A Fresh Look at Bayesian Cramér-Rao Bounds for Discrete-Time Nonlinear Filtering

Carsten Fritsche†, Emre Özkan, Lennart Svensson*, and Fredrik Gustafsson
Department of Electrical Engineering, Linköping University,SE-581 83 Linköping, Sweden
*Department of Signals and Systems, Chalmers University of Technology, SE-412 96 Göteborg, Sweden
† IFEN GmbH, 85586 Poing, Germany

***Abstract*—In this paper, we aim to relate different Bayesian Cramér-Rao bounds which appear in the discrete-time nonlinear filtering literature in a single framework. A comparative theoretical analysis of the bounds is provided in order to relate their tightness. The results can be used to provide a lower bound on the mean square error in nonlinear filtering. The findings are illustrated and verified by numerical experiments where the tightness of the bounds are compared.**

I. Introduction

The Cramér-Rao Bound (CRB) has become one of the most popular tools to provide a lower bound on estimation performance. For a vector of non-random (deterministic) parameters, it is given by the inverse of the Fisher information matrix and it can be used to provide a lower bound on the mean square estimation error (MSE) matrix of any unbiased estimator [1]. For a random parameter, Van Trees presented an analogous bound [2], which is often referred to in the literature as the Bayesian CRB (BCRB), posterior CRB or the Van Trees bound. The BCRB for a vector of random variables is defined as the inverse of the Bayesian Information matrix and it generally provides a lower bound on the MSE matrix of any estimator [2], [3].

In discrete-time nonlinear filtering the parameter vector of interest, which is modeled as randomly evolving over time and is also known as state vector, is estimated from a sequence of measurements available up to the current time. In [4], Tichavský et al. presented an elegant approach to recursively compute the BCRB for the general nonlinear filtering problem. During the same period of time, Bergman developed independently similar results, but which are applicable to a larger class of nonlinear models [5]. The BCRBs proposed therein explore the information contained in the entire state and measurement sequence up to the current time. This is reflected in building up the Bayesian information matrix of the joint (unconditional) distribution of the measurements and states, from which the lower bound for estimating the current state (i.e. the nonlinear filtering bound) can be found by extracting the lower-right submatrix of the inverse of the Bayesian information matrix [4], [5]. The solution proposed by Tichavský et al. and Bergman can be considered as the state-of-the-art for computing the BCRB in a nonlinear filtering context.

Recently, the concept of conditional BCRB for discrete-time nonlinear filtering was introduced [6], which can be employed especially in adaptive sensor management applications. The idea of the conditional BCRB is to evaluate the Bayesian information matrix of a joint distribution of the state sequence and the current measurement, but conditioned on the past measurement sequence. As a result, the conditional BCRB gives a bound on the conditional MSE matrix which is different to the BCRB of [4], [5], which holds for the unconditional MSE matrix. In [7], the concept of conditional BCRB has been further extended by introducing alternative approximations to compute the bound introduced in [6], and by introducing a new type of conditional BCRB which is based on evaluating the Bayesian information matrix of the marginal distribution of the current state and current measurement, conditioned on the past measurement sequence. Even though the authors of [6], [7] relate their work to existing approaches available in the literature, the relation of the proposed conditional BCRBs in terms of tightness among each other and with respect to the BCRB of [4], [5] for the unconditional MSE matrix is missing.

An early attempt to relate different versions of BCRBs (but not in the nonlinear filtering context) was performed by the excellent work of Bobrovsky et. al [8]. The ideas presented therein were picked-up again in the book by Van Trees but again in a more general context [3]. In [9], different versions of BCRBs were explored and their computation in graphical models via factor graphs and message passing algorithms were suggested.

The aim of this paper is to relate different versions of BCRBs appearing in the nonlinear filtering literature to each other in a single framework. In total, four different versions of BCRBs are identified to provide a lower bound on the unconditional MSE of nonlinear filtering. The relation among the different BCRBs in terms of tightness is assessed theoretically, and it is shown that in the special case of linear Gaussian systems all bounds coincide. The theoretical findings are then verifed in numerical examples to illustrate the tightness of the different bounds.

II. Problem Statement

Consider the following discrete-time nonlinear system

$$\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}, \mathbf{v}_k), \tag{1a}$$

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{w}_k), \tag{1b}$$

where $\mathbf{z}_k \in \mathbb{R}^{n_z}$ is the measurement vector at discrete time k, $\mathbf{x}_k \in \mathbb{R}^{n_x}$ is the state vector and $\mathbf{f}_k(\cdot)$ and $\mathbf{h}_k(\cdot)$ are arbitrary nonlinear mappings of appropriate dimensions. The noise vectors $\mathbf{v}_k \in \mathbb{R}^{n_v}$, $\mathbf{w}_k \in \mathbb{R}^{n_w}$ and the initial state $\mathbf{x}_0$ are assumed mutually independent white processes with arbitrary but known probability density functions (pdfs). We further introduce $\mathbf{X}_k = [\mathbf{x}_0^\mathsf{T}, \ldots, \mathbf{x}_k^\mathsf{T}]^\mathsf{T}$ and $\mathbf{Z}_k = [\mathbf{z}_1^\mathsf{T}, \ldots, \mathbf{z}_k^\mathsf{T}]^\mathsf{T}$ which denote the collection of augmented states and measurement vectors up to time k, and where T stands for matrix transpose. In nonlinear filtering, one is interested in estimating the current state $\mathbf{x}_k$ from the sequence of available noisy measurements $\mathbf{Z}_k$. The corresponding estimator is denoted as $\hat{\mathbf{x}}_k(\mathbf{Z}_k)$, which is a function of the measurement sequence $\mathbf{Z}_k$. The performance of any estimator $\hat{\mathbf{x}}_k(\mathbf{Z}_k)$ is commonly measured by the mean-square error (MSE) matrix,

$$\boldsymbol{\mathcal{M}}(\hat{\mathbf{x}}_k) = \mathbb{E}_{p(\mathbf{x}_k,\mathbf{Z}_k)}\left\{(\hat{\mathbf{x}}_k(\mathbf{Z}_k) - \mathbf{x}_k)(\cdot)^\mathsf{T}\right\}, \tag{2}$$

where $\mathbb{E}_{p(\mathbf{x}_k,\mathbf{Z}_k)}\{\cdot\}$ denotes expectation with respect to the joint density $p(\mathbf{x}_k, \mathbf{Z}_k)$. The minimum MSE (MMSE) estimator is

$$\hat{\mathbf{x}}^{\text{MSE}} \triangleq \hat{\mathbf{x}}_k^{\text{MSE}}(\mathbf{Z}_k) \triangleq \mathbb{E}_{p(\mathbf{x}_k|\mathbf{Z}_k)}\{\mathbf{x}_k\}, \tag{3}$$

where $p(\mathbf{x}_k|\mathbf{Z}_k)$ denotes the filtering density. The corresponding MSE matrix $\boldsymbol{\mathcal{M}}(\hat{\mathbf{x}}^{\text{MSE}})$ represents the optimal performance and thus gives the tightest lower bound on the performance of any estimator.

