Cramér Rao-Type Bounds for Sparse Bayesian Learning

Ranjitha P

February 11, 2012

1 System Model of SBL

2 Compressible Signal
- Examples

3 Bayesian Lower Bounds
- Contributions
- Noise Variance known: Bounds from the Joint pdf
- General Marginalized Bounds
- Noise Variance unknown: Bounds from the Joint pdf

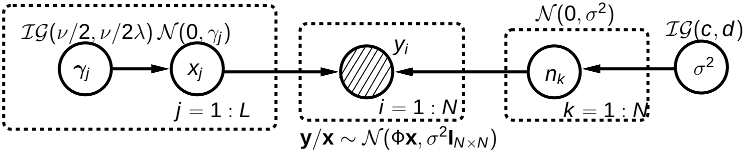
4 Simulation Results
- Lower Bounds on the MSE Performance of $\hat{\mathbf{x}}(\mathbf{y})$
- Lower Bounds on the MSE Performance of $\hat{\gamma}(\mathbf{y})$
- Lower Bounds on the MSE Performance of $\hat{\xi}(\mathbf{y})$

SMV-SBL

- Linear Single Measurement Vector (SMV) SBL model

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n},$$

$\mathbf{y} \in \mathbb{R}^N$, the measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{N \times L}$: known and $N < L$, $\mathbf{x} \in \mathbb{R}^L$: unknown compressible vector, $\mathbf{n} \sim \mathcal{N}(0, \sigma^2)$, σ^2 may be known or unknown.



- A vector $\mathbf{x}$ is p-compressible if $|x_i| \leq R i^{-1/p}$ for $i = 1, \ldots, L$
- Q: Is it possible to obtain such a compressible signal by drawing samples from a distribution?
- Answer: Yes, such priors are known as compressible priors.

- Laplace distribution is NOT compressible
- Generalized Compressible Prior: $\mathbf{x}$

$$p_{\mathbf{X}}(\mathbf{x}) \propto \prod_{i=1}^{L} \left(1 + \frac{|x_i|^{\tau}}{\nu}\right)^{-(\nu+1)/\tau},$$

where $x_i \in (-\infty, \infty), \quad \tau, \nu > 0$.

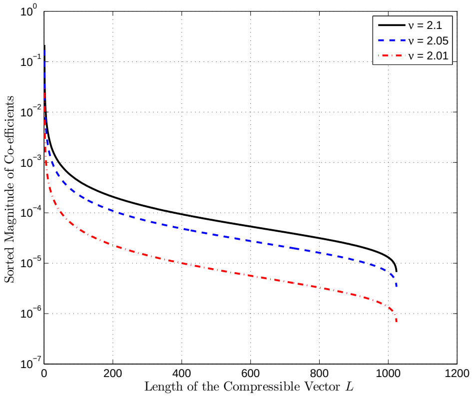


Figure: Decay profile of the sorted magnitudes of *i.i.d.* samples drawn from a Student-*t* distribution.

BCRB, HCRB and MCRB

- The MSE matrix $\mathbf{E}^{\boldsymbol{\theta}}$ is defined as

$$\mathbf{E}^{\boldsymbol{\theta}} \triangleq \mathbb{E}_{\mathbf{Y},\Theta_r}\left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{y}))(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{y}))^T\right],$$

  where Θ_r denotes the random parameters to be estimated (whose realization is given by $\boldsymbol{\theta}_r$).
- $\mathbf{I}^{\boldsymbol{\theta}}$ is expressed in terms of the individual blocks of submatrices, where the $(ij)^{th}$ block is given by

$$\mathbf{I}^{\boldsymbol{\theta}}_{ij} = -\mathbb{E}_{\mathbf{Y},\Theta_r}[\nabla_{\boldsymbol{\theta}_i}\nabla^T_{\boldsymbol{\theta}_j}\log p_{\mathbf{Y},\Theta_r;\Theta_d}(\mathbf{y},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)].$$

- A lower bound on the MSE matrix $\mathbf{E}^{\boldsymbol{\theta}}$ is given by the inverse of the FIM:

$$\mathbf{E}^{\boldsymbol{\theta}} \succeq \left(\mathbf{I}^{\boldsymbol{\theta}}\right)^{-1}.$$

Known Noise Variance

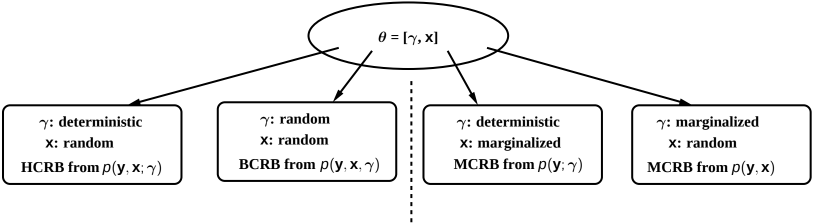


Figure: Summary of the lower bounds derived in this work when noise variance is assumed to be known.

