ON THE CRAMÉR-RAO LOWER BOUND UNDER MODEL MISMATCH

Carsten Fritsche[†], *Umut Orguner*[*], *Emre Özkan*[†], *and Fredrik Gustafsson*[†]

[†]Linköping University, Department of Electrical Engineering, Linköping, Sweden
e-mail:{carsten,emre,fredrik@isy.liu.se}
[*] Middle East Technical University, Department of Electrical & Electronics Engineering, Ankara, Turkey
e-mail:{umut@metu.edu.tr}

ABSTRACT

Cramér-Rao lower bounds (CRLBs) are proposed for deterministic parameter estimation under model mismatch conditions where the assumed data model used in the design of the estimators differs from the true data model. The proposed CRLBs are defined for the family of estimators that may have a specified bias (gradient) with respect to the assumed model. The resulting CRLBs are calculated for a linear Gaussian measurement model and compared to the performance of the maximum likelihood estimator for the corresponding estimation problem.

Index Terms— Statistical Signal Processing, Cramér-Rao Lower bound, Parameter Estimation, Model mismatch

1. INTRODUCTION

Evaluating the performance of estimators generally relies on the achievable accuracy for the considered problem. When the mean square error (MSE) is used in performance evaluation, the lower bounds for the achievable MSE are utilized to answer questions such as: 1) Are the performance requirements set for the estimator feasible? 2) Has the estimator under evaluation sufficiently close performance to what is achievable for the problem? 3) Is there a large gap between the estimator's performance and the best achievable performance suggesting that there might be improvements if alternative estimators are designed?

The most well-known and popular lower bound for assessing MSE performance is the Cramér-Rao lower bound (CRLB) [1, 2]. CRLB can be defined for both deterministic [3–7] and random parameter estimation [3, 8] problems for both unbiased [3–5] and biased estimators [3,4]. It is well known that the CRLB is generally achieved by estimators under high SNR conditions and if the CRLB is achievable, the maximum likelihood (ML) estimator achieves it.

In this work, we consider CRLB type lower bounds for deterministic parameter estimation under model mismatch conditions, where the assumed data model used in designing the estimator differs from the true model. Although the literature on CRLB under model-match conditions is vast, there are very few studies devoted to the model mismatch case [9, 10]. The most relevant contribution to our work in the literature is the recent work by Richmond and Horowitz [10] where a CRLB type bound is computed for the MSE of the estimators having a specified bias with respect to (w.r.t.) the *true* model. The fundamental difference between our approach and [10] is that, in our contribution, CRLBs are derived for estimators that are unbiased or that have a specified bias (gradient) w.r.t. the *assumed* model. Moreover, the two approaches propose different score functions. The CRLB derived here can be considered to be more meaningful, as it is not restricted to the estimators for which the bias w.r.t. the true model has to be known.

2. CRLB UNDER MODEL MISMATCH

In parameter estimation, we are interested in inferring a deterministic parameter $x \in \mathbb{R}^n$ from a set of noisy measurements $y \in \mathbb{R}^m$. The corresponding estimator $\hat{x}(y)$ often requires a suitable model that relates the data to the unknown parameter.In general, the true model is not known and hence a model mismatch appears which has to be accounted for. In the sequel, CRLBs under model mismatch conditions are developed that can be used to assess the fundamental performance limits of estimators which are influenced by model mismatch.

