Bayesian Techniques for Joint-Sparse Signal Recovery: Theory and Algorithms

Saurabh Khanna

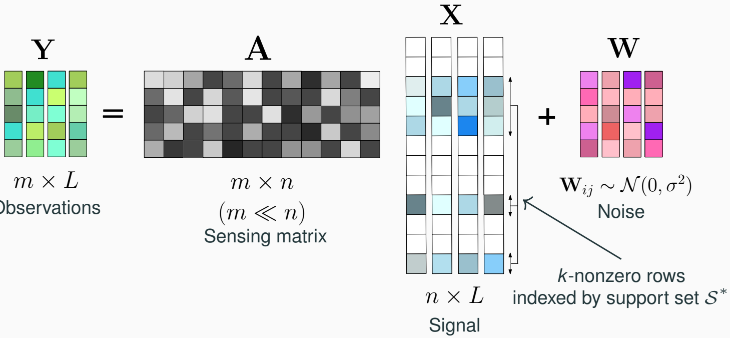
Advisor: Prof. Chandra R. Murthy
Electrical Communication Engineering Dept.,
Indian Institute of Science, Bangalore

Outline

- Joint sparse signal/support recovery problems
- Sparse Bayesian Learning (SBL) framework
  - New theoretical results
  - Covariance matching principle
  - Khatri-Rao product - restricted isometry and null space
- Rényi divergence based support recovery algorithm
- Distributed extensions of SBL
- Conclusions and future research

Canonical problem

- Consider the simultaneous linear equations: $\mathbf{y}_j = \mathbf{A}\mathbf{x}_j + \mathbf{w}_j,\ j \in [L]$.

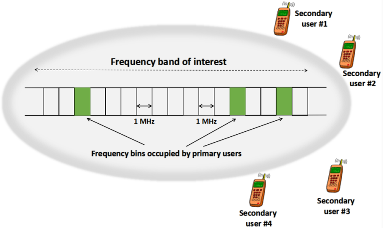


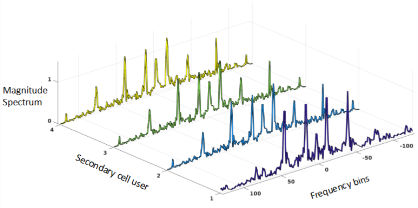
- Columns of **X** are **jointly sparse** with common nonzero support.

Multiple Measurement Vector problem	Joint Sparse Support Recovery
Reconstruct entire **X** from $\{\mathbf{Y}, \mathbf{A}\}$	Reconstruct support(**X**) from $\{\mathbf{Y}, \mathbf{A}\}$

Multi-sensor signal processing

- Spectrum sensing in cognitive radio network

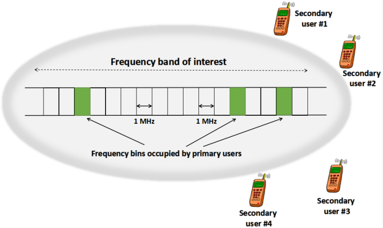


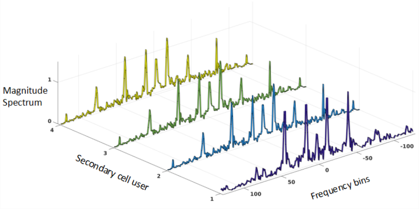


Joint sparsity in frequency domain

Multi-sensor signal processing

- Spectrum sensing in cognitive radio network



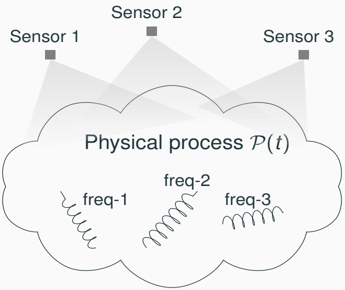


Joint sparsity in frequency domain

- Multi-sensor data is typically **highly structured or correlated** due to
  - overlapped sensing regions/common sensory target.

- [Tropp, 04], [Duarte, 05] proposed joint sparsity based data models for structured/correlated multi-sensor data.

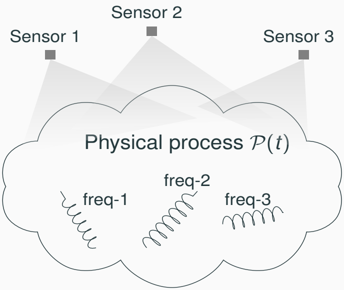
A generative model for multi-sensor data



Data from different sensors have *overlapping signal subspaces*.

approximate as *different linear combinations of the same elementary signals*

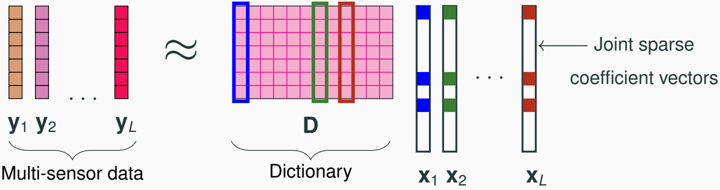
A generative model for multi-sensor data



Data from different sensors have *overlapping signal subspaces*.

approximate as *different linear combinations of the same elementary signals*

- Simultaneous Sparse Approximation (SSA) Model: [Tropp, 04]



A compression scheme for multi-sensor data

- Encoder:

$$\underbrace{[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_L]}_{\text{low dim. sketch}} = \underset{\substack{m\times n \\ \text{frame} \\ (m<<n)}}{\mathbf{A}} \underbrace{[\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_L]}_{\substack{\text{high dimensional} \\ \text{data vectors from } L \text{ sensors}}}$$

- Decoder:

$$[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_L] \approx \mathbf{A} \; \mathbf{D} \underbrace{[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L]}_{\text{joint sparse coefficients}} \quad \text{(SSA approx.)}$$

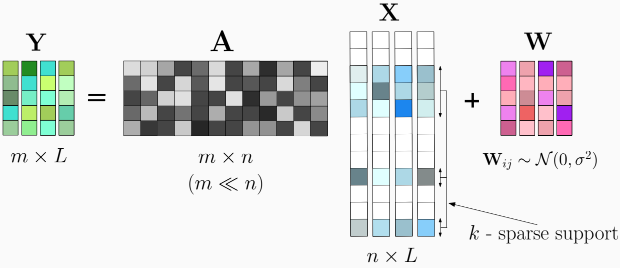
  - Step 1: First recover **joint-sparse** coefficients $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L\}$.
  - Step 2: Then reconstruct multi-sensor data as $\hat{\mathbf{s}}_j = \mathbf{D}\hat{\mathbf{x}}_j$.

Joint sparse recovery - applications

- Anomaly/sparse event localization [Jiang, 13], [Adler, 13], [Lagunas, 16]
- Cooperative spectrum sensing [Bazerque, 10], [Fanzi, 11]
- Distributed source coding [Baron, 09]
- Magnetoencephalography (MEG) [Fornasier, 08]
- Direction of arrival estimation [Tan, 14]
- MIMO wireless channel estimation [Prasad, 15], [Masood, 15]
- Hyperspectral imaging [Iordache, 14]

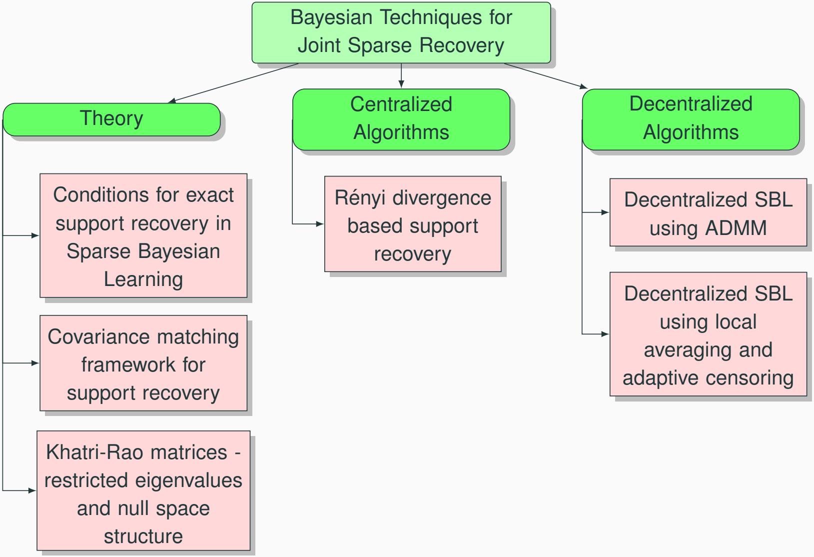
Main challenges and goals

Recover **X** or supp(**X**) from **Y**.



- Conditions for exact support recovery in Sparse Bayesian Learning.
  - What values of (m, n, L) allow perfect k-sparse support recovery?
  - Design guidelines for sensing matrix **A**.
- Algorithms for efficient estimation of **X** or supp(**X**)?
  - Handling extremely large signal dimensions.
  - Distributed/parallel implementation.

Thesis contributions



ℓ_0 bound and beyond..

Fundamental limits on support recovery

The ℓ_0 bound

$$\mathbf{L_0}: \quad \min_{\mathbf{X}\in\mathbb{R}^{n\times L}} \underbrace{\mathcal{R}(\mathbf{X})}_{\substack{\text{no. of nonzero} \\ \text{rows in } \mathbf{X}}} \quad \text{subject to } \mathbf{Y} = \mathbf{AX}.$$

Unique solution when... [Chen & Huo, 06]

A k-sparse $\mathbf{X}$ is uniquely recoverable via L_0 if

$$k < \frac{\text{spark}(\mathbf{A}) - 1 + \text{rank}(\mathbf{Y})}{2}. \qquad (\ell_0\text{-bound})$$

spark($\mathbf{A}$):= minimum no. of linearly dependent columns in $\mathbf{A}$.

