

Bounds on the Optimal Performance for Jump Markov Linear Gaussian Systems

Carsten Fritsche and Fredrik Gustafsson

Linköping University Post Print

N.B.: When citing this work, cite the original article.

©2013 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Carsten Fritsche and Fredrik Gustafsson, Bounds on the Optimal Performance for Jump Markov Linear Gaussian Systems, 2013, IEEE Transactions on Signal Processing, (61), 1, 92-98.

<http://dx.doi.org/10.1109/TSP.2012.2223690>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-89524>

Bounds on the Optimal Performance for Jump Markov Linear Gaussian Systems

Carsten Fritsche, *Member, IEEE*, and
Fredrik Gustafsson, *Fellow, IEEE*

Abstract—The performance of an optimal filter is lower bounded by the Bayesian Cramér-Rao Bound (BCRB). In some cases, this bound is tight (achieved by the optimal filter) asymptotically in information, i.e. high signal-to-noise ratio (SNR). However, for jump Markov linear Gaussian systems (JMLGS) the BCRB is not necessarily achieved for any SNR. In this paper, we derive a new bound which is tight for all SNRs. The bound evaluates the expected covariance of the optimal filter which is represented by one deterministic term and one stochastic term that is computed with Monte Carlo methods. The bound relates to and improves on a recently presented BCRB and an enumeration BCRB for JMLGS. We analyze their relations theoretically and illustrate them on a couple of examples.

Index Terms—Jump Markov linear Gaussian systems, performance bounds, statistical signal processing.

I. INTRODUCTION

In recent years, there has been an increased interest in developing performance bounds that theoretically predict the best achievable performance for multiple model filtering, i.e., jump Markov systems. Multiple models are used in various applications ranging from change detection, sensor fault detection or tracking of maneuvering targets in air traffic control, see for instance [1], [2].

In multiple model filtering, a discrete parameter representing the mode of the system is introduced, that evolves according to a jump Markov process. This discrete state is treated as unknown and either is estimated together with the continuous valued state vector, or it is treated as a nuisance parameter, which results in a filter, whose computational complexity increases exponentially with time. Various approaches have been proposed to solve both types of estimation problems, and perhaps the most prominent suboptimal solution that is widely used in practice today is the interacting multiple model (IMM) algorithm [3]. While the area of developing filters for such estimation problems has become relatively mature, the development of performance bounds has emerged during the last few years and is mostly related to the Bayesian Cramér-Rao bound (BCRB).

The BCRB for jump Markov systems has been proposed recently in [4]. The idea is based on evaluating the Bayesian information matrix for the complete trajectory of state vectors, using a Monte Carlo (MC) approach. This matrix is then constructed in an elegant way by marginalizing the discrete mode variables from all the densities involved, so that it is not necessary to evaluate derivatives with respect to discrete valued parameters, which cannot be handled. The BCRB for the current state vector is finally determined from the inverse of the Bayesian information matrix.

In some cases, the BCRB is tight (achieved by the optimal filter) asymptotically in information, i.e. high signal-to-noise ratio (SNR), see Example 1 in [5]. However, for jump Markov systems the BCRB is not necessarily achieved for any SNR. In [6, p. 96], it is proven that the BCRB holds only with equality, if and only if the posterior density is a multivariate Gaussian density. For jump Markov systems,

however, this condition is generally violated, since the posterior density is composed of a mixture of densities (Gaussian mixture in the linear Gaussian case). Thus, the corresponding BCRB must be lower and might not be tight for such systems.

In [7], a recursive formulation of a BCRB conditioned on a specific mode sequence is developed, that uses results from [8]. The unconditional BCRB is then found by computing a weighted average of the conditional BCRB over all possible mode sequences. Even though this bound, herein after referred to as Enumer-BCRB, will give a lower bound on the optimal performance, it might not be tight since for its derivation the mode sequence is implicitly assumed known.

A tight performance bound for jump Markov linear Gaussian system (JMLGS) models can be obtained basically by MC integration of the optimal filter covariance over different realizations of the state and measurement sequences. A standard approach to do this is to compute the sample covariance from the error between the optimal filter's posterior mean and the true state. However, this will need coarse approximations, and in any case has a large variance. We call this the direct approach. Our bound (or rather numerical implementation of the expected value of the optimal filter covariance) is computed as a sum of two terms. The first one is deterministic and the same as the Enumer-BCRB, and can be computed efficiently. We point out that this term can be lower bounded by linear in time approximations. The other term is related to the spread of the means contribution in variance computations, and it is computed with MC methods. The sum of these terms gets a lower total variance than the direct approach, in particular when the first term dominates. We illustrate our new bound on a couple of examples, and compare it to the Enumer-BCRB, BCRB and the actual performance of the optimal filter and a common approximative filter.

II. SYSTEM MODEL

Consider the following discrete-time JMLGS, that is described by the following process and measurement equation

$$\mathbf{x}_k = \mathbf{F}_k(r_k) \mathbf{x}_{k-1} + \mathbf{v}_k(r_k), \quad (1a)$$

$$\mathbf{z}_k = \mathbf{H}_k(r_k) \mathbf{x}_k + \mathbf{w}_k(r_k), \quad (1b)$$

where $\mathbf{z}_k \in \mathbb{R}^{n_z}$ is the measurement vector at discrete time k and $\mathbf{x}_k \in \mathbb{R}^{n_x}$ is the state vector and \mathbf{F} and \mathbf{H} are arbitrary linear mapping matrices of appropriate size. The process and measurement noise vectors $\mathbf{v}_k \in \mathbb{R}^{n_v}$ and $\mathbf{w}_k \in \mathbb{R}^{n_w}$ are mutually independent white processes distributed as $\mathbf{v}_k(r_k) \sim \mathcal{N}(\boldsymbol{\mu}_k(r_k), \mathbf{Q}_k(r_k))$ and $\mathbf{w}_k(r_k) \sim \mathcal{N}(\mathbf{m}_k(r_k), \mathbf{R}_k(r_k))$. The mode variable r_k denotes a discrete-time Markov chain with s states and transition probability matrix $\Pr\{r_k|r_{k-1}\}$. At times $k = 0$ and $k = 1$, prior information about the state \mathbf{x}_0 and mode r_1 is available in terms of the probability density function (pdf) $p(\mathbf{x}_0)$ and probability mass function (pmf) $\Pr\{r_1\}$. The initial state \mathbf{x}_0 is assumed to be Gaussian distributed with mean $\hat{\mathbf{x}}_0$ and covariance matrix \mathbf{P}_0 . In JMLGS, the estimation of \mathbf{x}_k is of primary concern, whereas the mode variable r_k is often seen as an unknown nuisance parameter that appears in the problem formulation.

In the following, let $\mathbf{x}_{0:k} = [\mathbf{x}_0^T, \dots, \mathbf{x}_k^T]^T$ and $\mathbf{z}_{1:k} = [\mathbf{z}_1^T, \dots, \mathbf{z}_k^T]^T$ denote the collection of states and measurement vectors up to time k . Furthermore, let the sequence of mode variables at time k be given by $r_{1:k}^i = (r_1^i, r_2^i, \dots, r_k^i)$, where $i = 1, \dots, s^k$. The gradient of a vector \mathbf{u} is defined as $\nabla_{\mathbf{u}} = [\partial/\partial u_1, \dots, \partial/\partial u_n]^T$ and the Laplace operator is defined as $\Delta_{\mathbf{u}}^t = \nabla_{\mathbf{u}}[\nabla_{\mathbf{u}}]^T$. The operator $\mathbb{E}_{p(\mathbf{x})}\{\cdot\}$ or equivalently $\mathbb{E}_{\mathbf{x}}\{\cdot\}$ denotes expectation and the subscript indicates the pdf (or pmf) that is used in the expectation.

III. PREVIOUS WORK

In this section, previous work on lower bounds for JMLGSs are briefly discussed. The optimal filter is revisited and it is shown, how this filter can be related to the computation of the BCRB.

