Sufficient conditions for sparse support recovery in M-SBL

Saurabh Khanna,

Signal Processing for Communication, ECE, IISc

Joint sparse signal recovery

<u>MMV model:</u> $\mathbf{y}_j = \mathbf{\Phi}\mathbf{x}_j + \mathbf{w}_j \qquad j = 1$ to L

$\mathbf{x}_j \in \mathbb{R}^n$ are unknown $k-$sparse vectors following JSM-2 signal model

- $\mathbf{x}_j$ have a common nonzero support
- Nonzero entries are uncorrelated

$\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ is the measurement matrix with $m \ll n$

$\mathbf{y}_j \in \mathbb{R}^m$ are the noisy linear measurements

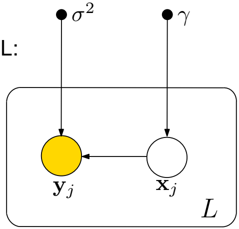
$\mathbf{w}_j \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m)$

Want to recover joint sparse vectors $\mathbf{x}_1, \mathbf{x}_2 \ldots \mathbf{x}_L$ from noisy linear measurements $\mathbf{y}_1, \mathbf{y}_2 \ldots \mathbf{y}_L$.

Examples of Joint sparse signal recovery algorithms: M-OMP, M-FOCUSS, Row-LASSO and M-SBL

M-SBL algorithm

Graphical model for MSBL:



$$\mathbf{x}_j \sim \mathcal{N}(0, \Gamma)$$
$$\Gamma = \mathrm{diag}(\gamma)$$
$$\mathbf{y}_j \sim \mathcal{N}(0, \sigma^2 \mathbf{I} + \Phi\Gamma\Phi^T)$$

Use of common γ induces joint sparsity in $\mathbf{x}_1, \mathbf{x}_2 \ldots \mathbf{x}_L$

For fixed $\boldsymbol{\gamma}$, we have $p(\hat{\mathbf{x}}_j | \mathbf{y}_j; \boldsymbol{\gamma}) \sim \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, where

$$\boldsymbol{\Sigma} = \Gamma^{-1} + \frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{\sigma^2}, \qquad \boldsymbol{\mu}_j = \frac{1}{\sigma^2} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Phi}^T \mathbf{y}_j$$

γ is found by solving Type-II ML estimation problem: $\hat{\gamma} = \underset{\boldsymbol{\gamma} \in \mathbb{R}_+^n}{\text{arg max}} \log p(\mathbf{Y}; \boldsymbol{\gamma})$

M-SBL algorithm

MSBL cost function:

$$
\begin{aligned}
\log p(\mathbf{Y};\boldsymbol{\gamma}) & \\
&= \log(\mathbf{y}_1, \mathbf{y}_2 \ldots \mathbf{y}_L; \boldsymbol{\gamma}) \\
&= \sum_{j=1}^{L} \log p(\mathbf{y}_j; \boldsymbol{\gamma}) \\
&= \sum_{j=1}^{L} \log \mathcal{N}(\mathbf{y}_j; 0, \sigma^2 \mathbf{I} + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^T) \\
&= -\frac{L}{2} \log 2\pi - \frac{L}{2} \log |\sigma^2 \mathbf{I} + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^T| - \frac{1}{2} \sum_{j=1}^{L} Tr\left((\sigma^2 \mathbf{I} + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^T)^{-1} \mathbf{y}_j \mathbf{y}_j^T \right) \\
&= -\frac{L}{2} \log 2\pi - \frac{L}{2} \log |\sigma^2 \mathbf{I} + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^T| - \frac{1}{2} Tr\left((\sigma^2 \mathbf{I} + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^T)^{-1} \mathbf{Y}\mathbf{Y}^T \right)
\end{aligned}
$$

M-SBL cost - interesting interpretation

MSBL cost function:

$$\log p(\mathbf{Y};\boldsymbol{\gamma}) = -\frac{L}{2}\log 2\pi - \frac{L}{2}\log|\sigma^2\mathbf{I} + \mathbf{\Phi\Gamma\Phi}^T| - \frac{1}{2}Tr\left((\sigma^2\mathbf{I} + \mathbf{\Phi\Gamma\Phi}^T)^{-1}\mathbf{YY}^T\right)$$

M-SBL cost - interesting interpretation

MSBL cost function:

$$\log p(\mathbf{Y};\boldsymbol{\gamma}) = -\frac{L}{2}\log 2\pi - \frac{L}{2}\log|\sigma^2\mathbf{I} + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^T| - \frac{1}{2}Tr\left((\sigma^2\mathbf{I} + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^T)^{-1}\mathbf{Y}\mathbf{Y}^T\right)$$