The ℓ_0 bound

$$\mathbf{L_0}: \quad \min_{\mathbf{X} \in \mathbb{R}^{n \times L}} \underbrace{\mathcal{R}(\mathbf{X})}_{\substack{\text{no. of nonzero} \\ \text{rows in } \mathbf{X}}} \quad \text{subject to } \mathbf{Y} = \mathbf{A}\mathbf{X}.$$

Unique solution when... [Chen & Huo, 06]

A k-sparse $\mathbf{X}$ is uniquely recoverable via L_0 if

$$k < \frac{\text{spark}(\mathbf{A}) - 1 + \text{rank}(\mathbf{Y})}{2}. \qquad (\ell_0\text{-bound})$$

spark($\mathbf{A}$):= minimum no. of linearly dependent columns in $\mathbf{A}$.

- Supports of size up to m are uniquely recoverable....when spark($\mathbf{A}$) $= m + 1$!

Towards ℓ_0 bound

Mixed norm regularization **[Chen & Huo, 06]**

$$\mathbf{L_{p,q}}: \quad \min_{\mathbf{X}} \sum_{i=1}^{m} \left(||\mathbf{X}(i,:)||_q \right)^p \text{ subject to } \mathbf{Y} = \mathbf{AX}.$$

- Joint sparse solution for $p \in [0,1]$ and $q \geq 1$.
- Unique k-sparse solution if $||\mathbf{A}_{\mathcal{S}}^{\dagger}\mathbf{a}_j|| < 1, \forall j \notin \mathcal{S}$.
- $k \left(\leq \frac{m}{2} \right)$ sparse $\mathbf{X}$ is uniquely recoverable.

Towards ℓ_0 bound

Mixed norm regularization **[Chen & Huo, 06]**

$$\mathbf{L_{p,q}} : \quad \min_{\mathbf{X}} \sum_{i=1}^{m} \left(||\mathbf{X}(i,:)||_q \right)^p \text{ subject to } \mathbf{Y} = \mathbf{AX}.$$

- Joint sparse solution for $p \in [0,1]$ and $q \geq 1$.
- Unique k-sparse solution if $||\mathbf{A}_{\mathcal{S}}^{\dagger}\mathbf{a}_j|| < 1, \forall j \notin \mathcal{S}$.
- $k \left(\leq \frac{m}{2} \right)$ sparse $\mathbf{X}$ is uniquely recoverable.

Iterative hard thresholding / greedy approach **[Blanchard, 14]**

- Examples: SOMP, Co-SAMP, SIHT.
- $k \leq \mathcal{O}\left(\frac{m}{\log n} \right)$ sparse supports are recoverable.

Meeting ℓ_0 bound

Multi Signal Classification (MUSIC) criterion [Peng & Bresler, 97]

- Index $j \in \text{support}(\mathbf{X}^*)$ iff

$$\mathbf{Q}^H \mathbf{a}_j = 0 \quad \text{or} \quad \mathbf{a}_j^H \mathbf{P}_{R(\mathbf{Q})} \mathbf{a}_j = 0,$$

  where the orthogonal columns of $\mathbf{Q}$ span the noise subspace.

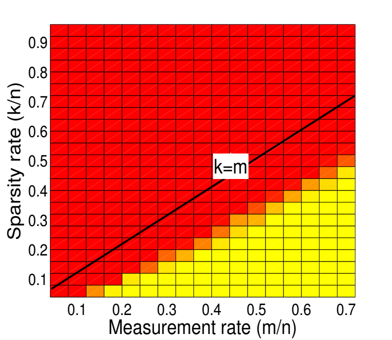
- MUSIC criterion recovers any $k(< m)$-sparse support when $\mathbf{A}$ has full spark!

- Algorithms: SA-MUSIC, CS-MUSIC.

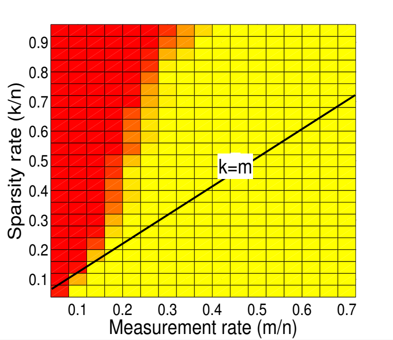
Beyond ℓ_0 bound?

Support recovery phase transition
($n = 200$, $L = 400$, SNR = 20 dB)

Simultaneous Orthgonal Matching Pursuit

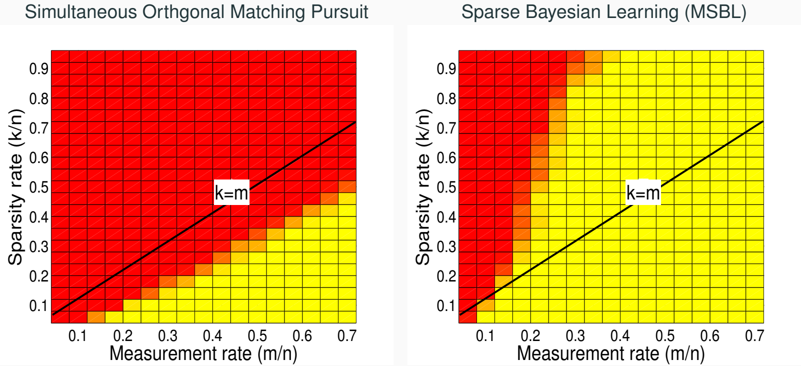


Sparse Bayesian Learning (MSBL)



Beyond ℓ_0 bound?

Support recovery phase transition
($n = 200$, $L = 400$, SNR = 20 dB)



Key Idea: Type-II estimation of **X** using correlation aware priors.

Sparse Bayesian Learning

Performance guarantees
&
connections to covariance matching

Sparse Bayesian Learning (SBL)

- $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W}$
  - $\mathbf{x}_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \boldsymbol{\Gamma}),\ \boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$ Correlation-aware prior!
  - $\mathrm{supp}(\mathbf{x}_j) = \mathrm{supp}(\gamma)$. Common covariance induces joint sparsity
  - Gaussian observations: $\mathbf{y}_j \sim \mathcal{N}(0, \sigma^2\mathbf{I} + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^T)$.

Sparse Bayesian Learning (SBL)

- $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$
  - $\mathbf{x}_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \boldsymbol{\Gamma}),\ \boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$ Correlation-aware prior!
  - supp($\mathbf{x}_j$) = supp(γ). Common covariance induces joint sparsity
  - Gaussian observations: $\mathbf{y}_j \sim \mathcal{N}(0, \sigma^2\mathbf{I} + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^T)$.

- Multiple Sparse Bayesian Learning (MSBL) [Wipf & Rao, 07]:

$$\begin{aligned}\hat{\gamma} &= \underset{\boldsymbol{\gamma}\in\mathbb{R}^n_+}{\mathrm{argmax}}\ \log p(\mathbf{Y}; \gamma) \\ &= \underset{\boldsymbol{\gamma}\in\mathbb{R}^n_+}{\mathrm{argmin}}\ L \log\left|\sigma^2\mathbf{I}_m + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^T\right| + \mathrm{tr}\left(\mathbf{Y}^T(\sigma^2\mathbf{I}_m + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^T)^{-1}\mathbf{Y}\right)\end{aligned}$$

  - Nonconvex objective, $\hat{\gamma}$ found via Expectation Maximization (EM).
  - Support($\hat{\gamma}$) declared as estimate of true support $\mathcal{S}^*$.

Support recovery in SBL (noiseless measurements)

Support error....a large deviation event

Let $\mathbf{x}_j \sim \mathcal{N}(0, \Gamma^*)$ and $\hat{\gamma}$ be a global maximizer of the MSBL objective, then

$$\mathbb{P}\left(\text{supp}(\hat{\gamma}) \neq \mathcal{S}^*\right) \leq \exp\left(-\frac{L\mathcal{D}_\alpha\left(p_{\hat{\boldsymbol{\gamma}}}, p_{\boldsymbol{\gamma}^*}\right)}{4}\right).$$

$\mathcal{D}_\alpha\left(p_{\hat{\boldsymbol{\gamma}}}, p_{\boldsymbol{\gamma}^*}\right)$:= α-Rényi Divergence between Gaussian densities:

$$p_{\hat{\boldsymbol{\gamma}}} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^T) \quad \text{and} \quad p_{\boldsymbol{\gamma}^*} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m + \mathbf{A}\boldsymbol{\Gamma}^*\mathbf{A}^T).$$

Support recovery in SBL (noiseless measurements)

Support error....a large deviation event

Let $\mathbf{x}_j \sim \mathcal{N}(0, \Gamma^*)$ and $\hat{\gamma}$ be a global maximizer of the MSBL objective, then

$$\mathbb{P}\left(\text{supp}(\hat{\gamma}) \neq \mathcal{S}^*\right) \leq \exp\left(-\frac{L\mathcal{D}_\alpha\left(p_{\hat{\gamma}}, p_{\gamma^*}\right)}{4}\right).$$

$\mathcal{D}_\alpha\left(p_{\hat{\gamma}}, p_{\gamma^*}\right)$:= α-Rényi Divergence between Gaussian densities:

$$p_{\hat{\gamma}} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T) \quad \text{and} \quad p_{\gamma^*} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma^*\mathbf{A}^T).$$

- For k-sparse vectors $\gamma^*, \hat{\gamma} \in \mathbb{R}_+^n$ with distinct supports,

$$\mathcal{D}_{1/2}\left(p_{\hat{\gamma}}, p_{\gamma^*}\right) \to \infty \;\text{ as }\; \sigma^2 \to 0,$$

  when $k < \text{spark}(\mathbf{A}) - 1$.