Optimal Filter: It is well known that the posterior pdf of JMLGSs is given by a Gaussian mixture

$$p(\mathbf{x}_k | \mathbf{z}_{1:k}) = \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i | \mathbf{z}_{1:k}\} p(\mathbf{x}_k | \mathbf{z}_{1:k}, r_{1:k}^i), \quad (2)$$

where $\Pr\{r_{1:k}^i | \mathbf{z}_{1:k}\} \propto p(\mathbf{z}_{1:k} | r_{1:k}^i) \cdot \Pr\{r_{1:k}^i\}$, and $p(\mathbf{x}_k | \mathbf{z}_{1:k}, r_{1:k}^i) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^{\text{opt},i}, \mathbf{P}_k^i)$ hold. Here, $\hat{\mathbf{x}}_k^{\text{opt},i}$ or equivalently $\hat{\mathbf{x}}_k^{\text{opt},i}(\mathbf{z}_{1:k})$ or $\hat{\mathbf{x}}_k^{\text{opt}}(\mathbf{z}_{1:k}, r_{1:k}^i)$ is referred to as the conditional optimal filter (or estimator). The (unconditional) optimal filter (or estimator) provides an analytical solution for computing (2) recursively, where the posterior moments $\hat{\mathbf{x}}_k^{\text{opt},i}$, \mathbf{P}_k^i as well as the model likelihood $p(\mathbf{z}_{1:k} | r_{1:k}^i)$ are calculated using a Kalman filter that is matched to the specific mode sequence $r_{1:k}^i$ [9], [10]. The conditional mean $\hat{\mathbf{x}}_k^{\text{opt}}$ and covariance matrix \mathbf{P}_k for \mathbf{x}_k , given the measurements $\mathbf{z}_{1:k}$, serve as optimal filter outputs and are given by

$$\hat{\mathbf{x}}_k^{\text{opt}} \equiv \hat{\mathbf{x}}_k^{\text{opt}}(\mathbf{z}_{1:k}) = \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i | \mathbf{z}_{1:k}\} \hat{\mathbf{x}}_k^{\text{opt},i}, \quad (3)$$

$$\mathbf{P}_k = \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i | \mathbf{z}_{1:k}\} \left[\mathbf{P}_k^i + [\hat{\mathbf{x}}_k^{\text{opt}} - \hat{\mathbf{x}}_k^{\text{opt},i}] [\hat{\mathbf{x}}_k^{\text{opt}} - \hat{\mathbf{x}}_k^{\text{opt},i}]^T \right] \quad (4)$$

We point out that in the optimal estimator the unknown mode sequence $r_{1:k}$ is not directly estimated. Instead, the optimal filter enumerates over all possible mode sequences $r_{1:k}^i$ which leads to a filter complexity that increases exponentially with time k .

Bayesian Cramér-Rao Bound: The BCRB provides a lower bound for any estimator $\hat{\mathbf{x}}_k(\mathbf{z}_{1:k})$ on the mean square error (MSE) matrix $\mathcal{M}(\hat{\mathbf{x}}_k)$. It is defined as the inverse of the Bayesian information submatrix \mathbf{J}_k ,

$$\mathcal{M}(\hat{\mathbf{x}}_k) \equiv \mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k})} \{ [\hat{\mathbf{x}}_k(\mathbf{z}_{1:k}) - \mathbf{x}_k][\cdot]^T \} \geq [\mathbf{J}_k]^{-1}, \quad (5)$$

where the matrix inequality $\mathbf{A} \geq \mathbf{C}$ means that the difference $\mathbf{A} - \mathbf{C}$ is a positive semidefinite matrix [6]. Recently, a recursive algorithm has been proposed to evaluate the BCRB for jump Markov nonlinear systems, where the mode r_k enters only into the process model, and which includes (1a) as special case [4]. The idea of [4] is to first evaluate the Bayesian information matrix $\mathbf{J}_{0:k}$ of the complete state trajectory $\mathbf{x}_{0:k}$ using an MC method, and then determine the BCRB matrix \mathbf{B}_1 , which is defined as the the $(n_x \times n_x)$ lower-right submatrix of $[\mathbf{J}_{0:k}]^{-1}$.

Alternatively, it is possible to evaluate the Bayesian information submatrix \mathbf{J}_k directly by using the optimal filter, yielding

$$\mathbf{J}_k = \mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k})} \left\{ \frac{[\nabla_{\mathbf{x}_k} p(\mathbf{x}_k | \mathbf{z}_{1:k})][\nabla_{\mathbf{x}_k} p(\mathbf{x}_k | \mathbf{z}_{1:k})]^T}{[p(\mathbf{x}_k | \mathbf{z}_{1:k})]^2} \right\}. \quad (6)$$

The inverse of \mathbf{J}_k will then give another BCRB matrix \mathbf{B}_2 , herein after referred to as marginalized BCRB (M-BCRB), which relates to \mathbf{B}_1 according to $\mathbf{B}_2 \geq \mathbf{B}_1$, see [11] for a proof. For JMLGS, analytical expressions for evaluating $p(\mathbf{x}_k | \mathbf{z}_{1:k})$ and its gradient are available, which can be computed from the optimal filter recursions, cf. (2). In this case, \mathbf{J}_k can be approximated numerically according to

$$\mathbf{J}_k \approx \frac{1}{N} \sum_{j=1}^N \frac{[\nabla_{\mathbf{x}_k} p(\mathbf{x}_k^{(j)} | \mathbf{z}_{1:k}^{(j)})][\nabla_{\mathbf{x}_k} p(\mathbf{x}_k^{(j)} | \mathbf{z}_{1:k}^{(j)})]^T}{[p(\mathbf{x}_k^{(j)} | \mathbf{z}_{1:k}^{(j)})]^2}, \quad (7)$$

where $\mathbf{x}_k^{(j)}$ and $\mathbf{z}_{1:k}^{(j)}$, $j = 1, \dots, N$ are independent and identically distributed (i.i.d.) samples, such that $(\mathbf{x}_k^{(j)}, \mathbf{z}_{1:k}^{(j)}) \sim p(\mathbf{x}_k, \mathbf{z}_{1:k})$. The gradient $\nabla_{\mathbf{x}_k} p(\mathbf{x}_k | \mathbf{z}_{1:k})$, required to evaluate (7) is given by

$$\nabla_{\mathbf{x}_k} p(\mathbf{x}_k | \mathbf{z}_{1:k}) = \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i | \mathbf{z}_{1:k}\} [\nabla_{\mathbf{x}_k} p(\mathbf{x}_k | \mathbf{z}_{1:k}, r_{1:k}^i)], \quad (8)$$

with $\nabla_{\mathbf{x}_k} p(\mathbf{x}_k | \mathbf{z}_{1:k}, r_{1:k}^i) = -\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^{\text{opt},i}, \mathbf{P}_k^i) [\mathbf{P}_k^i]^{-1} [\mathbf{x}_k - \hat{\mathbf{x}}_k^{\text{opt},i}]$.