2.1. Unbiased Estimators

We introduce an unbiased estimator $\hat{x}(y)$ that is not aware of the true measurement model. Hence, unbiasedness has to be defined w.r.t. an assumed model as follows

$$\mathbb{E}_{p(y|x)}\{\hat{x}(y)\} = \int \hat{x}(y)p(y|x)\,\mathrm{d}y = x, \tag{1}$$

where $p(y|x)$ is the assumed likelihood function. The mean square error matrix P under model mismatch is given as

$$\begin{aligned} P &= \mathbb{E}_{p_0(y|x)}\{(\hat{x}(y)-x)(\hat{x}(y)-x)^T\} \\ &= \int (\hat{x}(y)-x)(\hat{x}(y)-x)^T p_0(y|x)\,\mathrm{d}y, \end{aligned} \tag{2}$$

where $p_0(y|x)$ is the true likelihood function. Note that $\hat{x}(y)$ is the estimator derived under the assumed likelihood function $p(y|x)$, while the expectation for mean square error is performed w.r.t. the true likelihood function. Then, the CRLB under model mismatch is given by the following theorem.

Theorem 1. *If $\hat{x}(y)$ is any unbiased estimator of x w.r.t. the assumed model, then the MSE matrix under model mismatch can be lower bounded as follows*

$$P \geq J_{MM}^{-1}(x), \tag{3}$$

where the matrix inequality $A \geq B$ is equivalent to stating that $(A - B)$ is positive semi-definite. The $n \times n$ Fisher information matrix (FIM) under model mismatch is given by

$$J_{MM}(x) = \mathbb{E}_{p_0(y|x)} \left\{ s(x, y) s^T(x, y) \right\}, \tag{4}$$

with $n \times 1$ score function

$$s(x, y) = \frac{p(y|x)}{p_0(y|x)} \cdot [\nabla_x \log p(y|x)]. \tag{5}$$

Proof. See Appendix 5.1. □

It is worth stressing that the CRLB under model mismatch provides a lower bound on the *MSE* matrix under model mismatch and not the corresponding *covariance* matrix. This in turn means that the derived CRLB holds also for estimators that are biased w.r.t. the true model, but need to be unbiased w.r.t. the assumed model. In case there is no model mismatch, i.e. $p(y|x) = p_0(y|x)$ the FIM reduces to the standard FIM. Of particular importance is the condition when the bound satisfies the equality, as it is often used to assess if an estimator is efficient [4, 5]. For the model mismatch case, an unbiased estimator w.r.t. the assumed model is called efficient if the estimator's MSE matrix P coincides with the CRLB, i.e. $P = J_{\mathrm{MM}}^{-1}(x)$ holds. The following proposition gives the necessary and sufficient condition under which the estimator efficiency is achieved.

Proposition 1. *An unbiased estimator $\hat{x}(y)$ w.r.t. the assumed model is efficient, i.e. $P = J_{MM}^{-1}(x)$ holds, if and only if*

$$s(x, y) = J_{MM}(x) \cdot (\hat{x}(y) - x), \ \forall y. \tag{6}$$

Proof. See Appendix 5.2. □

In case there is no model mismatch, i.e. $p(y|x) = p_0(y|x)$ holds, the equality condition reduces to the well known equality condition for the standard CRLB, see [4, 5]. As a result, in order to test an estimator for efficiency requires only the knowledge of $s(x, y)$ and $J_{\mathrm{MM}}(x)$, which can be determined from the true likelihood $p_0(y|x)$ and the estimator's assumed likelihood $p(y|x)$, and the estimator $\hat{x}(y)$ w.r.t. the assumed model.

2.2. Biased Estimators

The results presented in Theorem 1, can be generalized to estimators $\hat{x}(y)$ that are biased w.r.t. the assumed model, i.e.

$$\mathbb{E}_{p(y|x)}\{\hat{x}(y)\} = x + b(x) \tag{7}$$

holds, where $b(x) = [b_1(x), b_2(x), \ldots, b_n(x)]^T$ denotes the bias vector that may depend on the unknown x. We further introduce the $n \times n$ bias Jacobian matrix $B(x) = \frac{\partial b(x)}{\partial x}$. Then, the CRLB under model mismatch for biased estimators can be stated in the following theorem.

