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:

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

$\mathbf{y} \in \mathbb{R}^m, \mathbf{x} \in \mathbb{R}^n$, and $\mathbf{A} \in \mathbb{R}^{m \times n}$

Overdetermined ($m > n$)

Unique or no solution

Least Squares

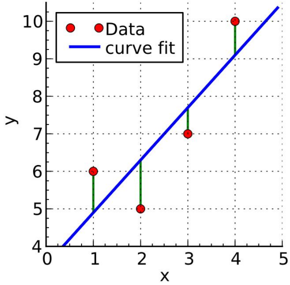
Linear system of equations:

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

$\mathbf{y} \in \mathbb{R}^m$, $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{A} \in \mathbb{R}^{m \times n}$

Overdetermined ($m > n$)

Unique or no solution



Line fitting

Least Squares

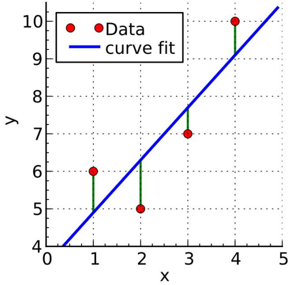
Linear system of equations:

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

$\mathbf{y} \in \mathbb{R}^m, \mathbf{x} \in \mathbb{R}^n$, and $\mathbf{A} \in \mathbb{R}^{m \times n}$

Overdetermined ($m > n$)

Unique or no solution



Line fitting

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

$$\hat{\mathbf{x}}_{\mathrm{LS}} = \arg\min_{\mathbf{x} \in \mathbb{R}^n} ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2$$

Least Squares

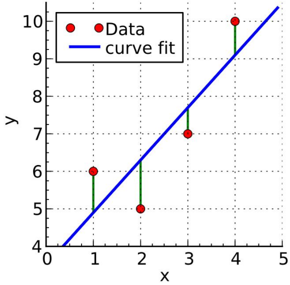
Linear system of equations:

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

$\mathbf{y} \in \mathbb{R}^m, \mathbf{x} \in \mathbb{R}^n$, and $\mathbf{A} \in \mathbb{R}^{m \times n}$

Overdetermined ($m > n$)

Unique or no solution



Line fitting

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

$$\hat{\mathbf{x}}_{\mathrm{LS}} = \arg\min_{\mathbf{x} \in \mathbb{R}^n} ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 \qquad \hat{\mathbf{x}}_{\mathrm{LS}} = \underbrace{(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}}_{\text{least squares solution}}$$

Least Squares

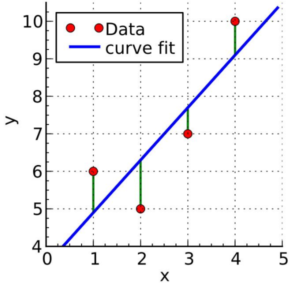
Linear system of equations:

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

$\mathbf{y} \in \mathbb{R}^m$, $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{A} \in \mathbb{R}^{m \times n}$

Overdetermined ($m > n$)

Unique or no solution



Line fitting

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

$$\hat{\mathbf{x}}_{\text{LS}} = \underset{\mathbf{x} \in \mathbb{R}^n}{\arg\min} \, ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 \qquad \hat{\mathbf{x}}_{\text{LS}} = \underbrace{(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}}_{\text{least squares solution}}$$

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

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

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

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

$$\begin{aligned} ||\hat{\mathbf{x}}_{\text{LS}} - \mathbf{x}^*||_2 &= \left|\left|\mathbf{A}^\dagger\tilde{\mathbf{y}} - \mathbf{x}^*\right|\right|_2 = \left|\left|(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{x}^* + \mathbf{e}\right) - \mathbf{x}^*\right|\right|_2 \\ &= \left|\left|\mathbf{A}^\dagger\mathbf{e}\right|\right|_2 \leq \left|\left|\left|\mathbf{A}^\dagger\right|\right|\right|_2 ||\mathbf{e}||_2 \end{aligned}$$

$$\left|\left|\left|\mathbf{A}^\dagger\right|\right|\right|_2 \leq \frac{1}{\lambda_{\min}(\mathbf{A}^T\mathbf{A})}\sqrt{\frac{\lambda_{\max}(\mathbf{A}^T\mathbf{A})}{\lambda_{\min}(\mathbf{A}^T\mathbf{A})}}$$

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

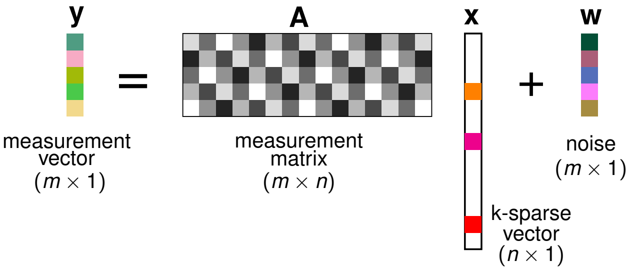
How far is $\hat{\mathbf{x}}_{\text{LS}}$ from $\mathbf{x}^*$?