Unknown Noise Variance

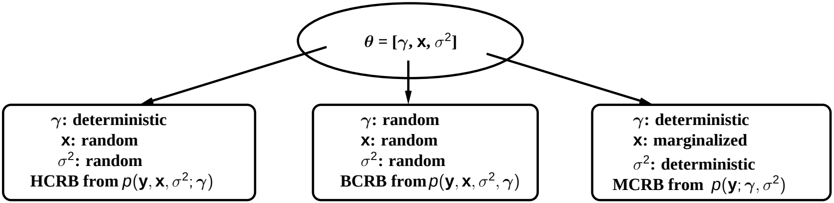


Figure: Different modeling assumptions and the corresponding bounds derived in this work when noise variance is assumed to be unknown.

HCRB for $\boldsymbol{\theta} = [\mathbf{x}, \gamma]$

Proposition

For the signal model in (3), the HCRB on the MSE matrix $\mathbf{E}^{\theta}$ of an unknown vector $\theta = [\mathbf{x}, \gamma]$, where the conditional distribution of the unknown compressible signal $\mathbf{x}/\gamma$ is $\mathcal{N}(0, \boldsymbol{\Upsilon})$ and γ is modeled as an unknown deterministic parameter, is given by $\mathbf{E}^{\theta} \succeq (\mathbf{H}^{\theta})^{-1}$, where

$$\mathbf{H}^{\theta} = \begin{bmatrix} \left(\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \boldsymbol{\Upsilon}^{-1}\right) & \mathbf{0}_{L\times L} \\ \mathbf{0}_{L\times L} & \text{diag}(2\gamma_1^2,\ 2\gamma_2^2,\ \ldots,\ 2\gamma_L^2)^{-1} \end{bmatrix}.$$

BCRB for $\boldsymbol{\theta} = [\mathbf{x}, \gamma]$

Proposition

For the signal model in (3), the BCRB on the MSE matrix $\mathbf{E}^{\boldsymbol{\theta}}$ *of an unknown random vector* $\boldsymbol{\theta} = [\mathbf{x}, \gamma]$, *where the conditional distribution of the unknown compressible signal* $\mathbf{x}/\gamma$ *is* $\mathcal{N}(0, \boldsymbol{\Upsilon})$, *the hyperprior distribution on* γ *is* $\prod_{i=1}^{L} \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2\lambda}\right)$, *is given by* $\mathbf{E}^{\boldsymbol{\theta}} \succeq (\mathbf{B}^{\boldsymbol{\theta}})^{-1}$, *where*

$$\mathbf{B}^{\boldsymbol{\theta}} = \begin{bmatrix} \left(\frac{\boldsymbol{\Phi}^T \boldsymbol{\Phi}}{\sigma^2} + \boldsymbol{\Upsilon}^{-1}\right) & \mathbf{0}_{L \times L} \\ \mathbf{0}_{L \times L} & \frac{\lambda^2(\nu+1)(\nu+7)}{2\nu} \mathbf{I}_{L \times L} \end{bmatrix}.$$

MCRB for $\boldsymbol{\theta} = [\gamma]$

Theorem

For the signal model in (3), the log likelihood function $\log p_{\mathbf{Y};\gamma}(\mathbf{y};\gamma)$ satisfies the regularity conditions. Further, the MCRB on the MSE matrix $\mathbf{E}^{\gamma}$ of the unknown deterministic vector $\boldsymbol{\theta} = [\gamma]$ is given by $\mathbf{E}^{\gamma} \succeq (\mathbf{M}^{\gamma})^{-1}$, where the ij^{th} element of $\mathbf{M}^{\gamma}$ is given by

$$\mathbf{M}_{ij}^{\gamma} = \frac{1}{2}(\Phi_j^T \Sigma_y^{-1} \Phi_i)^2,$$

for $1 \leq i, j \leq L$, where Φ_i is the i^{th} column of Φ, and $\Sigma_y = \sigma^2 \mathbf{I}_{N \times N} + \Phi \Upsilon \Phi^T$, as defined earlier.