For discrete-time linear systems with additive Gaussian noise, the MMSE estimator is given by the celebrated Kalman filter and the (minimum) MSE matrix is equivalent to the covariance matrix of the Kalman filter. For nonlinear systems defined as in (1), closed-form expressions for the MMSE estimator and its MSE matrix generally do not exist. In this case, it is rather difficult to evaluate the optimal performance bound and one has to resort to other techniques providing a lower bound on the MSE matrix. In the literature, different types of Bayesian bounds have been proposed for lower bounding the MSE matrix, see [3] for an excellent overview. For nonlinear filtering, however, the BCRB is identified as the perhaps most popular tool to provide a lower bound on the performance of any estimator. In this paper, we investigate how to provide a lower bound for the MSE matrix $\boldsymbol{\mathcal{M}}(\hat{\mathbf{x}}_k)$ by different versions of the BCRB.

III. Different Versions of the BCRB

A. The joint unconditional BCRB

The idea of the joint unconditional BCRB presented in [4], [5] is to provide a lower bound on the MSE matrix of the sequence of states $\mathbf{X}_k$. Let $\hat{\mathbf{X}}_k(\mathbf{Z}_k) = [\hat{\mathbf{x}}_0^\mathsf{T}(\mathbf{Z}_k), \ldots, \hat{\mathbf{x}}_k^\mathsf{T}(\mathbf{Z}_k)]^\mathsf{T}$ denote the collection of estimators up to time k based on the measurement sequence $\mathbf{Z}_k$ and a-priori known initial pdf $p(\mathbf{x}_0)$. Then, the MSE matrix for estimating the state sequence $\mathbf{X}_k$ can be bounded as follows

$$\begin{aligned}\boldsymbol{\mathcal{M}}(\hat{\mathbf{X}}_k) &= \mathbb{E}_{p(\mathbf{X}_k,\mathbf{Z}_k)}\left\{(\hat{\mathbf{X}}_k(\mathbf{Z}_k) - \mathbf{X}_k)(\cdot)^\mathsf{T}\right\} \\ &\geq [\mathbf{J}_{0:k}]^{-1}, \end{aligned}\tag{4}$$

where the matrix inequality $\mathbf{A} \geq \mathbf{C}$ means that the difference $\mathbf{A} - \mathbf{C}$ is a positive semi-definite matrix. The matrix $\mathbf{J}_{0:k}$ is known as Bayesian information matrix of the state sequence $\mathbf{X}_k$ and its inverse gives the BCRB for estimating $\mathbf{X}_k$. Let us introduce the gradient and Laplace operators

$$\nabla_{\mathbf{s}} = \left[\frac{\partial}{\partial s_1}, \ldots, \frac{\partial}{\partial s_n}\right]^\mathsf{T}, \tag{5}$$

$$\triangle_{\mathbf{s}}^{\mathbf{t}} = \nabla_{\mathbf{s}}\nabla_{\mathbf{t}}^\mathsf{T}, \tag{6}$$

for any vectors $\mathbf{s}$ and $\mathbf{t}$. Then, $\mathbf{J}_{0:k}$ can be expressed as

$$\mathbf{J}_{0:k} = \mathbb{E}_{p(\mathbf{x}_k,\mathbf{Z}_k)}\left\{-\triangle_{\mathbf{X}_k}^{\mathbf{X}_k}\log p(\mathbf{X}_k, \mathbf{Z}_k)\right\}. \tag{7}$$

The MSE matrix for estimating the state $\mathbf{x}_k$ can be found by taking the $(n_x \times n_x)$ lower-right submatrix of $\boldsymbol{\mathcal{M}}(\hat{\mathbf{X}}_k)$. This can be expressed mathematically by introducing a mapping matrix

$$\mathbf{U} = [\mathbf{0}, \cdots, \mathbf{0}, \mathbf{I}_{n_x}], \tag{8}$$

such that

$$\begin{aligned}\boldsymbol{\mathcal{M}}(\hat{\mathbf{x}}_k) &= \mathbf{U}\,\boldsymbol{\mathcal{M}}(\hat{\mathbf{X}}_k)\,\mathbf{U}^\mathsf{T} \\ &\geq \mathbf{U}\,[\mathbf{J}_{0:k}]^{-1}\,\mathbf{U}^\mathsf{T} \triangleq [\tilde{\mathbf{J}}_k]^{-1} = \mathbf{B}_1, \end{aligned}\tag{9}$$

holds, where $\mathbf{I}_{n_x}$ is the $(n_x \times n_x)$ identity matrix and $\mathbf{0}$ is a matrix of zeros of appropriate size. A recursive formula for computing $\tilde{\mathbf{J}}_k$, which does not require the inversion of large matrices such as $\mathbf{J}_{0:k}$, has been derived in Tichavský et al. and Bergman [4], [5]. In the following, $\mathbf{B}_1$ is referred to as the joint unconditional BCRB (JU-BCRB), since its computation is based on the evaluation of the Bayesian information matrix of the joint density $p(\mathbf{X}_k, \mathbf{Z}_k)$.

B. The marginal unconditional BCRB

Naturally, the marginal density

$$p(\mathbf{x}_k, \mathbf{Z}_k) = \int p(\mathbf{X}_k, \mathbf{Z}_k)\,\mathrm{d}\,\mathbf{X}_{k-1} \tag{10}$$

can also be used to define a lower bound. The resulting bound

$$\boldsymbol{\mathcal{M}}(\hat{\mathbf{x}}_k) \geq [\mathbf{J}_k]^{-1} = \mathbf{B}_2, \tag{11}$$

where

$$\mathbf{J}_k = \mathbb{E}_{p(\mathbf{x}_k,\mathbf{Z}_k)}\left\{-\triangle_{\mathbf{x}_k}^{\mathbf{x}_k}\log p(\mathbf{x}_k, \mathbf{Z}_k)\right\} \tag{12}$$

is here called the marginal unconditional BCRB (MU-BCRB). Bobrovsky et al. showed that the BCRB derived from the marginal density is always greater than or equal to the BCRB which is obtained from the joint density, see Proposition 1 in [8] for a proof. Thus, we can conclude that

$$\mathbf{B}_2 \geq \mathbf{B}_1 \tag{13}$$

must generally hold, i.e. the marginal unconditional BCRB is at least as tight as the joint unconditional BCRB. Note that a larger lower bound provides a stronger result, and 0 is always the smallest lower bound. This result is rather intuitive since the computation of the joint unconditional BCRB relies on evaluating the information contained in the whole state and measurement sequence, while the marginal unconditional BCRB extracts information only from the most recent state we are interested in, and the whole measurement sequence.