Enumeration Bayesian Cramér-Rao Bound: The enumeration method [7], [12] provides a lower bound on the MSE matrix for any conditional estimator $\hat{\mathbf{x}}_k(\mathbf{z}_{1:k}, r_{1:k})$. The idea of this method is to lower bound the conditional MSE matrix by the following expression:

$$\mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k} | r_{1:k}^i)} \{ [\hat{\mathbf{x}}_k^i(\mathbf{z}_{1:k}) - \mathbf{x}_k][\cdot]^T \} \geq [\mathbf{J}_k^i]^{-1}, \quad (9)$$

where $\hat{\mathbf{x}}_k^i(\mathbf{z}_{1:k})$ or equivalently $\hat{\mathbf{x}}_k(\mathbf{z}_{1:k}, r_{1:k}^i)$ denotes the conditional estimator of a particular mode sequence $r_{1:k}^i$, and where \mathbf{J}_k^i or equivalently $\mathbf{J}_k(r_{1:k}^i)$ denotes the conditional Bayesian information submatrix [7], [12]. By taking into account that the unconditional MSE matrix is related to the conditional MSE matrix through the smoothing property of expectations, i.e.

$$\mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k})} \{ [\hat{\mathbf{x}}_k(\mathbf{z}_{1:k}, r_{1:k}) - \mathbf{x}_k][\cdot]^T \} = \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i\} \mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k} | r_{1:k}^i)} \{ [\hat{\mathbf{x}}_k^i(\mathbf{z}_{1:k}) - \mathbf{x}_k][\cdot]^T \}, \quad (10)$$

the authors in [7], [12] develop an unconditional BCRB for any conditional estimator $\hat{\mathbf{x}}_k(\mathbf{z}_{1:k}, r_{1:k})$, which is given by the following expression:

$$\mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k})} \{ [\hat{\mathbf{x}}_k(\mathbf{z}_{1:k}, r_{1:k}) - \mathbf{x}_k][\cdot]^T \} \geq \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i\} [\mathbf{J}_k^i]^{-1}. \quad (11)$$

The Enumer-BCRB given in (11) can be seen as an average bound for conditional estimators, since it averages the bound given in (9) over all possible mode sequences. It is stated in [7], [12] without giving a proof that the Enumer-BCRB will be overly optimistic, i.e. it cannot be reached by any unconditional estimator $\hat{\mathbf{x}}_k(\mathbf{z}_{1:k})$. We note, that this fact can be proven, i.e. it can be shown that

$$\begin{aligned} \mathcal{M}(\hat{\mathbf{x}}_k(\mathbf{z}_{1:k})) &\geq \mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k})} \{ [\hat{\mathbf{x}}_k(\mathbf{z}_{1:k}, r_{1:k}) - \mathbf{x}_k][\cdot]^T \} \\ &\geq \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i\} [\mathbf{J}_k^i]^{-1}. \end{aligned} \quad (12)$$

holds, by interpreting (12) as an instance of the bound by Miller and Chang, where $r_{1:k}$ is treated as a nuisance parameter appearing in the problem formulation of JMLGSs, see [11] and references therein. The proof is then given by Lemma 2 of [11].

Even though the bounds presented in this section will lower bound the performance of any estimator $\hat{\mathbf{x}}_k(\mathbf{z}_{1:k})$, it is not clear when these bounds are tight. The aim of this paper is to develop a bound that is always tight. It will be shown that the new bound is related to the optimal filter covariance and that it generalizes the Enumer-BCRB.

IV. THE NEW PERFORMANCE BOUND

In this section, the main result is presented which is given by the following theorem:

Theorem 1. A bound on the MSE matrix is given by

$$\mathcal{M}(\hat{\mathbf{x}}_k) \geq \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i\} [\mathbf{J}_k^i]^{-1} + \mathbb{E}_{p(\mathbf{z}_{1:k})} \left\{ \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i | \mathbf{z}_{1:k}\} [\mathbf{m}_k^i(\mathbf{z}_{1:k})][\cdot]^\top \right\}, \quad (13)$$

where $\mathbf{m}_k^i(\mathbf{z}_{1:k}) = \hat{\mathbf{x}}_k^{\text{opt}}(\mathbf{z}_{1:k}) - \hat{\mathbf{x}}_k^{\text{opt},i}(\mathbf{z}_{1:k})$.

Proof: See Appendix. ■

Remark 1. The second sum component in (13) is known as the expected value of the “spread of the means term”. If this term is neglected we arrive at the Enumer-BCRB, cf. (12). Since the expected value of the spread of the means term yields a positive semidefinite matrix, it follows (without proof) that the new bound is tighter than the Enumer-BCRB.

Corollary 1. The expected value of the conditional covariance matrix \mathbf{P}_k of the optimal filter, cf. (4), satisfies the bound given in Theorem 1 with equality and is given by

$$\mathcal{M}(\hat{\mathbf{x}}_k^{\text{opt}}) \equiv \mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k})} \{ [\hat{\mathbf{x}}_k^{\text{opt}}(\mathbf{z}_{1:k}) - \mathbf{x}_k][\cdot]^\top \} = \mathbb{E}_{p(\mathbf{z}_{1:k})} \{ \mathbf{P}_k \}. \quad (14)$$

Numerical Approximation of the Performance Bound: This section is devoted to the evaluation of the expression given in (13). For the model given in (1), the conditional Bayesian information submatrix \mathbf{J}_k^i can be computed for $k \geq 1$ from the well-known recursive relationship [7], [12]:

$$\mathbf{J}_k^i = \left[\mathbf{Q}_k(r_k^i) + \mathbf{F}_k(r_k^i) [\mathbf{J}_{k-1}^i]^{-1} \mathbf{F}_k^\top(r_k^i) \right]^{-1} + \mathbf{H}_k^\top(r_k^i) [\mathbf{R}_k(r_k^i)]^{-1} \mathbf{H}_k(r_k^i), \quad (15)$$

where $\mathbf{J}_0 = \mathbf{P}_0^{-1}$. Note, that the above expression can be rewritten with the matrix inversion lemma into several different forms depending on the sizes n_x and n_z , where $\mathbf{Q}_k(r_k)$ might need to be factorized if $n_v < n_x$, see for instance [12], [13]. Since the computational complexity of evaluating the first sum in (13) grows exponentially with time k , the above described approach is feasible only for small values of k . For large values of k , an MC approach should be used which was suggested in [14]. The evaluation of the expected value of the spread of the means term is more demanding, since it involves the computation of a complicated expectation involving a high-dimensional integral. Here, an MC integration approach is proposed to approximate the expectation numerically, yielding

$$\mathbb{E}_{p(\mathbf{z}_{1:k})} \left\{ \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i | \mathbf{z}_{1:k}\} [\mathbf{m}_k^i(\mathbf{z}_{1:k})][\cdot]^\top \right\} \approx \frac{1}{N} \sum_{j=1}^N \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i | \mathbf{z}_{1:k}^{(j)}\} [\mathbf{m}_k^i(\mathbf{z}_{1:k}^{(j)})][\cdot]^\top, \quad (16)$$

where $\mathbf{z}_{1:k}^{(j)}$, $j = 1, \dots, N$ are i.i.d. samples such that $\mathbf{z}_{1:k}^{(j)} \sim p(\mathbf{z}_{1:k})$. A pseudocode for the computation of the newly proposed bound is given in Algorithm 1.

Computational Complexity Comparison: The computational complexity of the new performance bound increases exponentially with s^k . Compared to the Enumer-BCRB, an additional stochastic term given by (16) has to be evaluated, whose complexity is proportional to $N \cdot s^k$. The complexity of the M-BCRB is similar to that of the new performance bound, since the M-BCRB also requires the evaluation of the optimal filter recursions. The Bayesian information matrix, whose inverse gives the BCRB, is constructed

recursively with a complexity that is proportional to $N \cdot s \cdot k$. However, this approach requires a costly computation of a matrix inverse, whose dimension increases linearly with time k .

Monte Carlo Variance Reduction: According to Corollary 1, the new performance bound approximates the MSE matrix of the optimal filter. Let $\mathcal{M}_1(\hat{\mathbf{x}}_k^{\text{opt}})$ denote an MC estimate of the MSE matrix using the proposed approach, and let $\mathcal{M}_2(\hat{\mathbf{x}}_k^{\text{opt}})$ denote an MC estimate of the MSE matrix using the conventional direct approach, given by

$$\mathcal{M}_2(\hat{\mathbf{x}}_k^{\text{opt}}) = \frac{1}{N} \sum_{j=1}^N [\hat{\mathbf{x}}_k^{\text{opt}}(\mathbf{z}_{1:k}^{(j)}) - \mathbf{x}_k^{(j)}][\hat{\mathbf{x}}_k^{\text{opt}}(\mathbf{z}_{1:k}^{(j)}) - \mathbf{x}_k^{(j)}]^\top, \quad (17)$$

where $\mathbf{x}_k^{(j)}$ denotes the true state vector at the j -th MC run. Furthermore, let $\mathcal{M}_{1,n} = [\mathcal{M}_1]_{n,n}$ and $\mathcal{M}_{2,n} = [\mathcal{M}_2]_{n,n}$ denote the corresponding MSE for estimating the n -th element of \mathbf{x}_k , denoted as $x_{k,n}$, and where $[\mathbf{A}]_{i,j}$ denotes the matrix element at the i -th row and j -th column. Then, the MC variance for computing $\mathcal{M}_{1,n}$ and $\mathcal{M}_{2,n}$ relate to each other according to the following proposition.