Bregman matrix divergence with respect to $\phi(.) = -\log|.|$, is defined as:

$$\mathcal{D}(\mathbf{X},\mathbf{Y}) = \text{trace}(\mathbf{X}\mathbf{Y}^{-1}) - \log|\mathbf{X}\mathbf{Y}^{-1}| - N$$

M-SBL cost - interesting interpretation

MSBL cost function:

$$\log p(\mathbf{Y};\boldsymbol{\gamma}) = -\frac{L}{2}\log 2\pi - \frac{L}{2}\log|\sigma^2\mathbf{I}+\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T| - \frac{1}{2}Tr\left((\sigma^2\mathbf{I}+\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T)^{-1}\mathbf{Y}\mathbf{Y}^T\right)$$

Bregman matrix divergence with respect to $\phi(.) = -\log|.|$, is defined as:

$$\mathcal{D}(\mathbf{X},\mathbf{Y}) = \text{trace}(\mathbf{X}\mathbf{Y}^{-1}) - \log|\mathbf{X}\mathbf{Y}^{-1}| - N$$

MSBL cost can be interpreted as a matrix divergence:

$$\log p(\mathbf{Y};\boldsymbol{\gamma}) = -\frac{L}{2}\mathcal{D}_\phi\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T\ ,\ \sigma^2\mathbf{I}+\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T\right)\underbrace{-m - \frac{L}{2}\log 2\pi + \frac{L}{2}\log|\frac{1}{L}\mathbf{Y}\mathbf{Y}^T|}_{\text{constant}}$$

Thus, MSBL is trying to minimize the matrix divergence between

(i) $\frac{1}{L}\mathbf{Y}\mathbf{Y}^T$ (empirical covariance matrix)

(ii) $\sigma^2\mathbf{I}+\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T$ (parameterized covariance matrix)

M-SBL cost - interesting interpretation

MSBL is trying to do “energy matching” or “second moment matching”

$$\log p(\mathbf{Y}|\boldsymbol{\gamma}) \quad \propto \quad \text{const} \; + \; \mathcal{D}_{-\log|.|}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T \, , \; \sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T\right)$$

M-SBL cost - interesting interpretation

MSBL is trying to do “energy matching” or “second moment matching”

$$\log p(\mathbf{Y}|\boldsymbol{\gamma}) \quad \propto \quad \text{const} \; + \; \mathcal{D}_{-\log|.|}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T \,,\; \sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T\right)$$

[P. Pal and P.P. Vaidyanathan, ’15] proposed the Cov-MMV problem:

$$\min_{\boldsymbol{\gamma}\in\mathbb{R}^n_+} \quad ||\boldsymbol{\gamma}||_1 \qquad \text{subject to} \quad \frac{1}{L}\mathbf{Y}\mathbf{Y}^T = \sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T$$

M-SBL cost - interesting interpretation

MSBL is trying to do “energy matching” or “second moment matching”

$$\log p(\mathbf{Y}|\boldsymbol{\gamma}) \quad \propto \quad \text{const} \; + \; \mathcal{D}_{-\log|.|}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T \;, \; \sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T\right)$$

[P. Pal and P.P. Vaidyanathan, ’15] proposed the Cov-MMV problem:

$$\min_{\boldsymbol{\gamma}\in\mathbb{R}_+^n} \quad ||\boldsymbol{\gamma}||_1 \qquad \text{subject to} \quad \frac{1}{L}\mathbf{Y}\mathbf{Y}^T = \sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T$$

Cov-MMV can rewritten in LASSO form as:

$$\min_{\boldsymbol{\gamma}\in\mathbb{R}_+^n} \quad ||\boldsymbol{\gamma}||_1 \qquad \text{subject to} \quad (\boldsymbol{\Phi}\odot\boldsymbol{\Phi})\boldsymbol{\gamma} = \text{vec}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T - \sigma^2\mathbf{I}\right)$$

MSBL vs. Cov-MMV

Concave $\log|\ .\ |$ penalty is a better regularizer compared to convex $||\ .\ ||_1$ penalty

A better regularization of the feasible set translates to fewer measurement for guaranteeing a unique sparse solution.

We seek necessary and sufficient conditions for sparse support recovery for the M-SBL algorithm.

