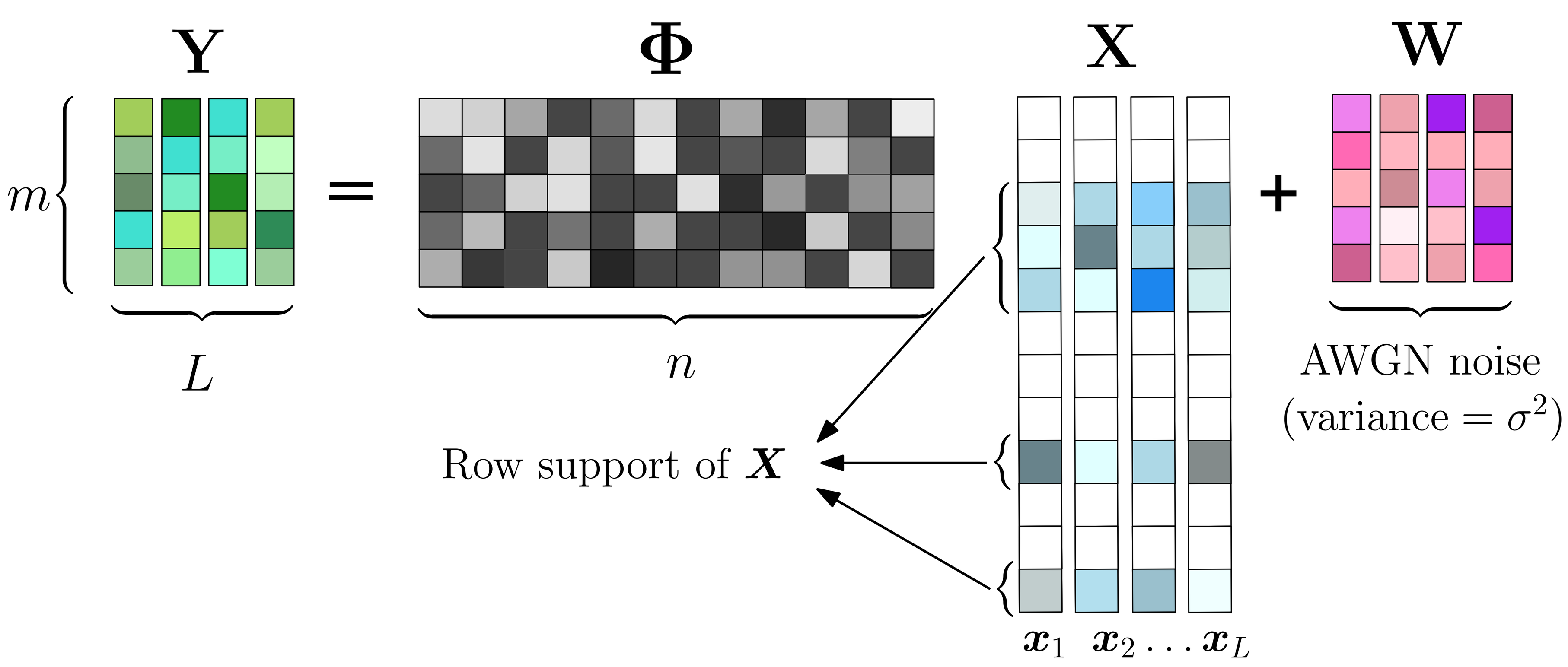
Covariance Matching Techniques for Joint Sparse Support Recovery

Saurabh Khanna Advisor: Chandra R. Murthy

Dept. of ECE, Indian Institute of Science, Bangalore, India

email: {sakhanna, cmurthy}@ece.iisc.ernet.in

Multiple Measurement Vector (MMV) model



Joint Sparse Support Recovery Problem

- Find the common nonzero support of $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L\}$ using their compressive, noisy linear measurements $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_L\}$.