C. The joint conditional BCRB

Another class of BCRBs can be found by decomposing the measurement vector into two parts, e.g., as follows

$$\mathbf{Z}_k = \begin{bmatrix} \mathbf{z}_k \\ \mathbf{Z}_{k-1} \end{bmatrix}. \tag{14}$$

The MSE matrix of any estimator $\hat{\mathbf{X}}_k(\mathbf{Z}_k)$ can be decomposed accordingly, yielding

$$\boldsymbol{\mathcal{M}}(\hat{\mathbf{X}}_k) = \mathbb{E}_{p(\mathbf{Z}_{k-1})}\mathbb{E}_{p_c}\left\{(\hat{\mathbf{X}}_k(\mathbf{Z}_k) - \mathbf{X}_k)(\cdot)^{\mathsf{T}}\right\} \tag{15}$$

with $p_c = p(\mathbf{X}_k, \mathbf{z}_k|\mathbf{Z}_{k-1})$. The inner expectation in (15) is the conditional MSE matrix denoted as:

$$\boldsymbol{\mathcal{M}}(\hat{\mathbf{X}}_k\big|\mathbf{Z}_{k-1}) = \mathbb{E}_{p_c}\left\{(\hat{\mathbf{X}}_k(\mathbf{Z}_k) - \mathbf{X}_k)(\cdot)^{\mathsf{T}}\right\}. \tag{16}$$

Similar to the proof for the unconditional MSE matrix given in [2], it can be shown that

$$\boldsymbol{\mathcal{M}}(\hat{\mathbf{X}}_k\big|\mathbf{Z}_{k-1}) \geq [\mathbf{J}_{0:k}(\mathbf{Z}_{k-1})]^{-1} \tag{17}$$

holds, see [10] for details, where $\mathbf{J}_{0:k}(\mathbf{Z}_{k-1})$ is the joint conditional Bayesian information matrix given by

$$\mathbf{J}_{0:k}(\mathbf{Z}_{k-1}) = \mathbb{E}_{p_c}\left\{-\Delta^{\mathbf{X}_k}_{\mathbf{X}_k}\log p(\mathbf{X}_k, \mathbf{z}_k|\mathbf{Z}_{k-1})\right\}. \tag{18}$$

Thus, the relation in (15) can be further lower bounded by

$$\begin{aligned}\boldsymbol{\mathcal{M}}(\hat{\mathbf{X}}_k) &= \mathbb{E}_{p(\mathbf{Z}_{k-1})}\{\boldsymbol{\mathcal{M}}(\hat{\mathbf{X}}_k\big|\mathbf{Z}_{k-1})\} \\ &\geq \mathbb{E}_{p(\mathbf{Z}_{k-1})}\{[\mathbf{J}_{0:k}(\mathbf{Z}_{k-1})]^{-1}\}. \end{aligned}\tag{19}$$

A lower bound for $\boldsymbol{\mathcal{M}}(\hat{\mathbf{x}}_k)$ then can be finally computed as follows:

$$\begin{aligned}\boldsymbol{\mathcal{M}}(\hat{\mathbf{x}}_k) &= \mathbf{U}\boldsymbol{\mathcal{M}}(\hat{\mathbf{X}}_k)\mathbf{U}^{\mathsf{T}} \\ &\geq \mathbb{E}_{p(\mathbf{Z}_{k-1})}\left\{\mathbf{U}\left[\mathbf{J}_{0:k}(\mathbf{Z}_{k-1})\right]^{-1}\mathbf{U}^{\mathsf{T}}\right\} \\ &\triangleq \mathbb{E}_{p(\mathbf{Z}_{k-1})}\left\{[\tilde{\mathbf{J}}_k(\mathbf{Z}_{k-1})]^{-1}\right\} = \mathbf{B}_3, \end{aligned}\tag{20}$$

where $[\tilde{\mathbf{J}}_k(\mathbf{Z}_{k-1})]^{-1}$ gives a lower bound for the conditional MSE matrix $\boldsymbol{\mathcal{M}}(\hat{\mathbf{x}}_k|\mathbf{Z}_{k-1})$.

The important result of (20) is that averaging the lower bound of the conditional MSE matrix over the past measurement sequence $\mathbf{Z}_{k-1}$ yields a lower bound $\mathbf{B}_3$ for the unconditional MSE matrix of any estimator. It has been shown in [6] that a recursive computation of the quantity $[\tilde{\mathbf{J}}_k(\mathbf{Z}_{k-1})]^{-1}$ is not possible without introducing further approximations. Hence, in order to obtain exact results it is necessary to directly compute the inverse of the $(k+1)n_x \times (k+1)n_x$ matrix $\mathbf{J}_{0:k}(\mathbf{Z}_{k-1})$, which eventually becomes impractical as time k increases.

Even though the bound presented in (20) is an unconditional BCRB, it is termed hereinafter the joint conditional BCRB (JC-BCRB) in order to highlight its dependency on the evaluation of the joint conditional Bayesian information matrix of the density $p(\mathbf{X}_k, \mathbf{z}_k|\mathbf{Z}_{k-1})$. The joint conditional BCRB can be further related to the joint unconditional BCRB defined in Section III-A according to the following theorem.

Theorem 1. *Assuming suitable regularity conditions are fulfilled [11], it holds that*

$$\mathbf{B}_3 \geq \mathbf{B}_1, \tag{21}$$

Proof: See Appendix ∎

From Theorem 1 we learn that the joint conditional BCRB is at least as tight as the joint unconditional BCRB.

D. The marginal conditional BCRB

It is also possible to define the MSE matrix for estimating $\mathbf{x}_k$ with respect to the conditional MSE matrix according to

$$\begin{aligned}\boldsymbol{\mathcal{M}}(\hat{\mathbf{x}}_k) &= \mathbb{E}_{p(\mathbf{Z}_{k-1})}\mathbb{E}_{p_m}\left\{(\hat{\mathbf{x}}_k(\mathbf{Z}_k) - \mathbf{x}_k)(\cdot)^{\mathsf{T}}\right\} \\ &= \mathbb{E}_{p(\mathbf{Z}_{k-1})}\left\{\boldsymbol{\mathcal{M}}(\hat{\mathbf{x}}_k\big|\mathbf{Z}_{k-1})\right\}, \end{aligned}\tag{22}$$

where $p_m = p(\mathbf{x}_k, \mathbf{z}_k|\mathbf{Z}_{k-1})$ and the conditional MSE matrix is given by

$$\boldsymbol{\mathcal{M}}(\hat{\mathbf{x}}_k\big|\mathbf{Z}_{k-1}) = \mathbb{E}_{p_m}\left\{(\hat{\mathbf{x}}_k(\mathbf{Z}_k) - \mathbf{x}_k)(\cdot)^{\mathsf{T}}\right\}. \tag{23}$$

A lower bound on the conditional MSE matrix is then given as follows

$$\boldsymbol{\mathcal{M}}(\hat{\mathbf{x}}_k\big|\mathbf{Z}_{k-1}) \geq [\mathbf{J}_k(\mathbf{Z}_{k-1})]^{-1}, \tag{24}$$

where $\mathbf{J}_k(\mathbf{Z}_{k-1})$ is the marginal conditional Bayesian information matrix given by

$$\mathbf{J}_k(\mathbf{Z}_{k-1}) = \mathbb{E}_{p_m}\left\{-\Delta^{\mathbf{x}_k}_{\mathbf{x}_k}\log p(\mathbf{x}_k, \mathbf{z}_k|\mathbf{Z}_{k-1})\right\}, \tag{25}$$

with $p_m = p(\mathbf{x}_k, \mathbf{z}_k|\mathbf{Z}_{k-1})$. The bound in (24) was first proposed in [7]. However, its relation to the bound derived from (17) has been overlooked. Following again the argumentation of Bobrovsky et al. [8], the bounds for the conditional MSE matrices can be further related to each other according to

$$[\mathbf{J}_k(\mathbf{Z}_{k-1})]^{-1} \geq [\tilde{\mathbf{J}}_k(\mathbf{Z}_{k-1})]^{-1}. \tag{26}$$

In this paper, however, we are interested in BCRBs for the unconditional MSE matrix. A lower bound for the unconditional MSE matrix for any estimator $\hat{\mathbf{x}}_k(\mathbf{Z}_k)$ is given as follows:

$$\boldsymbol{\mathcal{M}}(\hat{\mathbf{x}}_k) \geq \mathbb{E}_{p(\mathbf{Z}_{k-1})}\{[\mathbf{J}_k(\mathbf{Z}_{k-1})]^{-1}\} = \mathbf{B}_4. \tag{27}$$