MCRB for $\theta = [\mathbf{x}]$

The Student-t prior,

$$p_{\mathbf{X}}(\mathbf{x}) = \left(\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\right)^L \left(\frac{\lambda}{\pi\nu}\right)^{L/2} \prod_{i=1}^{L} \left(1 + \frac{\lambda x_i^2}{\nu}\right)^{-(\nu+1)/2},$$

where $x_i \in (-\infty, \infty)$, $\nu, \lambda > 0$, ν: number of degrees of freedom, λ: inverse variance.

Theorem

For the signal model in (3), the MCRB on the MSE matrix $\mathbf{E}^{\mathbf{x}}$ of the unknown compressible random vector $\theta = [\mathbf{x}]$ distributed as (1), is given by $\mathbf{E}^{\mathbf{x}} \succeq (\mathbf{M}^{\mathbf{x}})^{-1}$, where

$$\mathbf{M}^{\mathbf{x}} \triangleq \frac{\Phi^T \Phi}{\sigma^2} + \frac{\lambda(\nu+1)}{(\nu+3)} \mathbf{I}_{L \times L}.$$

GCP on $\mathbf{x}$:

$$p_{\mathbf{X}}(\mathbf{x}) = \left(\frac{\tau}{2}\left(\frac{\lambda}{\nu}\right)^{1/\tau}\frac{\Gamma((\nu+1)/\tau)}{\Gamma(1/\tau)\Gamma(\nu/\tau)}\right)^{L}\prod_{i=1}^{L}\left(1+\frac{\lambda\,|x_i|^{\tau}}{\nu}\right)^{-(\nu+1)/\tau} \quad (1)$$

Theorem

For the signal model in (3), the MCRB on the MSE matrix $\mathbf{E}_{\tau}^{\boldsymbol{\theta}}$ *of the unknown random vector* $\boldsymbol{\theta} = [\mathbf{x}]$*, where* $\mathbf{x}$ *is distributed by a GCP in (1) is given by* $\mathbf{E}_{\tau}^{\boldsymbol{\theta}} \succeq (\mathbf{M}_{\tau}^{\boldsymbol{\theta}})^{-1}$*, where*

$$\mathbf{M}_{\tau}^{\boldsymbol{\theta}} = \frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \frac{\tau^2(\nu+1)}{(\nu+\tau+1)}\left(\frac{\lambda}{\nu}\right)^{2/\tau}\frac{\Gamma\left(\frac{\nu+2}{\tau}\right)\Gamma\left(2-\frac{1}{\tau}\right)}{\Gamma\left(\frac{1}{\tau}\right)\Gamma\left(\frac{\nu}{\tau}\right)}\mathbf{I}_{L\times L}, \quad (2)$$

Preliminaries

- In the Bayesian formulation, the unknown noise variance is associated with a prior, $\sigma^2 \sim \mathcal{IG}(c, d)$,

$$p_{\Xi}(\xi) = \frac{d^c}{\Gamma(c)} \xi^{(-c-1)} \exp\left\{-\frac{d}{\xi}\right\}; \quad \xi \in (0, \infty),\ c, d > 0. \tag{3}$$

- Under this assumption, one can marginalize the unknown noise variance and obtain the marginalized likelihood $p(\mathbf{y}/\mathbf{x})$ as,

$$p(\mathbf{y}/\mathbf{x}) = \frac{(2d)^c \Gamma(N/2 + c)}{\Gamma(c)(\pi)^{N/2}} \left((\mathbf{y} - \Phi\mathbf{x})^T(\mathbf{y} - \Phi\mathbf{x}) + 2d\right)^{-\left(\frac{N}{2}+c\right)}, \tag{4}$$

which is a multivariate Student-t distribution.