Multiple Sparse Bayesian Learning

- Proposed by Wipf & Rao in 2007.
- Joint sparsity inducing prior on columns of $\mathbf{X}$.

$$\mathbf{x}_j \overset{iid}{\sim} \mathcal{N}(0, \boldsymbol{\Gamma}), \quad \boldsymbol{\Gamma} = \mathrm{diag}(\gamma_1, \gamma_2, \ldots, \gamma_n).$$

- Induced observation density:

$$\mathbf{y}_j \overset{iid}{\sim} \mathcal{N}(0, \sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi})$$

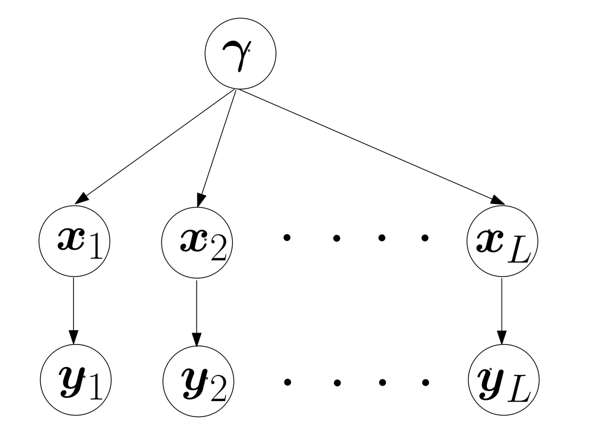


Figure: Graphical representation

- Maximum likelihood estimate of $\boldsymbol{\gamma}$ is obtained via Expectation Maximization (EM) procedure.

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

Sufficient Conditions for Exact Support Recovery

- **For Gaussian sources...**
  - $\mathbf{x}_j \overset{iid}{\sim} \mathcal{N}(0, \boldsymbol{\Gamma}^*)$, for some $\boldsymbol{\Gamma}^* = \mathrm{diag}(\boldsymbol{\gamma}^*)$.
  - $\boldsymbol{\gamma}^*$ is at most k-sparse with support $\mathcal{S}^*$.
  - $\gamma_{\text{min}} \preceq \boldsymbol{\gamma}^*_{\mathcal{S}^*} \preceq \gamma_{\text{max}}$.
- **MSBL recovers exact support...**

$$\mathbb{P}\left(\text{Support}(\hat{\boldsymbol{\gamma}}) = \mathcal{S}^*\right) \geq 1 - \exp\left(-\frac{\eta}{4}L\right),$$

where $\hat{\boldsymbol{\gamma}}$ is the k or less sparse output of MSBL, if

- **Condition 1:** Self Khatri-Rao matrix $\boldsymbol{\Phi} \odot \boldsymbol{\Phi}$ satisfies restricted isometry property of order $2k$, i.e., $\exists$ $\delta_{2k}^{\odot} \in (0,1)$ such that

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

holds for all $2k$ or less sparse vectors $\mathbf{z}$.

- **Condition 2:** $L \geq \left(\frac{c_1 k \log n}{\eta}\right)$, where $\eta = c_2\left(\frac{\gamma_{\text{min}}}{\sigma^2 + \gamma_{\text{max}}}\right)^2\left(1 - \delta_{2k}^{\odot}\right)$, c_1, c_2 are universal constants.

More support from fewer measurements

	k < Spark($\boldsymbol{\Phi}$) − 1	(Spark($\boldsymbol{\Phi}$) − 1) ≤ k ≤ (Krank ($\boldsymbol{\Phi} \odot \boldsymbol{\Phi}$) /2)
Noiseless meas.	$L \geq 1$	$L \geq \frac{c_1}{\eta}(k \log n)$
Noisy meas.	$L \geq \frac{c_1}{\eta}(k \log n)$	$L \geq \frac{c_1}{\eta}(k \log n)$

- For Gaussian measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}$, $(n > \frac{m^2+m}{2})$
  - Spark($\boldsymbol{\Phi}$) $= m$ w.h.p.
  - Kruskal rank($\boldsymbol{\Phi}$) $= \mathcal{O}(m^2)$ w.h.p.

