Conditional Posterior Cramér-Rao Lower Bounds for Nonlinear Recursive Filtering

Long Zuo, Ruixin Niu and Pramod K. Varshney
Dept. of Electrical Engineering and Computer Science, Syracuse University
Syracuse, NY, 13244, U.S.A.
{lzuo, rniu, varshney}@syr.edu

Abstract – Posterior Cramér Rao lower bounds (PCRLBs) [1] for sequential Bayesian estimators provide performance bounds for general nonlinear filtering problems and have been used widely for sensor management in tracking and fusion systems. However, the unconditional PCRLB [1] is an off-line bound that is obtained by taking the expectation of the Fisher information matrix (FIM) with respect to the measurement and the state to be estimated. In this paper, we introduce a new concept of conditional PCRLB, which is dependent on the observation data up to the current time, and adaptive to a particular realization of the system state. Therefore, it is expected to provide more accurate and effective performance evaluation than the conventional unconditional PCRLB. However, analytical computation of this new bound is, in general, intractable except when the system is linear and Gaussian. In this paper, we present a sequential Monte Carlo solution to compute the conditional PCRLB for nonlinear non-Gaussian sequential Bayesian estimation problems.

Keywords: Posterior Cramér Rao Lower Bounds, Bayesian Estimation, Particle Filters, Kalman Filters.

1 Introduction

The conventional Cramér-Rao lower bound (CRLB) [2] on the mean-squared error (MSE) provides the performance limit for any unbiased estimator of a fixed parameter. For a random parameter, Van Trees presented an analogous bound, the posterior CRLB (PCRLB) [2], or Bayesian CRLB, which shows that

\[ E[ \hat{x}(z) - x] (\hat{x}(z) - x)^T \geq J^{-1} \]  

(1)

where state \( x \) is a random vector to be estimated and \( z \) is the observation. \( n_x \) and \( n_z \) are the dimensions for each vector. \( \hat{x}(z) \) is an estimator of \( x \), which is a function of \( z \). The inequality means that \( E[ \hat{x}(z) - x] (\hat{x}(z) - x)^T - J^{-1} \) is a positive semidefinite matrix. \( J \) is the Fisher information matrix (FIM)

\[ J = E[ -\Delta x^T \log p(x, z)] \]  

(2)

And all the above expectations are taken with respect to \( p(x, z) \), which is the joint PDF of the pair \( (x, z) \). \( \Delta \) denotes the second derivative operator, namely

\[ \Delta x = \nabla_x \nabla_y^T \]  

(3)

in which \( \nabla \) denotes the gradient operator

\[ \nabla_x = \left[ \frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \right]^T \]  

(4)

The sequential Bayesian estimation problem is to find the estimate of the state from the measurements (observations) over time. The evolution of the state sequence \( x_k \) is assumed to be an unobserved first order Markov process, and the equation for such discrete-time stochastic process is often given by

\[ x_{k+1} = f_k(x_k, u_k) \]  

(5)

where \( f_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_x} \) is in general a nonlinear function of state \( x \), and \( \{u_k, k \in \{0\} \cup \mathbb{N}\} \) is an independent white process noise. \( n_u \) is the dimensions of the noise vector, and \( \mathbb{N} \) is the set of the natural numbers. The PDF of the initial state \( x_0 \) is assumed to be known.

The observations about the state are obtained from the measurement equation

\[ z_k = h_k(x_k, v_k) \]  

(6)

where \( h_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \rightarrow \mathbb{R}^{n_z} \) is in general a nonlinear function, and \( \{v_k, k \in \mathbb{N}\} \) is the measurement noise sequence, which is independent of \( x_k \) as well as \( u_k \).