$$\begin{aligned} ||\hat{\mathbf{x}}_{\text{LS}} - \mathbf{x}^*||_2 &= \left|\left|\mathbf{A}^\dagger\tilde{\mathbf{y}} - \mathbf{x}^*\right|\right|_2 = \left|\left|(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{x}^* + \mathbf{e}) - \mathbf{x}^*\right|\right|_2 \\ &= \left|\left|\mathbf{A}^\dagger\mathbf{e}\right|\right|_2 \le \left|\left|\left|\mathbf{A}^\dagger\right|\right|\right|_2 ||\mathbf{e}||_2 \end{aligned}$$

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

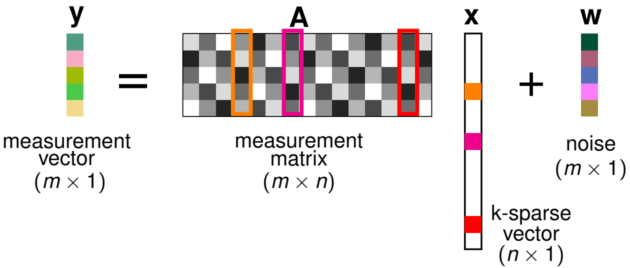
Smaller condition number of $\mathbf{A}^T\mathbf{A}$ implies lesser sensitivity to perturbations

Sparse Signal Recovery



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

Sparse Signal Recovery



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

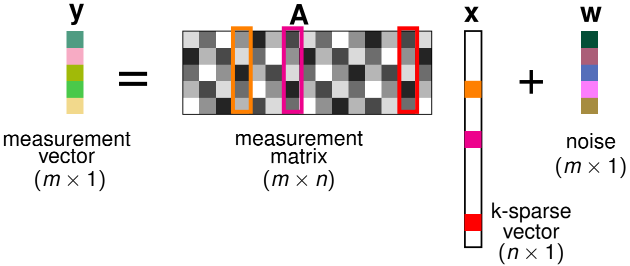
Two step recovery:

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

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

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

Sparse Signal Recovery



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

Two step recovery:

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

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

$$\mathbf{y} = \mathbf{A}_{\mathcal{S}}\mathbf{x}_{\mathcal{S}} + \mathbf{w}$$

overdetermined if $k > m$

Stable recovery of $\mathbf{x}_{\mathcal{S}}$ if condition no. of $\mathbf{A}_{\mathcal{S}}^T\mathbf{A}_{\mathcal{S}} \approx 1$

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 δ 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

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

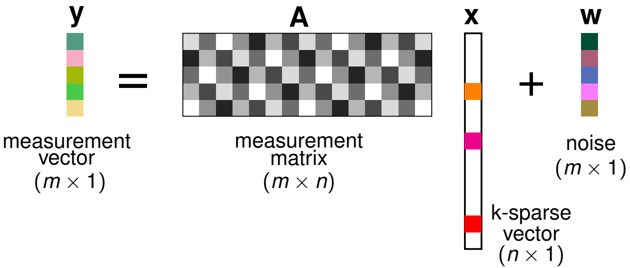
for all k-sparse vectors $\mathbf{z} \in \mathbb{R}^n$.

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

Alternate interpretations:

- $1-\delta_k^{\mathbf{A}} \;\leq\; \dfrac{\mathbf{z}^T\mathbf{A}^T\mathbf{A}\mathbf{z}}{\mathbf{z}^T\mathbf{z}} \;\leq\; 1+\delta_k^{\mathbf{A}} \qquad \forall\;$ k-sparse $\mathbf{z}$
- Eigenvalues of $\mathbf{A}_{\mathcal{S}}^T\mathbf{A}_{\mathcal{S}}$ lie in $[1-\delta_k^{\mathbf{A}}, 1+\delta_k^{\mathbf{A}}]$ for all supports $\mathcal{S}$, $|\mathcal{S}| \leq k$
- Condition no. of $\mathbf{A}_{\mathcal{S}}^T\mathbf{A}_{\mathcal{S}}$ is at most $\dfrac{1+\delta_k^{\mathbf{A}}}{1-\delta_k^{\mathbf{A}}}$ for all supports $\mathcal{S}$, $|\mathcal{S}| \leq k$

Sparse Signal Recovery



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

Two step recovery:

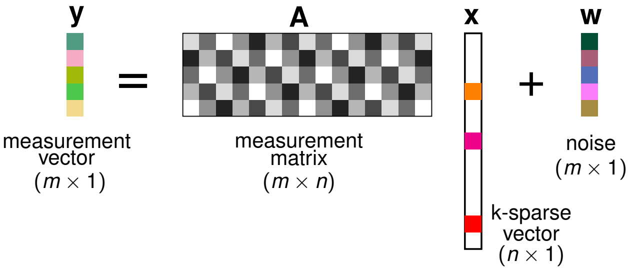
(i) Recover support $\mathcal{S}$ (indices of nonzero entries in $\mathbf{x}$)

