Covariance Matching techniques for Sparsity Pattern Recovery using Compressive Measurements

Saurabh Khanna

Electrical Communication Engineering Dept.
Indian Institute of Science, Bangalore
Outline

- Overview of **sparse signal recovery**
  - Least squares problem
  - Stable solution for a linear system of equations
  - Restricted Isometry Property

- Joint Sparse Signal Recovery
  - Motivation
  - Sparse Bayesian Learning - new results
  - Covariance matching framework
  - Restricted isometry of Khatri-Rao matrices
PART I
Sparse Signal Recovery - An Overview
Least Squares

Linear system of equations:

\[ y = Ax \]

\[ y \in \mathbb{R}^m, \quad x \in \mathbb{R}^n, \quad \text{and} \quad A \in \mathbb{R}^{m \times n} \]

Overdetermined \((m > n)\)

Unique or no solution
Least Squares

Linear system of equations:

\[ y = Ax \]

\( y \in \mathbb{R}^m, \ x \in \mathbb{R}^n, \) and \( A \in \mathbb{R}^{m \times n} \)

Overdetermined \((m > n)\)

Unique or no solution
Least Squares

Linear system of equations:

\[ y = Ax \]

\( y \in \mathbb{R}^m, x \in \mathbb{R}^n, \text{ and } A \in \mathbb{R}^{m \times n} \)

Overdetermined \((m > n)\)

Unique or no solution

An approximate solution minimizes the residual error, i.e.,

\[ \hat{x}_{LS} = \arg \min_{x \in \mathbb{R}^n} \|y - Ax\|_2^2 \]
Least Squares

Linear system of equations:
\[ y = Ax \]

\( y \in \mathbb{R}^m, x \in \mathbb{R}^n, \text{ and } A \in \mathbb{R}^{m \times n} \)

Overdetermined \((m > n)\)

Unique or no solution

An approximate solution minimizes the residual error, i.e.,

\[
\hat{x}_{LS} = \arg \min_{x \in \mathbb{R}^n} ||y - Ax||_2^2
\]

\[
\hat{x}_{LS} = \left( A^T A \right)^{-1} A^T y
\]

least squares solution
Least Squares

Linear system of equations:

\[ y = Ax \]

\( y \in \mathbb{R}^m, x \in \mathbb{R}^n, \text{ and } A \in \mathbb{R}^{m \times n} \)

Overdetermined \((m > n)\)

Unique or no solution

An approximate solution minimizes the residual error, i.e.,

\[ \hat{x}_{\text{LS}} = \arg \min_{x \in \mathbb{R}^n} ||y - Ax||_2^2 \]

\[ \hat{x}_{\text{LS}} = (A^T A)^{-1} A^T y \]

Least squares solution is unique and exists if \( A \) has full column rank
Least Squares using Perturbed Measurements

Let $\mathbf{x}^*$ be the ground truth, i.e., $\mathbf{y} = \mathbf{A}\mathbf{x}^*$
Least Squares using Perturbed Measurements

Let $\mathbf{x}^*$ be the ground truth, i.e., $\mathbf{y} = \mathbf{A}\mathbf{x}^*$

Perturbed measurements: $\tilde{\mathbf{y}} = \mathbf{A}\mathbf{x}^* + \mathbf{e}$
Least Squares using Perturbed Measurements

Let \( \mathbf{x}^* \) be the ground truth, i.e., \( \mathbf{y} = \mathbf{A}\mathbf{x}^* \)

Perturbed measurements: \( \tilde{\mathbf{y}} = \mathbf{A}\mathbf{x}^* + \mathbf{e} \)

Least squares estimate: \( \hat{\mathbf{x}}_{\text{LS}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\tilde{\mathbf{y}} = \mathbf{A}^\dagger\tilde{\mathbf{y}} \)
Least Squares using Perturbed Measurements

Let $x^*$ be the ground truth, i.e., $y = Ax^*$

Perturbed measurements: $\tilde{y} = Ax^* + e$

Least squares estimate: $\hat{x}_{LS} = (A^T A)^{-1} A^T \tilde{y} = A^\dagger \tilde{y}$

How far is $\hat{x}_{LS}$ from $x^*$?
Least Squares using Perturbed Measurements

Let $x^*$ be the ground truth, i.e., $y = Ax^*$

Perturbed measurements: $\tilde{y} = Ax^* + e$

Least squares estimate: $\hat{x}_{LS} = (A^T A)^{-1} A^T \tilde{y} = A^\dagger \tilde{y}$

How far is $\hat{x}_{LS}$ from $x^*$?