Objectives

- Analyze the support recovery performance of the Multiple Sparse Bayesian Learning (MSBL) algorithm.
- Devise a low complexity covariance matching based joint sparse support recovery algorithm.

Rényi Divergence based Covariance Matching Pursuit (RD-CMP)

- **Simplified Gaussian prior on sources**

$$\mathbf{x}_j \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \boldsymbol{\gamma} \times \mathrm{diag}(\mathbf{1}_{\mathcal{S}})\right),$$

  - Set $\mathcal{S}$ models the common support.
  - $\boldsymbol{\gamma}$ models the common variance of nonzero elements in $\mathbf{X}$
- **Induced observation density**

$$\Rightarrow \mathbf{y}_j \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma^2\mathbf{I}_m + \boldsymbol{\gamma}\boldsymbol{\Phi}_{\mathcal{S}}\boldsymbol{\Phi}_{\mathcal{S}}^T\right).$$

- Estimate $\hat{\mathcal{S}}$ via **generalized reverse $\mathcal{I}$-projection**,

$$\hat{\mathcal{S}} = \underset{\mathcal{S} \subseteq [n]}{\arg\min}\ \mathcal{D}_\alpha\left(\prod_{j=1}^{L}\mathcal{N}(\mathbf{y}_j; 0, \mathbf{R}_{\mathbf{Y}}),\ p(\mathbf{Y}; \mathcal{S})\right),$$

where $\mathcal{D}_\alpha$ is the Rényi divergence of order $\alpha \in (0,1)$.

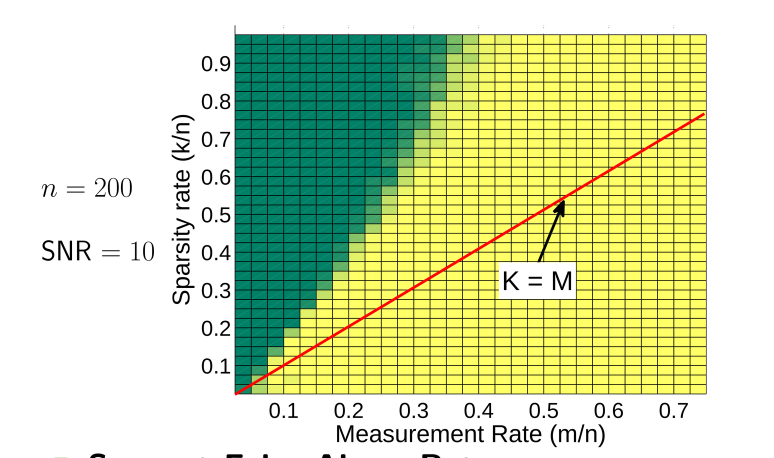
- **RD-CMP algorithm:**

$$\hat{\mathcal{S}} = \underset{\mathcal{S} \subseteq [n]}{\arg\min}\ \underbrace{\log\left|(1-\alpha)\mathbf{R}_{\mathbf{Y}} + \alpha\left(\sigma^2\mathbf{I}_m + \boldsymbol{\gamma}\boldsymbol{\Phi}_{\mathcal{S}}\boldsymbol{\Phi}_{\mathcal{S}}^T\right)\right|}_{\text{submodular in } \mathcal{S}} - \underbrace{\alpha\log\left|\sigma^2\mathbf{I}_m + \boldsymbol{\gamma}\boldsymbol{\Phi}_{\mathcal{S}}\boldsymbol{\Phi}_{\mathcal{S}}^T\right|}_{\text{submodular in } \mathcal{S}}.$$