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

$$\mathbf{y} = \mathbf{A}_\mathcal{S}\mathbf{x}_\mathcal{S} + \mathbf{w}$$ overdetermined if $k > m$

Condition no. of $\mathbf{A}_\mathcal{S}^T\mathbf{A}_\mathcal{S} \approx 1$ guarantees stable recovery of $\mathbf{x}_\mathcal{S}$

Sparse Signal Recovery



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

Two step recovery:

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

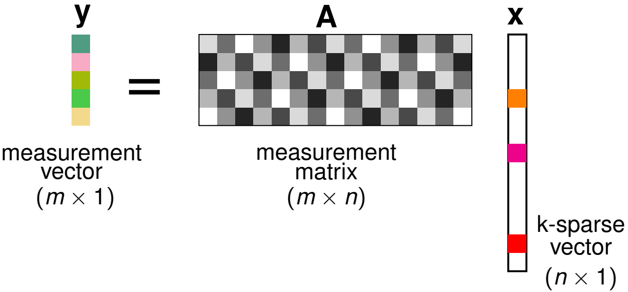
(ii) Recover $\mathbf{x}_\mathcal{S}$ using least squares on the reduced system:

$$\mathbf{y} = \mathbf{A}_\mathcal{S}\mathbf{x}_\mathcal{S} + \mathbf{w}$$ overdetermined if $k > m$

$\delta_k^{\mathbf{A}} \ll 1$

Condition no. of $\mathbf{A}_\mathcal{S}^T\mathbf{A}_\mathcal{S} \approx 1$ guarantees stable recovery of $\mathbf{x}_\mathcal{S}$

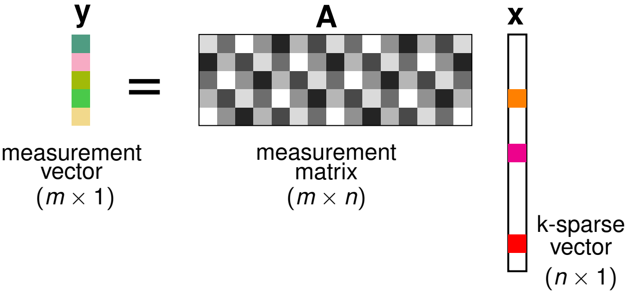
Uniqueness under noiseless measurements



RIP based guarantee for unique solution

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

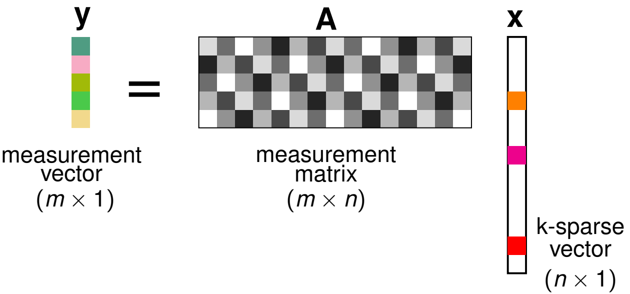
Uniqueness under noiseless measurements



RIP based guarantee for unique solution

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

Uniqueness under noiseless measurements

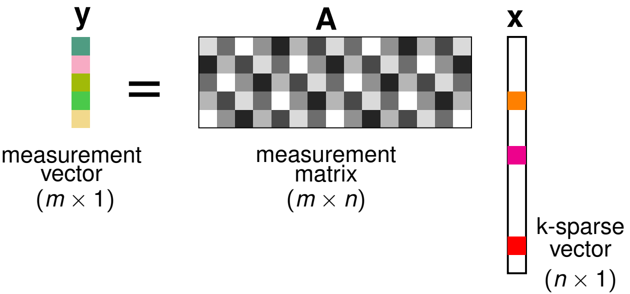


RIP based guarantee for unique solution

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

- $\mathbf{z}^T \mathbf{A}^T \mathbf{A} \mathbf{z} \geq (1 - \delta_{2k}^{\mathbf{A}}) ||\mathbf{z}||_2^2 > 0$ for all $2k$-sparse $\mathbf{z}$, ($\implies$ $2k$ **sparse vectors NOT allowed in Null(A)!**)

Uniqueness under noiseless measurements



RIP based guarantee for unique solution

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

- $\mathbf{z}^T \mathbf{A}^T \mathbf{A} \mathbf{z} \geq (1 - \delta_{2k}^{\mathbf{A}}) ||\mathbf{z}||_2^2 > 0$ for all $2k$-sparse $\mathbf{z}$, ($\implies$ **$2k$ sparse vectors NOT allowed in Null(A)!**)
- Let $\mathbf{x}_1, \mathbf{x}_2$ be distinct k-sparse solutions, then $\mathbf{y} = \mathbf{A}\mathbf{x}_1 = \mathbf{A}\mathbf{x}_2$. Thus, $\mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2) = 0$. Contradiction!.