If we denote the states and measurements up to time \( k \) as \( x_{0:k} \) and \( z_{1:k} \), then the joint pdf of \( (x_{0:k}, z_{1:k}) \) can be determined from (5) and (6) with known initial pdf \( p(x_0) \) and noise models for \( u_k \) and \( v_k \)

\[ p(x_{0:k}, z_{1:k}) = p(x_0) \prod_{i=1}^{k} p(x_{i}|x_{i-1}) \prod_{j=1}^{k} p(z_j|x_j) \]  

(7)
If we consider $x_{0:k}$ as a vector with dimension $(k + 1)n_x$, and define $J(x_{0:k})$ to be the $(k + 1)n_x \times (k + 1)n_x$ FIM of $x_{0:k}$ derived from the joint PDF $p(x_{0:k}, z_{1:k})$, (1) becomes

$$E[x_{0:k}(z_{1:k}) - x_{0:k}]^T (x_{0:k}(z_{1:k}) - x_{0:k}) \geq J^{-1}(x_{0:k})$$

Let us define $J_k$ as the information submatrix of $x_k$, and the inverse of $J_k$ is equal to the $n_x \times n_x$ lower-right submatrix of $J^{-1}(x_{0:k})$. Then, the MSE of the estimate for $x_k$ is bounded by $J_k^{-1}$. $J_k$ can be obtained directly from the computed inverse of the $(k + 1)n_x \times (k + 1)n_x$ matrix $J(x_{0:k})$. However, this is not an efficient approach. In [1], Tichavsky et al. provide an elegant recursive approach to calculate $J_k$ without manipulating the large matrices at each time $k$.

$$J_{k+1} = D_k^{21} - D_k^{12} (J_k + D_k^{11})^{-1} D_k^{12}$$

where

$$D_k^{11} = E\{-\Delta x_k \log p(x_{k+1}|x_k)\}$$
$$D_k^{12} = E\{-\Delta x_k \log p(x_{k+1}|x_k)\}$$
$$D_k^{21} = E\{-\Delta x_k \log p(x_{k+1}|x_k)\} = (D_k^{12})^T$$
$$D_k^{22} = E\{-\Delta x_k \log p(x_{k+1}|x_k) + \log p(z_{k+1}|x_{k+1})\}$$

The PCRLB can be used to solve the sensor management problems in multi-sensor systems. In [3], based on the PCRLB, a sensor deployment approach is developed to achieve better tracking accuracy while at the same time uses the limited sensor resources more efficiently. Further, PCRLB based criterion has been employed to manage sensor arrays for multi-target tracking problems [4, 5].

However, for the unconditional PCRLB, the FIM is derived by taking the expectation with respect to the joint distribution of the measurements and the target states up to the current time $k$. As a result, the very useful measurement information is averaged out and the unconditional PCRLB becomes an off-line bound. And the unconditional PCRLB can not reflect the target tracking performance for a particular track realization very faithfully. This is especially true when the uncertainty in the state model (or equivalently the state process noise) is high and thus the prior knowledge regarding the target state at the initial time quickly becomes irrelevant as the target state evolves over time.

To take advantage of the available measurements, we have derived a conditional PCRLB that is dependent on the past data. The conditional PCRLB provides a bound on the conditional MSE of the target state estimate, based on the measurements up to the current time.

2 Conditional PCRLB for the Nonlinear Dynamic Systems

Conditional PCRLB sets a bound on the performance of estimating $x_{0:k+1}$ when the new measurement $z_{k+1}$ becomes available given that the past measurements up to time $k$ are all known.

**Definition 1** $\hat{x}_{0:k+1}(z_{1:k+1}|z_{1:k})$ is a conditional estimator, and defined as a function of the observed data $z_{k+1}$ given the existing measurements $z_{1:k}$.

**Definition 2** Mean squared error of the conditional estimator at time $k + 1$ is defined as follows

$$MSE(\hat{x}_{0:k+1}(z_{1:k+1}|z_{1:k})) = E\{\hat{x}_{0:k+1}^T(z_{1:k+1}|z_{1:k})\}$$

where $\hat{x}_{0:k+1} = E\{p(x_{0:k+1}, z_{1:k+1}|z_{1:k})dxdz_{k+1}$

**Definition 3** Let $I(x_{0:k+1}|z_{1:k})$ be the $(k + 1)n_x \times (k + 1)n_x$ conditional Fisher information matrix of the state vector $x_{0:k+1}$ from time 0 to $k + 1$:

