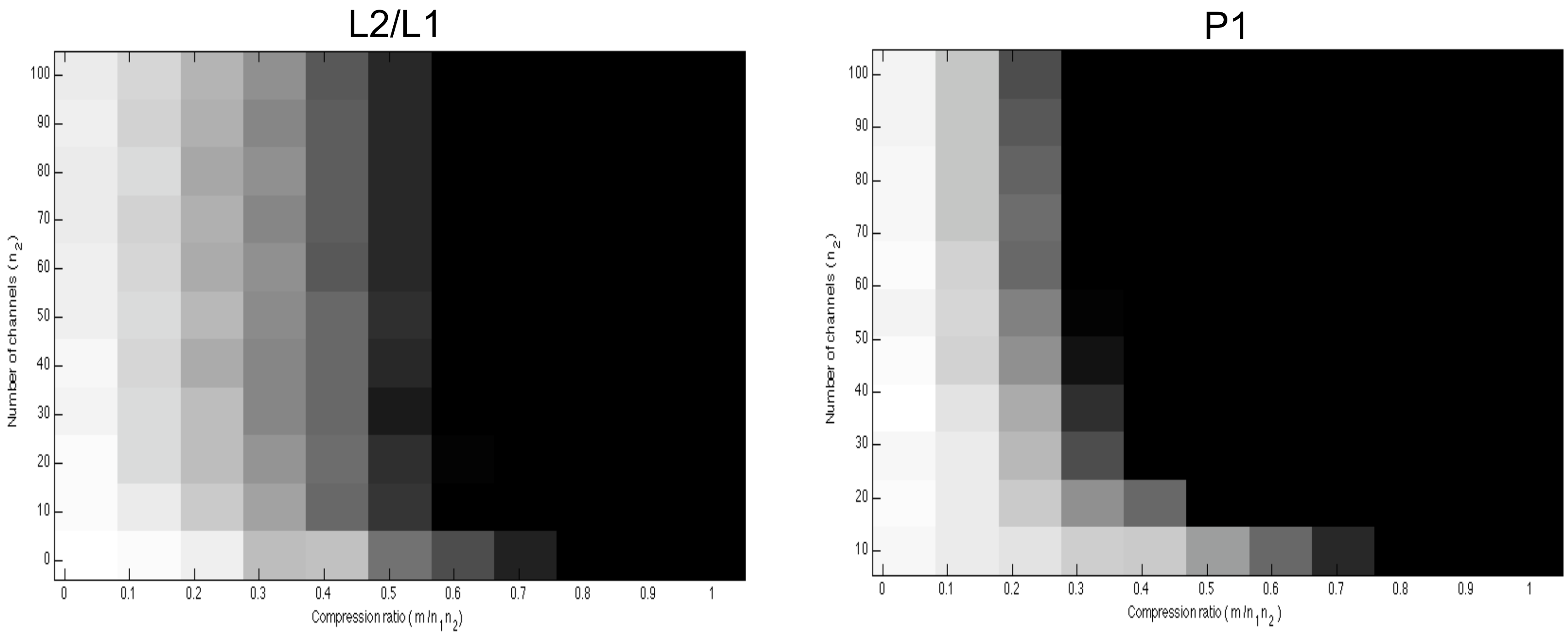


- 40x40 random data matrix, **Rank**(X) = 2, and Gaussian $\mathcal{A}$ $\quad \frac{r(n_1+n_2-r)}{n_1 n_2} \sim 0.1$
- P1: recovery for compression rates below sparsity ratio!