$$||\hat{x}_{LS} - x^*||_2 = ||A^\dagger \tilde{y} - x^*||_2 = ||(A^T A)^{-1} A^T (Ax^* + e) - x^*||_2$$

$$= ||A^\dagger e||_2 \leq ||A^\dagger||_2 ||e||_2$$

$$||A^\dagger||_2 \leq \frac{1}{\lambda_{\min}(A^T A)} \sqrt{\frac{\lambda_{\max}(A^T A)}{\lambda_{\min}(A^T A)}}$$
Least Squares using Perturbed Measurements

Let $\mathbf{x}^*$ be the ground truth, i.e., $\mathbf{y} = \mathbf{A}\mathbf{x}^*$

Perturbed measurements: $\tilde{\mathbf{y}} = \mathbf{A}\mathbf{x}^* + \mathbf{e}$

Least squares estimate: $\hat{\mathbf{x}}_{LS} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \tilde{\mathbf{y}} = \mathbf{A}^\dagger \tilde{\mathbf{y}}$

How far is $\hat{\mathbf{x}}_{LS}$ from $\mathbf{x}^*$?

$$
||\hat{\mathbf{x}}_{LS} - \mathbf{x}^*||_2 = ||\mathbf{A}^\dagger \tilde{\mathbf{y}} - \mathbf{x}^*||_2 = ||(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T (\mathbf{A}\mathbf{x}^* + \mathbf{e}) - \mathbf{x}^*||_2 \\
= ||\mathbf{A}^\dagger \mathbf{e}||_2 \leq ||\mathbf{A}^\dagger||_2 ||\mathbf{e}||_2
$$

$$
||\mathbf{A}^\dagger||_2 \leq \frac{1}{\lambda_{\min}(\mathbf{A}^T\mathbf{A})} \sqrt{\frac{\lambda_{\max}(\mathbf{A}^T\mathbf{A})}{\lambda_{\min}(\mathbf{A}^T\mathbf{A})}} \rightarrow \text{condition no. of } \mathbf{A}^T\mathbf{A}
$$

Smaller condition number of $\mathbf{A}^T\mathbf{A}$ implies lesser sensitivity to perturbations
**Sparse Signal Recovery**

\[
y = A x + w
\]

**Goal:** Recover unknown \( k \)-sparse vector \( x \) from \( y \)

- \( y \) : measurement vector \((m \times 1)\)
- \( A \) : measurement matrix \((m \times n)\)
- \( x \) : \( k \)-sparse vector \((n \times 1)\)
- \( w \) : noise \((m \times 1)\)
Sparse Signal Recovery

\[
\begin{align*}
\mathbf{y} & = \mathbf{A} \mathbf{x} + \mathbf{w} \\
\text{measurement vector} & (m \times 1) \\
\text{measurement matrix} & (m \times n) \\
\text{k-sparse vector} & (n \times 1) \\
\text{noise} & (m \times 1)
\end{align*}
\]

**Goal:** Recover unknown \(k\)-sparse vector \(\mathbf{x}\) from \(\mathbf{y}\)

Two step recovery:

(i) Recover support \(S\) (indices of nonzero entries in \(\mathbf{x}\))

(ii) Recover \(\mathbf{x}_S\) using least squares on the reduced system:

\[
\mathbf{y} = \mathbf{A}_S \mathbf{x}_S + \mathbf{w} \quad \text{overdetermined if } k > m
\]
Sparse Signal Recovery

Goal: Recover unknown $k$-sparse vector $\mathbf{x}$ from $\mathbf{y}$

Two step recovery:

(i) Recover support $S$ (indices of nonzero entries in $\mathbf{x}$)
(ii) Recover $\mathbf{x}_S$ using least squares on the reduced system:

$$\mathbf{y} = \mathbf{A}_S \mathbf{x}_S + \mathbf{w}$$

overdetermined if $k > m$

Stable recovery of $\mathbf{x}_S$ if condition no. of $\mathbf{A}_S^T \mathbf{A}_S \approx 1$
Candes and Tao, 2004

A matrix $\mathbf{A}$ is said to satisfy the Restricted Isometry Property (RIP) of order $k$, if there exists a constant $\delta \in (0, 1)$ such that

$$(1 - \delta) \|\mathbf{z}\|_2^2 \leq \|\mathbf{Az}\|_2^2 \leq (1 + \delta) \|\mathbf{z}\|_2^2$$

for all $k$-sparse vectors $\mathbf{z} \in \mathbb{R}^n$.

The smallest $\delta$ is the $k^{th}$ order restricted isometry constant ($k$-RIC) of $\mathbf{A}$. 
Restricted Isometry Property

Candes and Tao, 2004

A matrix $\mathbf{A}$ is said to satisfy the Restricted Isometry Property (RIP) of order $k$, if there exists a constant $\delta \in (0, 1)$ such that

$$
(1 - \delta) \| \mathbf{z} \|_2^2 \leq \| \mathbf{A} \mathbf{z} \|_2^2 \leq (1 + \delta) \| \mathbf{z} \|_2^2
$$

for all $k$-sparse vectors $\mathbf{z} \in \mathbb{R}^n$.