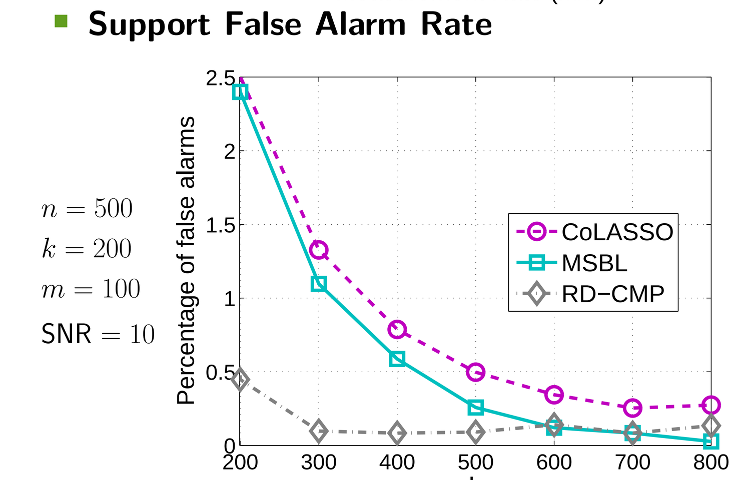
- **Key idea:** Rényi divergence objective is a difference of two submodular functions.
  - Iterative **majorization-minimization** procedure for finding $\hat{\mathcal{S}}$:
    Majorization: Replace 1st log-det term with its modular upper bound.
    Minimization: Minimize the resulting supermodular objective via greedy search.

Simulation results

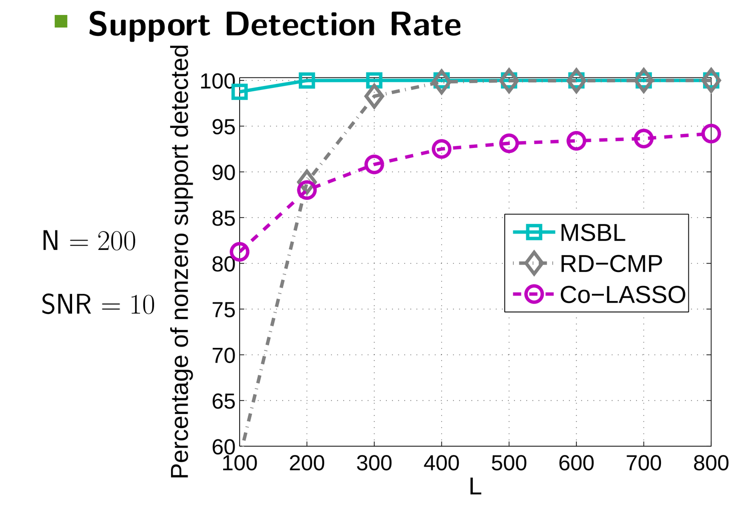
- RD-CMP Support Recovery Phase Transition



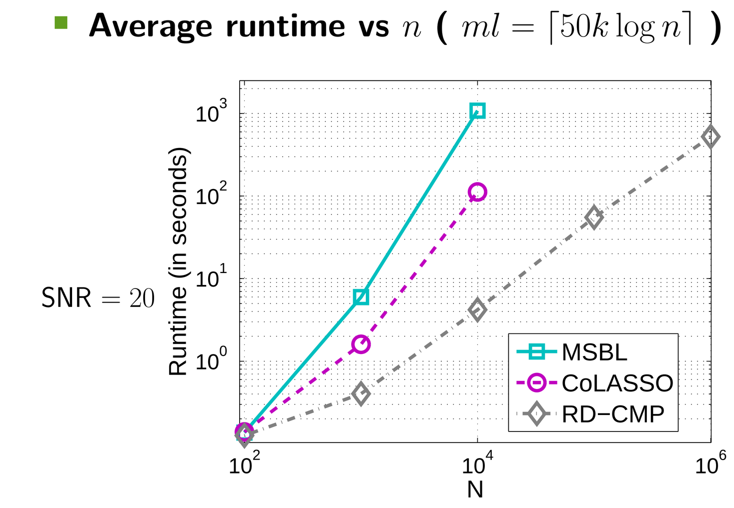
- Support False Alarm Rate



- Support Detection Rate



- Average runtime vs n ($ml = \lceil 50k\log n \rceil$)



References

- 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, *Rényi Divergence Based Covariance Matching Pursuit of Joint Sparse Support*, submitted to SPAWC-2017.