Limited degrees of freedom

P1 v.s. L2/L1

Reconstruction error

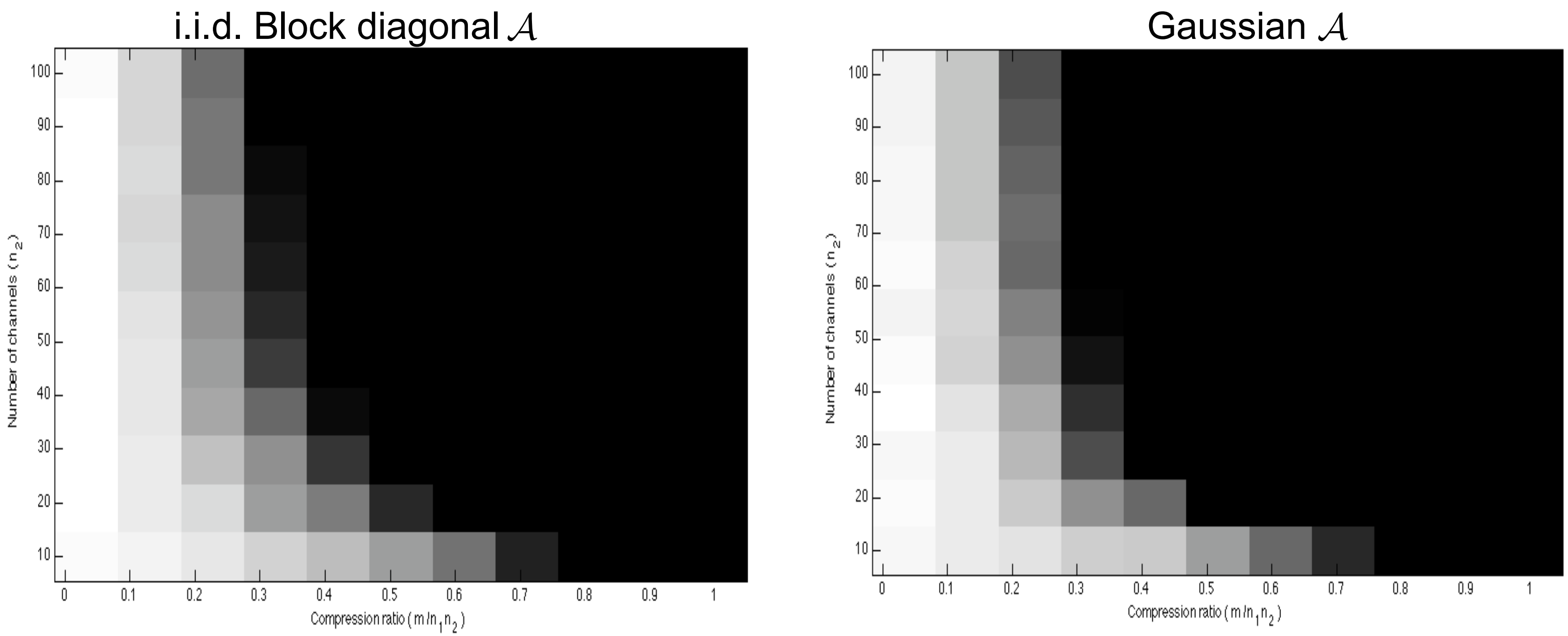


- 30xn_2 random data matrix, k=10, r = 3, and Gaussian $\mathcal{A}$
- For large n_2 , (P1) requires less measurements per channel than L2/L1.