Theorem 2. *If $\hat{x}(y)$ is a biased estimate of x w.r.t. the assumed model, then the MSE matrix under model mismatch can be lower bounded as follows*

$$P \geq [I_n + B(x)]\, J_{MM}^{-1}(x)\, [I_n + B(x)]^T \tag{8}$$

Proof. See Appendix 5.1. □

Note that the above inequality holds irrespective of whether the estimators are biased w.r.t. the true model or not.

3. APPLICATION TO LINEAR MODELS

The theoretical results of the previous section are validated on a couple of examples. It is assumed that the measurements are generated from the following true linear model

$$y = C_0 x + v_0, \tag{9}$$

where y is an $m \times 1$ observation vector, C_0 is a $m \times n$ observation matrix of rank n satisfying $m > n$, x is a $n \times 1$ vector of parameters to be estimated, and v_0 is an $m \times 1$ noise vector with pdf $p(v_0) = \mathcal{N}(v_0; 0, R_0)$. The true likelihood function is then given by $p_0(y|x) = \mathcal{N}(y; C_0 x, R_0)$. The estimator $\hat{x}(y)$ is generally not aware of the true model and subsequently has to introduce model assumptions. In the following it is assumed that the linear structure and the noise pdf is known, but C_0 and R_0 are unknown and are replaced by $C \neq C_0$ and $R \neq R_0$, respectively. Hence, the estimator's assumed likelihood function is given by $p(y|x) = \mathcal{N}(y; Cx, R)$.

3.1. FIM under model mismatch

The FIM under model mismatch, cf. (4), is given as follows:

$$\begin{aligned} J_{\mathrm{MM}}(x) = \sqrt{\frac{|R_0|}{|R|}} \sqrt{\frac{|\tilde{R}|}{|R|}} \exp\left\{\frac{1}{2} \bar{v}^T (R_0 - R/2)^{-1} \bar{v}\right\} \\ \times C^T R^{-1} \left[\tilde{R} + \tilde{v}\tilde{v}^T\right] R^{-1} C, \end{aligned} \tag{10a}$$

with

$$\tilde{R} = R/2 - R/2(R/2 - R_0)^{-1} R/2 > 0, \tag{10b}$$

$$\bar{v} = (C_0 - C)x, \tag{10c}$$

$$\tilde{v} = R/2(R/2 - R_0)^{-1} \bar{v}, \tag{10d}$$

under the assumption that $R_0 > R/2$. If this assumption is not satisfied J_{MM} goes to infinity. From the above expression, a couple of special cases can be derived. If $C_0 = C$, then

$$J_{\text{MM}}(x) = \sqrt{\frac{|R_0|}{|R|}}\sqrt{\frac{|\tilde{R}|}{|R|}}\, C^T R^{-1} \tilde{R} R^{-1} C. \qquad (11)$$

If $R_0 = R$, then we arrive at

$$J_{\text{MM}}(x) = \exp\left\{\bar{v}^T R^{-1} \bar{v}\right\} C^T R^{-1} \left[R + \bar{v}\bar{v}^T\right] R^{-1} C. \quad (12)$$

Clearly, if $C = C_0$ and $R = R_0$ are known, we arrive at the FIM for the true model, given by $J_{\text{TM}} = C_0^T R_0^{-1} C_0$. Similarly, the FIM for the assumed model is given by $J_{\text{AM}} = C^T R^{-1} C$.

3.2. MLE under model mismatch

For performance comparison, we introduce the ML estimator (MLE) w.r.t. the assumed model, which is given by

$$\hat{x}_{\text{ML}} = (C^T R^{-1} C)^{-1} C^T R^{-1} y. \qquad (13)$$

It can be easily shown that the MLE is unbiased w.r.t. the assumed model and its MSE matrix is equivalent to the CRLB for the assumed model, which is given by $\text{MSE}(\hat{x}_{\text{ML}}) = J_{\text{AM}}^{-1} = (C^T R^{-1} C)^{-1}$. The expected MSE performance of the MLE under model mismatch is of particular importance. The ML estimator bias and covariance w.r.t. the true model $p_0(y|x)$ is

$$\begin{aligned} b_0(\hat{e}_{\text{ML}}) =& [(C^T R^{-1} C)^{-1} C^T R^{-1} C_0 - I_n]\, x, \qquad (14)\\ \text{Cov}_0(\hat{e}_{\text{ML}}) =& (C^T R^{-1} C)^{-1} C^T R^{-1} R_0 R^{-1} C \\ & \times (C^T R^{-1} C)^{-1}. \qquad (15) \end{aligned}$$