Covariance Matching Techniques for Joint Sparse Support Recovery

Saurabh Khanna, PhD Student

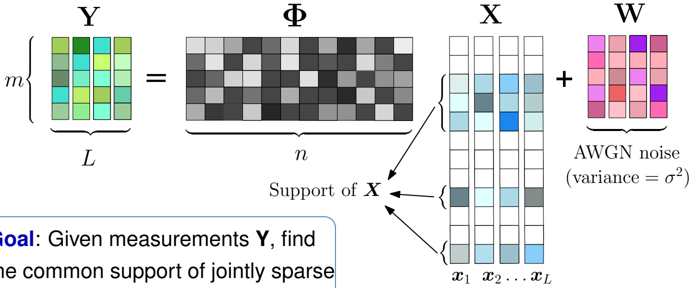
Advisor: Prof. Chandra R. Murthy

{**sakhanna, cmurthy**}**@ece.iisc.ernet.in**

Signal Processing for Communications Lab,
ECE Department,
Indian Institute of Science, Bangalore

Joint Sparse Support Recovery

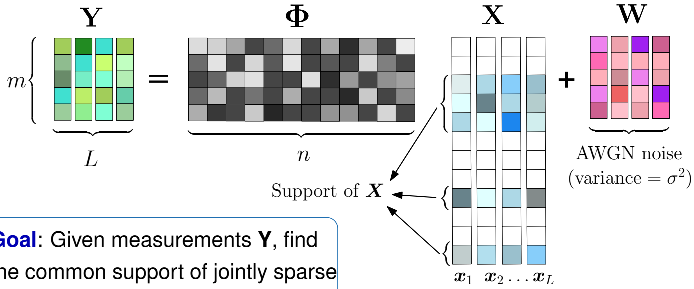
- Underdetermined multiple measurement vector (MMV) model:



Goal: Given measurements **Y**, find the common support of jointly sparse columns of **X**.

Joint Sparse Support Recovery

- Underdetermined multiple measurement vector (MMV) model:



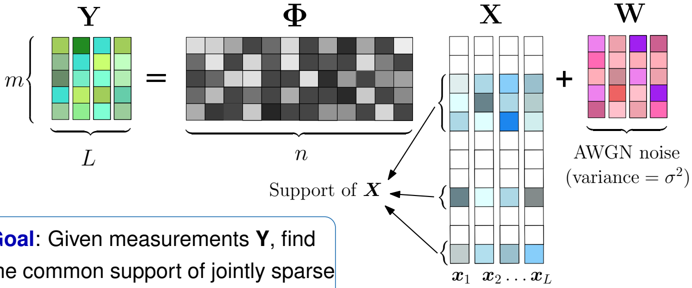
Goal: Given measurements **Y**, find the common support of jointly sparse columns of **X**.

Applications:

- Spectrum sensing
- MIMO channel estimation
- Direction of arrival estimation
- Event detection and localization

Joint Sparse Support Recovery

- Underdetermined multiple measurement vector (MMV) model:



Goal: Given measurements **Y**, find the common support of jointly sparse columns of **X**.

Applications:

- Spectrum sensing
- MIMO channel estimation
- Direction of arrival estimation
- Event detection and localization

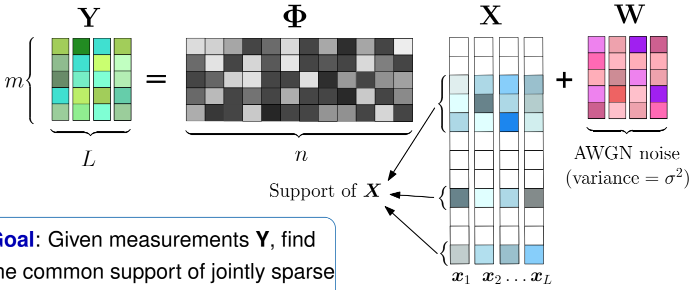
Popular recovery algorithms:

- Simultaneous OMP
- Row LASSO
- CS-MuSiC
- AMP-MMV

Can recover supports of size $\leq$ **m**.