Support recovery in M-SBL

Constrained MSBL problem:

$$\hat{\gamma} = \underset{\boldsymbol{\gamma} \in \Theta_K}{\text{arg max}} \, \mathcal{L}(\mathbf{Y}; \boldsymbol{\gamma})$$

$$\mathcal{L}(\mathbf{Y}; \boldsymbol{\gamma}) = -\frac{L}{2} \log |\sigma^2 \mathbf{I} + \boldsymbol{\Phi} \boldsymbol{\Gamma} \boldsymbol{\Phi}^T| - \frac{1}{2} Tr\left((\sigma^2 \mathbf{I} + \boldsymbol{\Phi} \boldsymbol{\Gamma} \boldsymbol{\Phi}^T)^{-1} \mathbf{Y} \mathbf{Y}^T \right)$$

$$\Theta_K = \left\{ \boldsymbol{\gamma} \in \mathbb{R}_+^n : ||\boldsymbol{\gamma}||_0 \leq K, ||\boldsymbol{\gamma}||_\infty \leq \gamma_{\text{max}} \right\}$$

Support recovery in M-SBL

Constrained MSBL problem:

$$\hat{\gamma} = \underset{\gamma \in \Theta_K}{\text{arg max}}\, \mathcal{L}(\mathbf{Y}; \boldsymbol{\gamma})$$

$$\mathcal{L}(\mathbf{Y}; \boldsymbol{\gamma}) = -\frac{L}{2} \log |\sigma^2 \mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T| - \frac{1}{2} Tr\left((\sigma^2 \mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T)^{-1} \mathbf{Y}\mathbf{Y}^T\right)$$

$$\Theta_K = \left\{\boldsymbol{\gamma} \in \mathbb{R}_+^n : ||\boldsymbol{\gamma}||_0 \leq K, ||\boldsymbol{\gamma}||_\infty \leq \gamma_{\text{max}}\right\}$$

Problem setup:

Let $\mathbf{x}_1, \mathbf{x}_2 \ldots \mathbf{x}_L$ be generated according to the true distribution $\mathcal{N}(0, \Gamma_0)$. Let $\mathcal{S}_0 = supp(\boldsymbol{\gamma}_0)$ and, $|\mathcal{S}_0| \leq K$.

Question:

What is the (m, L) measurement rate region which guarantees that the solution of constrained MSBL problem, denoted by $\hat{\gamma}$ satisfies $supp(\hat{\gamma}) = supp(\boldsymbol{\gamma}_0)$ with arbitrarity high probability.

Support recovery in M-SBL

Constrained MSBL:

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma} \in \Theta_K}{\text{arg max}}\, \mathcal{L}(\mathbf{Y}; \boldsymbol{\gamma})$$

$$\mathcal{L}(\mathbf{Y}; \boldsymbol{\gamma}) = \log p(\mathbf{Y}; \boldsymbol{\gamma})$$

$$\Theta_K = \{\boldsymbol{\gamma} \in \mathbb{R}_+^n : ||\boldsymbol{\gamma}||_0 \leq K, , ||\boldsymbol{\gamma}||_\infty \leq \gamma_{\text{max}}\}$$

Sucess event $\mathcal{A}$:

$$\begin{aligned} \mathcal{A} &\triangleq \{supp(\hat{\boldsymbol{\gamma}}) = supp(\boldsymbol{\gamma}_0) = \mathcal{S}_0\} \\ &= \left\{ \sup_{\boldsymbol{\gamma} \in \Theta_{\mathcal{S}_0}} \mathcal{L}(\mathbf{Y}, \boldsymbol{\gamma}) \geq \sup_{\boldsymbol{\gamma} \in \Theta_{\mathcal{S}}} \mathcal{L}(\mathbf{Y}, \boldsymbol{\gamma}), \quad \forall \mathcal{S} \neq \mathcal{S}_0 \right\} \end{aligned}$$

Support recovery in M-SBL

Constrained MSBL:

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma} \in \Theta_K}{\text{arg max}} \, \mathcal{L}(\mathbf{Y}; \boldsymbol{\gamma})$$

$$\mathcal{L}(\mathbf{Y}; \boldsymbol{\gamma}) = \log p(\mathbf{Y}; \boldsymbol{\gamma})$$

$$\Theta_K = \left\{ \boldsymbol{\gamma} \in \mathbb{R}_+^n : ||\boldsymbol{\gamma}||_0 \leq K, , ||\boldsymbol{\gamma}||_\infty \leq \gamma_{\text{max}} \right\}$$

Sucess event $\mathcal{A}$:

$$\begin{aligned} \mathcal{A} &\triangleq \{supp(\hat{\boldsymbol{\gamma}}) = supp(\boldsymbol{\gamma}_0) = \mathcal{S}_0\} \\ &= \left\{ \sup_{\boldsymbol{\gamma} \in \Theta_{\mathcal{S}_0}} \mathcal{L}(\mathbf{Y}, \boldsymbol{\gamma}) \geq \sup_{\boldsymbol{\gamma} \in \Theta_{\mathcal{S}}} \mathcal{L}(\mathbf{Y}, \boldsymbol{\gamma}), \;\; \forall \mathcal{S} \neq \mathcal{S}_0 \right\} \end{aligned}$$

Question:
Identify the measurement rate region (m, L) such that $\mathbb{P}(\mathcal{A}) \geq 1 - \delta$.