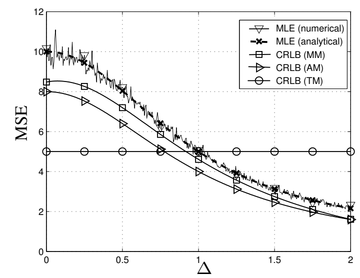
where we have defined $\hat{e}_{\text{ML}} = \hat{x}_{\text{ML}} - x$. Then, the MSE for the MLE under model mismatch can be expressed as follows:

$$\text{MSE}_0(\hat{e}_{\text{ML}}) = \text{Cov}_0(\hat{e}_{\text{ML}}) + b_0(\hat{e}_{\text{ML}}) b_0^T(\hat{e}_{\text{ML}}). \qquad (16)$$

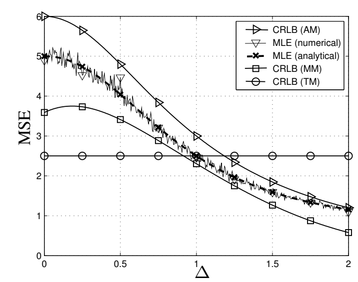
Again, a couple of special cases can be derived. If $C_0 = C$, then the MLE under model mismatch is unbiased, and $\text{MSE}_0(\hat{e}_{\text{ML}})$ equals $\text{Cov}_0(\hat{e}_{\text{ML}})$. If $R_0 = R$, then the MLE under model mismatch is biased, but the covariance reduces to $\text{Cov}_0(\hat{e}_{\text{ML}}) = (C^T R^{-1} C)^{-1}$.

3.3. Examples

In the following, the tightness of the CRLB under model mismatch is evaluated using different examples. For ease of exposition, we assume that $C_0 = [1, 1]^T$ and $C = [1, \Delta]^T$ where Δ is varied in the interval $[0, 2]$, and let $x = 1$. In the first example, we assume $R_0 = 10\, I_2$ and $R = 0.8\, R_0$, and compare the performance of the MLE under model mismatch (analytically using (16) and numerically using (13) from 2000 Monte Carlo runs) with the CRLB under model mismatch (CRLB (MM) $= J_{\text{MM}}^{-1}(x)$), the CRLB of the true model (CRLB (TM) $= J_{\text{TM}}^{-1}$), and the CRLB of the assumed model (CRLB (AM) $= J_{\text{AM}}^{-1}$). The results in Fig. 1 (a) show that both the CRLB (MM) and the CRLB (AM) provide a lower bound for all values Δ. For the case that $\Delta = 1$, there is no model mismatch in C and CRLB (TM) coincides with the MLE, which is a result of the special structure of R. While the CRLB (MM) is guaranteed to provide a lower bound for any unbiased estimator under model mismatch, this property generally does not hold for CRLB (TM) and CRLB (AM). In Fig. 1 (b), a second example is shown where we



(a) $R_0 = 10\, I_2$, $R = 0.8\, R_0$



(b) $R_0 = 5\, I_2$, $R = 1.2\, R_0$

Fig. 1. MSE vs. Δ of (a) Example 1 and (b) Example 2

assume $R_0 = 5\, I_2$ and $R = 1.2\, R_0$, i.e. the MLE is using a larger covariance than the true one. It can be observed that the CRLB (AM) no longer provides a lower bound on estimation performance, due to the increased uncertainty resulting from the choice of R. The CRLB (MM) however, is not affected by this and still provides a lower bound on the estimation performance.