Support recovery in SBL (noiseless measurements)

Support error....a large deviation event

Let $\mathbf{x}_j \sim \mathcal{N}(0, \Gamma^*)$ and $\hat{\gamma}$ be a global maximizer of the MSBL objective, then

$$\mathbb{P}\left(\text{supp}(\hat{\gamma}) \neq \mathcal{S}^*\right) \leq \exp\left(-\frac{L\mathcal{D}_\alpha\left(p_{\hat{\gamma}}, p_{\gamma^*}\right)}{4}\right).$$

$\mathcal{D}_\alpha\left(p_{\hat{\gamma}}, p_{\gamma^*}\right)$:= α-Rényi Divergence between Gaussian densities:

$$p_{\hat{\gamma}} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^T) \quad \text{and} \quad p_{\gamma^*} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m + \mathbf{A}\boldsymbol{\Gamma}^*\mathbf{A}^T).$$

- For k-sparse vectors $\gamma^*, \hat{\gamma} \in \mathbb{R}_+^n$ with distinct supports,

  $$\mathcal{D}_{1/2}\left(p_{\hat{\gamma}}, p_{\gamma^*}\right) \to \infty \;\text{ as }\; \sigma^2 \to 0,$$

  when $k < \text{spark}(\mathbf{A}) - 1$.

- **Implication:** If $|\mathcal{S}^*|, ||\hat{\gamma}||_0 < \text{spark}(\mathbf{A}) - 1$, then $\text{supp}(\hat{\gamma}) = \mathcal{S}^*$ almost surely!

Support recovery in SBL (noisy measurements)

Support error probability in MSBL

1. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L$ be i.i.d zero mean Gaussian vectors with support $\mathcal{S}^*$, $|\mathcal{S}^*| \leq k$.
2. and... variance of nonzero entries in $\mathbf{X}$ lie in $[\gamma_{\text{min}}, \gamma_{\text{max}}]$.

For any MSBL solution $\hat{\gamma}$ with nonzero coefficients in $[\gamma_{\text{min}}, \gamma_{\text{max}}]$,

$$\mathbb{P}\left(\text{supp}(\hat{\gamma}) \neq \mathcal{S}^*\right) \leq 2e^{-L\left(\frac{\eta}{8} - \frac{c_1 k \log n}{L}\right)},$$

where c_1 is a dimension free constant, and

$$\eta \triangleq \min_{\mathcal{S} \subseteq [n] \setminus \mathcal{S}^*} \min_{\substack{\boldsymbol{\gamma} \in \mathbb{R}_+^n, \\ \text{supp}(\boldsymbol{\gamma}) = \mathcal{S}}} \frac{||(\mathbf{A} \odot \mathbf{A})(\gamma - \gamma^*)||_2^2}{\left(|\mathcal{S} \setminus \mathcal{S}^*| + |\mathcal{S}^* \setminus \mathcal{S}|\right)\left(\sigma^2 + 2\gamma_{\text{max}} \sigma_{\text{max}}^2(\mathbf{A}_{\mathcal{S} \cup \mathcal{S}^*})\right)^2}.$$

- Support error probability vanishes for $\eta > 0$ and $L \geq O\left(\frac{k \log n}{\eta}\right)$.

Null Space of $\mathbf{A} \odot \mathbf{A}$

Strong Null Space Property

Suppose the ℓ_2-norm columns in $\mathbf{A} \in \mathbb{R}^{m \times n}$ lie in $[1-\alpha, 1+\alpha]$ for some $\alpha \in (0,1)$, then

$$||(\mathbf{A} \odot \mathbf{A})\mathbf{v}||_2^2 \geq \frac{(1-\alpha)^2}{2m} \left(||\mathbf{v}_+||_1^2 + ||\mathbf{v}_-||_1^2 \right)$$

for all $\mathbf{v} \in \mathbb{R}^n$ such that $\dfrac{||\mathbf{v}_+||_1}{||\mathbf{v}_-||_1} \geq 4 \left(\dfrac{1+\alpha}{1-\alpha} \right)^2$. Here, $\mathbf{v}_+$ and $\mathbf{v}_-$ are nonneg. vectors in $\mathbb{R}^n$ retaining only pos. and neg. entries of $\mathbf{v}$.

Null Space of $\mathbf{A} \odot \mathbf{A}$

Strong Null Space Property

Suppose the ℓ_2-norm columns in $\mathbf{A} \in \mathbb{R}^{m \times n}$ lie in $[1-\alpha, 1+\alpha]$ for some $\alpha \in (0,1)$, then

$$||(\mathbf{A} \odot \mathbf{A})\mathbf{v}||_2^2 \geq \frac{(1-\alpha)^2}{2m} \left(||\mathbf{v}_+||_1^2 + ||\mathbf{v}_-||_1^2 \right)$$

for all $\mathbf{v} \in \mathbb{R}^n$ such that $\frac{||\mathbf{v}_+||_1}{||\mathbf{v}_-||_1} \geq 4 \left(\frac{1+\alpha}{1-\alpha} \right)^2$. Here, $\mathbf{v}_+$ and $\mathbf{v}_-$ are nonneg. vectors in $\mathbb{R}^n$ retaining only pos. and neg. entries of $\mathbf{v}$.

- **Implication 1:** Null space of $\mathbf{A} \odot \mathbf{A}$ is devoid of vectors like

$$\Delta\gamma = \underbrace{\gamma}_{\text{dense nonnegative}} - \underbrace{\gamma^*}_{\text{sparse nonnegative}} .$$

Null Space of $\mathbf{A} \odot \mathbf{A}$

Strong Null Space Property

Suppose the ℓ_2-norm columns in $\mathbf{A} \in \mathbb{R}^{m \times n}$ lie in $[1-\alpha, 1+\alpha]$ for some $\alpha \in (0,1)$, then

$$||(\mathbf{A} \odot \mathbf{A})\mathbf{v}||_2^2 \geq \frac{(1-\alpha)^2}{2m}\left(||\mathbf{v}_+||_1^2 + ||\mathbf{v}_-||_1^2\right)$$

for all $\mathbf{v} \in \mathbb{R}^n$ such that $\frac{||\mathbf{v}_+||_1}{||\mathbf{v}_-||_1} \geq 4\left(\frac{1+\alpha}{1-\alpha}\right)^2$. Here, $\mathbf{v}_+$ and $\mathbf{v}_-$ are nonneg. vectors in $\mathbb{R}^n$ retaining only pos. and neg. entries of $\mathbf{v}$.

- **Implication 1:** Null space of $\mathbf{A} \odot \mathbf{A}$ is devoid of vectors like

$$\Delta\gamma = \underbrace{\gamma}_{\text{dense nonnegative}} - \underbrace{\gamma^*}_{\text{sparse nonnegative}}.$$

- **Implication 2:** For subgaussian $\mathbf{A}$ with $m \geq O(\log n)$ rows, and large enough L, MSBL solution is only $O(|\mathcal{S}^*|)$ sparse!

No dense MSBL solutions!

MSBL optimization - a closer look

- MSBL's log-likelihood objective:

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

$$\propto \log|\sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T| + \text{trace}\left(\left(\sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right)^{-1}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T\right)\right)$$

MSBL optimization - a closer look

- MSBL's log-likelihood objective:

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

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

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

MSBL optimization - a closer look

- MSBL's log-likelihood objective:

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

$$\propto \log|\sigma^2\mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T| + \text{trace}\left(\left(\sigma^2\mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right)^{-1}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T\right)\right)$$

$$\propto \underbrace{\mathcal{D}^{\text{Bregman}}_{-\log\det}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T,\ \sigma^2\mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right)}_{\text{Log Det Bregman Matrix Div.}} + \text{constant terms}$$

MSBL optimization - a closer look

- MSBL's log-likelihood objective:

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

$$\propto \log|\sigma^2\mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T| + \text{trace}\left(\left(\sigma^2\mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right)^{-1}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T\right)\right)$$

$$\propto \underbrace{\mathcal{D}^{\text{Bregman}}_{-\log\det}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T,\ \sigma^2\mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right)}_{\text{Log Det Bregman Matrix Div.}} + \text{constant terms}$$

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

MSBL optimization - a closer look

- MSBL's log-likelihood objective:

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

$$\propto \log|\sigma^2\mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T| + \text{trace}\left(\left(\sigma^2\mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right)^{-1}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T\right)\right)$$

$$\propto \underbrace{\mathcal{D}^{\text{Bregman}}_{-\log\det}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T,\ \sigma^2\mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T\right)}_{\text{Log Det Bregman Matrix Div.}} + \text{ constant terms}$$

- MSBL 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 other matrix divergences?