RIP based recovery guarantees

Recovery algorithm	**RIP based sufficient conditions for successful signal reconstruction**
ℓ_1-norm minimization	$\delta_k(\mathbf{A}) \leq 0.307$
OMP	$\delta_{k+1}(\mathbf{A}) \leq \frac{\sqrt{4k+1}-1}{2k}$
Co-SAMP	$\delta_{4k}(\mathbf{A}) \leq 0.1$
IHT	$\delta_{3k}(\mathbf{A}) \leq \frac{1}{\sqrt{32}}$
Basis Pursuit	$\delta_{2k}(\mathbf{A}) + 3\delta_{3k} \leq 1$
Subspace Pursuit	$\delta_{3k} \leq 0.139$

Finding exact k-RIC is NP hard

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

$$1 - \delta \;\; \leq \;\; \lambda_i\left(\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}}\right) \;\; \leq \;\; 1 + \delta,$$

for all supports $\mathcal{S}$, $|\mathcal{S}| \leq k$.

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 $\mathbf{A}$

Restricted Isometry of Gaussian matrices

Gaussian RIP condition by Candès and Tao, 2005

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

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

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,

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

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

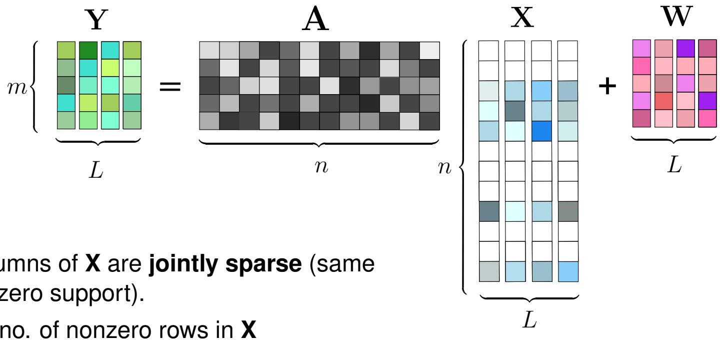
- Stability of sparse solution in noisy measurement case

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

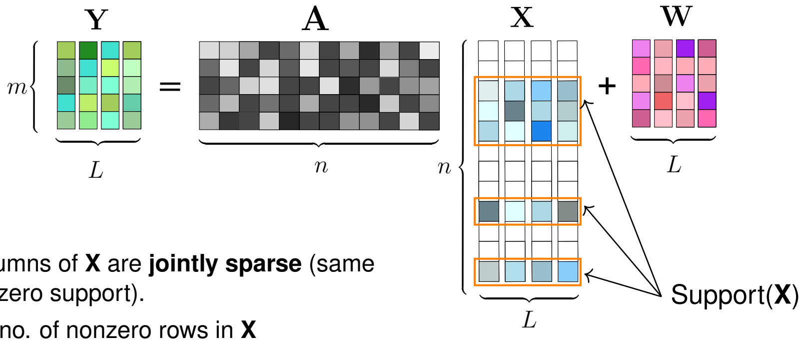
- Measurement model: $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$



- 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

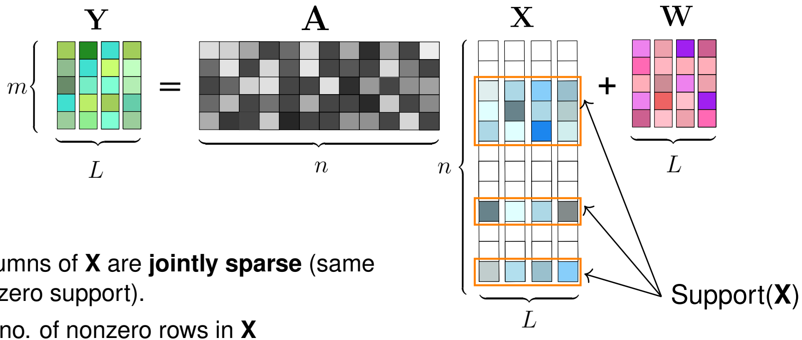
- Measurement model: $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$