The smallest $\delta$ is the $k^{th}$ order restricted isometry constant ($k$-RIC) of $\mathbf{A}$.

Alternate interpretations:

- $1 - \delta^A_k \leq \frac{\mathbf{z}^T \mathbf{A}^T \mathbf{A} \mathbf{z}}{\mathbf{z}^T \mathbf{z}} \leq 1 + \delta^A_k \quad \forall \ k$-sparse $\mathbf{z}$

- Eigenvalues of $\mathbf{A}^T_S \mathbf{A}_S$ lie in $[1 - \delta^A_k, 1 + \delta^A_k]$ for all supports $S$, $|S| \leq k$

- Condition no. of $\mathbf{A}^T_S \mathbf{A}_S$ is at most $\frac{1 + \delta^A_k}{1 - \delta^A_k}$ for all supports $S$, $|S| \leq k$
Sparse Signal Recovery

\[ y = A x + w \]

\[ \begin{align*}
    &\text{y} \quad \text{measurement vector} \\
    &\text{A} \quad \text{measurement matrix} \\
    &\text{x} \quad \text{k-sparse vector} \\
    &\text{w} \quad \text{noise}
\end{align*} \]

**Goal**: Recover unknown \( k \)-sparse vector \( x \) from \( y \)

Two step recovery:

(i) Recover support \( S \) (indices of nonzero entries in \( x \))

(ii) Recover \( x_S \) using least squares on the reduced system:

\[ y = A_S x_S + w \quad \text{overdetermined if } k > m \]

Condition no. of \( A_S^T A_S \approx 1 \) guarantees stable recovery of \( x_S \)
Sparse Signal Recovery

Goal: Recover unknown $k$-sparse vector $\mathbf{x}$ from $\mathbf{y}$

Two step recovery:

(i) Recover support $S$ (indices of nonzero entries in $\mathbf{x}$)

(ii) Recover $\mathbf{x}_S$ using least squares on the reduced system:

$$\mathbf{y} = \mathbf{A}_S \mathbf{x}_S + \mathbf{w} \quad \text{overdetermined if } k > m$$

Condition no. of $\mathbf{A}_S^T \mathbf{A}_S \approx 1$ guarantees stable recovery of $\mathbf{x}_S$
Uniqueness under noiseless measurements

\[ y = A \times x \]

**RIP based guarantee for unique solution**

If \( A \) satisfies \( \delta_{2k}^A < 1 \), then the noiseless sparse signal recovery problem has a unique \( k \)-sparse solution.
Uniqueness under noiseless measurements

\[
y = A x
\]

measurement vector \((m \times 1)\)

measurement matrix \((m \times n)\)

k-sparse vector \((n \times 1)\)

RIP based guarantee for unique solution

If \(A\) satisfies \(\delta_{2k}^A < 1\), then the noiseless sparse signal recovery problem has a unique \(k\)-sparse solution.
Uniqueness under noiseless measurements

\[ y = A \begin{bmatrix} \mathbf{x} \end{bmatrix} \]

If \( A \) satisfies \( \delta^A_{2k} < 1 \), then the noiseless sparse signal recovery problem has a unique \( k \)-sparse solution.

\[ z^T A^T A z \geq (1 - \delta^A_{2k}) \|z\|_2^2 > 0 \text{ for all } 2k\text{-sparse } z, \quad (\implies 2k \text{ sparse vectors NOT allowed in } \operatorname{Null}(A)! \)
Uniqueness under noiseless measurements

\[ y = \text{measurement vector} \quad (m \times 1) \]
\[ A = \text{measurement matrix} \quad (m \times n) \]
\[ x = \text{k-sparse vector} \quad (n \times 1) \]

RIP based guarantee for unique solution

If \( A \) satisfies \( \delta_{2k}^A < 1 \), then the noiseless sparse signal recovery problem has a unique \( k \)-sparse solution.

- \( z^T A^T A z \geq (1 - \delta_{2k}^A) \|z\|_2^2 > 0 \) for all \( 2k \)-sparse \( z \), \( \implies \ 2k \text{ sparse vectors NOT allowed in Null}(A)! \)
- Let \( x_1, x_2 \) be distinct \( k \)-sparse solutions, then \( y = Ax_1 = Ax_2 \). Thus, \( A(x_1 - x_2) = 0 \). \text{Contradiction!}.
RIP based recovery guarantees