Support recovery in M-SBL

We define expected log likelihood, $\bar{\mathcal{L}}(\mathbf{Y}, \boldsymbol{\gamma}) = \mathbb{E}\mathcal{L}(\mathbf{Y}, \boldsymbol{\gamma})$

Expectation is w.r.t $\mathbf{Y} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 \mathbf{I} + \boldsymbol{\Phi}\Gamma_0\boldsymbol{\Phi}^T)$

Then, we have

$$
\begin{aligned}
\bar{\mathcal{L}}(\mathbf{Y}, \boldsymbol{\gamma}_0) &= \mathbb{E} \log p(\mathbf{Y}; \boldsymbol{\gamma}_0) \\
\bar{\mathcal{L}}(\mathbf{Y}, \boldsymbol{\gamma}) &= \mathbb{E} \log p(\mathbf{Y}; \boldsymbol{\gamma}) \\
\bar{\mathcal{L}}(\mathbf{Y}, \boldsymbol{\gamma}_0) - \bar{\mathcal{L}}(\mathbf{Y}, \boldsymbol{\gamma}) &= \mathbb{E}_{\mathbf{Y} \sim p(\mathbf{Y}; \boldsymbol{\gamma}_0)} \log \frac{p(\mathbf{Y}; \boldsymbol{\gamma}_0)}{p(\mathbf{Y}; \boldsymbol{\gamma})} \\
&= L \cdot \underbrace{\mathcal{D}_{\text{KL}} \left(p(\mathbf{y}; \boldsymbol{\gamma}_0), p(\mathbf{y}; \boldsymbol{\gamma})\right)}_{\text{non-negative !}}
\end{aligned}
$$

Tradeoff between m and L

We want... $\mathbb{P}\left\{ \sup_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}_0}} \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}) \geq \sup_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}}} \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}), \quad \forall \mathcal{S} \neq \mathcal{S}_0 \right\} \geq 1-\delta.$

Tradeoff between m and L

We want... $$\mathbb{P}\left\{\sup_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}_0}} \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}) \geq \sup_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}}} \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}), \quad \forall \mathcal{S} \neq \mathcal{S}_0\right\} \geq 1-\delta.$$

$$L \cdot \mathcal{D}_{KL}(p(\mathbf{y};\gamma_0), p(\mathbf{y};\gamma))$$

Increases linearly with L

Increases with m

OR

$$\mathbb{E}\mathcal{L}(\mathbf{Y};\gamma_0) - \mathbb{E}\mathcal{L}(\mathbf{Y};\gamma)$$

$L \to \infty$ $L \to \infty$

$\mathcal{L}(\mathbf{Y};\gamma_0)$ $\mathcal{L}(\mathbf{Y};\gamma)$

Tradeoff between m and L

We want... $\mathbb{P}\left\{ \sup_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}_0}} \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}) \geq \sup_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}}} \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}), \quad \forall \mathcal{S} \neq \mathcal{S}_0 \right\} \geq 1-\delta.$

$$L \cdot \mathcal{D}_{KL}(p(\mathbf{y};\gamma_0), p(\mathbf{y};\gamma))$$

Increases linearly with L

Increases with m

OR

$$\mathbb{E}\mathcal{L}(\mathbf{Y};\gamma_0) - \mathbb{E}\mathcal{L}(\mathbf{Y};\gamma)$$

$L \to \infty$ $L \to \infty$

$\mathcal{L}(\mathbf{Y};\gamma_0)$ $\mathcal{L}(\mathbf{Y};\gamma)$

$$\left[\sup_{\gamma\in\Theta_{\mathcal{S}_0}} \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}) - \sup_{\gamma\in\Theta_{\mathcal{S}}} \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}) \right] = L \cdot \mathcal{D}_{\mathrm{KL}}\left(p(\mathbf{y};\boldsymbol{\gamma}_0), p(\mathbf{y};\boldsymbol{\gamma})\right) - \text{(conc. terms)}$$

$$
\begin{aligned}
\sup_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}_0}} \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}) &\geq \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}_0) = \bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}_0) - \left(\bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}_0) - \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}_0)\right) \\
&\geq \bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}_0) - |\bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}_0) - \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}_0)| \\
&\geq \bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}_0) - \sup_{\boldsymbol{\gamma}\in\Theta_K} |\bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}) - \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma})| \qquad (1)
\end{aligned}
$$

$$\begin{aligned}
\sup_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}_0}} \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}) \quad &\geq \quad \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}_0) \quad = \quad \bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}_0) - \big(\bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}_0) - \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}_0)\big) \\
&\geq \quad \bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}_0) - |\bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}_0) - \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}_0)| \\
&\geq \quad \bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}_0) - \sup_{\boldsymbol{\gamma}\in\Theta_K} |\bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}) - \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma})| \qquad (1)
\end{aligned}$$