Better n_2 v.s. $\widehat{m}$ tradeoff

Distributed v.s. Collaborative CS

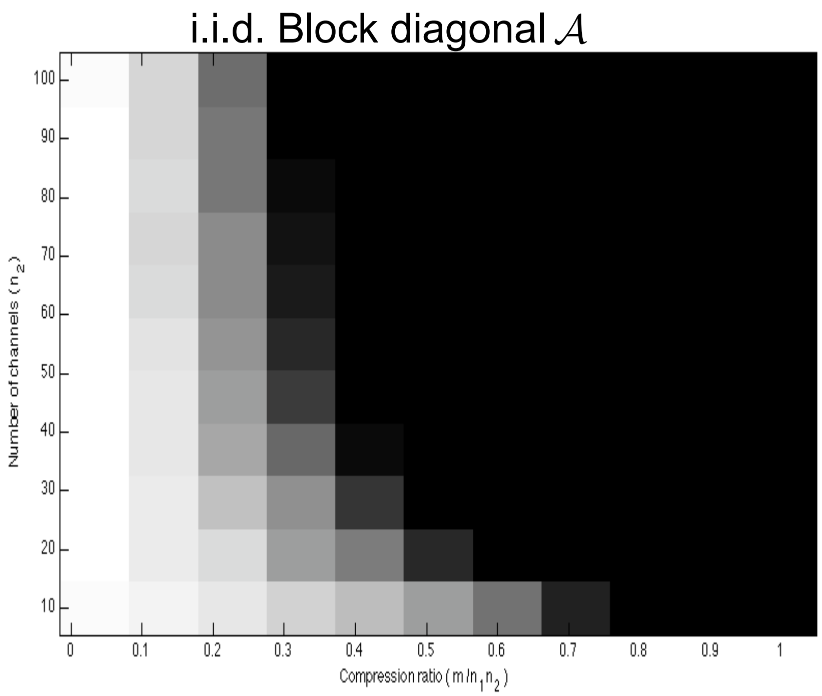
Reconstruction error

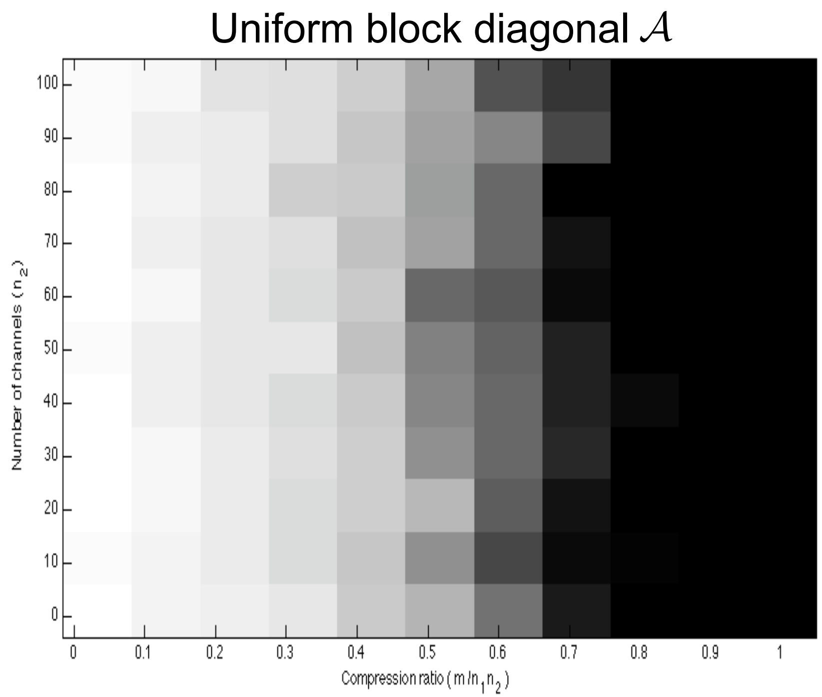


- $30 \times n_2$ random data matrix, k=10, r = 3 and P1 recovery
- Distributed sensing performs similar to dense/collaborative CS!! (e.g. good for sensor networks)

DCS v.s. MMV

Reconstruction error





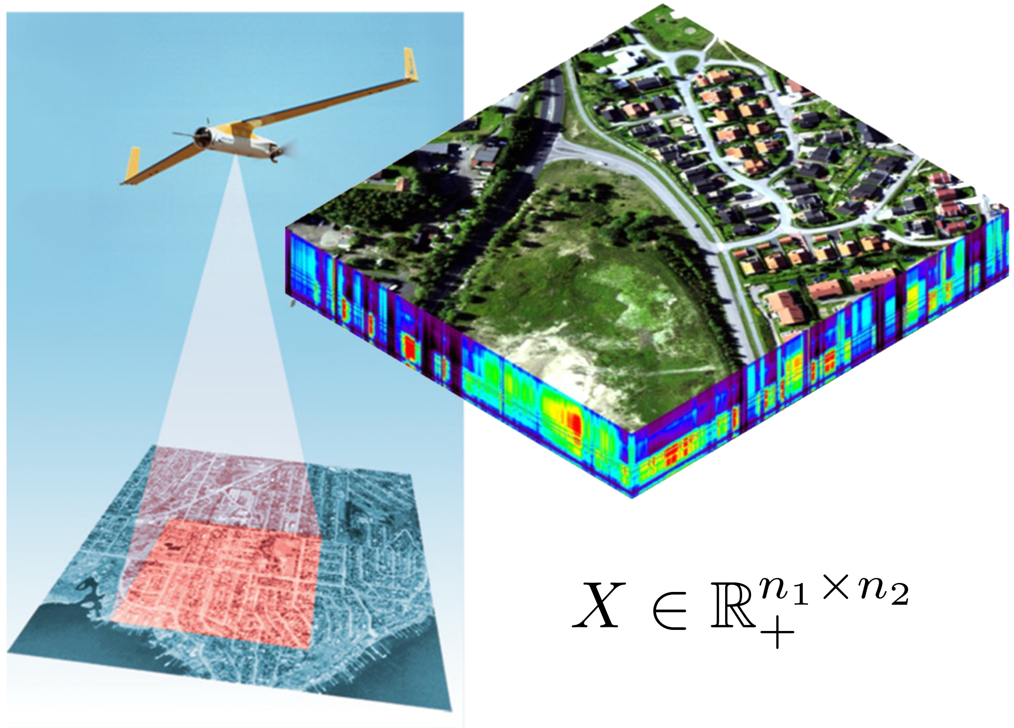
- $30 \times n_2$ random data matrix, k=10, r = 3 and P1 recovery
- For low-rank data , MMV doesn't improve by increasing the channels, as uniform sampling doesn't give many "independent measurements".