Again, averaging the lower bound for the conditional MSE matrix over the past measurement sequence $\mathbf{Z}_{k-1}$ yields a lower bound $\mathbf{B}_4$ for the unconditional MSE matrix of any estimator. The unconditional bound $\mathbf{B}_4$ is termed hereinafter the marginal conditional BCRB (MC-BCRB) in order to emphasize its dependency on the density $p(\mathbf{x}_k, \mathbf{z}_k|\mathbf{Z}_{k-1})$. From the inequality preservation property of expectations it finally follows that if (26) holds, then

$$\mathbf{B}_4 \geq \mathbf{B}_3 \tag{28}$$

must hold, i.e. the marginal conditional BCRB is at least as tight as the joint conditional BCRB. Lastly, it is possible to relate the marginal conditional BCRB to the marginal unconditional BCRB according to

$$\mathbf{B}_4 \geq \mathbf{B}_2. \tag{29}$$

The proof of this inequality is omitted here, but it can be easily checked that it follows from a slight modification of the proof for Theorem 1.

TABLE I
RELATIONSHIP BETWEEN THE BCRBS

Name, Eq.	Density of BIM	BIM	Bound
JU-BCRB, (9)	$p(\mathbf{X}_k, \mathbf{Z}_k)$	$\mathbf{J}_{0:k}$	$\mathbf{B}_1$
MU-BCRB, (11)	$p(\mathbf{x}_k, \mathbf{Z}_k)$	$\mathbf{J}_k$	$\mathbf{B}_2$
JC-BCRB, (20)	$p(\mathbf{X}_k, \mathbf{z}_k \vert \mathbf{Z}_{k-1})$	$\mathbf{J}_{0:k}(\mathbf{Z}_{k-1})$	$\mathbf{B}_3$
MC-BCRB, (27)	$p(\mathbf{x}_k, \mathbf{z}_k \vert \mathbf{Z}_{k-1})$	$\mathbf{J}_k(\mathbf{Z}_{k-1})$	$\mathbf{B}_4$

IV. RELATIONSHIP BETWEEN THE DIFFERENT BCRBS

A. Nonlinear Systems

In the previous section it was shown that different versions of BCRBs exist that all can be used for predicting the best achievable performance for nonlinear filtering. The BCRBs differ from each other in the amount of information they extract, which is resembled by the evaluation of different Bayesian information matrices. The amount of information extraction also determines the tightness of the bound, i.e. the capability to predict the best achievable nonlinear filtering performance. Generally, the bounds can be ordered in terms of tightness as follows

$$\mathbf{B}_4 \geq \mathbf{B}_2 \geq \mathbf{B}_1 \tag{30a}$$

and

$$\mathbf{B}_4 \geq \mathbf{B}_3 \geq \mathbf{B}_1. \tag{30b}$$

Thus, the BCRB proposed by Tichavský et al. and Bergman [4], [5], which can be considered as the state-of-the-art today, provides the least tight bound. This means, that the MSE predicted by this bound might be far away from the achievable MSE of the optimal filter. The three other bounds compared in this paper are all tighter, where the marginal conditional BCRB is the tightest bound. The most important properties of the different BCRBs are summarized in Table II.

B. Linear Additive Gaussian Systems

An important special case occurs if the underlying system is linear additive Gaussian, i.e.

$$\mathbf{x}_k = \mathbf{F}_k \cdot \mathbf{x}_{k-1} + \mathbf{v}_k, \tag{31a}$$

$$\mathbf{z}_k = \mathbf{H}_k \cdot \mathbf{x}_k + \mathbf{w}_k, \tag{31b}$$

where $\mathbf{F}_k$ and $\mathbf{H}_k$ are arbitrary linear mapping matrices of proper size, and where the noise densities are Gaussian distributed according to $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ and $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$. The pdf of the initial state is also Gaussian and given by $p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0; \mathbf{0}, \mathbf{P}_{0|0})$. For the system given by (31), the following theorem holds:

Theorem 2. *For linear additive Gaussian systems, the JU-BCRB, MC-BCRB, JC-BCRB and MC-BCRB are equal, i.e.*

$$\mathbf{B}_1 = \mathbf{B}_2 = \mathbf{B}_3 = \mathbf{B}_4 \tag{32}$$

holds.

Proof: See Appendix. ∎

V. NUMERICAL APPROXIMATION OF THE MU-BCRB

Algorithms for computing the information matrices $\tilde{\mathbf{J}}_k$, $\tilde{\mathbf{J}}_k(\mathbf{Z}_{k-1})$ and $\mathbf{J}_k(\mathbf{Z}_{k-1})$ have been developed in [4]–[7]. With these methods, it is relatively easy to compute the corresponding BCRBs $\mathbf{B}_1$, $\mathbf{B}_3$ and $\mathbf{B}_4$. In this section, we devise a method on how $\mathbf{J}_k$ and thus the MU-BCRB $\mathbf{B}_2$ can be computed.

In the following, we suggest a particle filter approximation to evaluate $\mathbf{J}_k$, see for instance [12]–[15] for an introduction to particle filters. The choice of this approach is originally inspired by [16] where they try to compute the mutual information from a particle filter. We take into account that the joint density can be decomposed as follows:

$$p(\mathbf{x}_k, \mathbf{Z}_k) = p(\mathbf{z}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{Z}_{k-1}) p(\mathbf{Z}_{k-1}) \tag{33}$$

Then, the information matrix $\mathbf{J}_k$ can be accordingly decomposed as:

$$\begin{aligned} \mathbf{J}_k &= \mathbb{E}_{p(\mathbf{x}_k,\mathbf{z}_k)} \left\{ -\triangle_{\mathbf{x}_k}^{\mathbf{x}_k} \log p(\mathbf{z}_k|\mathbf{x}_k) \right\} \\ &\quad + \mathbb{E}_{p(\mathbf{x}_k,\mathbf{Z}_{k-1})} \left\{ -\triangle_{\mathbf{x}_k}^{\mathbf{x}_k} \log p(\mathbf{x}_k|\mathbf{Z}_{k-1}) \right\} \\ &\triangleq \mathbf{J}_k^I + \mathbf{J}_k^{II}, \end{aligned} \tag{34}$$

where the term containing the pdf $p(\mathbf{Z}_{k-1})$ disappears as it does not depend on $\mathbf{x}_k$. The first term $\mathbf{J}_k^I$ can be easily approximated using Monte Carlo integration. In order to avoid the computation of the Hessian, it is more convenient to Monte Carlo approximate the following expression

$$\begin{aligned} \mathbf{J}_k^I &= \mathbb{E}_{p(\mathbf{x}_k,\mathbf{z}_k)} \left\{ \frac{[\nabla_{\mathbf{x}_k} p(\mathbf{z}_k|\mathbf{x}_k)][\cdot]^\mathsf{T}}{[p(\mathbf{z}_k|\mathbf{x}_k)]^2} \right\} \\ &\approx \frac{1}{N_{\text{mc}}} \sum_{l=1}^{N_{\text{mc}}} \left[\frac{[\nabla_{\mathbf{x}_k} p(\mathbf{z}_k^{(l)}|\mathbf{x}_k^{(l)})][\cdot]^\mathsf{T}}{[p(\mathbf{z}_k^{(l)}|\mathbf{x}_k^{(l)})]^2} \right], \end{aligned} \tag{35}$$