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

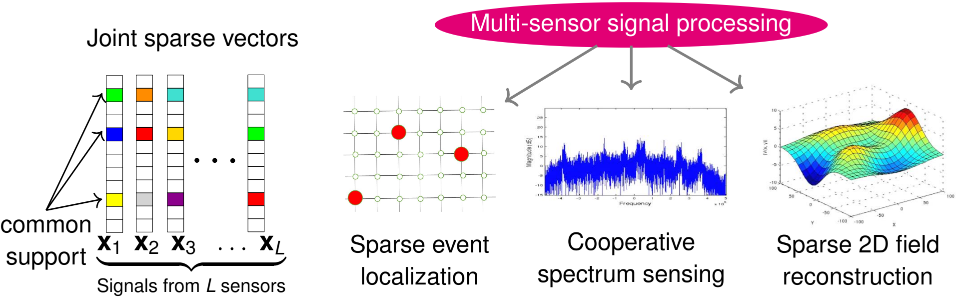


- Columns of **X** are **jointly sparse** (same nonzero support).
- k = no. of nonzero rows in **X**
- No inter/intra vector correlations in **X**

Multiple Measurement Vector (MMV) problem	Joint Sparse Support Recovery (JSSR)
Recover entire **X** from $\{\mathbf{Y}, \mathbf{A}, \sigma^2\}$	Recover support(**X**) from $\{\mathbf{Y}, \mathbf{A}, \sigma^2\}$

Applications

Joint sparse signals frequently arise in multi-sensor signal processing



Support Recovery via Sparse Bayesian Learning

- $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$

Support Recovery via Sparse Bayesian Learning

- $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$

Support Recovery via Sparse Bayesian Learning

- $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W}$
  - $\mathbf{x}_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \boldsymbol{\Gamma}),\ \boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$

Support Recovery via Sparse Bayesian Learning

- $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$
  - $\mathbf{x}_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \boldsymbol{\Gamma}),\ \boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$
  - Support(γ) = 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

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

Support Recovery via Sparse Bayesian Learning

- $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W}$
  - $\mathbf{x}_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \boldsymbol{\Gamma}),\ \boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$
  - Support(γ) = 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:

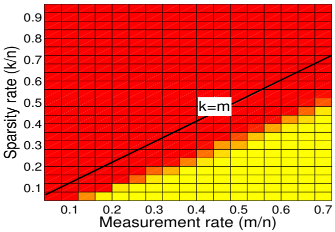
$$\hat{\gamma} = \underset{\boldsymbol{\gamma} \in \mathbb{R}_+^n}{\text{argmax}} \ \log p(\mathbf{Y}; \gamma)$$

  - $\hat{\gamma}$ found using Expectation Maximization (EM) procedure

Support Recovery via Sparse Bayesian Learning

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

SOMP support recovery phase transition

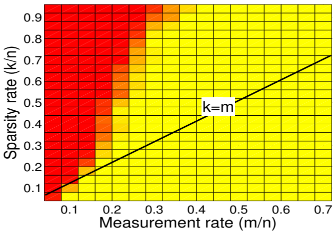


- MSBL algorithm:

$$\hat{\gamma} = \underset{\boldsymbol{\gamma} \in \mathbb{R}^n_+}{\mathrm{argmax}}\ \log p(\mathbf{Y}; \gamma)$$

  - $\hat{\gamma}$ found using Expectation Maximization (EM) procedure

MSBL support recovery phase transition



Recoverable support size k grows as $\mathcal{O}(m^2)$ 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 $\mathcal{S}^*$, $|\mathcal{S}^*| \leq k$, and with variances of the nonzero entries in $[\gamma_{\min}, \gamma_{\max}]$. Then,

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

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

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

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

Condition 2: $L \geq \dfrac{c_1 k \log n}{\eta}$, where $\eta = \dfrac{m}{8k}\left(\dfrac{\gamma_{\min}}{\sigma^2 + \gamma_{\max}}\right)^2 \dfrac{\left(1 - \delta_{2k}^{\odot}\right)}{\sup\limits_{\mathcal{S}:|\mathcal{S}|=2k} \left|\left|\left|\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}}\right|\right|\right|_2}$,

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:

$$-\log p(\mathbf{Y}; \gamma) \;=\; -\sum_{j=1}^{L} \log \mathcal{N}\left(\mathbf{y}_j; 0, \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right)$$

New interpretation of MSBL cost function

- MSBL's log-likelihood cost:

$$
\begin{aligned}
-\log p(\mathbf{Y};\gamma) &= -\sum_{j=1}^{L} \log \mathcal{N}\left(\mathbf{y}_j; 0, \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right) \\
&\propto \log |\sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T| + \text{trace}\left(\left(\sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right)^{-1}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T\right)\right)
\end{aligned}
$$

New interpretation of MSBL cost function

- MSBL's log-likelihood cost:

$$
\begin{aligned}
-\log p(\mathbf{Y};\gamma) &= -\sum_{j=1}^{L} \log \mathcal{N}\left(\mathbf{y}_j; 0, \sigma^2\mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right) \\
&\propto \log|\sigma^2\mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T| + \text{trace}\left(\left(\sigma^2\mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right)^{-1}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T\right)\right)
\end{aligned}
$$

