On Parametric Misspecified Bayesian Cramér-Rao bound: An application to linear/Gaussian systems

Shuo Tang (*tang.shu@ece.neu.edu*), Gerald LaMountain, Tales Imbiriba and Pau Closas
Electrical & Computer Engineering Department, Northeastern University, Boston, MA, USA

Background

Bayesian Cramér-Rao bound (BCRB)

Consider a general statistical model with unknown and random vector parameter $\boldsymbol{\psi} \in \mathbb{R}^{n_\psi \times 1}$, such that the model is characterized by its *prior* and likelihood distributions

$$\mathcal{M}_* = \{\boldsymbol{x}|\boldsymbol{\psi} \sim p_*(\boldsymbol{x}|\boldsymbol{\psi}),\ \boldsymbol{\psi} \sim p(\boldsymbol{\psi}) : \ \boldsymbol{\psi} \in \Psi \subset \mathbb{R}^{n_\psi}\}$$

For any unbiased estimator $\hat{\boldsymbol{\psi}}(\boldsymbol{x})$, the BCRB states that [1, 2]

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\left\{\left(\hat{\boldsymbol{\psi}}(\boldsymbol{x}) - \boldsymbol{\psi}\right)\left(\hat{\boldsymbol{\psi}}(\boldsymbol{x}) - \boldsymbol{\psi}\right)^\top\right\} - \boldsymbol{J}^{-1} \geq \boldsymbol{0}\,,$$

where $\boldsymbol{J} \in \mathbb{R}^{n_\psi \times n_\psi}$ denotes the so-called Bayesian Fisher Information Matrix (BFIM)

$$\boldsymbol{J} = \underbrace{\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\left\{\left(\frac{\partial}{\partial \boldsymbol{\psi}} \ln p(\boldsymbol{x}|\boldsymbol{\psi})\right)\left(\frac{\partial}{\partial \boldsymbol{\psi}} \ln p(\boldsymbol{x}|\boldsymbol{\psi})\right)^\top\right\}}_{\boldsymbol{J}_D} + \underbrace{\mathbb{E}_{\boldsymbol{\psi}}\left\{\left(\frac{\partial}{\partial \boldsymbol{\psi}} \ln p(\boldsymbol{\psi})\right)\left(\frac{\partial}{\partial \boldsymbol{\psi}} \ln p(\boldsymbol{\psi})\right)^\top\right\}}_{\boldsymbol{J}_P}$$

which is composed of the Fisher Information Matrix $\boldsymbol{J}_D$ (accounting for the information on $\boldsymbol{\psi}$ from the data) and the prior information matrix $\boldsymbol{J}_P$ (accounting for the *prior* on $\boldsymbol{\psi}$).

References

[1] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice-Hall, Inc., USA, 1993.

[2] Harry L Van Trees. *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory.* John Wiley & Sons, 2004.

[3] Quang H Vuong. Cramér-Rao bounds for misspecified models. Technical report, California Institute of Technology, 1986.

[4] Christ D Richmond and Larry L Horowitz. Parameter bounds on estimation accuracy under model misspecification. *IEEE Transactions on Signal Processing*, 63(9):2263–2278, 2015.

[5] Stefano Fortunati, Fulvio Gini, Maria S Greco, and Christ D Richmond. Performance bounds for parameter estimation under misspecified models: Fundamental findings and applications. *IEEE Signal Processing Magazine*, 34(6):142–157, 2017.

[6] Christ D Richmond. On constraints in parameter estimation and model misspecification. In *2018 21st International Conference on Information Fusion (FUSION)*, pages 1080–1085. IEEE, 2018.

Misspecified BCRB (MBCRB)

Assumed Model: $\mathcal{M} = \{\boldsymbol{x}|\boldsymbol{\theta} \sim f(\boldsymbol{x}|\boldsymbol{\theta}),\ \boldsymbol{\theta} \sim f(\boldsymbol{\theta})\ :\ \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^{n_\theta}\}$

where $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta \times 1}$ denotes the unknown random parameter which the estimator is attempting to infer from the available data $\boldsymbol{x} \in \mathbb{R}^{n_x \times 1}$. **Note that the assumed model $\mathcal{M}$ can differ from the true $\mathcal{M}_*$.**

Pseudotrue:
$$\boldsymbol{\theta}_0(\boldsymbol{\psi}) = \arg\min_{\boldsymbol{\theta}} \mathcal{D}\Big(p(\boldsymbol{x}|\boldsymbol{\psi})||f(\boldsymbol{x},\boldsymbol{\theta})\Big) = \arg\min_{\boldsymbol{\theta}}\Big(-\mathbb{E}_{\boldsymbol{x}|\boldsymbol{\psi}}\Big\{\ln f(\boldsymbol{x},\boldsymbol{\theta})\Big\}\Big),$$

which is slightly different from the pseudotrue parameter defined in other MCRB works [3–6], since we introduce the prior information in the assumed model.

