A NUMERICAL METHOD TO COMPUTE CRAMÉR-RAO-TYPE BOUNDS FOR CHALLENGING ESTIMATION PROBLEMS

Justin Dauwels[*]

RIKEN Brain Science Institute
Hirosawa 2-1, Wako-shi, Saitama, Japan
Email: justin@dauwels.com

Sascha Korl

Phonak AG
CH-8712 Staefa, Switzerland
Email: sascha.korl@phonak.ch

ABSTRACT

A numerical algorithm is proposed to compute Cramér-Rao-type bounds. The Cramér-Rao-type bounds are derived from information matrices of *marginals* of the joint pdf of the system at hand. The key ingredient is message-passing on a factor graph of the system. The method can be applied to a wide class of estimation problems. As an illustration, the problem of estimating the parameters of an AR model is considered.

1. INTRODUCTION

For many practical estimation problems, popular estimators such as the maximum likelihood estimator (ML), the maximum a posteriori estimator (MAP) or the minimum mean squared error estimator (MMSE) are infeasible. Therefore, one often needs to resort to suboptimal techniques.

Suboptimal estimators are typically compared based on their mean squared estimation error (MSE). However, the MSE is not an *absolute* performance measure; in order to determine whether a suboptimal algorithm is close to optimal (in terms of MSE), the MSE of the minimum mean squared error (MMSE) estimator is required. Unfortunately, the minimum achievable MSE can often not be computed (neither analytically, nor numerically), and one needs to resort to *bounds* on the minimum achievable MSE, typically *lower* bounds. A well-known family of such lower bounds are the Cramér-Rao-type bounds.

For the estimation of *parameters*, a commonly used lower bound for the MSE is the Cramér-Rao bound (CRB), given by the inverse of the Fisher information matrix [1] [2] ("standard CRB"). The CRB has been computed in a wide variety of contexts, ranging from communications (e.g., [3]), to signal processing (e.g., [4]) and beyond.

For some applications, a closed-form expression for the CRB is available; in other applications, e.g., estimation in AR(MA) models, the derivation of CRBs is involved. For example, the CRB has been derived for AR(MA)-models *without* observation noise [4], but for AR(MA)-models *with* observation noise, the CRB seems to be intractable.

Van Trees derived an analogous bound to the CRB for *random variables*, referred to as "Bayesian CRB" (BCRB) or "posterior CRB" or "Van Trees bound" [1]. Rather surprisingly, far less attention has been given to the BCRB than to the standard CRB. Tichavský et al. derived the BCRB for filtering in state-space models with *freely* evolving state [5]. In recent work [6], we have derived a message-passing algorithm to compute BCRBs for general estimation problems.

In addition, so-called *hybrid* Cramér-Rao bounds have been proposed [7]; they apply to the joint estimation of parameters and random variables.

There are two general strategies to obtain Cramér-Rao-type bounds for a given estimation problem. One may derive Cramér-Rao-type bounds from the information matrix of the *joint* probability density function (pdf) of the system at hand; alternatively, one may derive such bounds from information matrices of *marginals* of the joint pdf.

The information matrix of the joint pdf is often sparse since the probability function at hand usually has structure, i.e., the probability function often factors. In [6], we have shown how this sparseness can be exploited to compute Cramér-Rao-type bounds.

In this paper, we propose an algorithm to compute Cramér-Rao-type bounds derived from information matrices of marginals of the joint pdf. It has been shown in [8] that Cramér-Rao-type bounds obtained from information matrices of marginals are *tighter* than the corresponding bounds derived from the information matrix of the joint pdf. Note that the information matrix of a marginal is usually dense, in contrast to the information matrix of the joint pdf. The proposed algorithm is applicable to standard CRBs, BCRBs, and hybrid CRBs.

This paper is structured as follows. First, we review Cramér-Rao-type bounds. Then, we present our algorithm to compute Cramér-Rao-type bounds. As an illustration, we consider estimation in state-space models. At the end of the paper, we provide a numerical example.