Proposition 1. The MC variances of $\mathcal{M}_{1,n}$ and $\mathcal{M}_{2,n}$ satisfy the following inequality

$$\text{Var}_{\mathbf{z}}(\mathcal{M}_{2,n}) \geq \text{Var}_{\mathbf{z}}(\mathcal{M}_{1,n}) \quad (18)$$

Proof: See Appendix. ■

Thus, with the same number of MC runs, the variance in the estimates using the proposed approach can be decreased as compared to the direct approach. As an alternative, it is also possible to compute an MC estimate of the MSE matrix from the relation

$$\mathcal{M}_3(\hat{\mathbf{x}}_k^{\text{opt}}) \approx \frac{1}{N} \sum_{j=1}^N \mathbf{P}_k^{(j)}, \quad (19)$$

see Corollary 1, where $\mathbf{P}_k^{(j)}$ denotes the covariance matrix estimate of the optimal filter at the j -th MC run, cf. (4). By letting $\mathcal{M}_{3,n} = [\mathcal{M}_3]_{n,n} = 1/N \sum_{j=1}^N [g_4(\mathbf{z}_{1:k}^{(j)}) + g_1(\mathbf{z}_{1:k}^{(j)})]$ with $g_4(\mathbf{z}_{1:k}) = \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i | \mathbf{z}_{1:k}\} [\mathbf{P}_k^i]_{n,n}$ and $g_1(\mathbf{z}_{1:k}) = \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i | \mathbf{z}_{1:k}\} [\hat{\mathbf{x}}_{k,n}^{\text{opt}}(\mathbf{z}_{1:k}) - \hat{\mathbf{x}}_{k,n}^{\text{opt},i}(\mathbf{z}_{1:k})]^2$, the MC variance for computing $\mathcal{M}_{1,n}$ and $\mathcal{M}_{3,n}$ relate to each other according to the following proposition.

Proposition 2. The MC variances of $\mathcal{M}_{1,n}$ and $\mathcal{M}_{3,n}$ satisfy the following inequality

$$\text{Var}_{\mathbf{z}}(\mathcal{M}_{3,n}) \geq \text{Var}_{\mathbf{z}}(\mathcal{M}_{1,n}) \quad (20)$$

if and only if $\text{Var}_{\mathbf{z}}(g_4(\mathbf{z}_{1:k})) + 2 \cdot \text{Cov}_{\mathbf{z}}(g_4(\mathbf{z}_{1:k}), g_1(\mathbf{z}_{1:k})) \geq 0$

Proof: See Appendix. ■

The above given constraint condition is satisfied if $g_1(\mathbf{z}_{1:k})$ and $g_4(\mathbf{z}_{1:k})$ are positively correlated. For the most general model, however, it might be hard to prove theoretically that the constraint condition always holds. As a practical approach, it is therefore suggested to approximate the constraint condition via the respective sample (co-)variance expressions.

V. PERFORMANCE EVALUATION

The newly proposed bound, cf. Theorem 1, is compared to the following bounds and filter performances: 1. The BCRB and M-BCRB, cf. Section III; 2. The conservative Enumer-BCRB, cf. Section III; 3. The optimal filter using the direct approach \mathcal{M}_2 , cf. Section III, which coincides with the newly proposed bound according to Corollary 1; 4. The IMM filter [3], which is a state-of-the-art algorithm for the chosen problem.

Algorithm 1 Pseudocode for the computation of the newly proposed bound

- (1) At time $k = 0$, generate $\mathbf{x}_0^{(j)} \sim \mathcal{N}(\mathbf{x}_0; \hat{\mathbf{x}}_0, \mathbf{P}_0)$ for $j = 1, \dots, N$, where \mathbf{P}_0 gives the new bound.
 - (2) For $k = 1, 2, \dots$, and $j = 1, \dots, N$ do:
 - If $k = 1$, then generate $r_1^{(j)} \sim \Pr\{r_1\}$. In all other cases generate $r_k^{(j)} \sim \Pr\{r_k | r_{k-1}^{(j)}\}$ and set $r_{1:k}^{(j)} = [r_{1:k-1}^{(j)}, r_k^{(j)}]$. Furthermore, generate $\mathbf{x}_k^{(j)} \sim p(\mathbf{x}_k | \mathbf{x}_{k-1}^{(j)}, r_k^{(j)})$, $\mathbf{z}_k^{(j)} \sim p(\mathbf{z}_k | \mathbf{x}_k^{(j)}, r_k^{(j)})$ and set $\mathbf{z}_{1:k}^{(j)} = [\mathbf{z}_{1:k-1}^{(j)}, \mathbf{z}_k^{(j)}]$.
 - Evaluate the RHS of (13) as follows to obtain the new bound:
 - * **First sum component:** For $i = 1, \dots, s^k$, evaluate $\Pr\{r_{1:k}^i\}$ in closed-form, see equation (5) presented in [7]. Evaluate \mathbf{J}_k^i according to (15).
 - * **Second sum component:** Evaluate $\hat{\mathbf{x}}_k^{\text{opt}}(\mathbf{z}_{1:k}^{(j)}, r_{1:k}^i)$, $\Pr\{r_{1:k}^i | \mathbf{z}_{1:k}\}$ and $\hat{\mathbf{x}}_k^{\text{opt}}(\mathbf{z}_{1:k}^{(j)})$, cf. (3). Approximate numerically the expected value of the spread of the means term by (16).
-

Simulation Scenario: The performance is evaluated by means of simulations. Here, the example proposed in [4] is used, where a maneuvering target tracking scenario is considered. The target's movement is assumed to be in one dimension with state vector $\mathbf{x}_k = [x_k, \dot{x}_k, \ddot{x}_k]^\top$, representing position, velocity and acceleration, respectively. Further, it is assumed that the target's movement can switch between a nearly constant velocity model and a nearly constant acceleration model. The state transition matrix and the process noise covariance matrix of the nearly constant velocity model are given by

$$\mathbf{F}_k(1) = \begin{bmatrix} 1 & T & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and

$$\mathbf{Q}_k(1) = s_1 \begin{bmatrix} T^3/3 & T^2/2 & 0 \\ T^2/2 & T & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

where $s_1 = 2$ is the power spectral density. The state transition matrix and the process noise covariance matrix of the nearly constant acceleration model are given by

$$\mathbf{F}_k(2) = \begin{bmatrix} 1 & T & T^2/2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$\mathbf{Q}_k(2) = s_2 \begin{bmatrix} T^5/20 & T^4/8 & T^3/6 \\ T^4/8 & T^3/3 & T^2/2 \\ T^3/6 & T^2/2 & T \end{bmatrix},$$

with power spectral density $s_2 = 0.4$. It is further assumed that every $T = 5$ s, noisy measurements of the position state x_k are available. The measurement model and the corresponding measurement noise variance are assumed to be $\mathbf{H}_k = [1, 0, 0]$ and $\mathbf{R}_k = 50$. The mean and covariance of the prior density $p(\mathbf{x}_0)$ are given by $\hat{\mathbf{x}}_0 = [2, 2, 2]^\top$ and $\mathbf{P}_0 = \text{diag}([1, 1, 1])$. The initial mode probabilities are set to $\Pr\{r_1 = 1\} = 0.5$ and $\Pr\{r_1 = 2\} = 0.5$ and the transition probabilities are given by $\Pr\{r_k = 1 | r_{k-1} = 1\} = 0.9$ and $\Pr\{r_k = 2 | r_{k-1} = 2\} = 0.9$, respectively.