4. CONCLUSION

In this article, we derive a novel set of CRLBs which account for the errors that occur from possible model mismatches when the estimator is unaware of the true model. We provide simulation results where these bounds are used to predict the performance of the ML estimator in case of a model mismatch.

5. APPENDIX

5.1. Proof of Theorem 1 and Theorem 2

We mainly follow the classical derivation of the CRLB such as the one in [5] and extend it to the case of model mismatch and biased estimators (for unbiased estimators, simply set $b(x) = 0$). We assume the classical regularity condition given as

$$\int \nabla_x p(y|x)\,\mathrm{d}y = 0 \Leftrightarrow \int \nabla_x \log p(y|x) p(y|x)\,\mathrm{d}y = 0 \quad (17)$$

is satisfied for all x, where ∇_x denotes the gradient w.r.t. vector x. In order to cover the vector parameter case, we define arbitrary vectors $a, b \in \mathbb{R}^n$. The biasedness condition for x under the assumed likelihood can be written as

$$\int \hat{x}(y) p(y|x)\,\mathrm{d}y = x + b(x). \quad (18)$$

Taking the derivative of both sides with respect to x_i (ith element of x), we get

$$\int \hat{x}(y) \nabla_{x_i} p(y|x)\,\mathrm{d}y = e_i + \nabla_{x_i} b(x) \quad (19)$$

which is equivalent to

$$\int \hat{x}(y) \nabla_{x_i} \log p(y|x) p(y|x)\,\mathrm{d}y = e_i + \nabla_{x_i} b(x) \quad (20)$$

for $i = 1, \ldots, n$ where e_i is a vector of all zeros except the ith element which is unity. We can write (20) for $i = 1, \ldots, n$ in a single matrix equation given as

$$\int \hat{x}(y) \left[\nabla_x \log p(y|x)\right]^T p(y|x)\,\mathrm{d}y = I_n + B(x) \quad (21)$$

where I_n is an identity matrix of size $n \times n$ and $B(x)$ is the bias Jacobian matrix. Since (17) is satisfied, we have

$$x \int \left[\nabla_x \log p(y|x)\right]^T p(y|x)\,\mathrm{d}y = 0_n. \quad (22)$$

where 0_n is a matrix of zeros with size $n \times n$. Subtracting both sides of (22) from those of (21), we get

$$\int \left(\hat{x}(y) - x\right) \left[\nabla_x \log(y|x)\right]^T p(y|x)\,\mathrm{d}y = I_n + B(x). \quad (23)$$

We can write (23) as

$$\int \left(\hat{x}(y) - x\right) s^T(x, y)\, p_0(y|x)\,\mathrm{d}y = I_n + B(x), \quad (24)$$

with score function $s(x, y)$ as introduced in (5). In order to invoke the Cauchy Schwarz inequality we multiply both sides by a^T and b from the left and the right respectively to get

$$\int a^T \left(\hat{x}(y) - x\right) s^T(x, y)\, b\, p_0(y|x)\,\mathrm{d}y = a^T (I_n + B(x)) b. \quad (25)$$

Now invoking the Cauchy Schwarz inequality under the inner product given as

$$\langle f(\cdot), g(\cdot)\rangle \triangleq \int f(y) g(y) p_0(y|x)\,\mathrm{d}y \quad (26)$$

for two functions $f(\cdot), g(\cdot)$, we obtain

$$\int a^T \left(\hat{x}(y) - x\right) \left(\hat{x}(y) - x\right)^T a\, p_0(y|x)\,\mathrm{d}y \times \int b^T s(x, y) s^T(x, y)\, b\, p_0(y|x)\,\mathrm{d}y \geq (a^T (I_n + B(x)) b)^2, \quad (27)$$

which is equivalent to

$$a^T P a \geq \frac{(a^T (I_n + B(x)) b)^2}{b^T J_{\mathrm{MM}}(x)\, b}, \quad (28)$$