---

[*]The authors performed the work while at the Dept. of Information Technology and Electrical Engineering, ETH, CH-8092 Zürich, Switzerland. This project was supported in part by the Swiss National Science Foundation grant 200021-101955.

2. REVIEW OF CRAMÉR-RAO-TYPE BOUNDS

We briefly review the standard CRB and the BCRB. We refer to [7] for more information on hybrid CRBs.

2.1. Standard Cramér-Rao bound

We start by introducing our notation. Let $\Theta = (\Theta_1, \ldots, \Theta_n)^T$ be a parameter vector, and let $Y = (Y_1, \ldots, Y_N)^T$ be a real random vector (the extension to complex random vectors is straightforward). Suppose that $p(y|\theta)$ is the probability density function (pdf) of Y, which is parameterized by Θ. We consider the problem of estimating Θ from an observation vector $y = (y_1, \ldots, y_N)^T$. Let the function $\hat{\theta}(y)$ be an estimator of Θ based on the observation y. We define the error matrix $\mathbf{E}(\theta)$ as:

$$\mathbf{E}(\theta) \triangleq \mathrm{E}_{Y|\Theta}\big[(\hat{\theta}(Y) - \theta)(\hat{\theta}(Y) - \theta)^T\big]. \tag{1}$$

The Fisher information matrix $\mathbf{F}(\theta)$ is given by [1][2]:

$$\mathbf{F}_{ij}(\theta) \triangleq \mathrm{E}_{Y|\Theta}\left[\nabla_{\theta_i} \log p(Y|\theta) \nabla^T_{\theta_j} \log p(Y|\theta)\right], \tag{2}$$

where $\mathbf{F}_{ij}(\theta)$ is the (i,j)-th element of $\mathbf{F}(\theta)$. Note that $\mathbf{F}_{ij}$ is a matrix, since the components Θ_k are in general vectors.
If the estimator $\hat{\theta}(y)$ is unbiased, and the estimation problem is regular, the inverse of the Fisher information matrix is a lower bound on the error matrix $\mathbf{E}(\theta)$ (see, e.g., [1, pp. 66–67], [2, pp. 301–303]):

$$\mathbf{E}(\theta) \succeq \mathbf{F}(\theta)^{-1}. \tag{3}$$

Note that for many practical estimation problems, all estimators are necessarily biased. This is the case for example if Θ takes values in an interval $[a, b]$ or $[a, \infty)$, with $a, b \in \mathbb{R}$ and $a < b$. Nevertheless, the CRB (3) is a "high-SNR" bound for *any* regular estimator $\hat{\theta}(y)$, i.e., also for biased estimators $\hat{\theta}(y)$. As is well known, the ML estimator achieves the bound (3) at high SNR (under certain regularity conditions).

2.2. Bayesian Cramér-Rao bound

Let $X = (X_1, \ldots, X_n)^T$ and $Y = (Y_1, \ldots, Y_N)^T$, where X_k and Y_k are real random vectors (the extension to complex random vectors is straightforward). Suppose $p(x, y)$ is the joint pdf of X and Y. We consider the problem of estimating X from an observation vector $y = (y_1, \ldots, y_N)^T$. Let the function $\hat{x}(y)$ be an estimator of X based on the observation y. We define the error matrix $\mathbf{E}$ of the estimator $\hat{x}(y)$ as

$$\mathbf{E} \triangleq \mathrm{E}_{XY}[(\hat{x}(Y) - X)(\hat{x}(Y) - X)^T]. \tag{4}$$

The Bayesian information matrix $\mathbf{J}$ is given by

$$\mathbf{J}_{ij} \triangleq \mathrm{E}_{XY}\left[\nabla_{x_i} \log p(X, Y) \nabla^T_{x_j} \log p(X, Y)\right]. \tag{5}$$