$$\begin{aligned}
\sup_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}}} \mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}) \quad &= \quad \sup_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}}} \big[\bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}) + \big(\mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}) - \bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma})\big)\big] \\
&\leq \quad \sup_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}}} \bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}) + \sup_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}}} |\mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}) - \bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma})| \\
&\leq \quad \sup_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}}} \bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma}) + \sup_{\boldsymbol{\gamma}\in\Theta_K} |\mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}) - \bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma})| \qquad (2)
\end{aligned}$$

$$\begin{aligned}\sup_{\gamma\in\Theta_{\mathcal{S}_0}} \mathcal{L}(\mathbf{Y},\gamma) &\geq \mathcal{L}(\mathbf{Y},\gamma_0) = \bar{\mathcal{L}}(\mathbf{Y},\gamma_0) - \left(\bar{\mathcal{L}}(\mathbf{Y},\gamma_0) - \mathcal{L}(\mathbf{Y},\gamma_0)\right) \\ &\geq \bar{\mathcal{L}}(\mathbf{Y},\gamma_0) - |\bar{\mathcal{L}}(\mathbf{Y},\gamma_0) - \mathcal{L}(\mathbf{Y},\gamma_0)| \\ &\geq \bar{\mathcal{L}}(\mathbf{Y},\gamma_0) - \sup_{\gamma\in\Theta_K} |\bar{\mathcal{L}}(\mathbf{Y},\gamma) - \mathcal{L}(\mathbf{Y},\gamma)| \qquad (1)\end{aligned}$$

$$\begin{aligned}\sup_{\gamma\in\Theta_{\mathcal{S}}} \mathcal{L}(\mathbf{Y},\gamma) &= \sup_{\gamma\in\Theta_{\mathcal{S}}} \left[\bar{\mathcal{L}}(\mathbf{Y},\gamma) + \left(\mathcal{L}(\mathbf{Y},\gamma) - \bar{\mathcal{L}}(\mathbf{Y},\gamma)\right)\right] \\ &\leq \sup_{\gamma\in\Theta_{\mathcal{S}}} \bar{\mathcal{L}}(\mathbf{Y},\gamma) + \sup_{\gamma\in\Theta_{\mathcal{S}}} |\mathcal{L}(\mathbf{Y},\gamma) - \bar{\mathcal{L}}(\mathbf{Y},\gamma)| \\ &\leq \sup_{\gamma\in\Theta_{\mathcal{S}}} \bar{\mathcal{L}}(\mathbf{Y},\gamma) + \sup_{\gamma\in\Theta_K} |\mathcal{L}(\mathbf{Y},\gamma) - \bar{\mathcal{L}}(\mathbf{Y},\gamma)| \qquad (2)\end{aligned}$$

Subtracting (2) from (1), we get:

$$\begin{aligned}&\left[\sup_{\gamma\in\Theta_{\mathcal{S}_0}} \mathcal{L}(\mathbf{Y},\gamma) - \sup_{\gamma\in\Theta_{\mathcal{S}}} \mathcal{L}(\mathbf{Y},\gamma)\right] \\ &\qquad \geq \inf_{\gamma\in\Theta_{\mathcal{S}}} \left[L \cdot \mathcal{D}_{\text{KL}}\left(p(\mathbf{y};\gamma_0), p(\mathbf{y};\gamma)\right)\right] - 2 \sup_{\gamma\in\Theta_K} |\mathcal{L}(\mathbf{Y},\gamma) - \bar{\mathcal{L}}(\mathbf{Y},\gamma)|\end{aligned}$$

Lower bounding the KL divergence

KL divergence between multivariate Gaussians:

$$\mathcal{D}_{\mathsf{KL}}\left(\mathcal{N}(0,\mathbf{\Sigma}_1),\mathcal{N}(0,\mathbf{\Sigma}_1)\right) \triangleq \frac{1}{2}\log\frac{|\mathbf{\Sigma}_2|}{|\mathbf{\Sigma}_1|} + \frac{1}{2}\text{trace}\left(\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1\right) - \frac{m}{2}.$$

- The ratio terms $\frac{|\mathbf{\Sigma}_2|}{|\mathbf{\Sigma}_1|}$ and trace $\left(\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1\right)$ can be difficult to handle.
- Norm based terms like $||\mathbf{\Sigma}_2 - \mathbf{\Sigma}_1||$ are preferred.

$\alpha-$Renyi divergence lower bounds KL divergence for $\alpha \in (0,1)$.

$$\mathcal{D}_{\alpha}\left(\mathcal{N}(0,\mathbf{\Sigma}_1),\mathcal{N}(0,\mathbf{\Sigma}_1)\right) = \frac{1}{2(1-\alpha)}\log\frac{|\mathbf{\Sigma}_1|^{1-\alpha}|\mathbf{\Sigma}_2|^{\alpha}}{|(1-\alpha)\mathbf{\Sigma}_1+\alpha\mathbf{\Sigma}_2|}.$$

- $\mathcal{D}_{1/2}$ can be lower bounded in terms of $||\mathbf{\Sigma}_2 - \mathbf{\Sigma}_1||$ like difference terms.