where $\mathbf{x}_k^{(l)}$, $\mathbf{z}_k^{(l)}$, $l = 1, \ldots, \mathbf{N}_{\text{mc}}$, are independent and identically distributed samples such that $(\mathbf{x}_k^{(l)}, \mathbf{z}_k^{(l)}) \sim p(\mathbf{x}_k, \mathbf{z}_k)$. The second term $\mathbf{J}_k^{II}$ is more difficult, since a closed-form representation of the prediction density $p(\mathbf{x}_k|\mathbf{Z}_{k-1})$ is generally not available for nonlinear non-Gaussian systems. The idea is to approximate this term using a particle filter. Assume that a particle filter approximation of the posterior density $p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1})$ at time step $k-1$ is available,

$$\hat{p}(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1}) = \sum_{j=1}^{N} w_{k-1}^{(j)}\, \delta(\mathbf{x}_{k-1} - \mathbf{x}_{k-1}^{(j)}), \tag{36}$$

with positive weights

$$w_{k-1}^{(j)} = \frac{p(\mathbf{x}_{k-1}^{(j)}|\mathbf{Z}_{k-1})}{q(\mathbf{x}_{k-1}^{(j)}|\mathbf{Z}_{k-1})}, \tag{37}$$

where $\delta(\cdot)$ is the Dirac delta function, $q(\mathbf{x}_{k-1}^{(j)}|\mathbf{Z}_{k-1})$ is the importance distribution and where $\sum_j w_{k-1}^{(j)} = 1$ holds. Then,

an approximation of the prediction density is given by

$$p(\mathbf{x}_k|\mathbf{Z}_{k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})\, p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1})\, \mathrm{d}\mathbf{x}_{k-1} \approx \sum_{j=1}^{N} w_{k-1}^{(j)}\, p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(j)}) \triangleq \hat{p}(\mathbf{x}_k|\mathbf{Z}_{k-1}). \quad (38)$$

As a result, the prediction density in the particle filter can be represented by a weighted mixture of transition densities, which has the appealing advantage that gradients can be easily computed. Reformulating $\mathbf{J}_k^{II}$ in terms of gradients and replacing the true density $p(\mathbf{x}_k|\mathbf{Z}_{k-1})$ with the corresponding particle filter approximation $\hat{p}(\mathbf{x}_k|\mathbf{Z}_{k-1})$, the term $\mathbf{J}_k^{II}$ can be finally approximated as

$$\mathbf{J}_k^{II} = \mathbb{E}_{p(\mathbf{x}_k,\mathbf{Z}_{k-1})}\left\{\frac{[\nabla_{\mathbf{x}_k} p(\mathbf{x}_k|\mathbf{Z}_{k-1})][\cdot]^{\mathsf{T}}}{[p(\mathbf{x}_k|\mathbf{Z}_{k-1})]^2}\right\} \approx \frac{1}{N_{\mathrm{mc}}}\sum_{l=1}^{N_{\mathrm{mc}}}\left[\frac{[\nabla_{\mathbf{x}_k}\hat{p}(\mathbf{x}_k^{(l)}|\mathbf{Z}_{k-1}^{(l)})][\cdot]^{\mathsf{T}}}{[\hat{p}(\mathbf{x}_k^{(l)}|\mathbf{Z}_{k-1}^{(l)})]^2}\right], \quad (39)$$

where $\mathbf{x}_k^{(l)}$, $\mathbf{Z}_{k-1}^{(l)}$, $l = 1,\ldots,N_{\mathrm{mc}}$, are independent and identically distributed samples such that $(\mathbf{x}_k^{(l)}, \mathbf{Z}_{k-1}^{(l)}) \sim p(\mathbf{x}_k, \mathbf{Z}_{k-1})$. The algorithm to compute the MU-BCRB for the most general model (1) is summarized in Algorithm 1.

Algorithm 1 Computation of the MU-BCRB

(1) At time $k = 0$, generate $\mathbf{x}_0^{(j)} \sim p(\mathbf{x}_0)$ and evaluate $\nabla_{\mathbf{x}_0} p(\mathbf{x}_0^{(j)})$ and $p(\mathbf{x}_0^{(i)})$ for $j = 1, ..., N_{\mathrm{mc}}$. Compute the initial Bayesian information matrix $\mathbf{J}_0$ from

$$\mathbf{J}_0 \approx \frac{1}{N_{\mathrm{mc}}}\sum_{j=1}^{N_{\mathrm{mc}}}\frac{[\nabla_{\mathbf{x}_0} p(\mathbf{x}_0^{(j)})][\nabla_{\mathbf{x}_0} p(\mathbf{x}_0^{(j)})]^{\mathsf{T}}}{[p(\mathbf{x}_0^{(j)})]^2}$$

(2) For $k = 1, 2, \ldots,$ and $l = 1, \ldots, N_{\mathrm{mc}}$ do:
- Sample $\mathbf{x}_k^{(l)} \sim p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(l)})$ and $\mathbf{z}_k^{(l)} \sim p(\mathbf{z}_k|\mathbf{x}_k^{(l)})$.
- Compute the gradient $\nabla \mathbf{x}_k\, p(\mathbf{z}_k^{(l)}|\mathbf{x}_k^{(l)})$ and $p(\mathbf{z}_k^{(l)}|\mathbf{x}_k^{(l)})$, and evaluate $\mathbf{J}_k^{\mathrm{I}}$ according to (35).
- Simulate a particle filter with N particles that approximates $p(\mathbf{x}_k|\mathbf{Z}_{k-1})$ according to (38).
- Compute approximations of the gradient $\nabla \mathbf{x}_k\, \hat{p}(\mathbf{x}_k^{(l)}|\mathbf{Z}_{k-1}^{(l)})$ and the density $\hat{p}(\mathbf{x}_k^{(l)}|\mathbf{Z}_{k-1}^{(l)})$, and evaluate $\mathbf{J}_k^{\mathrm{II}}$ according to (39).
- Evaluate $\mathbf{J}_k$ using (34) and compute the MU-BCRB from (11).

VI. Computational Complexity

In this section we compare the computational requirements of the different BCRBs. For the subsequent complexity calculations we assume that the state vector dimension n_x is much smaller than the number of Monte Carlo runs N_{mc} and the number of particles N in the particle filter, i.e., $n_x << N_{\mathrm{mc}}$ and $n_x << N$.