Note that $\mathbf{J}_{ij}$ is a matrix, since the components X_k are in general vectors.
Van Trees proved a Cramér-Rao-type bound for random variables [1, pp. 72–73]:

$$\mathbf{E} \succeq \mathbf{J}^{-1}. \tag{6}$$

The inequality (6) is often referred to as the "Bayesian Cramér-Rao bound" (BCRB), "posterior CRB" or "Van Trees bound". It holds if the prior $p(x)$ is zero at the boundary of its support ("weak unbiasedness condition"), in addition to some regularity conditions. If the joint pdf $p(x, y)$ is Gaussian, the bound (6) holds with equality. Note that the BCRB also holds for *biased* estimators, in contrast to the CRB. The weak unbiasedness condition, however, is not necessarily fulfilled. On the other hand, the Bayesian Cramér-Rao bound (6) holds at high SNR for *any* regular joint pdf $p(x, y)$, i.e., also for a pdf $p(x, y)$ for which the weak unbiasedness condition is not met. In addition, the MAP estimator achieves the bound (6) at high SNR (under certain regularity conditions).

3. COMPUTING CRAMÉR-RAO-TYPE BOUNDS FROM MARGINALS

In practice, one is often interested in bounding the MSE for a particular variable X_k, i.e., for a particular *component* of the vector $X = (X_1, \ldots, X_n)$. For example, one may wish to compute a (standard unconditional) BCRB for the MSE:

$$\mathrm{E}_{X_k Y}[(\hat{x}_k(Y) - X_k)(\hat{x}_k(Y) - X_k)^T] \triangleq \mathbf{E}_{kk}, \tag{7}$$

which is the k-th diagonal element of the error matrix $\mathbf{E}$.
There are several ways to obtain a (standard unconditional) BCRB for (7). One may derive a (standard unconditional) BCRB from the information matrix of the *joint* pdf $p(x, y)$. For example, from the standard unconditional BCRB (6), it follows:

$$\mathbf{E}_{kk} \succeq [\mathbf{J}^{-1}]_{kk}, \tag{8}$$

where the unconditional Bayesian information matrix $\mathbf{J}$ is computed from the joint pdf $p(x, y)$.
Alternatively, instead of deriving the BCRB from the information matrix of $p(x, y)$ (cf. (8)), one may first *marginalize* over some variables X_ℓ ($\ell \neq k$), and compute the BCRB from the information matrix of the resulting *marginal* of $p(x, y)$.
Let us have a look at a simple example. Let $X \triangleq (X_1, X_2)^T$, and hence $p(x, y) \triangleq p(x_1, x_2, y)$. Suppose that we wish to obtain a BCRB for X_1. This can be done in two ways. One may compute the unconditional Bayesian information matrix of $p(x_1, y)$; the inverse of that matrix is a standard unconditional BCRB for X_1:

$$\begin{aligned} &\mathrm{E}_{X_1 Y}[(\hat{x}_1(Y) - X_1)(\hat{x}_1(Y) - X_1)^T] \\ &\qquad \succeq \mathrm{E}_{X_1 Y}\left[\nabla_{x_1} \nabla^T_{x_1} \log p(X_1, Y)\right]^{-1}, \end{aligned} \tag{9}$$

where:

$$p(x_1, y) \triangleq \int_{x_2} p(x_1, x_2, y) dx_2. \quad (10)$$

Alternatively, one may derive a standard unconditional BCRB from the unconditional Bayesian information matrix $\mathbf{J}$ of the joint pdf $p(x, y)$, which is a 2×2 block matrix. More precisely, the first diagonal element of the inverse of that matrix is a standard unconditional BCRB for X_1:

$$\mathrm{E}_{X_1 Y}[(\hat{x}_1(Y) - X_1)(\hat{x}_1(Y) - X_1)^T] \succeq \left[\mathbf{J}^{-1}\right]_{11}. \quad (11)$$