Log Det Bregman matrix divergence between matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{S}_{++}^m$ is defined as

$$
\mathcal{D}_\phi(\mathbf{X}, \mathbf{Y}) \triangleq \text{trace}(\mathbf{X}\mathbf{Y}^{-1}) - \log|\mathbf{X}\mathbf{Y}^{-1}| - m
$$

New interpretation of MSBL cost function

- MSBL's log-likelihood cost:

$$
\begin{aligned}
-\log p(\mathbf{Y};\gamma) &= -\sum_{j=1}^{L} \log \mathcal{N}\left(\mathbf{y}_j; 0, \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right) \\
&\propto \log |\sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T| + \text{trace}\left(\left(\sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right)^{-1}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T\right)\right) \\
&\propto \underbrace{\mathcal{D}^{\text{Bregman}}_{-\log\det}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T, \;\; \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right)}_{\text{Log Det Bregman Matrix Div.}} + \text{ constant terms}
\end{aligned}
$$

New interpretation of MSBL cost function

- MSBL's log-likelihood cost:

$$
\begin{aligned}
-\log p(\mathbf{Y};\gamma) &= -\sum_{j=1}^{L} \log \mathcal{N}\left(\mathbf{y}_j; 0, \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right) \\
&\propto \log|\sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T| + \text{trace}\left(\left(\sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right)^{-1}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T\right)\right) \\
&\propto \underbrace{\mathcal{D}^{\text{Bregman}}_{-\log\det}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T,\ \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right)}_{\text{Log Det Bregman Matrix Div.}} + \text{ constant terms}
\end{aligned}
$$

- MSBL optimization minimizes $\mathcal{D}^{\text{Bregman}}_{-\log\det}\left(\underbrace{\frac{1}{L}\mathbf{Y}\mathbf{Y}^T}_{\text{emp. cov mat}},\ \underbrace{\sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T}_{\text{param. cov mat}}\right)$

- Can we use some other matrix divergence?

Covariance Matching Framework for Support Recovery

- MMV model: $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$
  - $\mathbf{x}_j \sim \mathcal{N}(0, \mathrm{diag}(\boldsymbol{\gamma}))$
  - $\mathbf{y}_j \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T)$

Covariance Matching Framework for Support Recovery

- MMV model: $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$
  - $\mathbf{x}_j \sim \mathcal{N}(0, \mathrm{diag}(\gamma))$
  - $\mathbf{y}_j \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T)$
- Covariance matrices:
  - Empirical $\mathbf{R}_\mathbf{Y} = \frac{1}{L}\mathbf{Y}\mathbf{Y}^T$
  - Parameterized $\mathbf{\Sigma}_\gamma = \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T$

Covariance Matching Framework for Support Recovery

- MMV model: $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$
  - $\mathbf{x}_j \sim \mathcal{N}(0, \mathrm{diag}(\boldsymbol{\gamma}))$
  - $\mathbf{y}_j \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T)$
- Covariance matrices:
  - Empirical $\mathbf{R}_\mathbf{Y} = \frac{1}{L}\mathbf{Y}\mathbf{Y}^T$
  - Parameterized $\boldsymbol{\Sigma}_\gamma = \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T$
- Covariance Matching Principle:

$$\hat{\gamma} = \underset{\boldsymbol{\gamma} \in \mathbb{R}_+^n}{\text{arg min}} \ \textbf{distance}\left(\underbrace{\mathbf{R}_\mathbf{Y}}_{\substack{\text{empirical} \\ \text{MMV covariance}}} , \underbrace{\sigma^2 \mathbf{I} + \mathbf{A}\Gamma\mathbf{A}^T}_{\substack{\text{parameterized} \\ \text{MMV covariance}}} \right)$$

$$\text{support}(\mathbf{X}) = \text{support}(\hat{\gamma})$$

Examples of covariance matching algorithms

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

$$\hat{\gamma} = \underset{\gamma \in \mathbb{R}_+^n}{\arg\min} \left|\left|\left| \frac{1}{L}\mathbf{Y}\mathbf{Y}^T - (\sigma^2\mathbf{I} + \mathbf{A}\Gamma\mathbf{A}^T) \right|\right|\right|_F^2 + \lambda \, ||\gamma||_1$$

- Main features of the Co-LASSO [Pal & Vaidyanathan, 2013]
  - ℓ_1 norm penalty promotes recovery of sparse γ
  - Convex objective
  - Very high memory requirements

Examples of covariance matching algorithms

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

$$\hat{\gamma} = \underset{\boldsymbol{\gamma} \in \mathbb{R}_+^n}{\arg\min} \; \log \left| \sigma^2 \mathbf{I} + \mathbf{A}\Gamma\mathbf{A}^T \right| + \mathrm{tr}\left(\left(\sigma^2 \mathbf{I} + \mathbf{A}\Gamma\mathbf{A}^T \right)^{-1} \left(\frac{1}{L} \mathbf{Y}\mathbf{Y}^T \right) \right)$$