Simulation Results: In this section the simulation results are presented. All bounds and filters have been initialized with $\hat{\mathbf{x}}_0$ and \mathbf{P}_0 . The new bound is computed using Algorithm 1. The MSE matrix of the IMM filter is approximated numerically using the conventional

direct approach, cf. (17). The Enumeration-BCRB has been evaluated analytically. For all other bounds and filters, a total of $N = 50\,000$ MC runs are performed.

The results for position, velocity and acceleration are given in Fig. 1 (a)-(c). It can be observed that the optimal filter using the direct approach and the computationally less complex IMM filter have approximately the same performance. The newly proposed bound matches the optimal performance as expected, while the Enumer-BCRB, the BCRB and M-BCRB yield too conservative bounds. As expected, the M-BCRB is tighter than the BCRB. The difference between the BCRB and the Enumer-BCRB is discussed in [4]. The difference between the new bound and the Enumer-BCRB is a result of the missing spread of the means term. The contribution of the spread of the means term will be small and thus the difference between the two bounds will be small, if the posterior pdf $p(\mathbf{x}_k | \mathbf{z}_{1:k})$, which is a Gaussian mixture for models of the form (1), can be well approximated by a single Gaussian density. However, if the contribution of the spread of the means term is large, then the difference between Enumer-BCRB and the optimal performance will be large. In Fig. 1 (d), the constraint condition of Proposition 2 has been evaluated, in order to check whether the newly proposed bound (\mathcal{M}_1) is superior to the optimal filter covariance averaging approach (\mathcal{M}_3) in terms of MC variance. From the results it can be concluded that the newly proposed bound has a lower MC variance for estimating position and velocity, while for the acceleration the optimal filter covariance averaging approach yields lower MC variance.

VI. DISCUSSION

In this section, the differences between the new bound, the M-BCRB and the Enumer-BCRB are discussed. While the difference between the new bound and the Enumer-BCRB can be explained by the missing spread of the means term, it is not clear how the M-BCRB is related to the new bound and the Enumer-BCRB. It has been shown in the previous section that the Enumer-BCRB may be tighter than the M-BCRB. However, this is not always the case as the following example illustrates.

A JMLGS with scalar state and measurement equation is investigated, where the mode variable $r_k \in \{1, 2\}$ enters only into the process noise. This system can be written as

$$x_k = x_{k-1} + v_k(r_k), \quad (21a)$$

$$z_k = x_k + w_k. \quad (21b)$$

The measurement noise is set to $w_k \sim \mathcal{N}(0, 5)$, the initial state is Gaussian distributed with $x_0 \sim \mathcal{N}(5, 10)$ and the initial mode probabilities are set to $\Pr\{r_1 = 1\} = 1/2$ and $\Pr\{r_1 = 2\} = 1/2$, respectively. The mode-dependent process noise is Gaussian distributed according to $v_k(r_k) \sim \mathcal{N}(\mu_k(r_k), Q_k(r_k))$. In the following simulation study, the process noise of mode $r_k = 1$ is set to $v_k(1) \sim \mathcal{N}(0, 5)$. The mean of the process noise of the second mode $r_k = 2$ is varied and the variance is chosen according to the following two experiments:

- Experiment 1: $v_k(2) \sim \mathcal{N}([0, 40], 5)$,
- Experiment 2: $v_k(2) \sim \mathcal{N}([0, 40], 20)$.

For the ease of explanation, the different bounds introduced in Section V are evaluated only for $k = 1$ and the results are compared to the optimal filter using the direct approach. In Fig. 2, the results for the two different experiments are shown. For both experiments, it can be observed that the MC variance of the new approach (\mathcal{M}_1) is smaller than the MC variance of the direct approach (\mathcal{M}_2). The results further show that for changing $\mu_1(2)$ the Enumer-BCRB is constant, which is a result of the fact that the computation of the Enumer-BCRB is independent of $\mu_1(r_1)$. The differences between the new

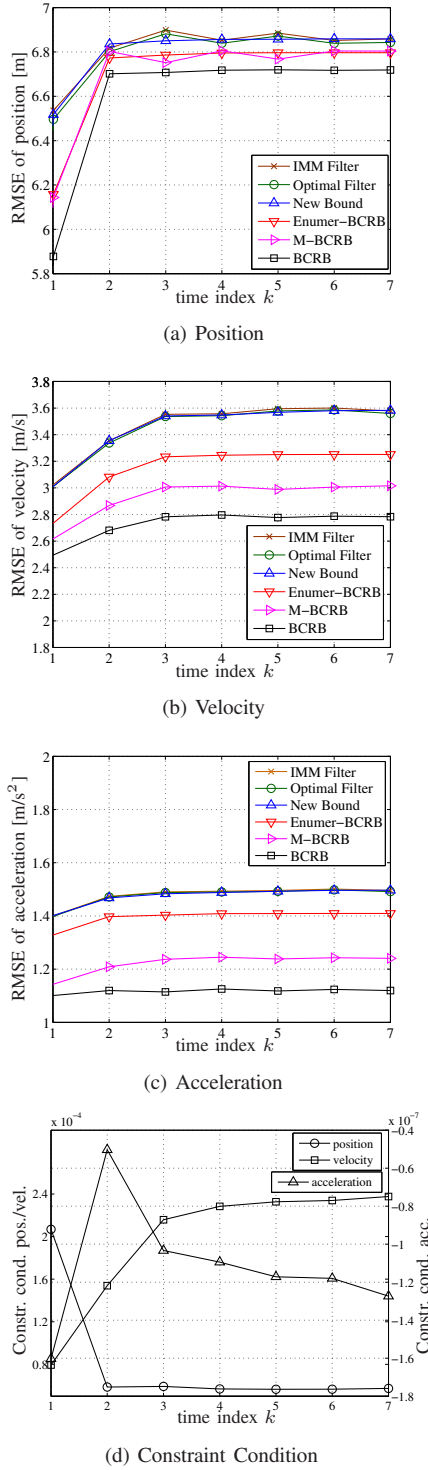


Fig. 1. RMSE vs. time step for (a) Position, (b) Velocity and (c) Acceleration based on $N = 50000$ MC runs, and (d) Constraint condition of Proposition 2

bound and the Enumer-BCRB are due to the missing spread of the means term. Especially for $5 \leq \mu_1(2) \leq 20$, the contribution of the spread of the means term is large and the Enumer-BCRB gives a poor prediction of the optimal performance. However, for $\mu_1(2) \geq 20$ the contribution of the spread of the means term becomes negligible and the Enumer-BCRB is able to predict the optimal filter performance. For Experiment 1, the M-BCRB is tighter than (or equal to) the

Enumer-BCRB. Especially for $\mu_1(2) \leq 7$, the M-BCRB is able to predict the optimal filter performance. However, as $\mu_1(2)$ increases further, this is no longer the case and at approximately $\mu_1(2) \geq 10$ the M-BCRB starts to decrease and approaches the Enumer-BCRB. This behavior can be explained as follows. As $\mu_1(2)$ increases further, the two models $p(z_1|r_1^i) = \mathcal{N}(z_1; 5 + \mu_1(r_1^i), 20)$ become separated in the sense that

$$p(z_1) = \sum_{r_1^n} \Pr\{r_1^n\} p(z_1|r_1^n) \approx \Pr\{r_1^i\} p(z_1|r_1^i), \quad (22)$$

which means that as long as $z_1 \sim p(z_1|r_1^i)$, then $p(z_1|r_1^i) \gg p(z_1|r_1^n)$ must hold for $r_1^n \neq r_1^i$. In this case, it follows that $\Pr\{r_1^i|z_1\} \approx 1$ and $\Pr\{r_1^n|z_1\} \approx 0$ for $r_1^n \neq r_1^i$ must hold, so that the posterior density can be approximated as

$$p(x_1|z_1) = \sum_{r_1} \Pr\{r_1^n|z_1\} p(x_1|z_1, r_1^n) \approx p(x_1|z_1, r_1^i). \quad (23)$$