Joint Sparse Support Recovery

- Underdetermined multiple measurement vector (MMV) model:



Goal: Given measurements **Y**, find the common support of jointly sparse columns of **X**.

Applications:

- Spectrum sensing
- MIMO channel estimation
- Direction of arrival estimation
- Event detection and localization

Popular recovery algorithms:

- Simultaneous OMP
- Row LASSO
- CS-MuSiC
- AMP-MMV

Can recover supports of size $\leq$ **m**.

Covariance matching approach can recover $\mathcal{O}(\mathbf{m^2})$ sparse support.

Contributions

- Derive sufficient conditions for exact support recovery in **Multiple Sparse Bayesian Learning (MSBL)** algorithm.
  - $\mathcal{O}(m^2)$ support recovery
  - Non-asymptotic guarantees

- Propose a new **low computational complexity** algorithm for support recovery based on covariance matching.

Sparse Bayesian Learning for MMV model (MSBL)

- A **parameterized**, **Gaussian prior** is imposed on the unknown signal matrix $\mathbf{X}$.

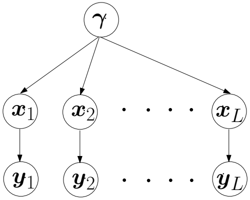
Signal prior:

$$\mathbf{x}_j \overset{i.i.d}{\sim} \mathcal{N}(0, \boldsymbol{\Gamma}),$$

where $\boldsymbol{\Gamma} = \mathrm{diag}(\gamma_1, \gamma_2, \ldots, \gamma_n)$.

Induced observation density:

$$\mathbf{y}_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2 \mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi})$$



Graphical representation

- Shared covariance matrix $\boldsymbol{\Gamma}$ induces joint sparsity in $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L$.

Sparse Bayesian Learning for MMV model (MSBL)

- A **parameterized**, **Gaussian prior** is imposed on the unknown signal matrix $\mathbf{X}$.

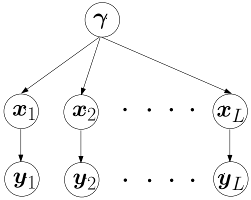
Signal prior:

$$\mathbf{x}_j \overset{i.i.d}{\sim} \mathcal{N}(0, \boldsymbol{\Gamma}),$$

where $\boldsymbol{\Gamma} = \mathrm{diag}(\gamma_1, \gamma_2, \ldots, \gamma_n)$.

Induced observation density:

$$\mathbf{y}_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi})$$



Graphical representation

- Shared covariance matrix $\boldsymbol{\Gamma}$ induces joint sparsity in $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L$.
- Maximum likelihood (ML) estimation of the covariance parameters γ, i.e.,

$$\hat{\gamma} = \underset{\gamma \in \mathbb{R}_+^n}{\arg\max} \ \log p(\mathbf{Y}; \gamma).$$

- Expectation Maximization (EM) procedure used to find the ML estimate $\hat{\gamma}$.

Sparse Bayesian Learning for MMV model (MSBL)

- A **parameterized**, **Gaussian prior** is imposed on the unknown signal matrix $\mathbf{X}$.

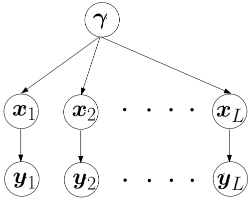
Signal prior:

$$\mathbf{x}_j \overset{i.i.d}{\sim} \mathcal{N}(0, \mathbf{\Gamma}),$$

where $\mathbf{\Gamma} = \mathrm{diag}(\gamma_1, \gamma_2, \ldots, \gamma_n)$.