HCRB for $\boldsymbol{\theta} = [\underbrace{\mathbf{x}, \gamma}_{\boldsymbol{\theta}'}, \xi]$

Proposition

For the signal model in (3), the HCRB on the MSE matrix $\mathbf{E}_{\xi}^{\boldsymbol{\theta}}$ of the unknown vector $\boldsymbol{\theta} = [\underbrace{\mathbf{x}, \gamma}_{\boldsymbol{\theta}'}, \xi]$, with the conditional distribution of the unknown compressible vector $\mathbf{x}/\gamma \sim \mathcal{N}(0, \boldsymbol{\Upsilon})$, and ξ modeled as an unknown deterministic parameter, is given by $(\mathbf{H}_{\xi}^{\boldsymbol{\theta}})^{-1}$, where

$$\mathbf{H}_{\xi}^{\boldsymbol{\theta}} = \begin{bmatrix} \mathbf{H}^{\boldsymbol{\theta}'} & \mathbf{0}_{L\times 1} \\ \mathbf{0}_{1\times L} & \dfrac{N}{2\xi^2} \end{bmatrix}. \quad (5)$$

BCRB for $\boldsymbol{\theta} = [\underbrace{\mathbf{x}, \gamma}_{\boldsymbol{\theta}'}, \xi]$

Proposition

For the signal model in (3), the HCRB on the MSE matrix $\mathbf{E}_{\xi}^{\boldsymbol{\theta}}$ of the unknown random vector $\boldsymbol{\theta} = [\underbrace{\mathbf{x}, \gamma}_{\boldsymbol{\theta}'}, \xi]$, with the conditional distribution of the unknown compressible vector given by $\mathbf{x}/\gamma$ is $\mathcal{N}(0, \boldsymbol{\Upsilon})$, where γ is modeled as an unknown deterministic or random parameter, and the unknown random parameter ξ is distributed as $\mathcal{IG}(c, d)$, is given by $(\mathbf{H}_{\xi}^{\boldsymbol{\theta}})^{-1}$, where

$$\mathbf{H}_{\xi}^{\boldsymbol{\theta}} = \begin{bmatrix} \mathbf{H}^{\boldsymbol{\theta}'} & \mathbf{0}_{L\times 1} \\ \mathbf{0}_{1\times L} & \frac{c(c+1)(N/2+c+3)}{d^2} \end{bmatrix}. \quad (6)$$

MCRB for $\boldsymbol{\theta} = [\gamma, \xi]$

Theorem

For the signal model in (3), the log likelihood function $\log p_{\mathbf{Y};\gamma,\xi}(\mathbf{y};\gamma,\xi)$ *satisfies the regularity condition. Further, the MCRB on the MSE matrix* $\mathbf{E}_{\xi}^{\boldsymbol{\theta}}$*, of the unknown deterministic vector* $\boldsymbol{\theta} = [\gamma, \xi]$ *is given by* $\mathbf{E}_{\xi}^{\boldsymbol{\theta}} \succeq (\mathbf{M}_{\xi}^{\boldsymbol{\theta}})^{-1}$*, where*

$$\mathbf{M}_{\xi}^{\boldsymbol{\theta}} = \begin{bmatrix} \mathbf{M}_{\xi}^{\boldsymbol{\theta}}(\gamma) & \mathbf{M}_{\xi}^{\boldsymbol{\theta}}(\gamma,\xi) \\ \mathbf{M}_{\xi}^{\boldsymbol{\theta}}(\xi,\gamma) & \mathbf{M}_{\xi}^{\boldsymbol{\theta}}(\xi) \end{bmatrix}, \tag{7}$$

$(\mathbf{M}_{\xi}^{\boldsymbol{\theta}}(\gamma))_{ij} = \frac{1}{2}\left\{(\Phi_j^T \Sigma_y^{-1} \Phi_i)^2\right\}$, $\mathbf{M}_{\boldsymbol{\theta}}^{\xi} = \frac{1}{2} Tr(\Sigma_y^{-2})$ *and*

$(\mathbf{M}_{\xi}^{\boldsymbol{\theta}}(\gamma,\xi))_i = (\mathbf{M}_{\xi}^{\boldsymbol{\theta}}(\xi,\gamma))_i = \frac{\Phi_i^T \Sigma_y^{-2} \Phi_i}{2}$.