Inserting (22) and (23) into the definition of \mathbf{J}_k , cf. (6), yields the scalar quantity

$$\begin{aligned} J_1 &= \mathbb{E}_{p(z_1)} \{ \mathbb{E}_{p(x_1|z_1)} \{ \Delta_{x_1}^{x_1} \log(p(x_1|z_1)) \} \} \\ &\approx \mathbb{E}_{\Pr\{r_1\}} \{ \mathbb{E}_{p(x_1|z_1, r_1)} \{ \Delta_{x_1}^{x_1} \log(p(x_1|z_1, r_1)) \} \} \\ &= \mathbb{E}_{\Pr\{r_1\}} \{ J_1(r_1) \}. \end{aligned} \quad (24)$$

Inversion of J_1 gives the M-BCRB, which can be related to the Enumer-BCRB according to Jensen's inequality,

$$[J_1]^{-1} \approx [\mathbb{E}_{\Pr\{r_1\}} \{ J_1(r_1) \}]^{-1} \leq \mathbb{E}_{\Pr\{r_1\}} \{ [J_1(r_1)]^{-1} \}. \quad (25)$$

For Experiment 1, the above relation holds with equality since both, the mode probabilities $\Pr\{r_1\}$ and the conditional Bayesian information submatrices $J_1(r_1)$ are chosen to be equal. In Experiment 2, however, the matrices $J_1(r_1)$ are different, so that the Enumer-BCRB will be tighter than the M-BCRB for large $\mu_1(2)$. For small $\mu_1(2)$, the reverse is true. A generalization of (25) in terms of Bayesian information matrices $\mathbf{J}_{0:k}$ can be found in [4]. A somewhat different interpretation in terms of Bayesian information submatrices \mathbf{J}_k will be given below for the sake of completeness. Assume that for $(\mathbf{x}_k, \mathbf{z}_{1:k}) \sim p(\mathbf{x}_k, \mathbf{z}_{1:k}|r_{1:k}^i)$, the joint density $p(\mathbf{x}_k, \mathbf{z}_{1:k})$ satisfies the following approximation

$$p(\mathbf{x}_k, \mathbf{z}_{1:k}) = \sum_{r_{1:k}^n} \Pr\{r_{1:k}^n\} p(\mathbf{x}_k, \mathbf{z}_{1:k}|r_{1:k}^n) \quad (26)$$

$$\approx \Pr\{r_{1:k}^i\} p(\mathbf{x}_k, \mathbf{z}_{1:k}|r_{1:k}^i), \quad (27)$$

where the conditional joint density is given by

$$p(\mathbf{x}_k, \mathbf{z}_{1:k}|r_{1:k}^i) = p(\mathbf{z}_{1:k}|r_{1:k}^i) p(\mathbf{x}_k|\mathbf{z}_{1:k}, r_{1:k}^i). \quad (28)$$

The approximation in (27) can be interpreted such that if $(\mathbf{x}_k, \mathbf{z}_{1:k}) \sim p(\mathbf{x}_k, \mathbf{z}_{1:k}|r_{1:k}^i)$, then the pdf $p(\mathbf{x}_k, \mathbf{z}_{1:k}|r_{1:k}^i)$ of that particular mode sequence must be much larger than the pdfs $p(\mathbf{x}_k, \mathbf{z}_{1:k}|r_{1:k}^n)$ of all other remaining mode sequences. Inserting (27) and (28) into the definition of the Bayesian information submatrix, cf. (6), yields

$$\begin{aligned} \mathbf{J}_k &\approx \mathbb{E}_{\Pr\{r_{1:k}\}} \{ \mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k}|r_{1:k})} \{ \Delta_{\mathbf{x}_k}^{\mathbf{x}_k} \log p(\mathbf{x}_k|\mathbf{z}_{1:k}, r_{1:k}) \} \} \\ &= \mathbb{E}_{\Pr\{r_{1:k}\}} \{ \mathbf{J}_k(r_{1:k}) \}. \end{aligned} \quad (29)$$

As a result, the M-BCRB and the Enumer-BCRB can be related via Jensen's inequality to each other, yielding

$$[\mathbf{J}_k]^{-1} \approx [\mathbb{E}_{\Pr\{r_{1:k}\}} \{ \mathbf{J}_k(r_{1:k}) \}]^{-1} \leq \mathbb{E}_{\Pr\{r_{1:k}\}} \{ [\mathbf{J}_k(r_{1:k})]^{-1} \}. \quad (30)$$

Note, that the example presented in this section is included in the formulation (27) as a special case. Another important special case occurs, when the measurements are mode-independent and the posterior densities $p(\mathbf{x}_k|\mathbf{z}_{1:k}, r_{1:k}^i)$ are substantially different, i.e.

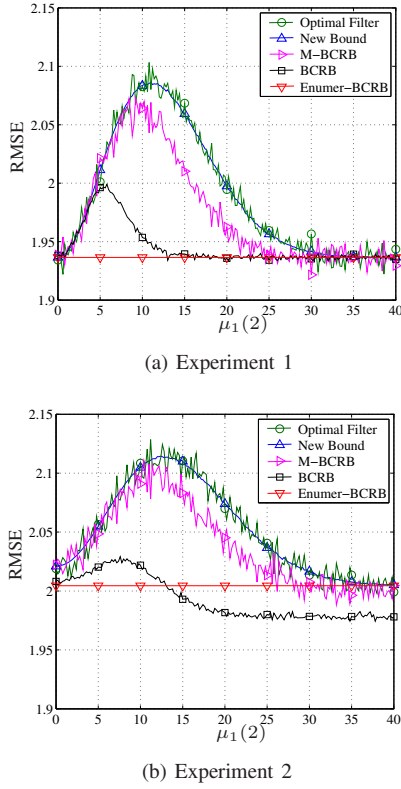


Fig. 2. RMSE at time step $k = 1$ vs. $\mu_1(2)$ assuming (a) $Q_1(2) = 5$ and (b) $Q_1(2) = 20$, based on $N = 50\,000$ MC runs.

if $(\mathbf{x}_k, \mathbf{z}_{1:k}) \sim p(\mathbf{x}_k, \mathbf{z}_{1:k} | r_{1:k}^i)$, then the pdf $p(\mathbf{x}_k | \mathbf{z}_{1:k}, r_{1:k}^i)$ of that particular mode sequence must be much larger than the pdfs $p(\mathbf{x}_k | \mathbf{z}_{1:k}, r_{1:k}^n)$ of all other remaining mode sequences. This case occurs, for instance, when the corresponding covariances differ from each other such that $\mathbf{P}_k^i \ll \mathbf{P}_k^n$ holds for all $r_{1:k}^n \neq r_{1:k}^i$, see [4] for an intuitive example.

VII. CONCLUSION

We have considered various performance bounds for the nonlinear filter problem. Our aim is to lower bound the MSE matrix of the state estimate using MC methods for the particular case of JMLGS models. For N MC runs, we can compute:

- The MSE matrix $\mathcal{M}_2(\hat{\mathbf{x}}_k)$ from any approximative linear filter $\hat{\mathbf{x}}_k$ using the direct approach, cf. (17). We used IMM for illustration.
- The MSE matrix $\mathcal{M}_2(\hat{\mathbf{x}}_k^{\text{opt}})$ for the optimal filter using the direct approach, cf. (17).
- The FIM \mathbf{J}_k using the optimal filter, cf. (6), whose inverse $\mathbf{B}_2 = [\mathbf{J}_k]^{-1}$ gives the M-BCRB.