$$I(x_{0:k+1}|z_{1:k}) = E\{\Delta z_{0:k+1} p(x_{0:k+1}, z_{1:k+1}|z_{1:k})\}$$

where $\Delta z_{0:k+1} = E\{\Delta z_{0:k+1} p(x_{0:k+1}, z_{1:k+1}|z_{1:k})\}$

Theorem 1 The conditional mean squared error of the state vector $x_{0:k+1}$ is lower bounded by the inverse of the conditional Fisher information matrix

$$E\{\hat{x}_{0:k+1}^T(z_{1:k+1}|z_{1:k})\} \geq I^{-1}(x_{0:k+1}|z_{1:k})$$

**Proof:** The following conditions are assumed to exist:

1. Assume $\partial p(x_{0:k+1}, z_{1:k+1}|z_{1:k})$ and $\partial^2 p(x_{0:k+1}, z_{1:k+1}|z_{1:k})$ exist and both are absolutely integrable with respect to $x_{0:k+1}$ and $z_{k+1}$. Then for any statistic $T$ such that $E\{p(x_{0:k+1}, z_{1:k+1}|z_{1:k})\} = \infty$, the operation of integration and differentiation by $x_i$ can be interchanged in $\int T p(x_{0:k+1}, z_{1:k+1}|z_{1:k})dx_{0:k+1}dz_{k+1}$. That is

$$\frac{\partial}{\partial x_i} \int T p(x_{0:k+1}, z_{1:k+1}|z_{1:k})dx_{0:k+1}dz_{k+1} = \int T \frac{\partial p(x_{0:k+1}, z_{1:k+1}|z_{1:k})}{\partial x_i} dx_{0:k+1}dz_{k+1}$$

2. $x_i$ is defined over the compact interval $[a_i, b_i]$, and for $i = 1, \ldots, (k + 2)n_x$