Covariance matching framework for support recovery

- Multiple measurement vectors (MMVs): $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W}$
  - $\mathbf{x}_j \sim \underbrace{\mathcal{N}(0, \mathrm{diag}(\gamma))}_{\text{correlation aware prior}}$
  - $\mathbf{y}_j \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m + \mathbf{A}\Gamma\mathbf{A}^T)$
- 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 Principle:

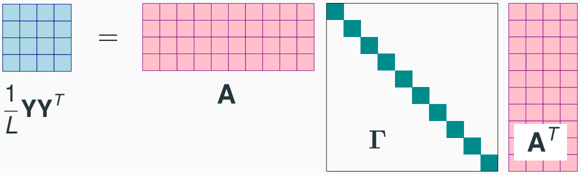
$$\hat{\gamma} = \underset{\gamma \in \mathbb{R}_+^n}{\text{argmin}} \ \textbf{distance}\Big(\underbrace{\mathbf{R}_{\mathbf{Y}}}_{\substack{\text{empirical} \\ \text{MMV covariance}}}, \underbrace{\sigma^2 \mathbf{I} + \mathbf{A}\Gamma\mathbf{A}^T}_{\substack{\text{parameterized} \\ \text{MMV covariance}}} \Big) + \lambda \underbrace{h(\gamma)}_{\substack{\text{optional} \\ \text{concave penalty}}}$$

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

Restricted Isometry of Khatri-Rao product

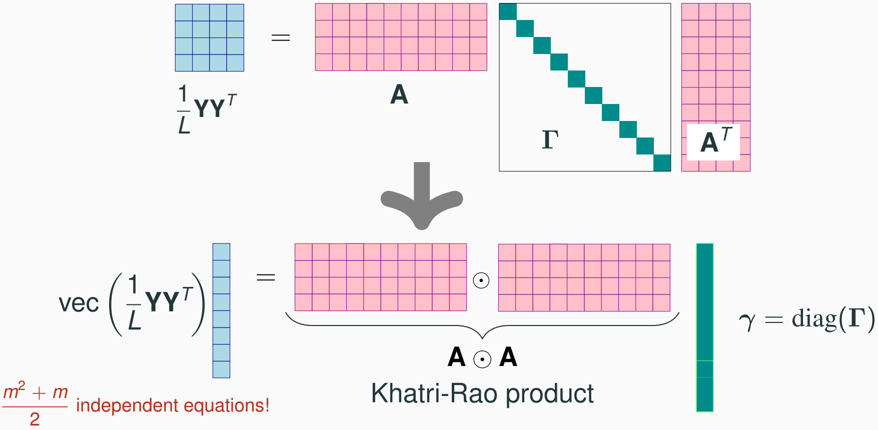
Covariance matching - a closer look

- Covariance matching constraint: $\frac{1}{L}\mathbf{Y}\mathbf{Y}^T \approx \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^T$



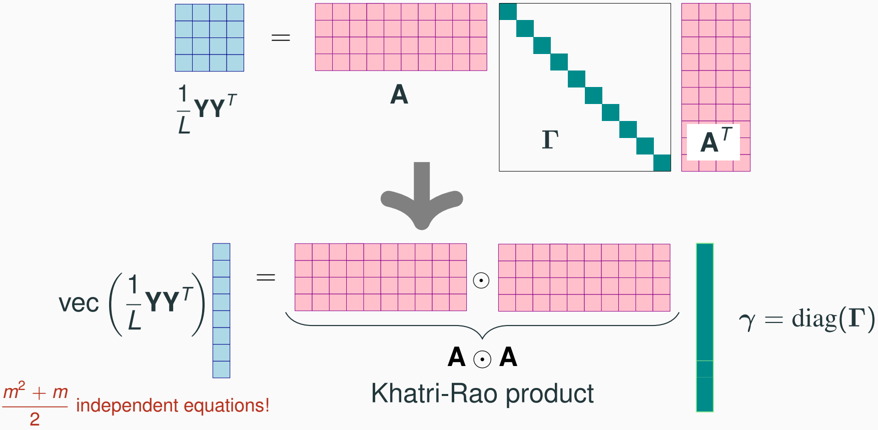
Covariance matching - a closer look

- Covariance matching constraint: $\frac{1}{L}\mathbf{Y}\mathbf{Y}^T \approx \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^T$



Covariance matching - a closer look

- Covariance matching constraint: $\frac{1}{L}\mathbf{Y}\mathbf{Y}^T \approx \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^T$



- Stable recovery of k-sparse γ.... possible if $\mathbf{A} \odot \mathbf{A}$ behaves like an isometry for all k-sparse nonnegative vectors.

Columnwise Khatri-Rao product

- Khatri-Rao product

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

$\otimes$ denotes Kronecker product

- Khatri-Rao product form arises naturally in
  - Sparsity pattern recovery (via covariance matching)
  - Direction of arrival estimation
  - PARAFAC based tensor decomposition
  - Estimation of power spectral density of stationary graph signals

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

Restricted Isometry Property (RIP)

Restricted Isometry Property of k^{th} order (k-RIP)

Matrix $\mathbf{A}$ satisfies k-RIP if there exists a constant $\delta_k \in (0, 1)$ such that

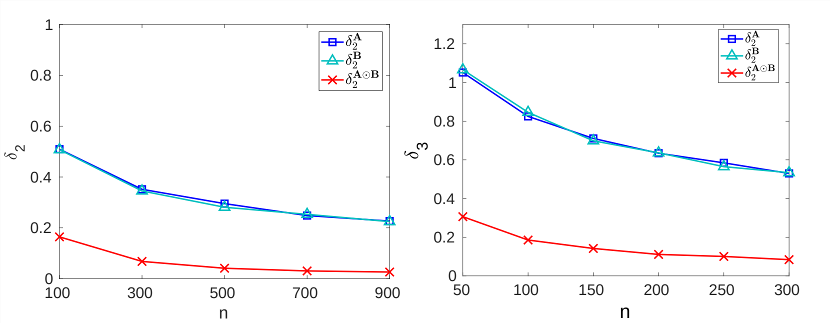
$$(1 - \delta_k) \, ||\mathbf{z}||_2^2 \leq ||\mathbf{A}\mathbf{z}||_2^2 \leq (1 + \delta_k) \, ||\mathbf{z}||_2^2 \, ,$$

for all k-sparse vectors $\mathbf{z}$.

- Smallest δ_k is called the k**-RIC** of $\mathbf{A}$.
- How small can k-RIC of a generic Khatri-Rao matrix $\mathbf{A} \odot \mathbf{B}$ be?

RIP of Khatri-Rao product - an empirical study

$$\mathbf{A}_{i,j}, \mathbf{B}_{i,j} \sim \mathcal{N}\left(0, \frac{1}{m}\right) \text{ and } m = 0.5n$$



RIP improved by taking Khatri-Rao product!

Deterministic RIC bound for $\mathbf{A} \odot \mathbf{B}$

Deterministic RIC bound

For $m \times n$ sized matrices $\mathbf{A}$ and $\mathbf{B}$ with unit ℓ_2-norm columns,

$$\delta_k(\mathbf{A} \odot \mathbf{B}) \leq [\max(\delta_k(\mathbf{A}), \delta_k(\mathbf{B}))]^2 \quad \text{for all } k \leq m.$$

- Mathematical tools:
  - $(\mathbf{A} \odot \mathbf{B})^T(\mathbf{A} \odot \mathbf{B}) = \mathbf{A}^T\mathbf{A} \circ \mathbf{B}^T\mathbf{B}$
  - Kantorovitch matrix inequalities
- Key features of the bound:
  - bound is expressed in terms of k-RICs of the input matrices
  - $\delta_k(\mathbf{A} \odot \mathbf{A}) \leq (\delta_k(\mathbf{A}))^2 < \delta_k(\mathbf{A})$. **RIP improves!**

Probabilistic RIC bound for Khatri-Rao product

Probabilistic RIC bound

Let $\mathbf{A}$ and $\mathbf{B}$ be $m \times n$ sized matrices with zero mean, unit variance, i.i.d. subgaussian entries satisfying $\left|\left|\mathbf{A}_{ij}\right|\right|_{\psi_2}, \left|\left|\mathbf{B}_{ij}\right|\right|_{\psi_2} \leq \beta$. Then,

$$\mathbb{P}\left(\delta_k\left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}}\right) \geq \delta\right) \leq \frac{10}{n^{2(\gamma-1)}}$$

for all $\gamma \geq 1$, provided

$$m \geq 4c\gamma\beta^2\left(\frac{k \log n}{\delta}\right).$$

Here, c is an absolute numerical constant.

- For $m \geq O\left(\frac{k \log n}{\delta}\right)$, one can have $\delta_k\left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{A}}{\sqrt{m}}\right) \leq \delta$ w.h.p.

Probabilistic RIC bound for Khatri-Rao product

Probabilistic RIC bound

Let **A** and **B** be $m \times n$ sized matrices with zero mean, unit variance, i.i.d. subgaussian entries satisfying $||\mathbf{A}_{ij}||_{\psi_2}, ||\mathbf{B}_{ij}||_{\psi_2} \leq \beta$. Then,

$$\mathbb{P}\left(\delta_k\left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}}\right) \geq \delta\right) \leq \frac{10}{n^{2(\gamma-1)}}$$

for all $\gamma \geq 1$, provided

$$m \geq 4c\gamma\beta^2\left(\frac{k \log n}{\delta}\right).$$

Here, c is an absolute numerical constant.

If A has i.i.d. Gaussian entries, $\mathbb{P}\left(\delta_k\left(\frac{\mathbf{A}}{\sqrt{m}}\right) > \delta\right) \leq \frac{1}{n^\alpha}$ provided $m \geq \frac{c}{\delta^2}(k+\alpha)\log n$ [Foucart & Rauhut, Thm. 9.27]

- For $m \geq O\left(\frac{k \log n}{\delta}\right)$, one can have $\delta_k\left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{A}}{\sqrt{m}}\right) \leq \delta$ w.h.p.