It has been proved in [8] that the tightest Bayesian Cramér-Rao bound for a variable X_k is obtained by first marginalizing over all variables X_ℓ ($\ell \neq k$), and by then computing the inverse information matrix of the resulting marginal $p(x_k, y)$ (or $p(x_k|y)$). For instance, the bound (9) is tighter than (11). It is typically easier, however, to derive Cramér-Rao bounds from the joint pdf (as in (11)) than from a marginal (as in (9)). In [6], we proposed a message-passing method for computing Cramér-Rao-type bounds from *joint* pdfs. In the following, we present such methods for computing Cramér-Rao-type bounds from *marginal* pdfs.

4. MESSAGE-PASSING ALGORITHM

In this section, we will consider standard CRBs. The extension to BCRBs and hybrid CRBs is straightforward.

We consider a system consisting of hidden random variables X and parameters Θ with joint pdf $p(x, y|\theta)$. We wish to compute the standard Cramér-Rao bound for Θ:

$$\mathbf{E}(\theta) \triangleq \mathrm{E}_{Y|\Theta}\Big[(\hat{\theta}(Y) - \theta)(\hat{\theta}(Y) - \theta)^T\Big] \succeq \mathbf{F}^{-1}(\theta) \triangleq \mathrm{E}_{Y|\Theta}\big[\nabla_\theta \log p(Y|\theta) \nabla_\theta^T \log p(Y|\theta)\big]^{-1}, \quad (12)$$

where

$$p(y|\theta) \triangleq \int_x p(x, y|\theta) dx. \quad (13)$$

The following lemma paves the way to a numerical algorithm for computing the bound (12).

Lemma 1 *If the integral $\int_x p(x, y|\theta)dx$ is differentiable under the integral sign (w.r.t. θ), then*

$$\nabla_\theta \log p(Y|\theta) = E_{X|\Theta Y}[\nabla_\theta \log p(X, Y|\theta)]. \quad (14)$$

An equality similar to (14) has been proved earlier in the context of code-aided synchronization [3]. The equality (14) is easily extended to standard CRBs and hybrid CRBs.

The expression in the RHS of (14) is usually as difficult to evaluate (analytically) as the expression in the LHS; in fact, both expressions are often intractable. One may then resort to *numerical* methods; the expression in the RHS of (14) suggests the following (numerical) algorithm to determine the bound (12):

1. Generate a list of samples $\{\hat{y}^{(j)}\}_{j=1}^N$ from $p(y|\theta)$.
2. Evaluate the expression:
$$\mathrm{E}_{X|\Theta Y}\Big[\nabla_\theta \log p(X, \hat{y}^{(j)}|\theta)\Big] \quad \text{for } j = 1, \ldots, N. \quad (15)$$
3. Compute the matrix $\hat{\mathbf{F}}^{(N)}(\theta)$:
$$\hat{\mathbf{F}}^{(N)}(\theta) \triangleq \frac{1}{N}\sum_{j=1}^N \Big[\mathrm{E}_{X|\Theta Y}\Big[\nabla_\theta \log p(X, \hat{y}^{(j)}|\theta)\Big] \cdot \mathrm{E}_{X|\Theta Y}\Big[\nabla_\theta \log p(X, \hat{y}^{(j)}|\theta)\Big]^T\Big]. \quad (16)$$

Eventually, we replace the Fisher information matrix $\mathbf{F}(\theta)$ in (12) by the approximation $\hat{\mathbf{F}}^{(N)}(\theta)$:

$$\mathbf{E}(\theta) \triangleq \mathrm{E}_{Y|\Theta}\Big[(\hat{\theta}(Y) - \theta)(\hat{\theta}(Y) - \theta)^T\Big] \succeq \Big[\hat{\mathbf{F}}^{(N)}(\theta)\Big]^{-1}. \quad (17)$$

Note that it is usually easy to sample from $p(y|\theta)$. The expression (15) can (in principle) be computed by sum(integral)-product message passing [9] on a cycle-free factor graph of $p(x, y|\theta)$. It sometimes results in closed-form expression (see Section 6). However, if the resulting expression (15) is intractable, one may use Monte-Carlo methods.