<table>
<thead>
<tr>
<th>Recovery algorithm</th>
<th>RIP based sufficient conditions for successful signal reconstruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell_1$-norm minimization</td>
<td>$\delta_k(A) \leq 0.307$</td>
</tr>
<tr>
<td>OMP</td>
<td>$\delta_{k+1}(A) \leq \frac{\sqrt{4k+1} - 1}{2k}$</td>
</tr>
<tr>
<td>Co-SAMP</td>
<td>$\delta_{4k}(A) \leq 0.1$</td>
</tr>
<tr>
<td>IHT</td>
<td>$\delta_{3k}(A) \leq \frac{1}{\sqrt{32}}$</td>
</tr>
<tr>
<td>Basis Pursuit</td>
<td>$\delta_{2k}(A) + 3\delta_{3k} \leq 1$</td>
</tr>
<tr>
<td>Subspace Pursuit</td>
<td>$\delta_{3k} \leq 0.139$</td>
</tr>
</tbody>
</table>
Finding exact $k$-RIC is NP hard

For any $A$, $k$-RIC of $A$ is the smallest $\delta \in (0, 1)$ such that

$$1 - \delta \leq \lambda_i \left( A_S^T A_S \right) \leq 1 + \delta,$$

for all supports $S$, $|S| \leq k$.

Unfortunately, finding the exact $k$-RIC of a matrix is NP hard! [Tillman & Pfetsch, 2013]

Hence, we look for upper bounds for $k$-RIC of $A$.
Restricted Isometry of Gaussian matrices

Gaussian RIP condition by Candès and Tao, 2005

Let $A$ be an $m \times n$ random matrix with i.i.d. $\mathcal{N}(0, 1)$ entries. Then,

$$
\mathbb{P} \left( \delta_k \left( \frac{A}{\sqrt{m}} \right) \geq \delta \right) \leq \frac{2}{(en/k)^k},
$$

provided $m \geq c \left( \frac{k \log \frac{en}{k}}{\delta^2} \right)$, where $c > 0$ is an absolute numerical constant.
Restricted Isometry of Gaussian matrices

Gaussian RIP condition by Candès and Tao, 2005

Let $A$ be an $m \times n$ random matrix with i.i.d. $\mathcal{N}(0, 1)$ entries. Then,

$$\mathbb{P} \left( \delta_k \left( \frac{A}{\sqrt{m}} \right) \geq \delta \right) \leq \frac{2}{(en/k)^k},$$

provided $m \geq c \left( \frac{k \log \frac{en}{k}}{\delta^2} \right)$, where $c > 0$ is an absolute numerical constant.

Result extends to subgaussian random matrices as well.
Recap - sparse signal recovery

Restricted Isometry Property (RIP) of the measurement matrix guarantees

- Stability of sparse solution in noisy measurement case

Gaussian random matrices of size $m \times n$ satisfy $k$-RIP with high probability if $m \geq \mathcal{O}\left(k \log \frac{n}{k}\right)$. 
PART II
Joint Sparse Signal Recovery
Joint Sparse Support Recovery

- **Measurement model:** \( Y = AX + W \)

- Columns of \( X \) are **jointly sparse** (same nonzero support).
- \( k = \) no. of nonzero rows in \( X \)
- No inter/intra vector correlations in \( X \)
Joint Sparse Support Recovery

- Measurement model: $Y = AX + W$

- Columns of $X$ are **jointly sparse** (same nonzero support).
- $k = \text{no. of nonzero rows in } X$
- No inter/intra vector correlations in $X$

**Support($X$)**
Joint Sparse Support Recovery

- Measurement model: $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$

  \[ m \begin{bmatrix} \mathbf{Y} \\ \mathbf{L} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{L} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{L} \end{bmatrix} + \begin{bmatrix} \mathbf{W} \\ \mathbf{L} \end{bmatrix} \]

- Columns of $\mathbf{X}$ are **jointly sparse** (same nonzero support).
- $k = \text{no. of nonzero rows in } \mathbf{X}$
- No inter/intra vector correlations in $\mathbf{X}$

**Multiple Measurement Vector (MMV) problem** vs. **Joint Sparse Support Recovery (JSSR)**

| Recover entire $\mathbf{X}$ from $\{\mathbf{Y}, \mathbf{A}, \sigma^2\}$ | Recover support($\mathbf{X}$) from $\{\mathbf{Y}, \mathbf{A}, \sigma^2\}$ |
Applications

Joint sparse signals frequently arise in multi-sensor signal processing.

Joint sparse vectors

Signals from $L$ sensors

Sparse event localization

Cooperative spectrum sensing

Sparse 2D field reconstruction
Support Recovery via Sparse Bayesian Learning

\[ Y = AX + W \]
Support Recovery via Sparse Bayesian Learning

\[ Y = AX + W \]
Support Recovery via Sparse Bayesian Learning

\[ \mathbf{Y} = \mathbf{A} \mathbf{X} + \mathbf{W} \]

- \( \mathbf{x}_j \) \text{i.i.d.} \( \mathcal{N}(0, \Gamma) \), \( \Gamma = \text{diag}(\gamma) \)
Support Recovery via Sparse Bayesian Learning

\[ \mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W} \]