Probabilistic RIC bound for Khatri-Rao product

Probabilistic RIC bound

Let **A** and **B** be $m \times n$ sized matrices with zero mean, unit variance, i.i.d. subgaussian entries satisfying $||\mathbf{A}_{ij}||_{\psi_2}, ||\mathbf{B}_{ij}||_{\psi_2} \leq \beta$. Then,

$$\mathbb{P}\left(\delta_k\left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}}\right) \geq \delta\right) \leq \frac{10}{n^{2(\gamma-1)}}$$

for all $\gamma \geq 1$, provided

$$m \geq 4c\gamma\beta^2\left(\frac{k \log n}{\delta}\right).$$

Here, c is an absolute numerical constant.

If A has i.i.d. Gaussian entries, $\mathbb{P}\left(\delta_k\left(\frac{\mathbf{A}}{\sqrt{m}}\right) > \delta\right) \leq \frac{1}{n^\alpha}$ provided $m \geq \frac{c}{\delta^2}(k+\alpha)\log n$

[Foucart & Rauhut, Thm. 9.27]

- For $m \geq O\left(\frac{k \log n}{\delta}\right)$, one can have $\delta_k\left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{A}}{\sqrt{m}}\right) \leq \delta$ w.h.p.

In MSBL, $O(k \log n)$ measurements per MMV are sufficient to guarantee exact recovery of any k-sparse support!

Rényi divergence based support recovery

Support recovery using Rényi Divergence

- MMV model: $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W}$
  - Let set $\mathcal{S}$ be the unknown support($\mathbf{X}$)
  - $\mathbf{x}_j \sim \mathcal{N}(0, \gamma \text{diag}(\mathbf{1}_{\mathcal{S}}))$
  - $\mathbf{y}_j \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m + \gamma \mathbf{A}_{\mathcal{S}} \mathbf{A}_{\mathcal{S}}^T)$

Covariance matrix parameterized by support $\mathcal{S}$!

Support recovery using Rényi Divergence

- MMV model: $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$
  - Let set $\mathcal{S}$ be the unknown support($\mathbf{X}$)
  - $\mathbf{x}_j \sim \mathcal{N}(0, \gamma \mathrm{diag}(\mathbf{1}_{\mathcal{S}}))$
  - $\mathbf{y}_j \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m + \gamma \mathbf{A}_{\mathcal{S}} \mathbf{A}_{\mathcal{S}}^T)$

Covariance matrix parameterized by support $\mathcal{S}$!

- Covariance matching using α-Rényi divergence

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

Support recovery using Rényi Divergence

- MMV model: $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W}$
  - Let set $\mathcal{S}$ be the unknown support($\mathbf{X}$)
  - $\mathbf{x}_j \sim \mathcal{N}(0, \gamma \text{diag}(\mathbf{1}_\mathcal{S}))$
  - $\mathbf{y}_j \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m + \gamma \mathbf{A}_\mathcal{S} \mathbf{A}_\mathcal{S}^T)$

Covariance matrix parameterized by support $\mathcal{S}$!

- Covariance matching using α-Rényi divergence

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

- **Rényi Divergence based Covariance Matching Pursuit (RD-CMP)**

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

Submodular functions - a primer

- Let $\mathcal{V}$ be the ground set of elements.
- Set function $f : \mathcal{V} \to \mathbb{R}_+$ is **submodular**, if for $\mathcal{S} \subseteq \mathcal{T} \subseteq \mathcal{V}$,
  - monotonicity

    $$f(\mathcal{S}) \leq f(\mathcal{T})$$

  - diminishing returns property

    $$f(\mathcal{T} \cup \{a\}) - f(\mathcal{T}) \leq f(\mathcal{S} \cup \{a\}) - f(\mathcal{S}) \quad \forall a \in \mathcal{V} \backslash \mathcal{T}$$

Submodular functions - a primer

- Let $\mathcal{V}$ be the ground set of elements.
- Set function $f : \mathcal{V} \rightarrow \mathbb{R}_+$ is **submodular**, if for $\mathcal{S} \subseteq \mathcal{T} \subseteq \mathcal{V}$,
  - monotonicity

    $$f(\mathcal{S}) \leq f(\mathcal{T})$$
  - diminishing returns property

    $$f(\mathcal{T} \cup \{a\}) - f(\mathcal{T}) \leq f(\mathcal{S} \cup \{a\}) - f(\mathcal{S}) \quad \forall a \in \mathcal{V} \backslash \mathcal{T}$$
- Examples: rank of matrix, joint entropy
  - $f(\mathcal{S}) = \log \left| \mathbf{A} + \gamma \mathbf{B}_{\mathcal{S}} \mathbf{B}_{\mathcal{S}}^T \right|$ is submodular in $\mathcal{S}$ for $\mathbf{A}, \gamma > 0$.

Submodular functions - a primer

- Let $\mathcal{V}$ be the ground set of elements.
- Set function $f : \mathcal{V} \to \mathbb{R}_+$ is **submodular**, if for $\mathcal{S} \subseteq \mathcal{T} \subseteq \mathcal{V}$,
  - monotonicity

    $$f(\mathcal{S}) \leq f(\mathcal{T})$$

  - diminishing returns property

    $$f(\mathcal{T} \cup \{a\}) - f(\mathcal{T}) \leq f(\mathcal{S} \cup \{a\}) - f(\mathcal{S}) \quad \forall a \in \mathcal{V} \backslash \mathcal{T}$$

- Examples: rank of matrix, joint entropy
  - $f(\mathcal{S}) = \log \left| \mathbf{A} + \gamma \mathbf{B}_{\mathcal{S}} \mathbf{B}_{\mathcal{S}}^T \right|$ is submodular in $\mathcal{S}$ for $\mathbf{A}, \gamma > 0$.
- Maximizing a submodular function subject to cardinality constraints
  - **Greedy algorithm maximizes submodular f to within $\left(1 - \frac{1}{e}\right) f_{\max}$**

Submodular functions - a primer

- Let $\mathcal{V}$ be the ground set of elements.
- Set function $f : \mathcal{V} \to \mathbb{R}_+$ is **submodular**, if for $\mathcal{S} \subseteq \mathcal{T} \subseteq \mathcal{V}$,
  - monotonicity

    $$f(\mathcal{S}) \leq f(\mathcal{T})$$

  - diminishing returns property

    $$f(\mathcal{T} \cup \{a\}) - f(\mathcal{T}) \leq f(\mathcal{S} \cup \{a\}) - f(\mathcal{S}) \quad \forall a \in \mathcal{V} \backslash \mathcal{T}$$

- Examples: rank of matrix, joint entropy
  - $f(\mathcal{S}) = \log \left| \mathbf{A} + \gamma \mathbf{B}_{\mathcal{S}} \mathbf{B}_{\mathcal{S}}^T \right|$ is submodular in $\mathcal{S}$ for $\mathbf{A}, \gamma > 0$.
- Maximizing a submodular function subject to cardinality constraints
  - **Greedy algorithm maximizes submodular f to within $\left(1 - \frac{1}{e}\right) f_{\text{max}}$**
- Submodular f admits a tight modular upper bound: [Nemhauser, 78]

$$f(\mathcal{S}) \;\leq\; f(\mathcal{X}) - \sum_{j \in \mathcal{X} \backslash \mathcal{S}} \left(f(\mathcal{X}) - f(\mathcal{X} \backslash \{j\})\right) + \sum_{j \in \mathcal{S} \backslash \mathcal{X}} \left(f(j) - f(\phi)\right)$$

Rényi Divergence based Covariance Matching Pursuit (RD-CMP)

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

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

- Majorization-minimization procedure for support set recovery.

Rényi Divergence based Covariance Matching Pursuit (RD-CMP)

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

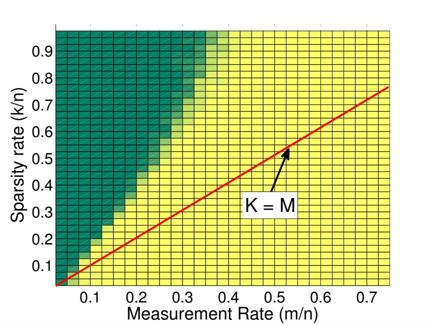
$$\hat{\mathcal{S}} = \underset{\mathcal{S}\subseteq[n]}{\text{argmin}} \ \underbrace{\log\left|(1-\alpha)\mathbf{R}_{\mathbf{Y}} + \alpha\left(\sigma^2\mathbf{I}_m + \boldsymbol{\gamma}\mathbf{A}_{\mathcal{S}}\mathbf{A}_{\mathcal{S}}^T\right)\right|}_{\text{submodular } f(\mathcal{S})} - \alpha \underbrace{\log\left|\sigma^2\mathbf{I}_m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

- Majorization-minimization procedure for support set recovery.
  - **Majorization step:** `[`k^{th} `iteration]`
    - Majorize objective by replacing 1^{st} log det term $f(\mathcal{S})$ with its modular upper bound $h^f_{\mathcal{S}_{t-1}}(\mathcal{S})$
  - **Minimization step:**
    - Minimize the majorized objective.

$$\mathcal{S}_{t+1} = \underset{\mathcal{S}\subseteq[n]}{\arg\min} \ \underbrace{h^f_{\mathcal{S}_t}(\mathcal{S}) - \alpha\log\left|\sigma^2\mathbf{I} + \boldsymbol{\gamma}\boldsymbol{\Phi}_{\mathcal{S}}\boldsymbol{\Phi}_{\mathcal{S}}^T\right|}_{\text{Supermodular func. minimized by greedy search}} .$$

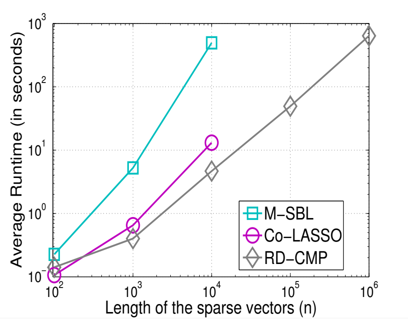
RD-CMP performance (1/2)

Support recovery phase transition
(SNR = 10 dB, n = 200, L = 200)



RD-CMP can recover k-sparse support from $m < k$ measurements per MMV!