$$\lim_{x_i \to a_i} p(x_{0:k+1}) = \lim_{x_i \to b_i} p(x_{0:k+1}) = 0$$

$$\lim_{x_i \to a_i} \partial p(x_{0:k+1}) = \lim_{x_i \to b_i} \partial p(x_{0:k+1}) = 0$$
Under the above assumptions, we have
\[
\frac{\partial}{\partial x_i} \int p(x_{0:k+1}, z_{k+1}|z_{1:k}) dx_{0:k+1} dz_{k+1} = 0 \quad (15)
\]
\[
\int \hat{x}_i \frac{\partial p(x_{0:k+1}, z_{k+1}|z_{1:k})}{\partial x_i} dx_{0:k+1} dz_{k+1} = 0 \quad (16)
\]
where \(\hat{x}_i\) stands for the estimate of \(x_i\). Also
\[
\int_{a_i}^{b_i} x_i \frac{\partial p(x_{0:k+1}, z_{k+1}|z_{1:k})}{\partial x_i} dx_i = -\int_{a_i}^{b_i} p(x_{0:k+1}, z_{k+1}|z_{1:k}) dx_i \quad (17)
\]
Then subtracting (18) from (16), it yields
\[
\int (\hat{x}_i - x_i) \frac{\partial p(x_{0:k+1}, z_{k+1}|z_{1:k})}{\partial x_i} dx_{0:k+1} dz_{k+1} = 1 \quad (19)
\]
Also we have the following identity
\[
\int (\hat{x}_i - x_i) \frac{\partial \log p(x_{0:k+1}, z_{k+1}|z_{1:k})}{\partial x_i} \times p(x_{0:k+1}, z_{k+1}|z_{1:k}) dx_{0:k+1} dz_{k+1} = 1 \quad (20)
\]
Plugging the above equation into (19)
\[
\frac{\partial \log p(x_{0:k+1}, z_{k+1}|z_{1:k})}{\partial x_i} = \frac{\frac{\partial p(x_{0:k+1}, z_{k+1}|z_{1:k})}{\partial x_i}}{p(x_{0:k+1}, z_{k+1}|z_{1:k})} \quad (21)
\]
Now consider for \(i \neq j\), we have
\[
\int (\hat{x}_i - x_i) \frac{\partial p(x_{0:k+1}, z_{k+1}|z_{1:k})}{\partial x_j} dx_{0:k+1} dz_{k+1} = 0 \quad (22)
\]
and
\[
\int (\hat{x}_i - x_i) \frac{\partial \log p(x_{0:k+1}, z_{k+1}|z_{1:k})}{\partial x_j} p(x_{0:k+1}, z_{k+1}|z_{1:k}) dx_{0:k+1} dz_{k+1} = 0 \quad (23)
\]
Combining (20) and (23) into matrix form, we have
\[
\int (x_{0:k+1} - x_{0:k+1}) \left[ \nabla^T_{x_{0:k+1}} \log p(x_{0:k+1}, z_{k+1}|z_{1:k}) \right] \times p(x_{0:k+1}, z_{k+1}|z_{1:k}) dx_{0:k+1} dz_{k+1} = I_{k+2} \quad (24)
\]
where \(I_{k+2}\) is an identity matrix with dimension \((k + 2) n_x\). Now pre-multiply by \(a^T\) and postmultiply by \(b\), where \(a\) and \(b\) are arbitrary column vectors with dimension \((k + 2) n_x\), we have
\[
\int a^T (x_{0:k+1} - x_{0:k+1}) \left[ \nabla^T_{x_{0:k+1}} \log p(x_{0:k+1}, z_{k+1}|z_{1:k}) \right] \times p(x_{0:k+1}, z_{k+1}|z_{1:k}) b dx_{0:k+1} dz_{k+1} = a^T b \quad (25)
\]
Applying the Cauchy-Schwarz inequality,
\[
(a^T b)^2 \leq a^T MSE(x_{0:k+1}|z_{1:k}) a \times b^T I(x_{0:k+1}|z_{1:k}) b
\]
where \(I(x_{0:k+1}|z_{1:k})\) is the conditional Fisher information, and is defined as
\[
I(x_{0:k+1}|z_{1:k}) = E \left\{ -\left. \nabla^T_{x_{0:k+1}} \log p(x_{0:k+1}, z_{k+1}|z_{1:k}) \right| z_{1:k} \right\}
\]
and \(MSE(x_{0:k+1}|z_{1:k})\) is the conditional mean squared error of the estimator \(\hat{x}_{0:k+1}\), and is defined as
\[
MSE(\hat{x}_{0:k+1}|z_{1:k}) = E \left\{ (\hat{x}_{0:k+1} - x_{0:k+1})(\hat{x}_{0:k+1} - x_{0:k+1})^T \right| z_{1:k} \right\}
\]
Note that in both (26) and (27), the expectation is taken with respect to \(p(x_{0:k+1}, z_{k+1}|z_{1:k})\). Since \(b\) is arbitrary, let
\[
b = I^{-1}(x_{0:k+1}|z_{1:k}) a
\]
then
\[
(a^T I^{-1}(x_{0:k+1}|z_{1:k}) a)^2 \leq a^T MSE(\hat{x}_{0:k+1}|z_{1:k}) a \times (a^T I^{-1}(x_{0:k+1}|z_{1:k}) a)
\]
Because \(I(x_{0:k+1}|z_{1:k})\) is real symmetric matrix, by applying Cauchy-Schwarz inequality again, it is easy to show that \(I(x_{0:k+1}|z_{1:k})\) is also positive definite, so \(I^{-1}(x_{0:k+1}|z_{1:k})\) is also positive definite. Then
\[
a^T I^{-1}(x_{0:k+1}|z_{1:k}) a \geq 0
\]
Therefor, we have
\[
a^T \left( MSE(\hat{x}_{0:k+1}|z_{1:k}) - I^{-1}(x_{0:k+1}|z_{1:k}) \right) a \geq 0
\]
Since vector \(a\) is arbitrary, \(MSE(\hat{x}_{0:k+1}|z_{1:k}) - I^{-1}(x_{0:k+1}|z_{1:k})\) is positive semidefinite. Q.E.D.