5. STATE-SPACE MODEL

As we pointed out, the expression (15) can be determined by sum-product message passing [9]. As an illustration, we consider briefly here estimation in general parameterized state-space models, which are ubiquitous in signal processing. We will assume for simplicity that the parameters are constant and that no prior is defined for the parameters. Our considerations are readily extended, however, to time-varying parameters and parameters with priors. The pdf of such a state-space model has the form:

$$p(x, y|\theta) \triangleq p_0(x_0) \prod_{k=1}^N p(x_k|x_{k-1}, \theta) p(y_k|x_k), \quad (18)$$

which is shown in Fig. 1. We can rewrite (15) as:

$$\mathrm{E}_{X|\Theta Y}\left[\nabla_\theta \log p(X, Y|\Theta)\right] = \sum_{k=1}^N \mathrm{E}_{X|\Theta Y}\left[\nabla_\theta \log p(X_k|X_{k-1}, \theta)\right]. \quad (19)$$

The expression $\mathrm{E}_{X|\Theta Y}\left[\nabla_\theta \log p(X_k|X_{k-1}, \theta)\right]$ in the RHS of (19) can be computed as:

$$\mathrm{E}_{X|\Theta Y}\left[\nabla_\theta \log p(X_k|X_{k-1}, \theta)\right] = \int_{x_{k-1}, x_k} p(x_k, x_{k-1}|\theta, y) \nabla_\theta \log p(x_k|x_{k-1}, \theta) dx_{k-1} dx_k, \quad (20)$$

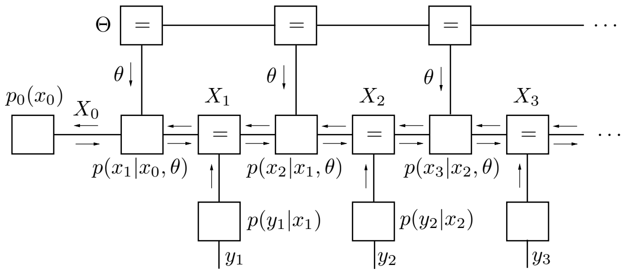

Fig. 1. Computation of (15) by message passing.

where the joint pdf $p(x_k, x_{k-1}|\theta, y)$ may be determined by the sum-product algorithm, as depicted in Fig. 1:

$$p(x_k, x_{k-1}|\theta, y) \propto \mu_{X_{k-1}\to p_k}(x_{k-1}) \\ \mu_{X_k\to p_k}(x_k)p(x_k|x_{k-1},\theta). \quad (21)$$

6. NUMERICAL EXAMPLE

We consider the following problem. Let $X_1, X_2, \ldots$ be a real random process ("auto-regressive (AR) model") defined by:

$$X_k = a_1X_{n-1} + a_2X_{k-2} + \cdots + a_MX_{k-M} + U_k, \quad (22)$$

where $a_1, \ldots, a_M$ are unknown real parameters, and $U_1, U_2, \ldots$ are real zero-mean Gaussian random variables with variance σ_U^2. We observe the real random variable Y_k given by:

$$Y_k = X_k + W_k, \quad (23)$$

where W_k is (real) zero-mean white Gaussian noise with variance σ_W^2. From the observation $y = (y_1, \ldots, y_N)$, one wishes to jointly estimate the coefficients $a_1, \ldots, a_M$, and the variances σ_U^2 and σ_W^2.

In this situation, the sum-product message passing depicted in Fig. 1 reduces to Kalman smoothing [9]; the expression (15) is available in closed-form.

We have computed the CRB for the estimation of a, σ_U^2 and σ_W^2 by means of the algorithm of Section 4 and 5. Fig. 2 shows the CRBs for the coefficients a together with the MSE of practical algorithms. From this figure, one can see that the CRBs are tight, especially as $N > 200$. It can also be seen that if the variances σ_U^2 and σ_W^2 are unknown, it becomes harder to estimate the coefficients.