For the computation of the different BCRBs, all approaches require the inversion of an information matrix. The recursive approach for computing the JU-BCRB ($\mathbf{B}_1$) [4], the MU-BCRB ($\mathbf{B}_2$) and the MC-BCRB ($\mathbf{B}_4$) are based on inverting an $(n_x \times n_x)$ information matrix. The computational complexity of the matrix inversion depends on its type and the specific technique used for the inversion, but can be roughly approximated with $\mathcal{O}(n_x^3)$ [17]. According to [6] the exact computation of the JC-BCRB ($\mathbf{B}_3$) requires the inversion of an $(k+1)n_x \times (k+1)n_x$ information matrix, yielding the computational complexity $\mathcal{O}(((k+1)n_x)^3)$. This complexity can be further reduced to $\mathcal{O}(n_x^3)$ by using the approximate recursive computations of the JC-BCRB ($\mathbf{B}_3$) that were suggested in [6], [7]. Besides the necessary matrix inversions, the computation of the different BCRBs require the averaging over different Monte Carlo runs. Furthermore, the BCRBs $\mathbf{B}_2$, $\mathbf{B}_3$ and $\mathbf{B}_4$ additionally require the cost of running particle filters. While the particle filter based computation of the information matrices of the MU-BCRB and the JC-BCRB has a complexity of $\mathcal{O}(N_{\mathrm{mc}}N)$ and $\mathcal{O}(N)$, the complexity to compute the information matrix of the MC-BCRB is $\mathcal{O}(N^2)$, see also [7]. The overall computational complexity for the different BCRBs is summarized in Table II. It can be concluded that the computation of the JU-BCRB has the lowest complexity, while the computation of the tightest bound $\mathbf{B}_4$ has the highest computational complexity. Additionally, it has to be noted that all bounds except the JU-BCRB rely on a particle filter approximation of the information matrix. The particle filter approximation generally suffers from the "curse of dimensionality" [18], and thus these bounds are expected to provide only acceptable results in state-space models with relatively low state vector dimension n_x. In scenarios with high state vector dimension n_x, the JU-BCRB shall be used as the preferred tool for providing a lower bound on the MSE matrix.

TABLE II
Computational Complexity of the BCRBs

Name	Reference	Complexity
JU-BCRB ($\mathbf{B}_1$)	[4]	$\mathcal{O}(N_{\mathrm{mc}})$
MU-BCRB ($\mathbf{B}_2$)	-	$\mathcal{O}(N_{\mathrm{mc}}N)$
JC-BCRB ($\mathbf{B}_3$), exact	[6]	$\mathcal{O}(N_{\mathrm{mc}}N) + \mathcal{O}(((k+1)n_x)^3)$
JC-BCRB ($\mathbf{B}_3$), approx.	[6], [7]	$\mathcal{O}(N_{\mathrm{mc}}N)$
MC-BCRB ($\mathbf{B}_4$)	[7]	$\mathcal{O}(N_{\mathrm{mc}}N^2)$

VII. Numerical Experiments

In this section, we conduct numerical experiments in order to compute the different bounds presented in the previous sections for two non-linear state space models. Consider the dynamic state space equations given below.

$$x_k = \alpha x_{k-1} + \beta\frac{x_{k-1}}{1+x_{k-1}^2} + \gamma\cos(1.2k) + v_k \quad (40a)$$

$$z_k = \kappa x_k^2 + w_k. \quad (40b)$$

where v_k and w_k are mutually independent zero mean Gaussian noise sequences with variances Q and R, and the initial state is zero mean Gaussian with variance $P_{0|0}$ respectively.

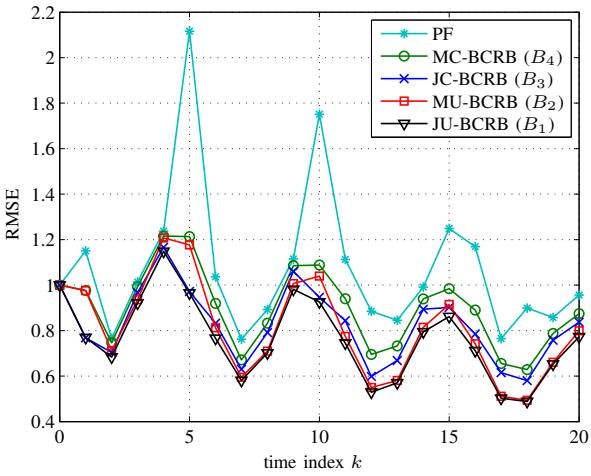

Fig. 1. PF RMS error compared with different BCRBs for Example-1

The model is known as univariate non-stationary growth model [10]. We will illustrate the differences of the bounds on two examples where the parameters $\alpha, \beta, \gamma, \kappa, Q, R, P_{0|0}$ are chosen differently.

A. Example-I

In the first example, we set the parameters as $\alpha = 1, \beta = 5, \gamma = 8, \kappa = 1/20, Q = 1, R = 1$ and $P_{0|0} = 1$. We compute the bounds by averaging over Monte-Carlo runs. 100 000 runs are done for the computation of B_1, B_2 and B_3 and 1000 MC runs are done for the computation of B_4. A bootstrap particle filter with 1000 particles is used for the computation of the bounds B_2, B_3 and B_4. In Figure 1, all the bounds are depicted together with the average RMSE of a particle filter which runs with 1000 particles. The average RMSE is computed over 100 000 runs. The results show that JC-BCRB (B_3) is tighter than JU-BCRB (B_1) and MC-BCRB (B_4) is tighter than MU-BCRB (B_2) which is consistent with our findings, see also (30). In addition, it can be seen that JC-BCRB (B_3) is crossing MU-BCRB (B_2) which means that they cannot be related in terms of tightness to each other. All the bounds are close to the average RMSE of the particle filter for this example. One realization of the true state is plotted together with the particles and their mean in Figure 2.

B. Example-II

In our second example, we illustrate a simplified model, where the parameters are chosen as $\alpha = 1, \beta = 0, \gamma = 0, \kappa = 1/20, Q = 1, R = 1$ and$P_{0|0} = 1$. For this choice of parameters, the posterior distribution of the state is bimodal. 100 000 MC runs are done for the computation of B_1, B_2 and B_3 and 1000 MC runs are done for the computation of B_4, where for the computation of the bounds B_2, B_3 and B_4 a bootstrap particle filter with 1000 particles is used. In Figure 3, one can observe that MC-BCRB (B_4) is the tightest bound. Both JC-BCRB (B_3) and MU-BCRB (B_2) are tighter

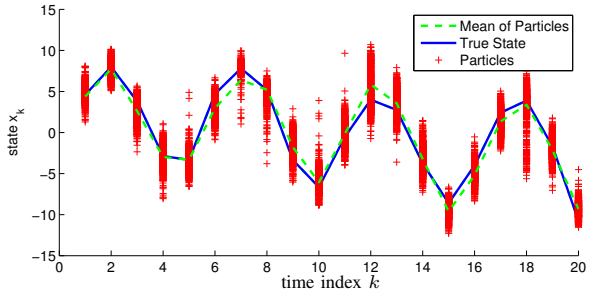

Fig. 2. One realization of the true state, the particles, and the mean of the particles for the model given in Example-I

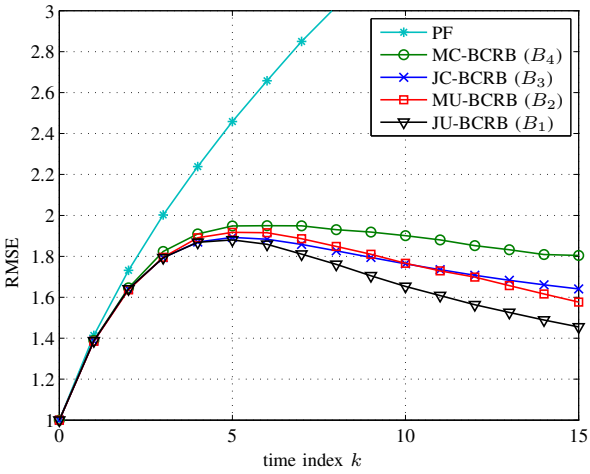

Fig. 3. PF RMS error and corresponding BCRBs for Example-2

than JU-BCRB (B_1), as expected. The average RMSE of the bootstrap particle filter is computed over 100 000 MC runs and the particle filter is using 1000 particles. Notice that, none of the bounds are close to the average RMSE of the particle filter. This is because the posterior distribution is bimodal and the BCRB does not account for the spread of the means. The average RMSE of the PF increases in time as the mean of the modes separate from each other and the mean of the particles is in the middle of the two modes or when the particles in one mode are depleted and the mean converges to a single mode (not necessarily to the correct one). For illustration, one realization of the true state is plotted together with the particles and their mean in Figure 4.