- \[ \mathbf{x}_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \Gamma), \quad \Gamma = \text{diag}(\mathbf{\gamma}) \]
- Support(\(\mathbf{\gamma}\)) = support(\(\mathbf{x}_j\))
- Common covariance induces joint sparsity in \(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L\)
Support Recovery via Sparse Bayesian Learning

- \( Y = AX + W \)
  - \( x_j \) i.i.d. \( \mathcal{N}(0, \Gamma) \), \( \Gamma = \text{diag}(\gamma) \)
  - Support(\( \gamma \)) = support(\( x_j \))
  - Common covariance induces joint sparsity in \( x_1, x_2, \ldots, x_L \)
  - \( y_j \sim \mathcal{N}(0, \sigma^2 I + A\Gamma A^T) \)

MSBL algorithm:
- \( \hat{\gamma} = \arg\max_{\gamma \in \mathbb{R}^n} \log p(Y; \gamma) \)
  - \( \hat{\gamma} \) found using Expectation Maximization (EM) procedure

SOMP support recovery phase transition

Recoverable support size \( k \) grows as \( O(m^2) \) in MSBL!
Support Recovery via Sparse Bayesian Learning

\[ \mathbf{Y} = \mathbf{A} \mathbf{X} + \mathbf{W} \]

- \( \mathbf{x}_j \) \( i.i.d. \sim \mathcal{N}(0, \Gamma), \; \Gamma = \text{diag}(\gamma) \)
- Support(\( \gamma \)) = support(\( \mathbf{x}_j \))
- Common covariance induces joint sparsity in \( \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L \)
- \( \mathbf{y}_j \sim \mathcal{N}(0, \sigma^2 \mathbf{I} + \mathbf{A} \Gamma \mathbf{A}^T) \)

MSBL algorithm:

\[ \hat{\gamma} = \arg\max_{\gamma \in \mathbb{R}^n_+} \log p(\mathbf{Y}; \gamma) \]

- \( \hat{\gamma} \) found using Expectation Maximization (EM) procedure
Support Recovery via Sparse Bayesian Learning

- \( \mathbf{Y} = \mathbf{A} \mathbf{X} + \mathbf{W} \)
  - \( \mathbf{x}_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \Gamma) \), \( \Gamma = \text{diag}(\gamma) \)
  - Support(\( \gamma \)) = support(\( \mathbf{x}_j \))
  - Common covariance induces joint sparsity in \( \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L \)
  - \( \mathbf{y}_j \sim \mathcal{N}(0, \sigma^2 \mathbf{I} + \mathbf{A} \Gamma \mathbf{A}^T) \)

- MSBL algorithm:
  \[ \hat{\gamma} = \arg\max_{\gamma \in \mathbb{R}_+^n} \log p(\mathbf{Y}; \gamma) \]
  - \( \hat{\gamma} \) found using Expectation Maximization (EM) procedure

\[ \text{Recoverable support size } k \text{ grows as } \mathcal{O}(m^2) \text{ in MSBL!} \]
Support Recovery via Sparse Bayesian Learning

Sufficient conditions for support recovery

Suppose $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L$ are i.i.d. zero mean Gaussian vectors with common support $S^*, |S^*| \leq k$, and with variances of the nonzero entries in $[\gamma_{\text{min}}, \gamma_{\text{max}}]$. Then,

$$
\mathbb{P} (\text{support}(\hat{\gamma}) \neq S^*) \leq \exp \left(-\frac{\eta}{8} L\right), \quad \text{if}
$$

**Condition 1:** Self Khatri-Rao product $\mathbf{A} \odot \mathbf{A}$ satisfies $2k$-RIP, i.e.,

$$(1 - \delta_{2k}^\odot) \|\mathbf{z}\|_2^2 \leq \|(\mathbf{A} \odot \mathbf{A})\mathbf{z}\|_2^2 \leq (1 + \delta_{2k}^\odot) \|\mathbf{z}\|_2^2$$

holds for all $2k$ or less sparse vectors $\mathbf{z}$, for some $\delta_{2k}^\odot \in (0, 1)$.

**Condition 2:** $L \geq \frac{c_1 k \log n}{\eta}$, where $\eta = \frac{m}{8k} \left(\frac{\gamma_{\text{min}}}{\sigma^2 + \gamma_{\text{max}}}\right)^2 \frac{(1 - \delta_{2k}^\odot)}{\sup_{S:|S|=2k} \|\mathbf{A}_S^T \mathbf{A}_S\|_2}$,

and $c_1$ is an absolute positive constants.