Lower bound for $\alpha-$Renyi divergence

Theorem

For $p_1 \sim \mathcal{N}(0, \mathbf{\Sigma}_1)$ *and* $p_2 \sim \mathcal{N}(0, \mathbf{\Sigma}_2)$, *the* α*-Renyi divergence can be lower bounded as*

$$D_{1/2}\left(p_1, p_2\right) \geq \frac{1}{2}\alpha(1-\alpha)m^*||\mathbf{\Sigma}_2 - \mathbf{\Sigma}_1||_F^2$$

where m^* *is the strong convexity constant of* $-\log|\ .\ |$ *function and, is evaluated as:*

$$m^* = \sup_{\mathbf{\Sigma}\in\Theta} \frac{1}{\left(\lambda_{max}(\mathbf{\Sigma})\right)^2}$$

where $\Theta = \{t\mathbf{\Sigma}_1 + (1-t)\mathbf{\Sigma}_2, t \in [0,1]\}$.

Upper bound for $|\mathcal{L}(\mathbf{Y};\boldsymbol{\gamma}) - \bar{\mathcal{L}}(\mathbf{Y};\boldsymbol{\gamma})|$

$$\begin{aligned}
|\mathcal{L}(\mathbf{Y};\boldsymbol{\gamma}) - \bar{\mathcal{L}}(\mathbf{Y};\boldsymbol{\gamma})| &= |\log p(\mathbf{Y};\boldsymbol{\gamma}) - \mathbb{E}\log p(\mathbf{Y};\boldsymbol{\gamma})| \\
&= \frac{1}{2}\left|\text{trace}\left((\sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T)^{-1}\left(\mathbf{Y}^T - \mathbb{E}\mathbf{Y}\mathbf{Y}^T\right)\right)\right| \\
&= \frac{L}{2}\left|\text{trace}\left((\sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T)^{-1}\left(\hat{\mathbf{R}}_{\mathbf{y}} - \mathbf{R}_{\mathbf{y}}\right)\right)\right| \\
&\leq \frac{L}{2}||(\sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T)^{-1}||_{\text{Tr}} \cdot |||\hat{\mathbf{R}}_{\mathbf{y}} - \mathbf{R}_{\mathbf{y}}|||_2 \\
&\leq \frac{Lm}{2\sigma^2}|||\hat{\mathbf{R}}_{\mathbf{y}} - \mathbf{R}_{\mathbf{y}}|||_2 \\
&\leq \frac{Lm\epsilon}{2\sigma^2}|||\mathbf{R}_{\mathbf{y}}|||_2 \qquad \text{[Vershyn, non-asymptotic RMT, '11]}
\end{aligned}$$

Last inequality holds with probability at least $1-\delta$, provided $L \geq \frac{C}{\epsilon^2}\log\frac{2}{\delta}$.

By eliminating ϵ, we get the following probabilistic bound:

$$\mathbb{P}\left(|\mathcal{L}(\mathbf{Y};\boldsymbol{\gamma}) - \bar{\mathcal{L}}(\mathbf{Y};\boldsymbol{\gamma})| \leq \frac{1}{2\sigma^2}\sqrt{Lm^2C\log\left(\frac{2}{\delta}\right)}|||\sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}_{\mathbf{0}}\boldsymbol{\Phi}^T|||_2\right) \geq 1-\delta.$$

Sufficient conditions

Recall, that for $\mathcal{A}$ to hold, we need (m, L) such that

$$\inf_{\boldsymbol{\gamma} \in \Theta_{\mathcal{S}}} \left[L \cdot \mathcal{D}_{\mathsf{KL}} \left(p(\mathbf{y}; \boldsymbol{\gamma}_0), p(\mathbf{y}; \boldsymbol{\gamma}) \right) \right] \geq 2 \sup_{\boldsymbol{\gamma} \in \Theta_K} |\mathcal{L}(\mathbf{Y}, \boldsymbol{\gamma}) - \bar{\mathcal{L}}(\mathbf{Y}, \boldsymbol{\gamma})|$$

Sufficient conditions

Recall, that for $\mathcal{A}$ to hold, we need (m, L) such that

$$\inf_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}}} \left[L \cdot \mathcal{D}_{\mathsf{KL}}\left(p(\mathbf{y};\boldsymbol{\gamma}_0), p(\mathbf{y};\boldsymbol{\gamma})\right)\right] \geq 2 \sup_{\boldsymbol{\gamma}\in\Theta_K} |\mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}) - \bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma})|$$