As $N \rightarrow \infty$, these obviously relate as $\mathcal{M}_2(\hat{\mathbf{x}}_k) > \mathcal{M}_2(\hat{\mathbf{x}}_k^{\text{opt}}) \geq \mathbf{B}_2$. The disadvantage of direct approaches is a huge computational burden and a large variance from the MC integration. There are two remedies suggested in literature. One is based on approximating the BCRB using MC methods, see Section III. This yields a quite conservative bound \mathbf{B}_1 for multimodal posterior distributions, which relates to the M-BCRB according to $\mathbf{B}_2 \geq \mathbf{B}_1$. The Enumer-BCRB computes one term of the expected covariance based on averaging covariances from Kalman filters conditioned on each of the possible mode sequences, see Section III. This term is deterministic and straightforward to compute. However, one positive semidefinite term is neglected in the derivation. We derive the proper expression, and suggest a MC based

approach to compute the second term, see Section IV. Simulation results show, that the newly proposed bound is tighter than the M-BCRB, BCRB and Enumer-BCRB. Furthermore, it is demonstrated that the new bound is able to predict the optimal performance with a much lower MC variance than this is possible using the optimal filter and the direct approach.

ACKNOWLEDGEMENT

This work was partly supported by the Linneaus Center for Control, Autonomy, and Decision-making in Complex Systems (CADICS) funded by the Swedish Research Council (VR).

APPENDIX

Proof of Theorem 1: For notational convenience the estimator's dependency on the measurements is omitted in the following. By making use of the smoothing property of expectations, cf. (10), the expression on the left hand side of (5) can be rewritten as

$$\begin{aligned} \mathcal{M}(\hat{\mathbf{x}}_k) &= \mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k})} \{ [\hat{\mathbf{x}}_k - \mathbf{x}_k][\cdot]^T \} \\ &= \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i\} \mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k} | r_{1:k}^i)} \{ [\hat{\mathbf{x}}_k - \mathbf{x}_k][\cdot]^T \}. \end{aligned} \quad (31)$$

This expression can be further manipulated by replacing $\hat{\mathbf{x}}_k$ on the RHS by the optimal estimator $\hat{\mathbf{x}}_k^{\text{opt}}$ and by additionally including the conditional optimal estimator $\hat{\mathbf{x}}_k^{\text{opt},i}$, yielding

$$\begin{aligned} \mathcal{M}(\hat{\mathbf{x}}_k) &\geq \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i\} \mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k} | r_{1:k}^i)} \left\{ \left[[\hat{\mathbf{x}}_k^{\text{opt}} - \hat{\mathbf{x}}_k^{\text{opt},i}] + [\hat{\mathbf{x}}_k^{\text{opt},i} - \mathbf{x}_k] \right] [\cdot]^T \right\} \\ &= \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i\} \mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k} | r_{1:k}^i)} \{ [\hat{\mathbf{x}}_k^{\text{opt},i} - \mathbf{x}_k][\cdot]^T \} \\ &\quad + \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i\} \mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k} | r_{1:k}^i)} \{ [\hat{\mathbf{x}}_k^{\text{opt}} - \hat{\mathbf{x}}_k^{\text{opt},i}][\hat{\mathbf{x}}_k^{\text{opt},i} - \mathbf{x}_k]^T \} \\ &\quad + \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i\} \mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k} | r_{1:k}^i)} \{ [\hat{\mathbf{x}}_k^{\text{opt},i} - \mathbf{x}_k][\hat{\mathbf{x}}_k^{\text{opt}} - \hat{\mathbf{x}}_k^{\text{opt},i}]^T \} \\ &\quad + \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i\} \mathbb{E}_{p(\mathbf{z}_{1:k} | r_{1:k}^i)} \{ [\hat{\mathbf{x}}_k^{\text{opt}} - \hat{\mathbf{x}}_k^{\text{opt},i}][\cdot]^T \}. \end{aligned} \quad (32)$$

Note, that the expectation w.r.t. $\mathbf{z}_{1:k}$ cannot be dropped, since the estimators depend on the measurements. Further simplification of (32) is possible by taking into account that

$$\begin{aligned} &\mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k} | r_{1:k}^i)} \{ [\hat{\mathbf{x}}_k^{\text{opt}} - \hat{\mathbf{x}}_k^{\text{opt},i}][\hat{\mathbf{x}}_k^{\text{opt},i} - \mathbf{x}_k]^T \} \\ &= \mathbb{E}_{p(\mathbf{z}_{1:k} | r_{1:k}^i)} \{ \mathbb{E}_{p(\mathbf{x}_k | \mathbf{z}_{1:k}, r_{1:k}^i)} \{ [\hat{\mathbf{x}}_k^{\text{opt}} - \hat{\mathbf{x}}_k^{\text{opt},i}][\hat{\mathbf{x}}_k^{\text{opt},i} - \mathbf{x}_k]^T \} \} \\ &= \mathbb{E}_{p(\mathbf{z}_{1:k} | r_{1:k}^i)} \{ \hat{\mathbf{x}}_k^{\text{opt}} \hat{\mathbf{x}}_k^{\text{opt},i,T} - \hat{\mathbf{x}}_k^{\text{opt}} \mathbb{E}_{p(\mathbf{x}_k | \mathbf{z}_{1:k}, r_{1:k}^i)} \{ \mathbf{x}_k^T \} \\ &\quad - \hat{\mathbf{x}}_k^{\text{opt},i} \hat{\mathbf{x}}_k^{\text{opt},i,T} + \hat{\mathbf{x}}_k^{\text{opt},i} \mathbb{E}_{p(\mathbf{x}_k | \mathbf{z}_{1:k}, r_{1:k}^i)} \{ \mathbf{x}_k^T \} \} \\ &= \mathbf{0} \end{aligned} \quad (33)$$

holds, where the last equality follows from the fact that

$$\hat{\mathbf{x}}_k^{\text{opt},i} = \mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k} | r_{1:k}^i)} \{ \mathbf{x}_k \} \quad (34)$$

is satisfied. In a similar manner, it can be shown that

$$\mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k} | r_{1:k}^i)} \{ [\hat{\mathbf{x}}_k^{\text{opt},i} - \mathbf{x}_k][\hat{\mathbf{x}}_k^{\text{opt}} - \hat{\mathbf{x}}_k^{\text{opt},i}]^T \} = \mathbf{0} \quad (35)$$

holds. Note, that the zero equalities in (33) and (35) follow from the projection theorem, which states that the estimate is orthogonal to the estimation error. By further taking into account that the last sum in

(32) can be rearranged using Bayes rule, (32) can be finally rewritten as

$$\begin{aligned} \mathcal{M}(\hat{\mathbf{x}}_k) &\geq \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i\} \mathbb{E}_{p(\mathbf{x}_k, \mathbf{z}_{1:k} | r_{1:k}^i)} \{[\hat{\mathbf{x}}_k^{\text{opt},i} - \mathbf{x}_k][\cdot]^\top\} \\ &\quad + \mathbb{E}_{p(\mathbf{z}_{1:k})} \left\{ \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i | \mathbf{z}_{1:k}\} [\hat{\mathbf{x}}_k^{\text{opt}} - \hat{\mathbf{x}}_k^{\text{opt},i}][\cdot]^\top \right\}. \end{aligned} \quad (36)$$