* The above result holds for column normalized $\mathbf{A}$.
New interpretation of MSBL cost function

- MSBL’s log-likelihood cost:

\[- \log p(Y; \gamma) = - \sum_{j=1}^{L} \log \mathcal{N}(y_j; 0, \sigma^2 I_m + A \Gamma A^T)\]
New interpretation of MSBL cost function

- MSBL's log-likelihood cost:

\[- \log p(Y; \gamma) = \sum_{j=1}^{L} \log \mathcal{N}(y_j; 0, \sigma^2 I_m + \mathbf{A} \Gamma \mathbf{A}^T)\]

\[\propto \log |\sigma^2 I_m + \mathbf{A} \Gamma \mathbf{A}^T| + \text{trace} \left( (\sigma^2 I_m + \mathbf{A} \Gamma \mathbf{A}^T)^{-1} \left( \frac{1}{L} YY^T \right) \right)\]
New interpretation of MSBL cost function

- MSBL’s log-likelihood cost:

\[ -\log p(Y; \gamma) = -\sum_{j=1}^{L} \log \mathcal{N}(y_j; 0, \sigma^2 I_m + A\Gamma A^T) \]

\[ \propto \log |\sigma^2 I_m + A\Gamma A^T| + \text{trace} \left( \left( \sigma^2 I_m + A\Gamma A^T \right)^{-1} \left( \frac{1}{L} YY^T \right) \right) \]

Log Det Bregman matrix divergence between matrices \( X, Y \in S^m_{++} \) is defined as

\[ \mathcal{D}_\phi(X, Y) \triangleq \text{trace}(XY^{-1}) - \log |XY^{-1}| - m \]
New interpretation of MSBL cost function

- **MSBL's log-likelihood cost:**

\[
- \log p(Y; \gamma) = - \sum_{j=1}^{L} \log \mathcal{N}(y_j; 0, \sigma^2 I_m + \mathbf{A} \Gamma \mathbf{A}^T)
\]

\[
\propto \log |\sigma^2 I_m + \mathbf{A} \Gamma \mathbf{A}^T| + \text{trace} \left( \left( \sigma^2 I_m + \mathbf{A} \Gamma \mathbf{A}^T \right)^{-1} \left( \frac{1}{L} YY^T \right) \right)
\]

\[
\propto D_{\text{Bregman}} \left( \frac{1}{L} YY^T, \sigma^2 I_m + \mathbf{A} \Gamma \mathbf{A}^T \right) + \text{constant terms}
\]

Log Det Bregman Matrix Div.
New interpretation of MSBL cost function

- MSBL’s log-likelihood cost:

\[- \log p(Y; \gamma) = - \sum_{j=1}^{L} \log \mathcal{N}(y_j; 0, \sigma^2 I_m + A\Gamma A^T)\]

\[\propto \log |\sigma^2 I_m + A\Gamma A^T| + \text{trace} \left( \left( \sigma^2 I_m + A\Gamma A^T \right)^{-1} \left( \frac{1}{L} YY^T \right) \right)\]

\[\propto D_{\text{Bregman}} \left( \frac{1}{L} YY^T, \sigma^2 I_m + A\Gamma A^T \right) + \text{constant terms} \]

Log Det Bregman Matrix Div.

- MSBL optimization minimizes \(D_{\text{Bregman}} \left( \frac{1}{L} YY^T, \sigma^2 I_m + A\Gamma A^T \right)\)

- Can we use some other matrix divergence?
Covariance Matching Framework for Support Recovery

- MMV model: \( \mathbf{Y} = \mathbf{AX} + \mathbf{W} \)
  - \( \mathbf{x}_j \sim \mathcal{N}(0, \text{diag}(\gamma)) \)
  - \( \mathbf{y}_j \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T) \)
Covariance Matching Framework for Support Recovery

- **MMV model:** \( Y = AX + W \)
  - \( x_j \sim \mathcal{N}(0, \text{diag}(\gamma)) \)
  - \( y_j \sim \mathcal{N}(0, \sigma^2 I_m + A\Gamma A^T) \)

- **Covariance matrices:**
  - Empirical \( R_Y = \frac{1}{L} YY^T \)
  - Parameterized \( \Sigma_{\gamma} = \sigma^2 I_m + A\Gamma A^T \)
Covariance Matching Framework for Support Recovery

- **MMV model:** $Y = AX + W$
  - $x_j \sim \mathcal{N}(0, \text{diag}(\gamma))$
  - $y_j \sim \mathcal{N}(0, \sigma^2 I_m + A\Gamma A^T)$

- **Covariance matrices:**
  - Empirical $R_Y = \frac{1}{L} Y Y^T$
  - Parameterized $\Sigma_\gamma = \sigma^2 I_m + A\Gamma A^T$

- **Covariance Matching Principle:**
  \[
  \hat{\gamma} = \arg \min_{\gamma \in \mathbb{R}_+^n} \text{distance}(R_Y, \sigma^2 I + A\Gamma A^T)
  \]

  \[
  \text{support}(X) = \text{support}(\hat{\gamma})
  \]
Examples of covariance matching algorithms