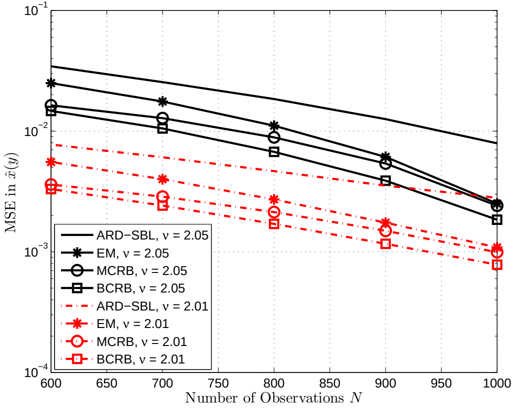


Figure: Plot of the MSE performance of $\hat{\mathbf{x}}(\mathbf{y})$, the corresponding MCRB and BCRB as a function of N, where SNR = 40dB.

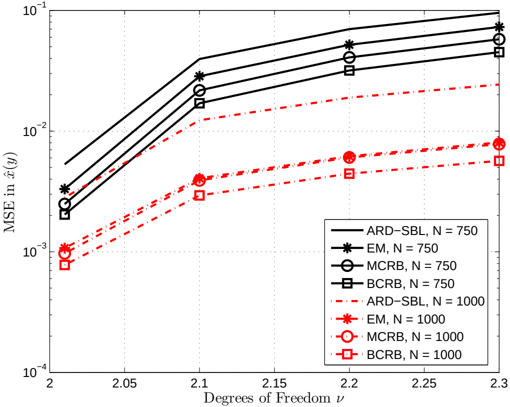


Figure: Plot of the MSE performance of $\hat{\mathbf{x}}(\mathbf{y})$, the corresponding MCRB and BCRB as a function of ν, where SNR = 40dB.

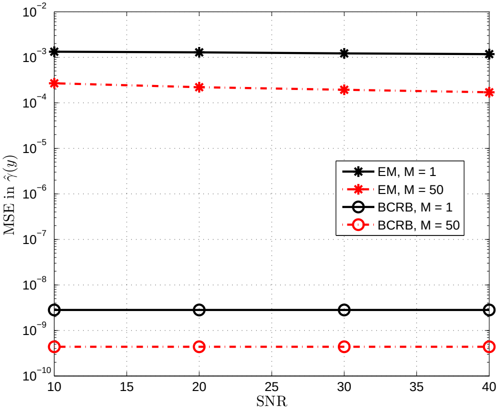


Figure: Plot of the MSE performance of $\hat{\gamma}(\mathbf{y})$ and the corresponding HCRB as a function of SNR, where $N = 1000$.

SNR(dB)		10	20	30
$M = 1$	MSE	0.05429	0.05270	0.05132
	MCRB	0.05218	0.05134	0.05070
	BCRB	0.04880	0.04880	0.04880
$M = 50$	MSE	0.04500	0.03923	0.03476
	MCRB	0.0012	0.0011	0.0010
	BCRB	9.766×10^{-4}	9.766×10^{-4}	9.766×10^{-4}

Table: MSE of the estimator $\hat{\gamma}(\mathbf{y})$, the MCRB and the BCRB as a function of SNR for $N = 1500$.

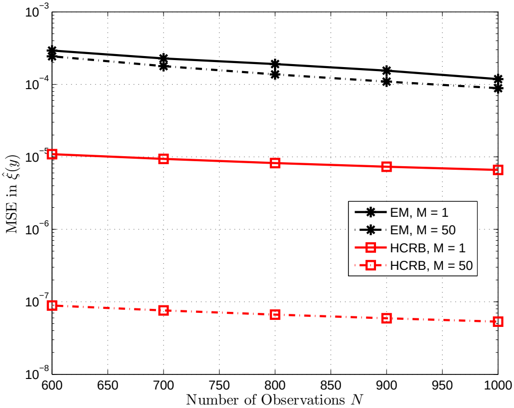


Figure: Plot of MSE performance of $\hat{\xi}(\mathbf{y})$ along with the HCRB as a function of *N*.

N		1500	1700
$M = 1$	MSE	0.7362×10^{-8}	0.6360×10^{-8}
	MCRB	0.3796×10^{-8}	0.3071×10^{-8}
	HCRB	0.1333×10^{-8}	0.1176×10^{-8}
$M = 50$	MSE	0.9304×10^{-9}	0.8661×10^{-9}
	MCRB	0.6803×10^{-10}	0.6142×10^{-10}
	HCRB	0.2666×10^{-10}	0.2352×10^{-10}

Table: MSE of the estimator $\hat{\xi}(\mathbf{y})$, the MCRB and the HCRB as a function of N.