Or equivalently,

$$\frac{L}{8}\left(\frac{||\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0||_F}{|||\boldsymbol{\Sigma}(t)|||_2}\right)^2 \geq \frac{1}{\sigma^2}\sqrt{Lm^2C\log\frac{2}{\delta}} \cdot |||\sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}_{\mathbf{0}}\boldsymbol{\Phi}^T|||_2$$

Sufficient conditions

Recall, that for $\mathcal{A}$ to hold, we need (m, L) such that

$$\inf_{\boldsymbol{\gamma}\in\Theta_{\mathcal{S}}} \left[L \cdot \mathcal{D}_{\text{KL}}\left(p(\mathbf{y};\boldsymbol{\gamma}_0), p(\mathbf{y};\boldsymbol{\gamma})\right)\right] \geq 2 \sup_{\boldsymbol{\gamma}\in\Theta_K} |\mathcal{L}(\mathbf{Y},\boldsymbol{\gamma}) - \bar{\mathcal{L}}(\mathbf{Y},\boldsymbol{\gamma})|$$

Or equivalently,

$$\frac{L}{8}\left(\frac{||\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_0||_F}{|||\boldsymbol{\Sigma}(t)|||_2}\right)^2 \geq \frac{1}{\sigma^2}\sqrt{Lm^2C\log\frac{2}{\delta}}\cdot|||\sigma^2\mathbf{I}+\boldsymbol{\Phi}\boldsymbol{\Gamma_0}\boldsymbol{\Phi}^T|||_2$$

Or equivalently,

$$L \geq \frac{64Cm^2}{\sigma^4}\cdot\log\frac{2}{\delta}\cdot|||\sigma^2\mathbf{I}+\boldsymbol{\Phi}\boldsymbol{\Gamma_0}\boldsymbol{\Phi}^T|||_2^2\cdot\left(\frac{|||\boldsymbol{\Sigma}(t)|||_2}{||\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_0||_F}\right)^4$$

Sufficient conditions

For $P(\mathcal{A}) \geq 1 - \delta$, we need (m, L) such that

$$L \geq \frac{64Cm^2}{\sigma^4} \cdot \log\frac{2}{\delta} \cdot |||\sigma^2\mathbf{I} + \mathbf{\Phi}\mathbf{\Gamma_0}\mathbf{\Phi}^T|||_2^2 \cdot \left(\frac{|||\mathbf{\Sigma}(t)|||_2}{||\mathbf{\Sigma} - \mathbf{\Sigma}_0||_F}\right)^4$$

Sufficient conditions

For $P(\mathcal{A}) \geq 1 - \delta$, we need (m, L) such that

$$L \geq \frac{64Cm^2}{\sigma^4} \cdot \log\frac{2}{\delta} \cdot |||\sigma^2\mathbf{I} + \mathbf{\Phi}\mathbf{\Gamma_0}\mathbf{\Phi}^T|||_2^2 \cdot \left(\frac{|||\mathbf{\Sigma}(t)|||_2}{||\mathbf{\Sigma} - \mathbf{\Sigma}_0||_F}\right)^4$$

Or equivalently,

$$L \geq 64Cm^2 \log\frac{2}{\delta} \cdot \left(1 + \frac{\gamma_{\text{max}}}{\sigma^2}(1 + \delta_K)\right)^2 \cdot \left(\frac{|||\mathbf{\Sigma}(t)|||_2}{||\mathbf{\Sigma} - \mathbf{\Sigma}_0||_F}\right)^4$$

Sufficient conditions

For $P(\mathcal{A}) \geq 1 - \delta$, we need (m, L) such that

$$L \geq \frac{64Cm^2}{\sigma^4} \cdot \log\frac{2}{\delta} \cdot |||\sigma^2\mathbf{I} + \mathbf{\Phi}\mathbf{\Gamma_0}\mathbf{\Phi}^T|||_2^2 \cdot \left(\frac{|||\mathbf{\Sigma}(t)|||_2}{||\mathbf{\Sigma} - \mathbf{\Sigma}_0||_F}\right)^4$$

Or equivalently,

$$L \geq 64Cm^2 \log\frac{2}{\delta} \cdot \left(1 + \frac{\gamma_{\text{max}}}{\sigma^2}(1 + \delta_K)\right)^2 \cdot \left(\frac{|||\mathbf{\Sigma}(t)|||_2}{||\mathbf{\Sigma} - \mathbf{\Sigma}_0||_F}\right)^4$$

Or equivalently,

$$L \geq 64Cm^2 \log\frac{2}{\delta} \cdot \left(1 + \frac{\gamma_{\text{max}}}{\sigma^2}(1 + \delta_K)\right)^2 \cdot \left(\frac{\sigma^2 + \gamma_{\text{max}}(1 + \delta_{2K})}{||(\mathbf{\Phi} \odot \mathbf{\Phi})(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)||_2}\right)^4$$