Hyperspectral Images

- A collection of hundreds of images acquired simultaneously in narrow and adjacent spectral bands/channels.

n_2: # spectral bands/channels
n_1: image resolution per channel

- HSI is generated from few "source images" based on a "*linear mixture*" model.



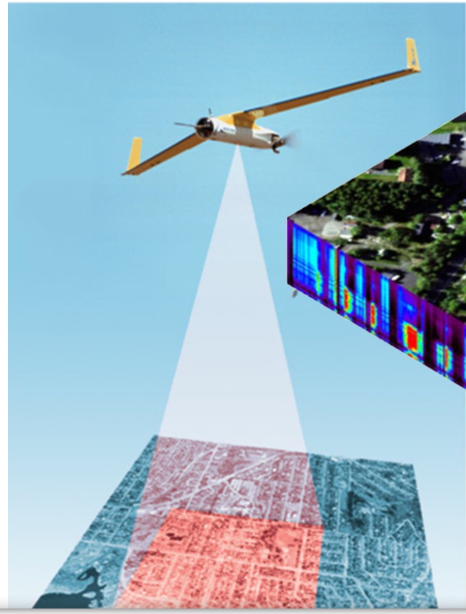
$$X \in \mathbb{R}_+^{n_1 \times n_2}$$

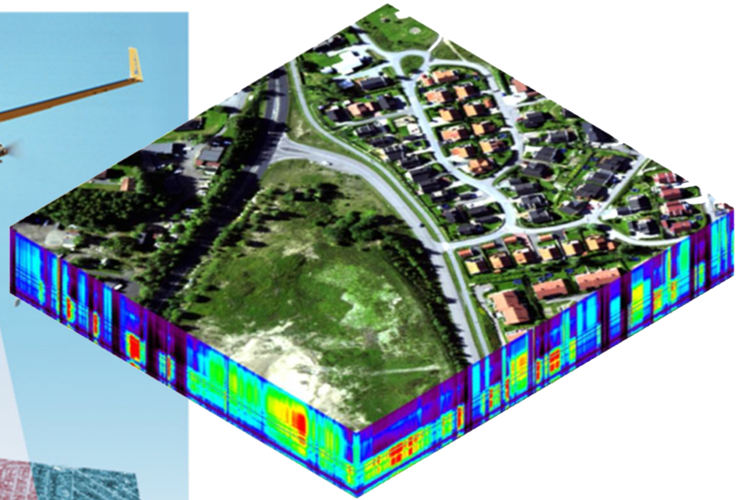
- Region is composed of few materials + Source images are sparse in Wavelet basis ⇒ HSI is "approximately" low-rank and joint-sparse.

Hyperspectral Images

- A collection of hundreds of images acquired simultaneously in narrow and adjacent spectral bands/channels.

n_2: # spectral bands/channels
n_1: image resolution per channel





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

- HSI is generated from few

As it is costly to acquire each pixel of HSI,
it becomes very interesting to use CS approach!