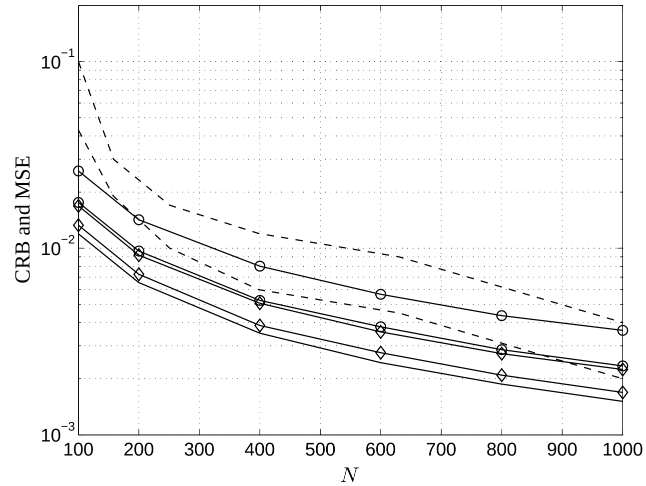

Fig. 2. CRB for coefficient estimation. Shown are (1) LPC-algorithm with known $\sigma_U^2 = 0.1$ and $\sigma_W^2 = 0$ (solid line); (2) Grid-based algorithms of [10] with unknown $\sigma_U^2 = 0.1$ and unknown $\sigma_W^2 = 0.01$, and 0.001 (dashed); (3) CRB for known/unknown $\sigma_U^2 = 0.1$ and known/unknown $\sigma_W^2 = 0.01$ (circles), 0.001 (diamands).

7. REFERENCES

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Part I,* Wiley, New York, 1968.

[2] M. J. Schervish, *Theory of Statistics,* Springer Verlag, 1995.

[3] N. Noels, H. Steendam, and M. Moeneclaey, "Carrier and Clock Recovery in (Turbo) Coded Systems: Cramér-Rao Bound and Synchronizer Performance," *EURASIP Journal on Applied Signal Processing, JASP, Special issue on Turbo Processing,* vol. 2005, pp. 972–980, May 2005.

[4] B. Friedlander and B. Porat, "The Exact Cramér-Rao Bound for Gaussian Autoregressive Processes," *IEEE Transactions on Aerospace and Electronic Systems,* vol. 25, Jan. 1989, pp. 3–7.

[5] P. Tichavský, C. H. Muravchik, and A. Nehorai, "Posterior Cramér-Rao Bounds for Discrete-Time Non-Linear Filtering," *IEEE Transactions on Signal Processing,* vol. 46, May 1998, pp. 1386–1396.

[6] J. Dauwels, "Computing Bayesian Cramér-Rao Bounds," *Proc. 2005 IEEE Int. Symp. Information Theory,* to appear.

[7] Y. Rockah, and P. Schultheiss, "Array Shape Calibration Using Sources in Unknown Locations. I. Far-Field Sources," IEEE Trans. Acoust., Speech, and Signal Proc., vol. 35, no. 3, pp. 286–299, 1987.

[8] B. Bobrovsky, E. Mayer-Wolf, and M. Zakai, "Some Classes of Global Cramér-Rao Bounds," *A. Statistics,* vol. 15, pp. 1421-1438, 1987.

[9] H.-A. Loeliger, "An Introduction to Factor Graphs," *IEEE Signal Processing Magazine*, Jan. 2004, pp. 28–41.

[10] S. Korl, H.-A. Loeliger, A.-G. Lindgren, "AR Model Parameter Estimation: From Factor Graphs to Algorithms," *Proc. ICASSP'04,* vol. 5, Montreal, Canada, May 17–21, 2004, pp. 509–512.