Induced observation density:

$$\mathbf{y}_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2\mathbf{I} + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi})$$



Graphical representation

- Shared covariance matrix $\mathbf{\Gamma}$ induces joint sparsity in $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L$.
- Maximum likelihood (ML) estimation of the covariance parameters γ, i.e.,

$$\hat{\gamma} = \underset{\gamma \in \mathbb{R}_+^n}{\arg\max} \; \log p(\mathbf{Y}; \gamma).$$

- Expectation Maximization (EM) procedure used to find the ML estimate $\hat{\gamma}$.

When does MSBL recovers the true support of X?

Support Recovery in MSBL

Theorem (Sufficient conditions for exact support recovery in MSBL[1])

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L$ *be i.i.d. zero mean jointly sparse Gaussian sources with support* $\mathcal{S}^*$, $|\mathcal{S}^*| \leq k$, *and with variances of the nonzero entries inside* $[\gamma_{min}, \gamma_{max}]$.

Let $\hat{\gamma}$ *be the* k*-sparse output of MSBL. Then,*

$$\mathbb{P}\left(Support(\hat{\gamma}) = \mathcal{S}^*\right) \geq 1 - \exp\left(-\frac{\eta}{4}L\right),$$

if conditions ***C1*** *and* ***C2*** *are satisfied.*

Condition 1: *Self Khatri-Rao matrix* $\mathbf{\Phi} \odot \mathbf{\Phi}$ *satisfies restricted isometry property of order* $2k$*, i.e., for all* $2k$ *or less sparse vectors* $\mathbf{z}$,

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

holds simultaneously for some $\delta_{2k}^{\odot} \in (0, 1)$.

Condition 2: $L \geq \left(\frac{c_1 k \log n}{\eta}\right)$, *where* $\eta = c_2 \left(\frac{\gamma_{min}}{\sigma^2 + \gamma_{max}}\right)^2 \left(1 - \delta_{2k}^{\odot}\right)$

[1] S. Khanna and C. R. Murthy, On the Support Recovery of Jointly Sparse Gaussian Sources using Sparse Bayesian Learning. arXiV preprint: arXiv:1703.04930.

Support Recovery in MSBL[2]

- For Gaussian measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ satisfying $\delta_{2k}^{\mathbf{\Phi} \odot \mathbf{\Phi}} < 1$, MSBL can perfectly recover any k-sparse support with high probability, if

Case I: $\mathbf{k} < \mathbf{m}$	
Noiseless measurements	$L \geq 1$
Noisy measurements	$L \geq \mathcal{O}(k \log n)$

Case II: $\mathbf{k} \in \left[\mathbf{m}\ ,\ (\mathbf{m}^2 + \mathbf{m})/4\right]$	
Noiseless measurements	$L \geq \mathcal{O}(k \log n)$
Noisy measurements	$L \geq \mathcal{O}(k \log n)$

[2] S. Khanna and C. R. Murthy, On the Support Recovery of Jointly Sparse Gaussian Sources using Sparse Bayesian Learning. arXiV preprint: arXiv:1703.04930.

Covariance matching in MSBL

- MSBL's log-likelihood objective:

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

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

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

Covariance matching in MSBL

- MSBL's log-likelihood objective:

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

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

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

MSBL minimizes $\mathcal{D}_\phi\left(\underbrace{\frac{1}{L}\mathbf{YY}^T}_{\text{emp. cov mat}} \;,\; \underbrace{\sigma^2\mathbf{I}_m + \mathbf{\Phi\Gamma\Phi}^T}_{\text{param. cov mat}}\right)$ to find γ.