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

Examples of covariance matching algorithms

- α-**Rényi divergence** based covariance matching (RD-CMP)

$$\hat{\gamma} = \underset{\mathcal{S} \subseteq [n]}{\arg\min}\, \mathcal{D}_\alpha \left(\mathcal{N}\left(0, \frac{1}{L}\mathbf{Y}\mathbf{Y}^T\right), \mathcal{N}\left(0, \sigma^2 \mathbf{I} + \gamma \mathbf{A}_{\mathcal{S}} \mathbf{A}_{\mathcal{S}}^T\right) \right)$$

$\mathbf{x}_j \sim \mathcal{N}(0, \gamma \mathrm{diag}(\mathbf{1}_{\mathcal{S}}))$

- RD-CMP objective is a **difference of two submodular functions**

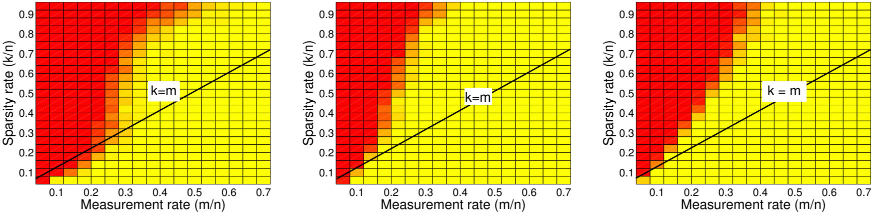
$$\hat{\mathcal{S}} = \underset{\mathcal{S} \subseteq [n]}{\operatorname{argmin}}\, \underbrace{\log \left| (1-\alpha)\mathbf{R}_{\mathbf{Y}} + \alpha \left(\sigma^2 \mathbf{I}_m + \gamma \mathbf{A}_{\mathcal{S}} \mathbf{A}_{\mathcal{S}}^T \right) \right|}_{f(\mathcal{S}),\ \text{submodular in } \mathcal{S}} - \underbrace{\alpha \log \left| \sigma^2 \mathbf{I}_m + \gamma \mathbf{A}_{\mathcal{S}} \mathbf{A}_{\mathcal{S}}^T \right|}_{g(\mathcal{S}),\ \text{submodular in } \mathcal{S}}$$

- 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

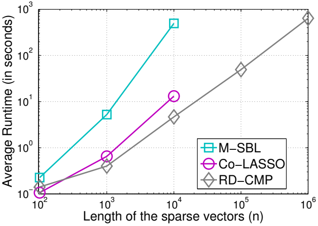
Performance

- 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:

$$\hat{\gamma} = \underset{\boldsymbol{\gamma} \in \mathbb{R}_+^n}{\arg\min} \ \textbf{distance}\Big(\underbrace{\mathbf{R_Y}}_{\substack{\text{empirical} \\ \text{MMV covariance}}} , \underbrace{\sigma^2 \mathbf{I} + \mathbf{A}\Gamma\mathbf{A}^T}_{\substack{\text{parameterized} \\ \text{MMV covariance}}} \Big)$$

Covariance Matching Framework for Support Recovery

- Covariance Matching Principle:

$$\hat{\gamma} = \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}_+^n} \textbf{distance}\left(\underbrace{\mathbf{R_Y}}_{\substack{\text{empirical}\\ \text{MMV covariance}}} , \underbrace{\sigma^2\mathbf{I} + \mathbf{A}\Gamma\mathbf{A}^T}_{\substack{\text{parameterized}\\ \text{MMV covariance}}} \right)$$

- A closer look at covariance matching constraint: $\mathbf{R_Y} \approx \sigma^2\mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T$

$$\text{vec}\left(\mathbf{R_Y} - \sigma^2\mathbf{I}_m\right) \approx (\mathbf{A} \odot \mathbf{A})\,\gamma$$

Covariance Matching Framework for Support Recovery

- Covariance Matching Principle:

$$\hat{\gamma} = \underset{\boldsymbol{\gamma} \in \mathbb{R}_+^n}{\arg\min}\ \textbf{distance}\left(\underbrace{\mathbf{R_Y}}_{\substack{\text{empirical} \\ \text{MMV covariance}}} , \underbrace{\sigma^2 \mathbf{I} + \mathbf{A}\Gamma\mathbf{A}^T}_{\substack{\text{parameterized} \\ \text{MMV covariance}}} \right)$$

- A closer look at covariance matching constraint: $\mathbf{R_Y} \approx \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T$

$$\text{vec}\left(\mathbf{R_Y} - \sigma^2 \mathbf{I}_m\right) \approx (\mathbf{A} \odot \mathbf{A})\,\gamma$$