MBCRB: **Theorem 1.** *Given the true model $\mathcal{M}_*$ parameterized by $\boldsymbol{\psi}$ and the assumed model $\mathcal{M}$ parameterized by $\boldsymbol{\theta}$, the error covariance of any MS-unbiased estimator satisfies that*

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\left\{\left(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}_0(\boldsymbol{\psi})\right)\left(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}_0(\boldsymbol{\psi})\right)^\top\right\} - \mathbb{E}_{\boldsymbol{\psi}}\left\{\frac{\partial \boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right\}\boldsymbol{J}^{-1}\mathbb{E}_{\boldsymbol{\psi}}\left\{\frac{\partial \boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right\}^\top \geq \boldsymbol{0}\,,$$

Extended Biased Bound: (when $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ belong to the same vector space $\Theta = \Psi$)

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\left\{\left(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\psi}\right)\left(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\psi}\right)^\top\right\} \geq \mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\left\{\frac{\partial \boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right\}\boldsymbol{J}^{-1}\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\left\{\frac{\partial \boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right\}^\top + \boldsymbol{r}\boldsymbol{r}^\top,$$

where the biased term is $\boldsymbol{r} = \boldsymbol{\theta}_0(\boldsymbol{\psi}) - \boldsymbol{\psi}$.

Experiments of Linear Gaussian System Application

True Model: $\boldsymbol{\psi} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\psi}}, \boldsymbol{\Sigma}_{\boldsymbol{\psi}}),\ \ \boldsymbol{x}_n|\boldsymbol{\psi} \sim \mathcal{N}(\boldsymbol{H}_*\boldsymbol{\psi}, \boldsymbol{\Sigma}_*),\ \ n = 1, \ldots, N$

Assumed Model: $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}),\ \ \boldsymbol{x}_n|\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{H}\boldsymbol{\theta}, \boldsymbol{\Sigma}),\ \ n = 1, \ldots, N$

Parameters Setting: Note that the above parameter can be arbitrary. In the experiments, we use

- $n_{\boldsymbol{\psi}} = n_{\boldsymbol{\theta}} = 3$, $\boldsymbol{\mu}_{\boldsymbol{\psi}} = [10,\ 20,\ 5]^\top$, $\boldsymbol{\Sigma}_{\boldsymbol{\psi}} = \sigma^2_{\boldsymbol{\psi}}\boldsymbol{I} = 0.5\boldsymbol{I}$
- $\boldsymbol{H} = h_*\boldsymbol{I}$ and $h_* = 1$, $\boldsymbol{\Sigma}_* = \sigma^2_*\boldsymbol{Q}$, with $\sigma^2_* = 0.04$ and $\boldsymbol{Q}_{i,j} = \rho^{|i-j|}$, controller by $\rho = 0.5$
- The parameters of the assumed model are the same, **except for those mentioned misspecified ones:**