**Definition 4** \(I(x_{k+1}|z_{1:k})\) is defined as the conditional Fisher information matrix for estimating \(x_{k+1}\), and \(I^{-1}(x_{k+1}|z_{1:k})\) is equal to the right-lower block of \(I^{-1}(x_{0:k+1}|z_{1:k})\).

By definition, \(I^{-1}(x_{k+1}|z_{1:k})\) is a bound on the MSE of the estimate for \(x_{k+1}\) given \(z_{1:k}\). And we propose an iterative approach to calculate \(I^{-1}(x_{k+1}|z_{1:k})\) without manipulating the large matrix \(I(x_{0:k+1}|z_{1:k})\).

**Definition 5** Auxiliary Fisher information matrix for the state vector from time 0 to \(k\) is defined as
\[
I_A(x_{0:k}|z_{1:k}) \triangleq E_{p(x_{0:k}|z_{1:k})} \left\{ -\Delta^T_{0:k} \log p(x_{0:k}|z_{1:k}) \right\} = -\int \left[ \Delta^T_{0:k} \log p(x_{0:k}|z_{1:k}) \right] p(x_{0:k}|z_{1:k}) dx_{0:k}
\]
Definition 6 We also define $I_A(x_k|z_{1:k})$ as the auxiliary Fisher information sub-matrix for $x_k$, and $I_A^{-1}(x_k|z_{1:k})$ is equal to the $n_x \times n_x$ lower-right block of $I_A^{-1}(x_0:k|z_{1:k})$.

The matrix inversion Lemma [6] is very useful for the derivations in this paper and it is provided as follows:

Lemma 1 The inversion equality for a symmetric matrix is
\[
\begin{bmatrix}
A & B \\
B^T & C
\end{bmatrix}^{-1} = \begin{bmatrix}
D^{-1} & -A^{-1}BE^{-1} \\
-E^{-1}BT & A^{-1}
\end{bmatrix} \tag{33}
\]
where $A$, $B$, and $C$ are sub-matrices with appropriate dimensions, and $D = A-BC^{-1}B^T$, $E = C-B^TA^{-1}B$.

The following theorem gives an approach for recursively computing $I_A(x_k|z_{1:k})$. The derivation as given in the proof requires the matrix inversion lemma (Lemma 1).

Theorem 2 The sequence of $I_A(x_k|z_{1:k})$ can be computed recursively as follows
\[
I_A(x_k|z_{1:k}) \approx S_k^{22} - S_k^{21} \left[ S_k^{11} + I_A(x_{k-1}|z_{1:k-1}) \right]^{-1} S_k^{12}
\] \tag{34}
where
\[
S_k^{11} = E_p(x_{0:k}|z_{1:k}) \left[ -\Delta x_k \log p(x_k|x_{k-1}) \right]
\]
\[
S_k^{12} = E_p(x_{0:k}|z_{1:k}) \left[ -\Delta x_k \log p(x_k|x_{k-1}) \right]
\]
\[
S_k^{21} = E_p(x_{0:k}|z_{1:k}) \left[ -\Delta x_k \log p(x_k|x_{k-1}) \right] = (S_k^{21})^T
\]
\[
S_k^{22} = E_p(x_{0:k}|z_{1:k}) \left[ -\Delta x_k \log p(x_k|x_{k-1}) + \log p(z_k|x_k) \right]
\]