Khatri-Rao product

Covariance Matching Framework for Support Recovery

- Covariance Matching Principle:

$$\hat{\gamma} = \underset{\boldsymbol{\gamma}\in\mathbb{R}_{+}^{n}}{\arg\min}\ \textbf{distance}\left(\underbrace{\mathbf{R_Y}}_{\substack{\text{empirical}\\ \text{MMV covariance}}}, \underbrace{\sigma^2\mathbf{I} + \mathbf{A}\Gamma\mathbf{A}^T}_{\substack{\text{parameterized}\\ \text{MMV covariance}}} \right)$$

- A closer look at covariance matching constraint: $\mathbf{R_Y} \approx \sigma^2\mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T$

$$\text{vec}\left(\mathbf{R_Y} - \sigma^2\mathbf{I}_m\right) \approx \underbrace{(\mathbf{A}\odot\mathbf{A})}_{\text{Khatri-Rao product}}\gamma$$

- For stable recovery of a k-sparse γ, $\mathbf{A}\odot\mathbf{A}$ must behave as an isometry for the restricted class of all k-sparse vectors

Covariance Matching Framework for Support Recovery

- Covariance Matching Principle:

$$\hat{\gamma} = \underset{\boldsymbol{\gamma} \in \mathbb{R}_+^n}{\arg\min} \ \textbf{distance}\left(\underbrace{\mathbf{R_Y}}_{\substack{\text{empirical} \\ \text{MMV covariance}}} , \underbrace{\sigma^2 \mathbf{I} + \mathbf{A}\Gamma\mathbf{A}^T}_{\substack{\text{parameterized} \\ \text{MMV covariance}}} \right)$$

- A closer look at covariance matching constraint: $\mathbf{R_Y} \approx \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T$

$$\text{vec}\left(\mathbf{R_Y} - \sigma^2 \mathbf{I}_m\right) \approx \underbrace{(\mathbf{A} \odot \mathbf{A})}_{\text{Khatri-Rao product}} \gamma$$

- For stable recovery of a k-sparse γ, $\mathbf{A} \odot \mathbf{A}$ 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

$$\underbrace{\begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_p \\ | & | & & | \end{bmatrix}}_{\mathbf{A}\ (m \times p)} \odot \underbrace{\begin{bmatrix} | & | & & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_p \\ | & | & & | \end{bmatrix}}_{\mathbf{B}\ (m \times p)} = \underbrace{\begin{bmatrix} | & | & & | \\ \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \dots & \mathbf{a}_p \otimes \mathbf{b}_p \\ | & | & & | \end{bmatrix}}_{\mathbf{A} \odot \mathbf{B}\ (m^2 \times p)}$$

$\otimes$ denotes Kronecker product

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

- When does $\mathbf{A} \odot \mathbf{B}$ satisfy the Restricted Isometry Property?

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,

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

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 $\delta_k\left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}}\right) \leq \delta$ 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,

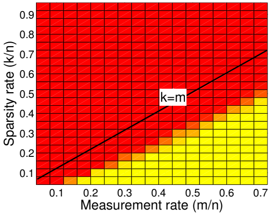
$$\mathbb{P}\left(\delta_k\left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}}\right) \geq \delta\right) \leq \frac{4e}{n^{2(\beta-1)}}$$

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 $\delta_k\left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}}\right) \leq \delta$ w.h.p.
- In MSBL, $\delta_{2k}\left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{A}}{\sqrt{m}}\right) < 1$ can guarantee perfect support recovery w.h.p., if $m \geq \mathcal{O}(\sqrt{k})$!

Conventional Support Recovery

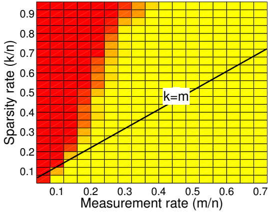
- Type-I estimation of $\mathbf{X}$
  - $\mathbf{X}$ = unknown deterministic
- Work with $\mathbf{Y}$ directly
- RIP of $\mathbf{A}$ plays a role
- No. of meas: $m \geq \mathcal{O}(k)$



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

Covariance Matching

- Type-II estimation of $\mathbf{X}$
  - $\mathbf{x}_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \mathrm{diag}(\gamma))$
- 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

Related Publications

- S. Khanna and C. R. Murthy, "*On the Support Recovery of Jointly Sparse Gaussian Sources using Sparse Bayesian Learning*," (arXiV preprint: arXiv:1703.04930).
- S. Khanna and C. R. Murthy, "*On the Restricted Isometry of the Columnwise Khatri-Rao Product*," (submitted to IEEE Trans. Signal Process.).
- S. Khanna and C. R. Murthy, "*Rényi Divergence based Covariance Matching Pursuit of Joint Sparse Support*," IEEE Workshop on Signal Processing SPAWC-17), Sapporo, Japan, 2017, pp. 1-6.