- **Frobenius matrix norm** based covariance matching (Co-LASSO)

\[
\hat{\gamma} = \arg \min_{\gamma \in \mathbb{R}^n_+} \left\| \frac{1}{L} YY^T - (\sigma^2 I + A \Gamma A^T) \right\|_F^2 + \lambda \|\gamma\|_1
\]

- **Main features of the Co-LASSO** [Pal & Vaidyanathan, 2013]
  - $\ell_1$ norm penalty promotes recovery of sparse $\gamma$
  - Convex objective
  - Very high memory requirements
Examples of covariance matching algorithms

- Log-Det Bregman matrix divergence based covariance matching (MSBL)

\[ \hat{\gamma} = \arg\min_{\gamma \in \mathbb{R}_+^n} \log \left| \sigma^2 I + A \Gamma A^T \right| + \text{tr} \left( \left( \sigma^2 I + A \Gamma A^T \right)^{-1} \left( \frac{1}{L} YY^T \right) \right) \]

- Main features of MSBL [Wipf & Rao, 2007]
  - Non-convex objective
  - Expectation Maximization based implementation (slow!)
  - Good performance
Examples of covariance matching algorithms

- $\alpha$-Rényi divergence based covariance matching (RD-CMP)

$$\hat{\gamma} = \arg \min_{S \subseteq [n]} D_\alpha \left( \mathcal{N} \left( 0, \frac{1}{L} YY^T \right), \mathcal{N} \left( 0, \sigma^2 I + \gamma A_S A^T_S \right) \right)$$

- RD-CMP objective is a \textbf{difference of two submodular functions}

$$\hat{S} = \arg\min_{S \subseteq [n]} \log \left| (1 - \alpha) R_Y + \alpha \left( \sigma^2 I_m + \gamma A_S A^T_S \right) \right| - \alpha \log \left| \sigma^2 I_m + \gamma A_S A^T_S \right|$$

Main features of RD-CMP algorithm [Khanna & Murthy, 2017]

- Generalizes the MSBL cost function
- Objective is difference of two submodular set functions (optimized via Majorization-Minimization)
- Very low computational complexity
Performance

- Support recovery phase transition for $n = 200$, $L = 400$ and SNR = 10 dB

Co-LASSO

MSBL

RD-CMP

- Average runtime vs signal dimension
  - SNR = 10 dB
  - $k = \lceil 50 \log_{10} n \rceil$
  - $m = \lceil 0.75k \rceil$
  - $mL = \lceil 50k \log_{10} n \rceil$
Covariance Matching Framework for Support Recovery

**Covariance Matching Principle:**

\[ \hat{\gamma} = \arg \min_{\gamma \in \mathbb{R}^n_+} \text{distance} \left( R_Y, \sigma^2 I + A\Gamma A^T \right) \]

- Empirical MMV covariance
- Parameterized MMV covariance

**A closer look at covariance matching constraint:**

\[ R_Y \approx \sigma^2 I_m + A\Gamma A^T \approx (A \odot A) \gamma \]

Khatri-Rao product

For stable recovery of a \( k \)-sparse \( \gamma \), \( A \odot A \) must behave as an isometry for the restricted class of all \( k \)-sparse vectors.
Covariance Matching Framework for Support Recovery

- Covariance Matching Principle:

$$\hat{\gamma} = \arg \min_{\gamma \in \mathbb{R}^n_+} \text{distance} \left( R_Y, \begin{pmatrix} R_Y & \sigma^2 I_m + A\Gamma A^T \end{pmatrix} \right)$$

- A closer look at covariance matching constraint:

$$R_Y \approx \sigma^2 I_m + A\Gamma A^T$$

$$\text{vec} \left( R_Y - \sigma^2 I_m \right) \approx (A \odot A) \gamma$$
Covariance Matching Framework for Support Recovery

- **Covariance Matching Principle:**

  \[
  \hat{\gamma} = \arg \min_{\gamma \in \mathbb{R}^n_+} \text{distance}(R_Y, \sigma^2 I_m + A\Gamma A^T) \]

  - Empirical MMV covariance
  - Parameterized MMV covariance

- A closer look at covariance matching constraint: \( R_Y \approx \sigma^2 I_m + A\Gamma A^T \)

  \[
  \text{vec}(R_Y - \sigma^2 I_m) \approx (A \odot A) \gamma
  \]

  - Khatri-Rao product
Covariance Matching Framework for Support Recovery

- Covariance Matching Principle:
  \[ \hat{\gamma} = \arg \min_{\gamma \in \mathbb{R}^n_+} \text{distance} \left( R_Y, \sigma^2 I_m + A\Gamma A^T \right) \]

A closer look at covariance matching constraint: \( R_Y \approx \sigma^2 I_m + A\Gamma A^T \)

\[ \text{vec} \left( R_Y - \sigma^2 I_m \right) \approx (A \odot A) \gamma \]