Average runtime vs signal dimension
(SNR $= 10$ dB, $k = \lceil 50 \log_{10} n \rceil$,
$m = \lceil 0.75k \rceil$, $mL = \lceil 50k \log_{10} n \rceil$)

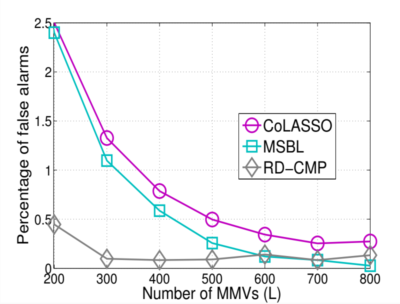


RD-CMP can solve a million variable problem in 10s of minutes.

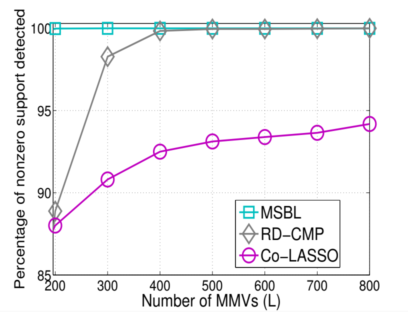
RD-CMP performance (2/2)

SNR = 10 dB, n = 500, k = 200, m=100

Support False Alarm Rate



Support Detection Rate

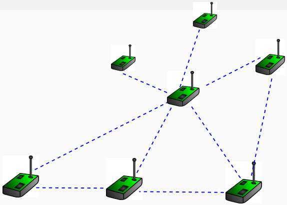


RD-CMP performs better than Co-LASSO
but slightly worse than MSBL

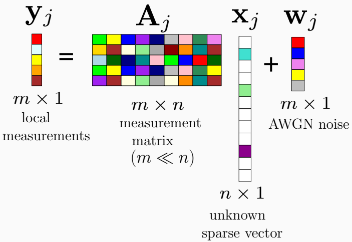
Distributed joint sparse signal recovery

Distributed Joint Sparse Signal Recovery

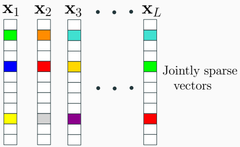
- Network of L sensor nodes
- Single hop communication between nodes



Measurement model at node-j



Network wide joint sparsity



$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L$ have the same support support

- **Goal:** Decentralized estimation of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L$.
- Exploit joint sparsity to reduce no. of local measurements.

Decentralized MSBL

- MSBL's EM updates:
  - **E-step:** Update the posterior $p(\mathbf{x}_j|\mathbf{y}_j;\gamma^k) \sim \mathcal{N}(\mu_j^{k+1}, \boldsymbol{\Sigma}_j^{k+1})$

$$\boldsymbol{\Sigma}_j^{k+1} = \left[(\boldsymbol{\Gamma}^k)^{-1} + \frac{\mathbf{A}_j^T\mathbf{A}_j}{\sigma_j^2}\right]^{-1}, \text{ and } \boldsymbol{\mu}_j^{k+1} = \sigma_j^{-2}\boldsymbol{\Sigma}_j^{k+1}\mathbf{A}_j^T\mathbf{y}_j$$

Decentralized MSBL

- MSBL's EM updates:
  - **E-step:** Update the posterior $p(\mathbf{x}_j|\mathbf{y}_j;\gamma^k) \sim \mathcal{N}(\mu_j^{k+1}, \mathbf{\Sigma}_j^{k+1})$

$$\mathbf{\Sigma}_j^{k+1} = \left[(\mathbf{\Gamma}^k)^{-1} + \frac{\mathbf{A}_j^T \mathbf{A}_j}{\sigma_j^2} \right]^{-1}, \text{ and } \boldsymbol{\mu}_j^{k+1} = \sigma_j^{-2} \mathbf{\Sigma}_j^{k+1} \mathbf{A}_j^T \mathbf{y}_j$$

locally computed at each node

Decentralized MSBL

- MSBL's EM updates:
  - **E-step:** Update the posterior $p(\mathbf{x}_j|\mathbf{y}_j;\gamma^k) \sim \mathcal{N}(\mu_j^{k+1}, \mathbf{\Sigma}_j^{k+1})$

$$\mathbf{\Sigma}_j^{k+1} = \left[(\mathbf{\Gamma}^k)^{-1} + \frac{\mathbf{A}_j^T\mathbf{A}_j}{\sigma_j^2}\right]^{-1}, \text{ and } \boldsymbol{\mu}_j^{k+1} = \sigma_j^{-2}\mathbf{\Sigma}_j^{k+1}\mathbf{A}_j^T\mathbf{y}_j$$

locally computed at each node

  - **M-step:** Maximize the tight lower bound on $\log p(\mathbf{Y};\gamma)$

$$\gamma^{k+1} = \underset{\gamma\in\mathbb{R}_+^n}{\text{argmax}}\ \mathbb{E}_{\mathbf{x}_j|\mathbf{y}_j,\gamma^k}\left[\log p(\mathbf{Y},\mathbf{X};\gamma)\right] = \frac{1}{L}\sum_{j=1}^{L}\underbrace{\left(\left(\boldsymbol{\mu}_j^k\right)^2 + \text{diag}\left(\mathbf{\Sigma}_j^{k+1}\right)\right)}_{\mathbf{a}_j^{k+1}}$$

Decentralized MSBL

- MSBL's EM updates:
  - **E-step:** Update the posterior $p(\mathbf{x}_j|\mathbf{y}_j;\gamma^k) \sim \mathcal{N}(\mu_j^{k+1}, \mathbf{\Sigma}_j^{k+1})$

$$\mathbf{\Sigma}_j^{k+1} = \left[(\mathbf{\Gamma}^k)^{-1} + \frac{\mathbf{A}_j^T\mathbf{A}_j}{\sigma_j^2}\right]^{-1}, \text{ and } \boldsymbol{\mu}_j^{k+1} = \sigma_j^{-2}\mathbf{\Sigma}_j^{k+1}\mathbf{A}_j^T\mathbf{y}_j$$

locally computed at each node

  - **M-step:** Maximize the tight lower bound on $\log p(\mathbf{Y};\gamma)$

$$\gamma^{k+1} = \underset{\gamma\in\mathbb{R}_+^n}{\text{argmax}} \; \mathbb{E}_{\mathbf{x}_j|\mathbf{y}_j,\gamma^k}\left[\log p(\mathbf{Y},\mathbf{X};\gamma)\right] = \frac{1}{L}\sum_{j=1}^{L}\underbrace{\left(\left(\boldsymbol{\mu}_j^k\right)^2 + \text{diag}\left(\mathbf{\Sigma}_j^{k+1}\right)\right)}_{\mathbf{a}_j^{k+1}}$$

  - **Decentralized M-step:** Each node maintains a local copy of γ.

$$\gamma^{k+1} = \underset{\gamma_1,\gamma_2,\ldots\gamma_L}{\text{argmin}} \sum_{j=1}^{L}\left|\left|\gamma_j - \mathbf{a}_j^{k+1}\right|\right|_2^2 \quad \text{subj. to } \gamma_1 = \gamma_2 = \ldots = \gamma_L$$

Decentralized MSBL

- MSBL's EM updates:
  - **E-step:** Update the posterior $p(\mathbf{x}_j|\mathbf{y}_j;\gamma^k) \sim \mathcal{N}(\mu_j^{k+1}, \mathbf{\Sigma}_j^{k+1})$

$$\mathbf{\Sigma}_j^{k+1} = \left[(\mathbf{\Gamma}^k)^{-1} + \frac{\mathbf{A}_j^T\mathbf{A}_j}{\sigma_j^2}\right]^{-1}, \text{ and } \boldsymbol{\mu}_j^{k+1} = \sigma_j^{-2}\mathbf{\Sigma}_j^{k+1}\mathbf{A}_j^T\mathbf{y}_j$$

locally computed at each node

  - **M-step:** Maximize the tight lower bound on $\log p(\mathbf{Y};\gamma)$

$$\gamma^{k+1} = \underset{\gamma\in\mathbb{R}_+^n}{\text{argmax}}\ \mathbb{E}_{\mathbf{x}_j|\mathbf{y}_j,\gamma^k}[\log p(\mathbf{Y},\mathbf{X};\gamma)] = \frac{1}{L}\sum_{j=1}^{L}\underbrace{\left(\left(\boldsymbol{\mu}_j^k\right)^2 + \text{diag}\left(\mathbf{\Sigma}_j^{k+1}\right)\right)}_{\mathbf{a}_j^{k+1}}$$

  - **Decentralized M-step:** Each node maintains a local copy of γ.

$$\gamma^{k+1} = \underset{\gamma_1,\gamma_2,\ldots\gamma_L}{\text{argmin}}\sum_{j=1}^{L}\left|\left|\gamma_j - \mathbf{a}_j^{k+1}\right|\right|_2^2 \quad \text{subj. to } \gamma_1 = \gamma_2 = \ldots = \gamma_L$$

Decentralized ADMM

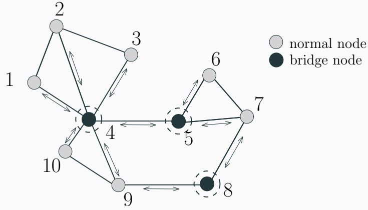
Local averaging

Consensus Based Distributed Sparse Bayesian Learning (CB-DSBL)

- Decentralized M-step (ADMM form)

$$\min_{\substack{\gamma_1,\gamma_2,\ldots,\gamma_L \\ \gamma_{b_1},\gamma_{b_2},\ldots,\gamma_{b_{|\mathcal{B}|}}}} \sum_{j=1}^{L} \left|\left|\gamma_j - \mathbf{a}_j^{k+1}\right|\right|^2$$

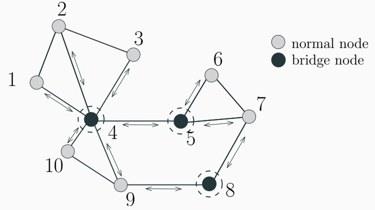
subj. to $\gamma_j = \gamma_b, \;\; j \in [L], b \in \mathcal{B}_j$



- Consensus enforced by using bridge variables.