VIII. Conclusion

In this paper, we aim at describing different BCRB bounds in a unifying framework for nonlinear filtering in order to provide a better perception of the existing bounds in the literature. Furthermore, we provided the basic relations, in means of tightness, between different bounds and provided simple numerical examples for the illustration of their performance.

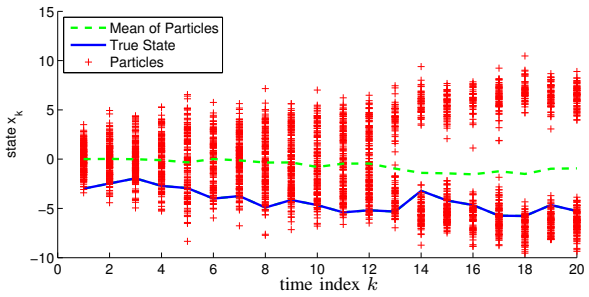

Fig. 4. One realization of the true state, the particles, and the mean of the particles for the model given in Example-2

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support from the Swedish Research Council under the frame project grant Extended Target Tracking (ETT), and the support from the Swedish Foundation for Strategic Research (SSF) under the frame project grant COOP-LOC.

APPENDIX

PROOF OF THEOREM 1

It is easy to show that

$$\boldsymbol{\mathcal{M}}(\hat{\mathbf{X}}_k) \geq \mathbb{E}_{p(\mathbf{Z}_{k-1})}\left\{[\mathbf{J}_{0:k}(\mathbf{Z}_{k-1})]^{-1}\right\} \tag{41}$$

holds. The joint conditional Bayesian information matrix is given by

$$\mathbf{J}_{0:k}(\mathbf{Z}_{k-1}) = \mathbb{E}_{p_c}\left\{-\Delta^{\mathbf{X}_k}_{\mathbf{X}_k}\log p(\mathbf{X}_k, \mathbf{z}_k|\mathbf{Z}_{k-1})\right\}, \tag{42}$$

which can be rewritten using Bayes' rule according to

$$p_c \triangleq p(\mathbf{X}_k, \mathbf{z}_k|\mathbf{Z}_{k-1}) = \frac{p(\mathbf{X}_k, \mathbf{Z}_k)}{p(\mathbf{Z}_{k-1})}. \tag{43}$$

Thus, we can further decompose

$$\begin{aligned} -\Delta^{\mathbf{X}_k}_{\mathbf{X}_k}\log p(\mathbf{X}_k, \mathbf{z}_k|\mathbf{Z}_{k-1}) &= -\Delta^{\mathbf{X}_k}_{\mathbf{X}_k}\log p(\mathbf{X}_k, \mathbf{Z}_k) \\ &\quad +\Delta^{\mathbf{X}_k}_{\mathbf{X}_k}\log p(\mathbf{Z}_{k-1}) \\ &= -\Delta^{\mathbf{X}_k}_{\mathbf{X}_k}\log p(\mathbf{X}_k, \mathbf{Z}_k), \end{aligned} \tag{44}$$

where the second equality holds since $p(\mathbf{Z}_{k-1})$ does not depend on $\mathbf{X}_k$. Inserting (44) into (42) and taking the expectation w.r.t. $\mathbf{Z}_{k-1}$ on both sides of (42) gives

$$\begin{aligned} \mathbb{E}_{p(\mathbf{Z}_{k-1})}\{\mathbf{J}_{0:k}(\mathbf{Z}_{k-1})\} &= \mathbb{E}_{p(\mathbf{X}_k,\mathbf{Z}_k)}\left\{-\Delta^{\mathbf{X}_k}_{\mathbf{X}_k}\log p(\mathbf{X}_k, \mathbf{Z}_k)\right\} \\ &= \mathbf{J}_{0:k}. \end{aligned} \tag{45}$$

Jensen's inequality now yields the relation

$$\mathbb{E}_{p(\mathbf{Z}_{k-1})}\left\{[\mathbf{J}_{0:k}(\mathbf{Z}_{k-1})]^{-1}\right\} \geq [\mathbf{J}_{0:k}]^{-1}. \tag{46}$$

Extracting the $(n_x \times n_x)$ lower-right submatrix on both sides of (46) does not alter the inequality, so that

$$\mathbf{B}_3 \geq \mathbf{B}_1 \tag{47}$$

must hold, which concludes our proof. □

PROOF OF THEOREM 2

In order to proof Theorem 2 it is sufficient to show that the different Bayesian information matrices are equal, i.e.

$$\tilde{\mathbf{J}}_k = \mathbf{J}_k = \tilde{\mathbf{J}}_k(\mathbf{Z}_{k-1}) = \mathbf{J}_k(\mathbf{Z}_{k-1}) \tag{48}$$

holds. In [4] it is shown that for linear Gaussian systems the matrix $\tilde{\mathbf{J}}_k$ can be computed from the following recursion

$$\tilde{\mathbf{J}}_k = [\mathbf{Q}_k + \mathbf{F}_k\tilde{\mathbf{J}}^{-1}_{k-1}\mathbf{F}^{\mathsf{T}}_k]^{-1} + \mathbf{H}^{\mathsf{T}}_k\mathbf{R}^{-1}_k\mathbf{H}_k, \tag{49}$$

which is initialized with $\tilde{\mathbf{J}}_0 = \mathbf{P}^{-1}_{0|0}$. Clearly, this expression is independent of $\mathbf{Z}_{k-1}$ which implies that the conditional Bayesian information matrices $\tilde{\mathbf{J}}_k(\mathbf{Z}_{k-1})$ and $\mathbf{J}_k(\mathbf{Z}_{k-1})$ must be independent of $\mathbf{Z}_{k-1}$. This is also the reason why it suffices to proof (48), since in this case the expectations for computing $\mathbf{B}_3$ and $\mathbf{B}_4$ can be dropped. The equivalence of $\tilde{\mathbf{J}}_k(\mathbf{Z}_{k-1})$ and $\tilde{\mathbf{J}}_k$ has been proven in [10, Theorem 2] and is not repeated here. Instead, we focus on showing that $\mathbf{J}_k$ and $\mathbf{J}_k(\mathbf{Z}_{k-1})$ reduce to the expression given in (49). By making use of Bayes' rule $p(\mathbf{x}_k, \mathbf{Z}_k) = p(\mathbf{x}_k|\mathbf{Z}_k)\cdot p(\mathbf{Z}_k)$ the expression in (12) can be rewritten as

$$\mathbf{J}_k = \mathbb{E}_{p(\mathbf{x}_k,\mathbf{Z}_k)}\left\{-\triangle^{\mathbf{x}_k}_{\mathbf{x}_k}\log p(\mathbf{x}_k|\mathbf{Z}_k)\right\}. \tag{50}$$