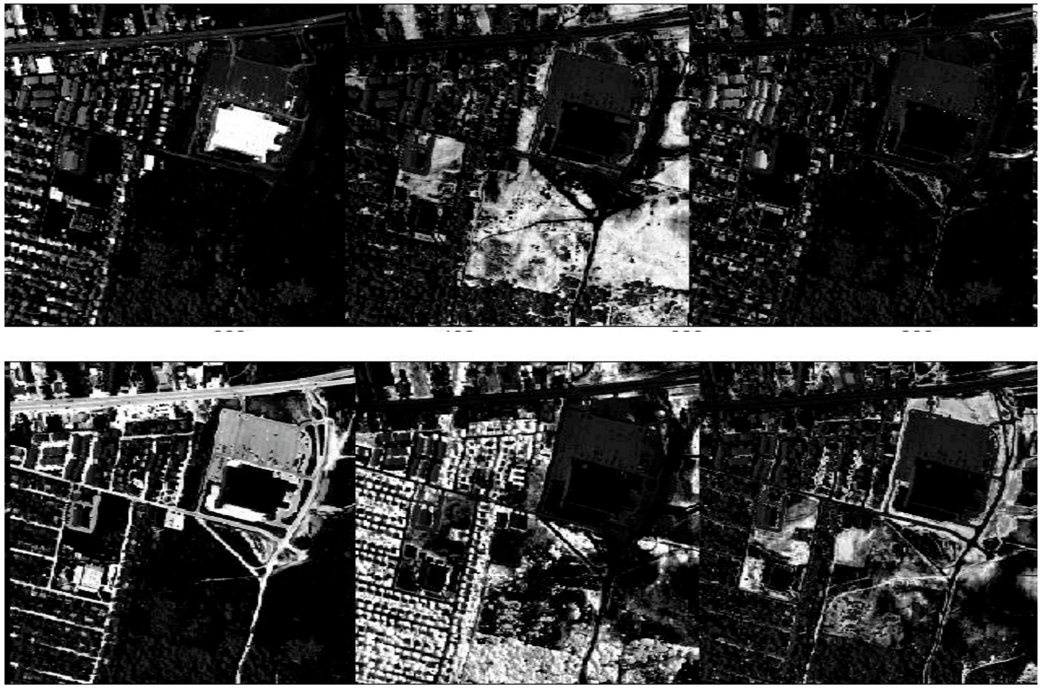
- Region is composed of few materials + Source images are sparse in Wavelet basis ⇒ HSI is “approximately” low-rank and joint-sparse.

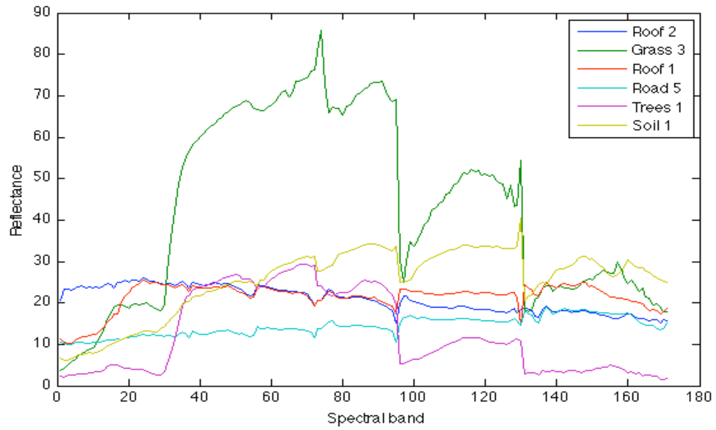
Real data with noisy measurements

- Hyperspectral Imaging (URBAN data set)

$n_1 = 256 \times 256,\ n_2 = 171,\ r \simeq 6$

Source Images





Few source images, all piecewise smooth

$\Rightarrow$ HSI cube is “approximately” LR-JS

Real data with noisy measurements (Cont.)

Original Image

Recovery for 40dB SNR

Recovery for 20dB SNR

HSI recovery from noisy CS samples using P1

$\mathcal{A}$: “random convolution” sampling op. *[Romberg 2009]*

Compression rate $m/(n_1 n2) = 1/16$

Sampling SNR	∞	40	20	10	0
Reconstruction SNR	42.7	**34.5**	**21.1**	14.1	6.6

Real data with noisy measurements (Cont.)

Original Image

Recovery for 40dB SNR

Recovery for 20dB SNR

HSI recovery from noisy CS samples using P1

$\mathcal{A}$: "random convolution" sampling op. *[Romberg 2009]*

Compression rate $m/(n_1 n2) = 1/16$

						$m/(n_1 n_2)=1/32$
Sampling SNR	∞	40	20	10	0	∞
Reconstruction SNR	42.7	**34.5**	**21.1**	14.1	6.6	**18.7**

Summary

- Joint sparse multichannel data are often forming a low-rank matrix (the nonzero coefficients are correlated).
- This model efficiently reduce degrees of freedom of data.
- A more advanced “joint-recovery” approach: The proposed convex minimizations are capturing both priors simultaneously.
- Theoretical guarantees for “stable” recovery indicate significant reduction in required number of CS measurements.
- This approach is applicable to distributed CS scenarios (no theoretical bounds yet)

Summary

- Joint sparse multichannel data are often forming a low-rank matrix (the nonzero coefficients are correlated).
- This model efficiently reduce degrees of freedom of data.
- A more advanced “joint-recovery” approach: The proposed convex minimizations are capturing both priors simultaneously.
- Theoretical guarantees for “stable” recovery indicate significant reduction in required number of CS measurements.
- This approach is applicable to distributed CS scenarios (no theoretical bounds yet)

Thnx!