Sufficient conditions

For $P(\mathcal{A}) \geq 1 - \delta$, we need (m, L) such that

$$L \geq 64Cm^2 \log \frac{2}{\delta} \cdot \left(1 + \frac{\gamma_{\text{max}}}{\sigma^2}(1 + \delta_K)\right)^2 \cdot \left(\frac{\sigma^2 + \gamma_{\text{max}}(1 + \delta_{2K})}{||(\mathbf{\Phi} \odot \mathbf{\Phi})(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)||_2}\right)^4$$

Sufficient conditions

For $P(\mathcal{A}) \geq 1 - \delta$, we need (m, L) such that

$$L \geq 64Cm^2 \log \frac{2}{\delta} \cdot \left(1 + \frac{\gamma_{\text{max}}}{\sigma^2}(1 + \delta_K)\right)^2 \cdot \left(\frac{\sigma^2 + \gamma_{\text{max}}(1 + \delta_{2K})}{||(\mathbf{\Phi} \odot \mathbf{\Phi})(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)||_2}\right)^4$$

Or equivalently,

$$L \geq 64Cm^2 \log \frac{2}{\delta} \cdot \left(1 + \frac{\gamma_{\text{max}}}{\sigma^2}(1 + \delta_K)\right)^2 \cdot \left(\frac{\sigma^2 + \gamma_{\text{max}}(1 + \delta_{2K})}{\gamma_{\text{min}}\sqrt{1 - \delta_{2K}(\mathbf{\Phi} \odot \mathbf{\Phi})}}\right)^4$$

Sufficient conditions

For $P(\mathcal{A}) \geq 1 - \delta$, we need (m, L) such that

$$L \geq 64Cm^2 \log \frac{2}{\delta} \cdot \left(1 + \frac{\gamma_{\text{max}}}{\sigma^2}(1 + \delta_K)\right)^2 \cdot \left(\frac{\sigma^2 + \gamma_{\text{max}}(1 + \delta_{2K})}{||(\mathbf{\Phi} \odot \mathbf{\Phi})(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)||_2}\right)^4$$

Or equivalently,

$$L \geq 64Cm^2 \log \frac{2}{\delta} \cdot \left(1 + \frac{\gamma_{\text{max}}}{\sigma^2}(1 + \delta_K)\right)^2 \cdot \left(\frac{\sigma^2 + \gamma_{\text{max}}(1 + \delta_{2K})}{\gamma_{\text{min}}\sqrt{1 - \delta_{2K}(\mathbf{\Phi} \odot \mathbf{\Phi})}}\right)^4$$

Or equivalently,

$$L \geq 64Cm^2 \log \frac{2}{\delta} \cdot \left(1 + \frac{\gamma_{\text{max}}}{\sigma^2}\right)^6 \left(\frac{\sigma^2}{\gamma_{\text{min}}}\right)^4 \cdot \frac{(1 + \delta_K)^2(1 + \delta_{2K})^4}{(1 - \delta_{2K}(\mathbf{\Phi} \odot \mathbf{\Phi}))^2}$$

Sparse support recovery in MSBL - sufficient conditions

Theorem

Let $\mathbf{x}_1, \mathbf{x}_2 \ldots \mathbf{x}_L$ *be generated according to the true distribution* $\mathcal{N}(0, \Gamma_0)$*. Let* $\mathcal{S}_0 = supp(\boldsymbol{\gamma}_0)$ *and,* $|\mathcal{S}_0| \leq K$*. Let* $\hat{\boldsymbol{\gamma}}$ *be the global maximizer of* $\log p(\mathbf{Y}; \boldsymbol{\gamma})$ *and,* $\hat{\boldsymbol{\gamma}} \in \Theta_K$*, where* $\Theta_K = \left\{\boldsymbol{\gamma} \in \mathbb{R}_+^n : ||\boldsymbol{\gamma}||_0 \leq K, , ||\boldsymbol{\gamma}||_\infty \leq \gamma_{max}\right\}$*. Then, the following holds.*

$$\mathbb{P}\left(supp(\hat{\boldsymbol{\gamma}}) = \mathcal{S}_0\right) \geq 1 - 2e^{-\eta m L}$$

where $\eta(m) = \left[64Cm^3 \cdot \left(1 + \frac{\gamma_{max}}{\sigma^2}\right)^6 \left(\frac{\sigma^2}{\gamma_{min}}\right)^4 \cdot \frac{(1+\delta_K)^2(1+\delta_{2K})^4}{(1-\delta_{2K}(\boldsymbol{\Phi} \odot \boldsymbol{\Phi}))^2}\right]^{-1}$.