It is well known that the posterior density in the linear Gaussian case is given by $p(\mathbf{x}_k|\mathbf{Z}_k) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k})$, where $\hat{\mathbf{x}}_{k|k}$ and $\mathbf{P}_{k|k}$ can be computed recursively using the Kalman filter. Thus, straightforward evaluation of (50) yields

$$\begin{aligned} \mathbf{J}_k &= [\mathbf{P}_{k|k}]^{-1} \\ &= [\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}^{\mathsf{T}}_k[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}^{\mathsf{T}}_k + \mathbf{R}_k]^{-1} \\ &\quad \times\mathbf{H}_k\mathbf{P}^{\mathsf{T}}_{k|k-1}]^{-1}. \\ &= \left[[\mathbf{P}^{-1}_{k|k-1} + \mathbf{H}^{\mathsf{T}}_k\mathbf{R}^{-1}_k\mathbf{H}_k]^{-1}\right]^{-1} \qquad (51) \\ &= [\mathbf{Q}_k + \mathbf{F}_k\mathbf{P}_{k-1|k-1}\mathbf{F}^{\mathsf{T}}_k]^{-1} + \mathbf{H}^{\mathsf{T}}_k\mathbf{R}^{-1}_k\mathbf{H}_k \qquad (52) \end{aligned}$$

where the third equality follows from using the matrix inversion lemma [19]. By further taking into account that $\mathbf{J}_{k-1} = \mathbf{P}^{-1}_{k-1|k-1}$ holds, the expression in (52) can be written as a recursion, yielding

$$\mathbf{J}_k = [\mathbf{Q}_k + \mathbf{F}_k\mathbf{J}^{-1}_{k-1}\mathbf{F}^{\mathsf{T}}_k]^{-1} + \mathbf{H}^{\mathsf{T}}_k\mathbf{R}^{-1}_k\mathbf{H}_k, \tag{53}$$

which is initialized with $\mathbf{J}_0 = \mathbf{P}^{-1}_{0|0}$. Since both recursions in (49) and (53) are initialized with the same matrix $\mathbf{P}^{-1}_{0|0}$ this yields that $\tilde{\mathbf{J}}_k = \mathbf{J}_k\ \forall k$ must hold. For the marginal conditional Bayesian information matrix $\mathbf{J}_k(\mathbf{Z}_{k-1})$ in (25), the decomposition $p(\mathbf{x}_k, \mathbf{z}_k|\mathbf{Z}_{k-1}) = p(\mathbf{z}_k|\mathbf{x}_k)\cdot p(\mathbf{x}_k|\mathbf{Z}_{k-1})$ yields

$$\begin{aligned} \mathbf{J}_k(\mathbf{Z}_{k-1}) &= \mathbb{E}_{p(\mathbf{x}_k,\mathbf{z}_k|\mathbf{Z}_{k-1})}\left\{-\triangle^{\mathbf{x}_k}_{\mathbf{x}_k}\log p(\mathbf{z}_k|\mathbf{x}_k)\right\} \\ &\quad +\mathbb{E}_{p(\mathbf{x}_k|\mathbf{Z}_{k-1})}\left\{-\triangle^{\mathbf{x}_k}_{\mathbf{x}_k}\log p(\mathbf{x}_k|\mathbf{Z}_{k-1})\right\} \end{aligned} \tag{54}$$

In linear Gaussian systems, the likelihood and the prediction density are given by $p(\mathbf{z}_k|\mathbf{x}_k) = \mathcal{N}(\mathbf{z}_k; \mathbf{H}_k, \mathbf{R}_k)$ and $p(\mathbf{x}_k|\mathbf{Z}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})$. Further evaluating the expression in (54) yields

$$\mathbf{J}_k(\mathbf{Z}_{k-1}) = \mathbf{H}^{\mathsf{T}}_k\mathbf{R}^{-1}_k\mathbf{H}_k + \mathbf{P}^{-1}_{k|k-1} \tag{55}$$

which is the same as the RHS of (51). Hence, by following the same argumentation as above we can conclude that the recursion

$$\mathbf{J}_k(\mathbf{Z}_{k-1}) = [\mathbf{Q}_k + \mathbf{F}_k \mathbf{J}_{k-1}^{-1} \mathbf{F}_k^{\mathsf{T}}]^{-1} + \mathbf{H}_k^{\mathsf{T}} \mathbf{R}_k^{-1} \mathbf{H}_k, \quad (56)$$

must hold, which is initialized with $\mathbf{J}_0(\mathbf{Z}_{-1}) = \mathbf{P}_{0|0}^{-1}$. As (56) is the same as (49) and both recursions are initialized with the same matrix $\mathbf{P}_{0|0}^{-1}$, we can conclude that $\tilde{\mathbf{J}}_k = \mathbf{J}_k(\mathbf{Z}_{k-1})$ $\forall k$ must hold. This completes the proof. □

REFERENCES

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

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

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

[4] P. Tichavský, C. H. Muravchik, and A. Nehorai, "Posterior Cramér-Rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1386–1396, May 1998.

[5] N. Bergman, "Recursive Bayesian estimation: Navigation and tracking applications," Ph.D. dissertation, Linköping University, Linköping, Sweden, 1999.

[6] L. Zuo, R. Niu, and P. Varshney, "Conditional posterior Cramér-Rao lower bounds for nonlinear sequential Bayesian estimation," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 1 –14, Jan. 2011.

[7] Y. Zheng, O. Ozdemir, R. Niu, and P. K. Varshney, "New conditional posterior Cramér-Rao lower bounds for nonlinear sequential Bayesian estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5549–5556, Oct. 2012.

[8] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, "Some classes of global Cramér-Rao bounds," *The Annals of Statistics*, vol. 15, no. 4, pp. 1421–1438, 1987.

[9] J. Dauwels, "On graphical models for communications and machine learning: Algorithms, bounds, and analog implementation," Ph.D. dissertation, ETH Zurich, 2006.

[10] L. Zuo, "Conditional PCRLB for target tracking in sensor networks," Ph.D. dissertation, Syracuse Universityx, Dept. Electr. Eng. Comput. Sci., Syracuse, NY, USA, Dec 2010.

[11] F. Gustafsson, *Statistical Sensor Fusion*. Lund, Sweden: Studentlitteratur AB, 2010.

[12] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proceedings on Radar and Signal Processing*, vol. 140, 1993, pp. 107–113.

[13] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential Monte Carlo methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[14] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. New York, NY, USA: Springer-Verlag, 2001.

[15] F. Gustafsson, "Particle filter theory and practice with positioning applications," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 53–82, Jul 2010.

[16] Y. Boers, H. Driessen, A. Bagchi, and P. Mandal, "Particle filter based entropy," in *Information Fusion (FUSION), 2010 13th Conference on*, Edinburgh, Scotland, Aug 2010, pp. 1–8.

[17] R. Karlsson, T. Schn, and F. Gustafsson, "Complexity analysis of the marginalized particle filter," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4408–4411, Nov. 2005.

[18] F. Daum and J. Huang, "Curse of dimensionality and particle filters," in *Proc. of IEEE Aerospace Conference*, vol. 4, Big Sky, MT, USA, Mar 2003, pp. 1979–1993.

[19] D. S. Bernstein, *Matrix mathematics: theory, facts and formulas*, 2nd ed. Princeton, NJ, USA: Princeton University Press, 2009.