A new covariance matching algorithm

- Simplified Gaussian prior on $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L$, parameterized directly by support $\mathcal{S}$:

$$\mathbf{x}_j \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \gamma \times \mathrm{diag}(\mathbf{1}_{\mathcal{S}})\right),$$
$$\mathbf{y}_j \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma^2 \mathbf{I}_m + \gamma \mathbf{\Phi}_{\mathcal{S}} \mathbf{\Phi}_{\mathcal{S}}^T\right)$$

- Estimate $\hat{\mathcal{S}}$ via generalized reverse $\mathcal{I}$-projection,

$$\hat{\mathcal{S}} = \underset{\mathcal{S} \subseteq [n]}{\arg\min}\ \mathcal{D}_\alpha \left(\prod_{j=1}^{L} \mathcal{N}(\mathbf{y}_j; 0, \mathbf{R}_{\mathbf{Y}}),\ p(\mathbf{Y}; \mathcal{S}) \right),$$

where $\mathcal{D}_\alpha$ is the Rényi divergence of order $\alpha \in (0, 1]$.

Rényi divergence based Covariance Matching Pursuit (RD-CMP):

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

A new covariance matching algorithm

- Simplified Gaussian prior on $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L$, parameterized directly by support $\mathcal{S}$:

$$\mathbf{x}_j \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \gamma \times \mathrm{diag}(\mathbf{1}_{\mathcal{S}})\right),$$
$$\mathbf{y}_j \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma^2 \mathbf{I}_m + \gamma \mathbf{\Phi}_{\mathcal{S}} \mathbf{\Phi}_{\mathcal{S}}^T\right)$$

- Estimate $\hat{\mathcal{S}}$ via generalized reverse $\mathcal{I}$-projection,

$$\hat{\mathcal{S}} = \underset{\mathcal{S} \subseteq [n]}{\arg\min}\ \mathcal{D}_\alpha \left(\prod_{j=1}^{L} \mathcal{N}(\mathbf{y}_j; 0, \mathbf{R}_{\mathbf{Y}}) \ ,\ p(\mathbf{Y}; \mathcal{S}) \right),$$

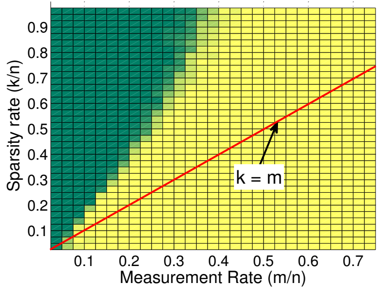
where $\mathcal{D}_\alpha$ is the Rényi divergence of order $\alpha \in (0, 1]$.

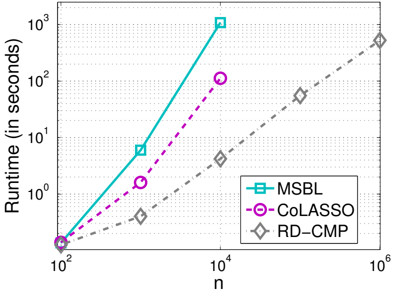
Rényi divergence based Covariance Matching Pursuit (RD-CMP):

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

- Rényi divergence objective is a **difference of two submodular set functions**.
  - A two step iterative Majorization-minimization procedure for finding $\hat{\mathcal{S}}$:
    1. Replace 1^{st} log-det term with its modular upper bound.
    2. Minimize the resulting supermodular objective via greedy search.

RD-CMP's performance[3]





Support recovery phase transition

($n = 200$, SNR = 10 dB)

Average runtime vs signal dimension (n)

SNR = 20 dB,

$k = \lceil 50 \log_{10} n \rceil$, $mL = \lceil 50k \log_{10} n \rceil$

[3] S. Khanna and C. R. Murthy, Renyi Divergence Based Covariance Matching Pursuit of Joint Sparse Support, submitted to SPAWC-2017.

Thank you... questions?

Saurabh Khanna

PhD Student,
Signal Processing for Communication (SPC) Lab,
Department of Electrical Communication Engineering,
Indian Institute of Science, Bangalore

sakhanna@ece.iisc.ernet.in