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

Mohammad Golbabaee
Pierre Vandergheynst
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: $\text{Rank}(X) = r \ll \min(k, n_2)$

\[ X : \text{Card}(\text{supp}(X)) \leq k \]

• 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: $\mathcal{A} \rightarrow A \in \mathbb{R}^{m \times n_1 n_2}$

\[ y = AX_{\text{vec}} + z \]
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: $\mathcal{A} \to A \in \mathbb{R}^{m \times n_1 n_2}$
  \[ y = AX_{\text{vec}} + z \]
  - "Gaussian operator " $\mathcal{A}(.) \to A$ 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:
  \[ \mathcal{A} \rightarrow A \in \mathbb{R}^{m \times n_1 n_2} \]
  \[ y = AX_{\text{vec}} + z \]
  - “Gaussian operator” $\mathcal{A}(.) \rightarrow A$ is i.i.d. Gaussian $\sim \mathcal{N}(0, 1/m)$
  - “i.i.d. Block-Diagonal” $\mathcal{A}$: Random block-diagonal $A$ with i.i.d blocks (for Distributed CS)
    \[ A_j \in \mathbb{R}^{{\hat{m}} \times n_1} \quad \forall j \in \{1, ..., n_2\} \]
    \[ \hat{m} = m/n_2 : \text{measurements per channel} \]
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.

CS idea:
• 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:

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

- Stable recovery guaranty by “Block-RIP”, for Gaussian $A$ :

$$m \gtrsim O \left( k \log (k/n_1) + kn_2 \right)$$

[Candes Plan, 2009]

• Nuclear norm minimization for low-rank data:

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

- Stable recovery guaranty by “Rank-RIP”, for Gaussian $A$ :

$$m \gtrsim O \left( r(n_1 + n_2) \right)$$

[Candes Plan, 2009]
Structure-Aware Recovery (Prior Arts)

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

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

  - Stable recovery guaranty by “Block-RIP”, for Gaussian $A$:

    $$m \gtrsim O\left(k \log(k/n_1) + kn_2\right)$$

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

- Nuclear norm minimization for low-rank data:

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

  - Stable recovery guaranty by “Rank-RIP”, for Gaussian $A$:

    $$m \gtrsim O\left(r(n_1 + n_2)\right)$$
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 $\mathcal{A}$:

\[
m \gtrsim \mathcal{O}\left(k \log(k/n_1) + kn_2\right)
\]

- Increasing # channels ($n_2 \gg k \log(k/n_1)$) support recovery improves, however, for decoding the sparse coefficients, it requires $\hat{m} \gtrsim \mathcal{O}(k)$ (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 $\mathcal{A}$:

\[
m \gtrsim \mathcal{O}\left(r(n_1 + n_2)\right)
\]

- Sparsity of data is neglected (performance degrades as $n_1 \gg n_2$)!
Our Approach
Convex Optimizations for LR-JS Recovery

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

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

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

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

• Solutions of P1-3 coincides for proper regularizations.
• “Low-rank and joint-sparse” matrix recovery by one of the following three convex minimizations:

\[ \text{arg min}_X \|X\|_* \]
\[ \text{subject to } \|y - \mathcal{A}(X)\|_2 \leq \epsilon, \]
\[ \|X\|_{2,1} \leq \gamma. \]

P1:

\[ \text{arg min}_X \|X\|_{2,1} \]
\[ \text{subject to } \|y - \mathcal{A}(X)\|_2 \leq \epsilon, \]
\[ \|X\|_* \leq \tau. \]

P2:

\[ \text{arg min}_X \|X\|_{2,1} + \lambda\|X\|_* \]
\[ \text{subject to } \|y - \mathcal{A}(X)\|_2 \leq \epsilon. \]

P3:

Here, we focus on P1.

• Solutions of P1-3 coincides for proper regularizations.
Theoretical Bounds
LR-JS Restricted Isometry Property

• **Definition:** *For integers* $k=1, 2, \ldots$ *and* $r = 1, 2, \ldots$, *$A$ satisfies the “restricted isometry property”, if for all* $k$-joint sparse and rank $r$ *matrices* $X$ *we have,*

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

*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 $0 < t < 1$,

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

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

$$m \geq \kappa_1 \left( k \log(n_1/k) + kr + n_2r \right),$$

$\kappa_0$ and $\kappa_1$ are fixed constant for a given $\delta_{r,k}$. 
**RIP Random Sampling Operators**

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

  \[ \mathcal{P} \left( \| A(X) \|_2^2 - \| X \|_F^2 > t \| X \|_F^2 \right) \leq C \exp (-cm), \]

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

  \[ m \geq \kappa_1 \left( k \log(n_1/k) + kr + n_2r \right), \]

  $\kappa_0$ and $\kappa_1$ are fixed constant for a given $\delta_{r,k}$.