Consensus Based Distributed Sparse Bayesian Learning (CB-DSBL)

- Decentralized M-step (ADMM form)



$$\min_{\substack{\gamma_1, \gamma_2, \ldots, \gamma_L \\ \gamma_{b_1}, \gamma_{b_2}, \ldots, \gamma_{b_{|\mathcal{B}|}}}} \sum_{j=1}^{L} \left|\left| \gamma_j - \mathbf{a}_j^{k+1} \right|\right|^2$$

subj. to $\gamma_j = \gamma_b, \;\; j \in [L], b \in \mathcal{B}_j$

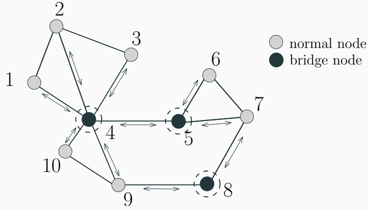
- Consensus enforced by using bridge variables.

Local variables:$\rightarrow$ γ_1 γ_2 γ_3 γ_4 γ_5 γ_6 γ_7 γ_8 γ_9 γ_{10}

Bridge variables:$\rightarrow$ γ_{b4} γ_{b5} γ_{b8}

Consensus Based Distributed Sparse Bayesian Learning (CB-DSBL)

- Decentralized M-step (ADMM form)



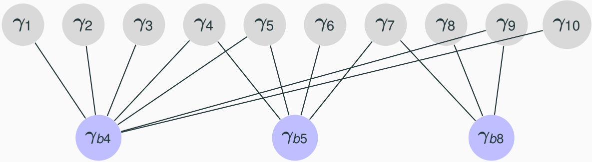
$$\min_{\substack{\gamma_1,\gamma_2,\ldots,\gamma_L \\ \gamma_{b_1},\gamma_{b_2},\ldots,\gamma_{b_{|\mathcal{B}|}}}} \sum_{j=1}^{L} \left|\left|\gamma_j - \mathbf{a}_j^{k+1}\right|\right|^2$$

subj. to $\gamma_j = \gamma_b, \;\; j \in [L], b \in \mathcal{B}_j$

- Consensus enforced by using bridge variables.

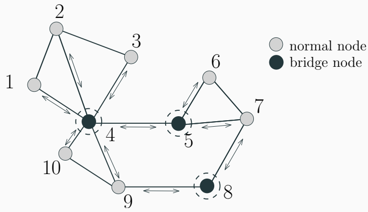
Local variables:→

Bridge variables:→



Consensus Based Distributed Sparse Bayesian Learning (CB-DSBL)

- Decentralized M-step (ADMM form)



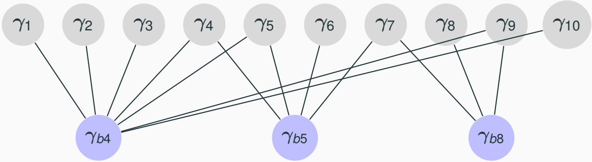
$$\min_{\substack{\gamma_1,\gamma_2,\ldots,\gamma_L\\ \gamma_{b_1},\gamma_{b_2},\ldots,\gamma_{b_{|\mathcal{B}|}}}} \sum_{j=1}^{L} \left|\left|\gamma_j - \mathbf{a}_j^{k+1}\right|\right|^2$$

$$\text{subj. to } \gamma_j = \gamma_b, \quad j \in [L], b \in \mathcal{B}_j$$

- Consensus enforced by using bridge variables.

Local variables:→

Bridge variables:→



- Augmented Lagrangian

extra quadratic penalty

$$L_\rho(\gamma_j, \gamma_b, \boldsymbol{\lambda}) \triangleq \sum_{j=1}^{L} \left|\left|\gamma_j - \mathbf{a}_j^{k+1}\right|\right|_2^2 + \sum_{j=1}^{L} \sum_{b \in \mathcal{B}_j} \lambda_{j,b}^T (\gamma_j - \gamma_b) + \frac{\rho}{2} \sum_{j=1}^{L} \sum_{b \in \mathcal{B}_j} ||\gamma_j - \gamma_b||_2^2$$

ADMM convergence - bridge node topology

- Decentralized ADMM iterations converge **R-linearly**
  - The primal optimality gap $\sum_{j=1}^{L} ||\gamma_j - \gamma_j^*||_2^2 \leq c_r$, where $c_r \to 0$ monotonically

ADMM convergence - bridge node topology

- Decentralized ADMM iterations converge **R-linearly**
  - The primal optimality gap $\sum_{j=1}^{L} \left|\left|\gamma_j - \gamma_j^*\right|\right|_2^2 \leq c_r$, where $c_r \to 0$ monotonically
- Optimal ADMM parameter ρ

$$\rho_{\text{opt}} = \frac{M_f}{\sigma_{\max}\sigma_{\min}} \left[\frac{\sqrt{(\kappa - 1)^2 + 4\kappa\kappa_f^2} + (\kappa - 1)}{\sqrt{(\kappa - 1)^2 + 4\kappa\kappa_f^2} - (\kappa - 1)} \right]^{\frac{1}{2}}$$

where

$$\kappa_f = \frac{M_f}{m_f} = \frac{\text{Lipschitz const. of } \nabla f}{\text{strong convexity const. of } f}$$

$$\kappa = \frac{\sigma_{\max}^2}{\sigma_{\min}^2} = \frac{\text{max \# bridge nodes per node}}{\text{min \# bridge nodes per node}}$$

ADMM convergence - bridge node topology

- Decentralized ADMM iterations converge **R-linearly**
  - The primal optimality gap $\sum_{j=1}^{L} ||\gamma_j - \gamma_j^*||_2^2 \leq c_r$, where $c_r \to 0$ monotonically

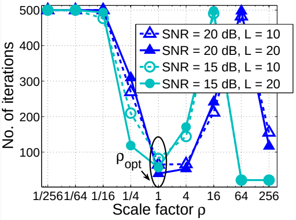
- Optimal ADMM parameter ρ

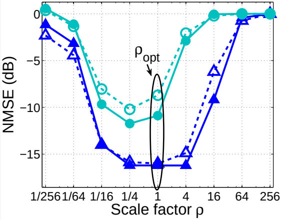
$$\rho_{\text{opt}} = \frac{M_f}{\sigma_{\max}\sigma_{\min}} \left[\frac{\sqrt{(\kappa-1)^2 + 4\kappa\kappa_f^2} + (\kappa-1)}{\sqrt{(\kappa-1)^2 + 4\kappa\kappa_f^2} - (\kappa-1)} \right]^{\frac{1}{2}}$$

where

$$\kappa_f = \frac{M_f}{m_f} = \frac{\text{Lipschitz const. of } \nabla f}{\text{strong convexity const. of } f}$$

$$\kappa = \frac{\sigma_{\max}^2}{\sigma_{\min}^2} = \frac{\text{max \# bridge nodes per node}}{\text{min \# bridge nodes per node}}$$



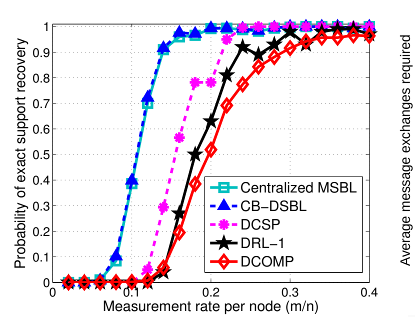


ADMM convergence is sensitive to ρ

CB-DSBL performance

Support recovery probability

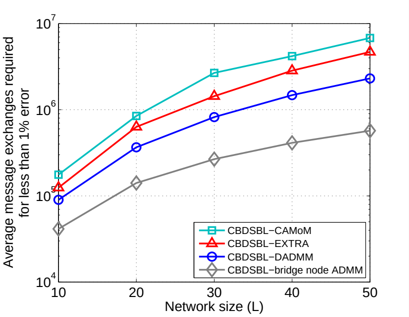
SNR = 15 dB, $n = 50$,
10% sparsity, network size = 10



Decentralized CB-DSBL matches the performance of MSBL

Communication complexity

SNR = 30 dB, $n = 50$, $m = 10$,
10% sparsity, no. of trials = 500



Bridge node ADMM has lower communication complexity than D-ADMM and EXTRA

Summarizing main contributions

- Derived new sufficient conditions for exact support recovery in Sparse Bayesian Learning.
- Proposed a new covariance matching framework for support recovery.
- Derived upper bounds for restricted isometry constants of generic Khatri-Rao product matrices.
- Proposed a novel Rényi divergence based support recovery algorithm suitable for big data applications.
- Proposed two new decentralized SBL extensions with focus on low communication complexity.