where P and $J_{\mathrm{MM}}(x)$ are defined as in (2) and (4). Since b is arbitrary, we can choose it as $b = J_{\mathrm{MM}}^{-1}(x) \cdot (I_n + B(x))^T a$, to give

$$a^T P a \geq \frac{(a^T (I_n + B(x)) J_{\mathrm{MM}}^{-1}(x) (I_n + B(x))^T a)^2}{a^T (I_n + B(x)) J_{MM}^{-1}(x) (I_n + B(x))^T a} = a^T \left(I_n + B(x)\right) J_{\mathrm{MM}}^{-1}(x) (I_n + B(x))^T a \quad (29)$$

Since the inequality (29) holds for arbitrary vectors a, the expression given in (8) holds (and (3) holds when $B(x) = 0$), which concludes our proof of Theorem 1 and Theorem 2. □

5.2. Proof of Proposition 1

The equality for the Cauchy-Schwarz inequality used in the derivation of the CRLB under model mismatch is obtained if and only if

$$a^T (\hat{x}(y) - x) = c(x) b^T s(x, y) \quad \forall y, \quad (30)$$

where $c(x)$ is a scalar which may depend on x but not on y. Since the selection $b = J_{MM}^{-1}(x) a$ is made, we have equality if and only if

$$a^T (\hat{x}(y) - x) = c(x) a^T J_{MM}^{-1}(x) s(x, y). \quad (31)$$

Since a is arbitrary, the equality is achieved if and only if

$$(\hat{x}(y) - x) = c(x) J_{MM}^{-1}(x) s(x, y). \quad (32)$$

We multiply both sides of the equation above by $s^T(x, y)$ from the right to obtain

$$(\hat{x}(y) - x) s^T(x, y) = c(x) J_{MM}^{-1}(x) s(x, y) s^T(x, y). \quad (33)$$

Taking expected value of both sides w.r.t. to the true model, we get

$$E_{p_0(y|x)} \left[(\hat{x}(y) - x) s^T(x, y)\right] = c(x) J_{MM}^{-1}(x) J_{MM}(x)$$
$$I_n = c(x) I_n, \quad (34)$$

where the second equality follows from the fact that (24) holds with $B(x) = 0$. Hence $c(x) = 1$ which, when substituted into (32), completes the proof. □

6. REFERENCES

[1] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, vol. 3, pp. 81–91, 1945.

[2] H. Cramér, "A contribution to the theory of statistical estimation," *Skand. Akt. Tidskr.*, vol. 29, pp. 85–94, 1946.

[3] Harry L. van Trees, *Detection, Estimation and Modulation Theory Part I*, John Wiley & Sons, New York, NY, USA, 1968.

[4] L.L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley Publishing Company, Boston, MA, USA, 1991.

[5] Steven M. Kay, *Fundamentals of statistical signal processing: Estimation theory*, Prentice-Hall, Upper Saddle River, NJ, USA, 1st edition, 1993.

[6] A. N. D'Andrea, U. Mengali, and R. Reggiannini, "The modified Cramér-Rao bound and its application to synchronization problems," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 1391–1399, Feb 1994.

[7] Philippe Forster and P. Larzabal, "On lower bounds for deterministic parameter estimation," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, May 2002, vol. 2, pp. II–1137–II–1140.

[8] H. L. van Trees and K. L. Bell, Eds., *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*, Wiley-IEEE Press, Piscataway, NJ, USA, 2007.

[9] Wen Xu, A. B. Baggeroer, and K. L. Bell, "A bound on mean-square estimation error with background parameter mismatch," *IEEE Trans. Inf. Theory*, vol. 50, no. 4, pp. 621–632, April 2004.

[10] Christ D. Richmond and Larry L. Horowitz, "Parameter bounds under misspecified models," in *2013 Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov 2013, pp. 176–180.