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

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

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

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

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

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

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

For sub-Gaussian measurement ensembles, if

$$m \geq \mathcal{O}(k \log(n_1/k) + kr + n_2r)$$

- 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 $r \ll k \ll n_1$
Implementation

P1:

\[
\begin{align*}
\text{arg min}_{X} & \quad \|X\|_*, \\
\text{subject to} & \quad \|y - A(X)\|_2 \leq \epsilon, \\
& \quad \|X\|_{2,1} \leq \gamma.
\end{align*}
\]
Implementation

\[ P_1 = \arg\min_X f_1(X) + i_C(X) \quad \text{where,} \quad i_C(X) = 0 \quad \text{if} \ X \in C \]
\[ +\infty \quad \text{otherwise} \]
Implementation

1. Soft thresholding of the singular values.
2. Dijkstra method for projection into intersection $C_{\epsilon,\tau}(y, A)$.

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

**P1:**

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

- Sum of two “lower semi-continuous” convex functions \([\text{Combettes, Pesquet'10}]

$\Rightarrow$ Douglas-Rachford algorithm \([\text{Douglas, Rachford,1956}]

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

\[ \arg\min_X \|X\|_* \quad \text{subject to} \quad \|y - A(X)\|_2 \leq \epsilon, \|X\|_{2,1} \leq \gamma. \]
Numerical Experiments
P1 v.s. L2/L1

Reconstruction error

- 40x40 random data matrix, \( \text{Rank}(X) = 2 \), and Gaussian \( \mathcal{A} \)
- P1: recovery for compression rates below sparsity ratio!
P1 v.s. L2/L1

Reconstruction error

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

Limited degrees of freedom
30xn₂ random data matrix, k=10, r = 3, and Gaussian \( A \)

For large \( n₂ \), (P1) requires less measurements per channel than L2/L1.

Better \( n₂ \) v.s. \( \hat{m} \) tradeoff
Distributed v.s. Collaborative CS

Reconstruction error

- 30xn2 random data matrix, k=10, r = 3 and P1 recovery
- Distributed sensing performs similar to dense/collaborative CS!! (e.g. good for sensor networks)
- 30xn^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.

  \[ n_2: \text{# spectral bands/channels} \]
  \[ n_1: \text{image resolution per channel} \]

• HSI is generated from few “source images” based on a “linear mixture” model.

• Region is composed of few materials \( \Rightarrow \) 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.

\[ n_2: \text{# spectral bands/channels} \]
\[ n_1: \text{image resolution per channel} \]

• HSI is generated from few

\[ X \in \mathbb{R}^{n_1 \times n_2} \]

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)
  \[ n_1 = 256 \times 256, \ n_2 = 171, \ r \approx 6 \]

Few source images, all piecewise smooth
⇒ HSI cube is “approximately” LR-JS
Real data with noisy measurements (Cont.)

HSI recovery from noisy CS samples using P1

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

Compression rate \( m/(n_1n_2) = 1/16 \)

<table>
<thead>
<tr>
<th>Sampling SNR</th>
<th>( \infty )</th>
<th>40</th>
<th>20</th>
<th>10</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reconstruction SNR</td>
<td>42.7</td>
<td>34.5</td>
<td>21.1</td>
<td>14.1</td>
<td>6.6</td>
</tr>
</tbody>
</table>
HSI recovery from noisy CS samples using P1

\(\mathcal{A} \) : “random convolution” sampling op. \cite{Romberg2009}

Compression rate \(\frac{m}{(n_1 n_2)} = 1/16\)

Sampling SNR

<table>
<thead>
<tr>
<th></th>
<th>(\infty)</th>
<th>40</th>
<th>20</th>
<th>10</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reconstruction SNR</td>
<td>42.7</td>
<td>34.5</td>
<td>21.1</td>
<td>14.1</td>
<td>6.6</td>
</tr>
</tbody>
</table>

\(\frac{m}{n_1 n_2} = 1/32\)

Recovery for 40dB SNR

Recovery for 20dB SNR
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!