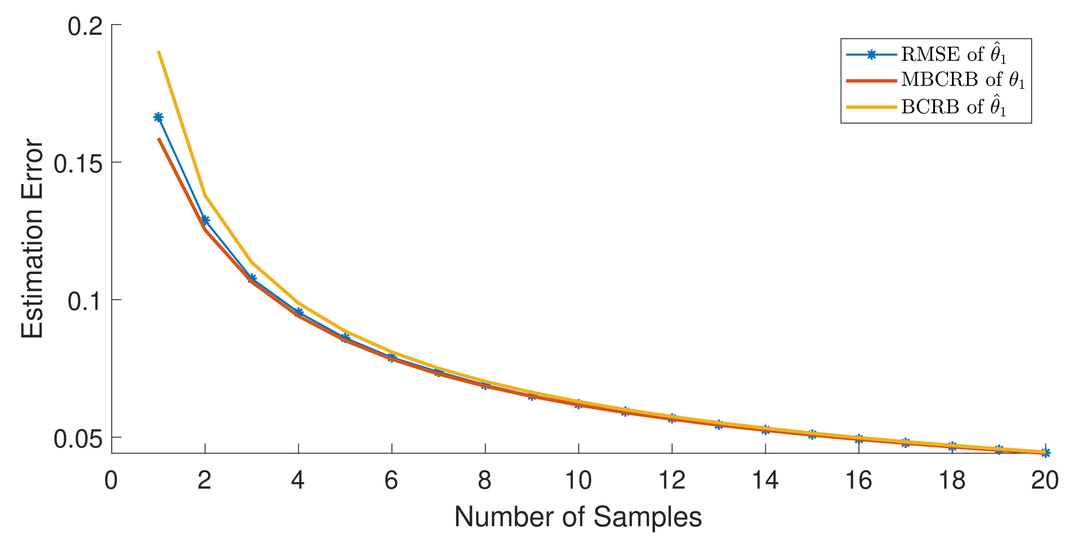

Figure 1: RMSE vs MBCRB and BCRB

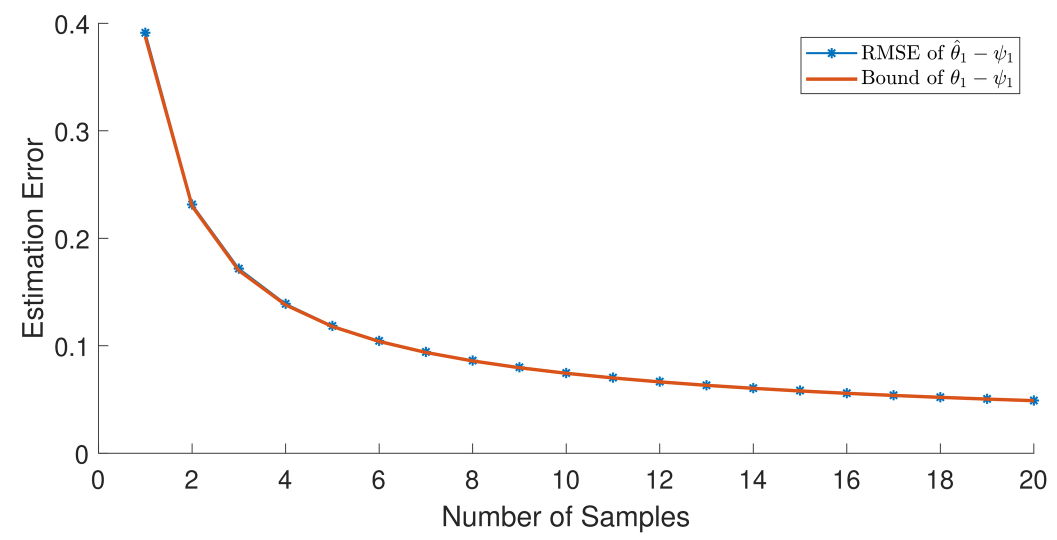

Figure 2: RMSE vs biased bound on $\hat{\boldsymbol{\theta}} - \boldsymbol{\psi}$

($\boldsymbol{\mu}_{\boldsymbol{\theta}} = [8,\ 18,\ 6]^\top$ and $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I} = 0.1\boldsymbol{I}$)