The first sum in (36) is the contribution of the conditional optimal estimator $\hat{\mathbf{x}}_k^{\text{opt},i}$ to the bound. The second sum in (36) is known as the expected value of the “spread of the means” term which takes into account the deviation between the optimal estimators $\hat{\mathbf{x}}_k^{\text{opt}}$ and $\hat{\mathbf{x}}_k^{\text{opt},i}$. By making use of (9), which holds with equality for the conditional optimal estimator, a new bound for any estimator $\hat{\mathbf{x}}_k$ is then given by:

$$\begin{aligned} \mathcal{M}(\hat{\mathbf{x}}_k) &\geq \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i\} [\mathbf{J}_k(r_{1:k}^i)]^{-1} \\ &\quad + \mathbb{E}_{p(\mathbf{z}_{1:k})} \left\{ \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i | \mathbf{z}_{1:k}\} [\hat{\mathbf{x}}_k^{\text{opt}} - \hat{\mathbf{x}}_k^{\text{opt},i}][\cdot]^\top \right\} \end{aligned} \quad (37)$$

which concludes our proof of Theorem 1. \square

Proof of Proposition 1: For the direct approach, the MC variance can be expressed as

$$\text{Var}_{x,\mathbf{z}}(\mathcal{M}_{2,n}) = \frac{\text{Var}_{x,\mathbf{z}}(g_2(x_{k,n}, \mathbf{z}_{1:k}))}{N}, \quad (38)$$

with $g_2(x_{k,n}, \mathbf{z}_{1:k}) = [\hat{x}_{k,n}^{\text{opt}}(\mathbf{z}_{1:k}) - x_{k,n}]^2$, and where $\text{Var}_{x,\mathbf{z}}(g_2(x_{k,n}, \mathbf{z}_{1:k}))$ is a short-hand notation for $\mathbb{E}_{p(x_{k,n}, \mathbf{z}_{1:k})} \{ (g_2(x_{k,n}, \mathbf{z}_{1:k}) - \mathbb{E}_{p(x_{k,n}, \mathbf{z}_{1:k})} \{g_2(x_{k,n}, \mathbf{z}_{1:k})\})^2 \}$. Similarly, the MC variance of the proposed approach is given by

$$\text{Var}_{\mathbf{z}}(\mathcal{M}_{1,n}) = \frac{\text{Var}_{\mathbf{z}}(g_1(\mathbf{z}_{1:k}))}{N}, \quad (39)$$

with $g_1(\mathbf{z}_{1:k}) = \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i | \mathbf{z}_{1:k}\} [\hat{x}_{k,n}^{\text{opt}}(\mathbf{z}_{1:k}) - \hat{x}_{k,n}^{\text{opt}}(\mathbf{z}_{1:k}, r_{1:k}^i)]^2$. The numerator of (38) can be rewritten by using the law of total variance, yielding

$$\begin{aligned} \text{Var}_{x,\mathbf{z}}(g_2(x_{k,n}, \mathbf{z}_{1:k})) &= \mathbb{E}_{\mathbf{z}} \{ \text{Var}_{x|\mathbf{z}}(g_2(x_{k,n}, \mathbf{z}_{1:k})) \} \\ &\quad + \text{Var}_{\mathbf{z}}(\mathbb{E}_{x|\mathbf{z}}(g_2(x_{k,n}, \mathbf{z}_{1:k}))). \end{aligned} \quad (40)$$

We note that $\mathbb{E}_{x|\mathbf{z}}(g_2(x_{k,n}, \mathbf{z}_{1:k}))$ is the optimal filter variance for estimating $x_{k,n}$. Hence, we can write

$$\text{Var}_{\mathbf{z}}(\mathbb{E}_{x|\mathbf{z}}(g_2(x_{k,n}, \mathbf{z}_{1:k}))) = \text{Var}_{\mathbf{z}}(g_3(\mathbf{z}_{1:k})), \quad (41)$$

where $g_3(\mathbf{z}_{1:k}) = g_4(\mathbf{z}_{1:k}) + g_1(\mathbf{z}_{1:k})$, with $g_4(\mathbf{z}_{1:k}) = \sum_{i=1}^{s^k} \Pr\{r_{1:k}^i | \mathbf{z}_{1:k}\} [\mathbf{P}_k^i]_{n,n}$. We further take into account that $g_4(\mathbf{z}_{1:k})$ can be replaced by the equivalent deterministic expression $\sum_{i=1}^{s^k} \Pr\{r_{1:k}^i\} [\mathbf{J}_k^i]_{n,n}$, yielding

$$\text{Var}_{x,\mathbf{z}}(g_2(x_{k,n}, \mathbf{z}_{1:k})) = \mathbb{E}_{\mathbf{z}} \{ \text{Var}_{x|\mathbf{z}}(g_2(x_{k,n}, \mathbf{z}_{1:k})) \} + \text{Var}_{\mathbf{z}}(g_1(\mathbf{z}_{1:k})). \quad (42)$$

Since $\mathbb{E}_{\mathbf{z}} \{ \text{Var}_{x|\mathbf{z}}(g_2(x_{k,n}, \mathbf{z}_{1:k})) \} \geq 0$, we can finally prove that (18) holds. \square

Proof of Proposition 2: The MC variance of $\mathcal{M}_{3,n}$ is given by

$$\text{Var}_{\mathbf{z}}(\mathcal{M}_{3,n}) = \frac{\text{Var}_{\mathbf{z}}(g_3(\mathbf{z}_{1:k}))}{N}, \quad (43)$$

where $g_3(\mathbf{z}_{1:k}) = g_4(\mathbf{z}_{1:k}) + g_1(\mathbf{z}_{1:k})$, with $g_4(\mathbf{z}_{1:k})$ and $g_1(\mathbf{z}_{1:k})$ defined in (41) and (39). The numerator of (43) can be equivalently

expressed as

$$\begin{aligned} \text{Var}_{\mathbf{z}}(g_4(\mathbf{z}_{1:k}) + g_1(\mathbf{z}_{1:k})) &= \text{Var}_{\mathbf{z}}(g_4(\mathbf{z}_{1:k})) + \text{Var}_{\mathbf{z}}(g_1(\mathbf{z}_{1:k})) \\ &\quad + 2 \cdot \text{Cov}_{\mathbf{z}}(g_4(\mathbf{z}_{1:k}), g_1(\mathbf{z}_{1:k})). \end{aligned} \quad (44)$$

Under the constraint that

$\text{Var}_{\mathbf{z}}(g_4(\mathbf{z}_{1:k})) + 2 \cdot \text{Cov}_{\mathbf{z}}(g_4(\mathbf{z}_{1:k}), g_1(\mathbf{z}_{1:k})) \geq 0$ is fulfilled, we can finally prove that (20) holds. \square

REFERENCES

- [1] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York, NY, USA: Wiley-Interscience, 2001.
- [2] F. Gustafsson, *Adaptive Filtering and Change Detection*. New York, NY, USA: John Wiley & Sons, 2000.
- [3] H. A. P. Blom and Y. Bar-Shalom, “The interacting multiple model algorithm for systems with Markovian switching coefficients,” *IEEE Trans. Autom. Control*, vol. 33, no. 8, pp. 780–783, 1988.
- [4] L. Svensson, “On the Bayesian Cramér-Rao bound for Markovian switching systems,” *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4507–4516, Sept. 2010.
- [5] 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.
- [6] H. L. van Trees, *Detection, Estimation and Modulation Theory Part I*. New York, NY, USA: John Wiley & Sons, 1968.
- [7] A. Bessell, B. Ristic, A. Farina, X. Wang, and M. S. Arulampalam, “Error performance bounds for tracking a manoeuvring target,” in *Proc. of the International Conference on Information Fusion*, vol. 1, Cairns, Queensland, Australia, Jul. 2003, pp. 903–910.
- [8] 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.
- [9] G. A. Ackerson and K. S. Fu, “On state estimation in switching environments,” *IEEE Trans. Autom. Control*, vol. 15, no. 1, pp. 10–17, 1970.
- [10] F. Gustafsson, *Statistical Sensor Fusion*. Lund, Sweden: Studentlitteratur AB, 2010.
- [11] 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.
- [12] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Boston, MA, USA: Artech-House, 2004.
- [13] N. Bergman, “Recursive Bayesian estimation: Navigation and tracking applications,” Ph.D. dissertation, Linköping University, Linköping, Sweden, 1999.
- [14] M. L. Hernandez, B. Ristic, and A. Farina, “A performance bound for manoeuvring target tracking using best-fitting Gaussian distributions,” in *Proc. of International Conference on Information Fusion*, Philadelphia, PA, USA, Jul. 2005, pp. 1–8.