Proof: Decompose $I_A(x_{0:k-1}|z_{1:k-1})$ as
\[
I_A(x_{0:k-1}|z_{1:k-1}) = -E_p(x_{0:k-1}|z_{1:k-1}) \begin{bmatrix}
\Delta x_{k-1} \\
\Delta x_k
\end{bmatrix} \log p_k^{a}
\]
where $p_k^{a} \triangleq p(x_{0:k-1}|z_{1:k-1})$. Then applying Lemma 1, we have
\[
I_A(x_{k-1}|z_{1:k-1}) = A_k^{22} - A_k^{21} (A_k^{-1})^{-1} A_k^{12}
\] \tag{35}
Now consider $I_A(x_0:k|z_{1:k})$. We have
\[
I_A(x_0:k|z_{1:k}) = E_p(x_{0:k}|z_{1:k}) \left[ -1 \right]
\]
\[
\begin{bmatrix}
\Delta x_{k-2} \\
\Delta x_{k-1} \\
\Delta x_k
\end{bmatrix} \log p_k^{a}
\]
where $0$s stand for blocks of zeros of appropriate dimensions. The top-left sub-matrix of $I_A(x_0:k|z_{1:k})$ is a function of $z_k$, which can be estimated by its expectation with respect to $p(z_k|x_{1:k-1})$, if we take $z_k$ and $z_{1:k-1}$ as random vector and measurement realizations respectively. So we have
\[
E_p(x_{0:k}|z_{1:k}) \left[ -\Delta x_{k-2} \log p(x_{0:k}|z_{1:k}) \right]
\]
\[
\approx E_p(x_{0:k}|z_{1:k-1}) \left\{ E_p(x_{0:k}|z_{1:k}) \left[ -\Delta x_{k-2} \log p_k^{a} \right] \right\}
\]
\[
= A_k^{11}
\] \tag{37}
where the use of (7) has been made. The other elements in the top-left sub-matrix can be derived following a similar procedure, and finally we have
\[
I_A(x_0:k|z_{1:k}) \approx \begin{bmatrix}
A_k^{11} & A_k^{12} & 0 \\
A_k^{21} & A_k^{22} + S_k^{11} & S_k^{12} \\
0 & S_k^{21} & S_k^{22}
\end{bmatrix}
\] \tag{38}
Because the auxiliary information sub-matrix $I_A(x_k|z_{1:k})$ is equal to the inverse of the right-lower block of $I_A^{-1}(x_0:k|z_{1:k})$, we have
\[
I_A(x_k|z_{1:k}) \approx S_k^{22} - \begin{bmatrix}
S_k^{21} & I_A(x_{k-1}|z_{1:k-1})
\end{bmatrix}^{-1} S_k^{12}
\]
Q.E.D.

The auxiliary Fisher information $I_A(x_k|z_{1:k})$ will be used for deriving the conditional Fisher information matrix, which is given in the following theorem.

Theorem 3 The sequence of conditional Fisher information $\{I(x_{k+1}|z_{1:k})\}$ for estimating state vectors $\{x_{k+1}\}$ can be recursively computed as follows
\[
I(x_{k+1}|z_{1:k}) = B_k^{12} - B_k^{21} \left[ B_k^{11} + I_A(x_k|z_{1:k}) \right]^{-1} B_k^{12}
\] \tag{39}
where
\[
B_k^{11} = E_p(x_{k+1}|z_{k+1}) \left[ -\Delta x_k \log p(x_{k+1}|x_k) \right]
\] \tag{40}
\[
B_k^{12} = E_p(x_{k+1}|z_{k+1}) \left[ -\Delta x_{k+1} \log p(x_{k+1}|x_k) \right]
\] \tag{41}
\[
B_k^{21} = E_p(x_{k+1}|z_{k+1}) \left[ -\Delta x_{k+1} \log p(x_{k+1}|x_k) \right] = (B_k^{12})^T
\] \tag{42}
\[
B_k^{22} = E_p(x_{k+1}|z_{k+1}) \left[ -\Delta x_{k+1} \log p(x_{k+1}|x_k) + \log p(z_{k+1}|x_k) \right]
\] \tag{43}
\[
p_k^{c} \triangleq p(x_{k+1},z_{k+1}|z_{1:k})
\] \tag{44}

Proof: The conditional Fisher information matrix can be decomposed as follows
\[
I(x_{0:k+1}|z_{1:k}) = E_p(x_{0:k+1}|z_{1:k}) \left[ -1 \right]
\]
\[
\begin{bmatrix}
\Delta x_{k-1} & \Delta x_k & 0 \\
\Delta x_{k-1} & \Delta x_k & 0 \\
\Delta x_k & \Delta x_k & \Delta x_k
\end{bmatrix} \log p_k^{a}
\]
where $0$s stand for blocks of zeros of appropriate dimensions. The top-left sub-matrix of $I(x_0:k+1|z_{1:k})$ is
where
\[ C_{k}^{11} = E_{p_{k}^{c+1}} \left[ -\Delta x_{0:k-1} \log p_{k+1}^{c+1} \right] = A_{k}^{11} \] (46)