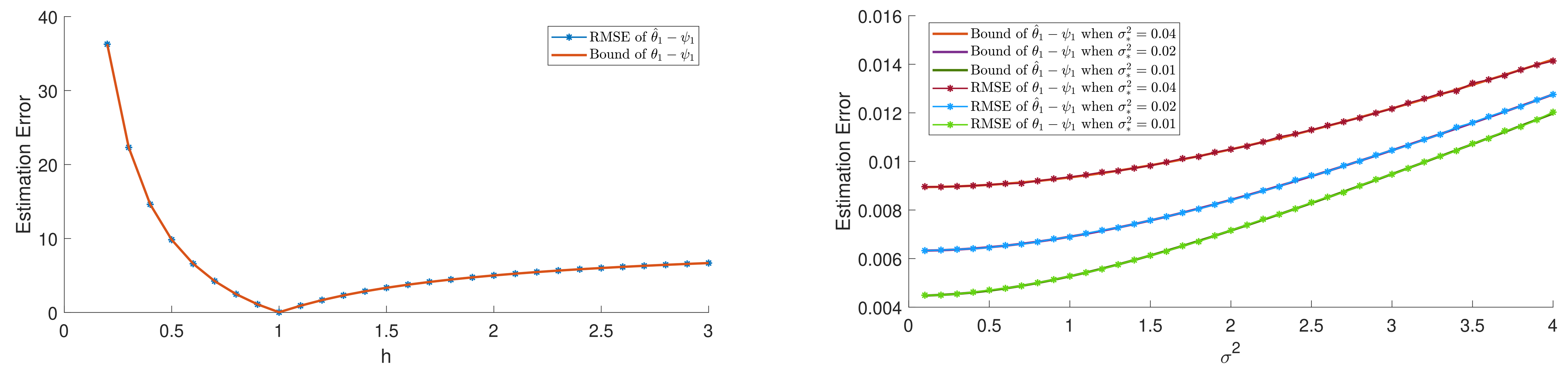

Figure 3: RMSE vs biased bound on $\hat{\boldsymbol{\theta}} - \boldsymbol{\psi}$ under different model misspecifications: when varying h (left panel) and varying σ^2 (right panel)

Assumptions and Derivations

Assumption 1: $\mathbb{E}_{\boldsymbol{x}|\boldsymbol{\psi}}\left\{\frac{\partial}{\partial \boldsymbol{\psi}} \ln p(\boldsymbol{x}|\boldsymbol{\psi})\right\} = \boldsymbol{0}$.

Assumption 2: $p(\psi_i = \psi_{i,\min}) = p(\psi_i = \psi_{i,\max}) = 0$, where $\boldsymbol{\psi} \in \Psi = \Psi_1 \times \cdots \times \Psi_{n_\psi}$ with $\Psi_i \triangleq [\psi_{i,\min}, \psi_{i,\max}]$ being the value range for each ψ_i, $i \in \{1, \ldots, n_\psi\}$, and integration limits $\psi_{i,\min}$ and $\psi_{i,\max}$ independent of $\boldsymbol{\psi}$.

Lemma 1: For $i \in \{1, \ldots, n_\psi\}$ and $j \in \{1, \ldots, n_\theta\}$

$$\int_\Psi \hat{\theta}_j(\boldsymbol{x})\frac{\partial}{\partial \psi_i}p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi} = 0$$

where $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)^\top \in \mathbb{R}^{n_\theta \times 1}$ is the estimator of $\boldsymbol{\theta}_0(\boldsymbol{\psi})$.

Lemma 2: Given pseudotrue $\boldsymbol{\theta}_0(\boldsymbol{\psi}) = (\theta_{0,1}(\boldsymbol{\psi}), \ldots, \theta_{0,n_\theta}(\boldsymbol{\psi}))^\top \in \mathbb{R}^{n_\theta \times 1}$,

$$\int_\Psi \theta_{0,j}(\boldsymbol{\psi})\frac{\partial}{\partial \psi_i}p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi} = -\int_\Psi \frac{\partial \theta_{0,j}(\boldsymbol{\psi})}{\partial \psi_i}p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi}.$$

With **Lemma 1 and 2**, we have

$$\iint \left(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}_0(\boldsymbol{\psi})\right)\left(\frac{\partial}{\partial \boldsymbol{\psi}} \ln p(\boldsymbol{x},\boldsymbol{\psi})\right)^\top p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{x} = \iint \frac{\partial \boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{x}$$

Please refer to our paper for the following details of leveraging Cauchy-Schwarz inequality and some tricky vector operations to reach the result in Theorem 1.

Conclusions

- This work proposes a new **Bayesian pseudotrue parameter**.
- We extend the existing works on CRB-type bounds for misspecified models to **a general Bayesian setting** by deriving the **MBCRB** lower-bounding $\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\left\{\left(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}_0(\boldsymbol{\psi})\right)\left(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}_0(\boldsymbol{\psi})\right)^\top\right\}$ when the *prior* is accounted for.
- A **biased bound** is derived based on MBCRB to bound the error centered on the **true parameter**.
- Experiments using **linear Gaussian systems** show that the proposed MCRB can lower-bound the errors **tightly at any level of misspecification**, where the traditional BCRB cannot.