Current and future research

- Restricted eigenvalues characterization for self Khatri-Rao product $\mathbf{A} \odot \mathbf{A}$
  - For $m = O(\sqrt{k})$ regime.
- Recovery of joint sparse vectors with inter/intra vector correlations.
  - Sample complexity of robust recovery from underdetermined measurements.
- Local minima of likelihood functions...
  - Is there a phase transition phenomenon that explains the existence of local minima?
- Design of new cost functions for covariance matching.
  - Which attributes of the cost function dictates the support recovery performance?
- Sample complexity of RD-CMP algorithm.
  - Role of α-parameter in Rényi divergence.

Publications

- Journal articles
  - **S. Khanna** and C. R. Murthy, "*Decentralized Joint-Sparse Signal Recovery: A Sparse Bayesian Learning Approach*," in IEEE Trans. Signal and Info. Process. over Netw., vol. 3, no. 1, pp. 29-45, March 2017.
  - **S. Khanna** and C. R. Murthy, "*Communication Efficient Decentralized Sparse Bayesian Learning of Joint Sparse Signals*," in IEEE Trans. Signal and Info. Process. over Netw. , vol.PP, no.99, pp.1-14.
  - **S. Khanna** and C. R. Murthy, "*On the Restricted Isometry of Column- wise Khatri-Rao Product*", IEEE Trans. on Sig. Proc., vol. 66, no. 5, Mar. 2018
  - **S. Khanna** and C. R. Murthy, "*On the Support Recovery of Jointly Sparse Gaussian Sources using Sparse Bayesian Learning*," (arXiv:1703.04930).

- Conference proceedings
  - **S. Khanna** and C. R. Murthy, "*Decentralized Bayesian learning of jointly sparse signals*," 2014 IEEE GLOBECOM Conference, Austin, TX, 2014, pp. 3103-3108.
  - **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.

Take home insights...

MSBL objective... a Bregman matrix divergence.

Beyond ℓ_0-bound support recovery via covariance matching.

MSBL exactly recovers any $k < \text{spark}(\mathbf{A}) - 1$ sparse support even from a single noiseless MMV.

For subgaussian $\mathbf{A}$, MSBL perfectly recovers any k sparse support from $m = O(k \log n)$ noisy measurements per MMV, provided $L = O(k^2 \log n)$.

Cost function design is the key to faster inference!

Decentralized ADMM iterations converges R-linearly in a bridge node based network topology.

Fusion Based Decentralized Sparse Bayesian Learning (FB-DSBL)

Step 1: Local SBL iteration.

- Update local posterior $q(\mathbf{x}_j | \mathbf{y}_j; \gamma_j^k)$.
- $\gamma_j^{k+1} = \arg\max_{\gamma \succeq 0} \mathbb{E}_{\mathbf{x}_j \sim q_j} \log p(\mathbf{y}_j, \mathbf{x}_j; \gamma)$.

Step 2: Support estimation via indexwise log-likelihood ratio tests.

- $\mathcal{H}_0 : \gamma_j(i) = 0, \quad \mathcal{H}_1 : \gamma_j(i) > 0$.
- $\mathcal{H}_1$ if $\log \dfrac{p(\mathbf{y}_j; \mathcal{H}_1)}{p(\mathbf{y}_j; \mathcal{H}_0)} \geq \theta$.

Fusion Based Decentralized Sparse Bayesian Learning (FB-DSBL)

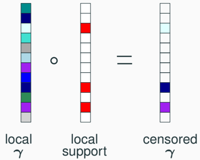
Step 1**:** Local SBL iteration.

- Update local posterior $q(\mathbf{x}_j|\mathbf{y}_j;\gamma_j^k)$.
- $\gamma_j^{k+1} = \arg\max_{\gamma \succeq 0} \mathbb{E}_{\mathbf{x}_j \sim q_j} \log p(\mathbf{y}_j, \mathbf{x}_j; \gamma)$.

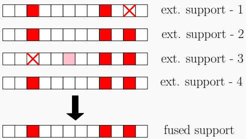
Step 2**:** Support estimation via indexwise log-likelihood ratio tests.

- $\mathcal{H}_0 : \gamma_j(i) = 0, \quad \mathcal{H}_1 : \gamma_j(i) > 0$.
- $\mathcal{H}_1$ if $\log \dfrac{p(\mathbf{y}_j;\mathcal{H}_1)}{p(\mathbf{y}_j;\mathcal{H}_0)} \geq \theta$.

Step-3: Broadcast censored copy of γ.



Step-4: Fuse support estimates from other nodes using **majority rule**.



Fusion Based Decentralized Sparse Bayesian Learning (FB-DSBL)

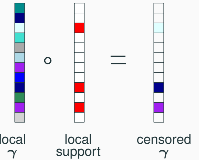
Step 1: Local SBL iteration.

- Update local posterior $q(\mathbf{x}_j|\mathbf{y}_j;\gamma_j^k)$.
- $\gamma_j^{k+1} = \arg\max_{\gamma \succeq 0} \mathbb{E}_{\mathbf{x}_j \sim q_j} \log p(\mathbf{y}_j, \mathbf{x}_j; \gamma)$.

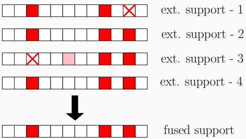
Step 2: Support estimation via indexwise log-likelihood ratio tests.

- $\mathcal{H}_0 : \gamma_j(i) = 0, \quad \mathcal{H}_1 : \gamma_j(i) > 0$.
- $\mathcal{H}_1$ if $\log \dfrac{p(\mathbf{y}_j; \mathcal{H}_1)}{p(\mathbf{y}_j; \mathcal{H}_0)} \geq \theta$.

Step-3: Broadcast censored copy of γ.



Step-4: Fuse support estimates from other nodes using **majority rule**.



Step 5: Assimilate shared information from neighboring nodes to refine local γ.

If the majority says i^{th} index is zero:

- $\gamma_j^{k+1}(i)$ = avg. of own estimate and received estimates (censored values replaced by zero)

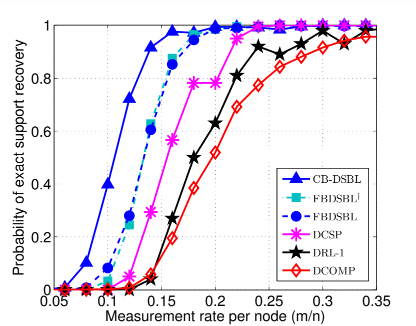
If the majority says i^{th} index is non-zero:

- $\gamma_j^{k+1}(i)$ = avg. of own estimate and received non-censored estimates

FB-DSBL performance

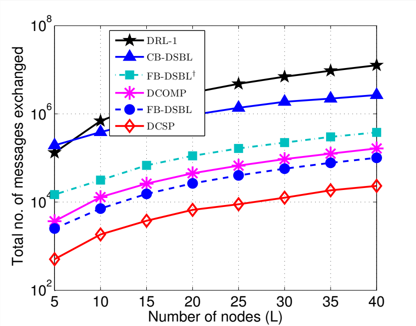
Support recovery probability

SNR = 15 dB, $n = 50$, 10% sparsity, no. of nodes (L) = 10.



Communication complexity

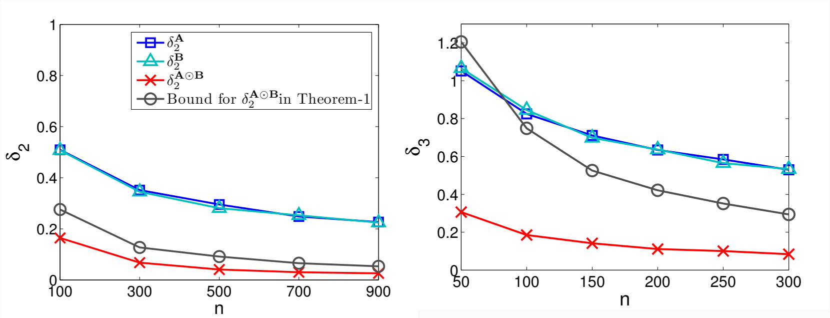
SNR = 30 dB, $n = 50$, $m = 10$, 10% sparsity, no. of trials = 500.



FB-DSBL has "Bayesian" like performance and "Greedy" like communication complexity

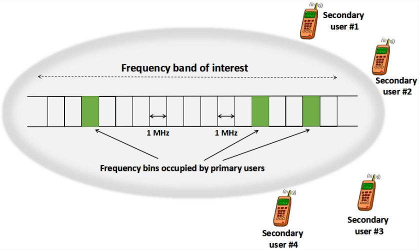
RIP of Khatri-Rao product - an empirical study (plots with bounds)

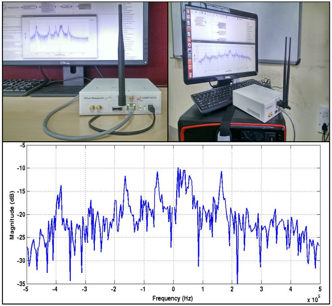
$$\mathbf{A}_{i,j}, \mathbf{B}_{i,j} \sim \mathcal{N}\left(0, \frac{1}{m}\right) \text{ and } m = 0.5n$$



RIP improved by taking Khatri-Rao product!

Wideband Spectrum Sensing





- Experimental setup
  - No. primary users = 5
  - No. secondary users = 10
  - 11 of total 128 frequency subbands are in use
  - SNR range: -2.4 to 7.8