Khatri-Rao product

For stable recovery of a \( k \)-sparse \( \gamma \), \( A \odot A \) must behave as an isometry for the restricted class of all \( k \)-sparse vectors.
Covariance Matching Framework for Support Recovery

- Covariance Matching Principle:

\[ \hat{\gamma} = \arg \min_{\gamma \in \mathbb{R}^n_+} \text{distance} \left( \begin{array}{l}
R_Y \\
\text{empirical MMV covariance}
\end{array} \right), \begin{array}{l}
\sigma^2 I + A \Gamma A^T \\
\text{parameterized MMV covariance}
\end{array} \]

- A closer look at covariance matching constraint: \( R_Y \approx \sigma^2 I_m + A \Gamma A^T \)

\[ \text{vec} \left( R_Y - \sigma^2 I_m \right) \approx (A \odot A) \gamma \]

Khatri-Rao product

- For stable recovery of a \( k \)-sparse \( \gamma \), \( A \odot A \) must behave as an isometry for the restricted class of all \( k \)-sparse vectors [When is this true?]
Columnwise Khatri-Rao product

- Columnwise Khatri-Rao product

\[
\begin{bmatrix}
a_1 & a_2 & \ldots & a_p \\
\end{bmatrix} \odot \begin{bmatrix}
b_1 & b_2 & \ldots & b_p \\
\end{bmatrix} = \begin{bmatrix}
a_1 \otimes b_1 & a_2 \otimes b_2 & \ldots & a_p \otimes b_p \\
\end{bmatrix}
\]

\( \odot \) denotes Kronecker product

- Khatri-Rao product arises naturally in
  - Sparsity pattern recovery (via covariance matching)
  - Direction of arrival estimation
  - Tensor decomposition

- When does \( A \odot B \) satisfy the Restricted Isometry Property?
Suppose $A$ and $B$ are $m \times n$ matrices with real i.i.d. $\mathcal{N}(0, 1)$ entries. Then,

$$
P \left( \delta_k \left( \frac{A}{\sqrt{m}} \odot \frac{B}{\sqrt{m}} \right) \geq \delta \right) \leq \frac{4e}{n^{2(\beta-1)}}$$

provided that $m \geq \left( \frac{c_1 \beta^{3/2}}{\delta} \right) \sqrt{k} (\log n)^{3/2}$. The results holds for all $\beta \geq 1$, and $c_1$ is an absolute positive numerical constant.

- For $m \geq \mathcal{O}\left( \frac{\sqrt{k} \log^{3/2} n}{\delta} \right)$, we have $\delta_k \left( \frac{A}{\sqrt{m}} \odot \frac{B}{\sqrt{m}} \right) \leq \delta$ w.h.p.
Restricted Isometry of Khatri-Rao Product

Suppose \( A \) and \( B \) are \( m \times n \) matrices with real i.i.d. \( \mathcal{N}(0, 1) \) entries. Then,

\[
P\left( \delta_k \left( \frac{A}{\sqrt{m}} \odot \frac{B}{\sqrt{m}} \right) \geq \delta \right) \leq \frac{4e}{n^2(\beta-1)}
\]

provided that \( m \geq \left( \frac{c_1 \beta^{3/2}}{\delta} \right) \sqrt{k} (\log n)^{3/2} \). The results holds for all \( \beta \geq 1 \), and \( c_1 \) is an absolute positive numerical constant.

- For \( m \geq \mathcal{O} \left( \frac{\sqrt{k} \log^{3/2} n}{\delta} \right) \), we have \( \delta_k \left( \frac{A}{\sqrt{m}} \odot \frac{B}{\sqrt{m}} \right) \leq \delta \) w.h.p.

- In MSBL, \( \delta_{2k} \left( \frac{A}{\sqrt{m}} \odot \frac{A}{\sqrt{m}} \right) < 1 \) can guarantee perfect support recovery w.h.p., if \( m \geq \mathcal{O}(\sqrt{k})! \)
Conventional Support Recovery

- Type-I estimation of $\mathbf{X}$
  - $\mathbf{X} = \text{unknown deterministic}$
- Work with $\mathbf{Y}$ directly
- RIP of $\mathbf{A}$ plays a role
- No. of meas: $m \geq \mathcal{O}(k)$

- Examples: SOMP, row-LASSO, M-FOCUSS

---

Covariance Matching

- Type-II estimation of $\mathbf{X}$
  - $\mathbf{x}_j \sim \mathcal{N}(0, \text{diag}(\gamma))$
- Work with $\frac{1}{L} \mathbf{Y}\mathbf{Y}^T$ (sample covariance)
- RIP of $\mathbf{A} \odot \mathbf{A}$ plays a role
- No. of meas: $m \geq \mathcal{O}(\sqrt{k})$

- Examples: Co-LASSO, MSBL, RD-CMP