Note that the second step comes from the identity of
\[ p_{k+1}^{c+1} = p(x_{0:k} | z_{1:k}) p(x_{k+1} | x_{k}) p(z_{k+1} | x_{k+1}) \]
In a similar manner, \( C_{k}^{22} \) can be derived as
\[ C_{k}^{22} = \left( C_{k}^{12} \right)^{T} = A_{k}^{21} \] (48)

The conditional Fisher information sub-matrix \( I(x_{k+1} | z_{1:k}) \) can be found as the inverse of the right-lower sub-matrix of \( I^{-1}(x_{0:k+1} | z_{1:k}) \). So
\[
I(x_{k+1} | z_{1:k}) = B_{k+1}^{22} - B_{k+1}^{21} [A_{k}^{11} - A_{k}^{12} A_{k}^{21}]^{-1} B_{k+1}^{12} \] (51)

Using (35), we finally have
\[
I(x_{k+1} | z_{1:k}) = B_{k+1}^{22} - B_{k+1}^{21} (I_{A}(x_{k} | z_{1:k}))^{-1} B_{k+1}^{12} \]

Q.E.D.

According to Theorem 1, the inverse of \( I(x_{k+1} | z_{1:k}) \), which is the lower-right block of \( I^{-1}(x_{0:k+1} | z_{1:k}) \), sets a lower bound on the MSE of estimating \( x_{k+1} \) conditioned on the past measurements \( z_{1:k} \).

### 3 A Sequential Monte Carlo solution for Conditional PCRLB

In Section 2, we have shown that given the available measurement data \( z_{1:k} \), the conditional Fisher information matrix \( I(x_{k+1} | z_{1:k}) \) can be recursively calculated according to Theorems 2 and 3. However, in most cases, direct computation of \( B_{k+1}^{21} \) involves high-dimensional integration, and in general analytical solutions do not exist. Here sequential Monte Carlo methods, or particle filters [7][8], are proposed to evaluate these terms for nonlinear/non-Gaussian dynamic systems. And the proposed particle filter based conditional PCRLB evaluation solution is very convenient, since the conditional bound can be evaluated online as a by-product of the particle filter state estimation process, as shown later in the paper.

Under the assumptions that the states evolve according to a first-order Markov process and the observations are conditionally independent given the states, the joint PDF up to time \( k+1 \) can be factorized as
\[
p(x_{0:k+1}, z_{1:k+1}) = p(z_{k+1} | x_{k+1}) p(x_{k+1} | x_{k}) \times p(x_{0:k} | z_{1:k}) p(z_{1:k}) \] (50)

Letting \( N \) denote the number of particles used in the particle filter, the posterior pdf \( p(x_{0:k} | z_{1:k}) \) at time \( k \) can be approximated by the particles
\[
p(x_{0:k} | z_{1:k}) \approx \frac{1}{N} \sum_{l=1}^{N} \delta(x_{0:k} - x_{0:k}^{l}) \] (51)

where we assume that the resampling has been performed at time \( k \), so that each particle has an identical weight \( \frac{1}{N} \). With (50) and (51),
\[
p(x_{0:k+1}, z_{k+1} | z_{1:k}) \approx \frac{1}{N} \sum_{l=1}^{N} \delta(x_{k+1} - x_{k+1}^{l}) p(x_{k+1} | x_{k}) \sum_{l=1}^{N} \delta(x_{0:k} - x_{0:k}^{l}) \] (52)

We also derive another approximation for \( p(x_{0:k+1}, z_{k+1} | z_{1:k}) \), which is given by the following lemma.

**Lemma 2**
\[
p(x_{0:k+1}, z_{k+1} | z_{1:k}) \approx \frac{1}{N} \sum_{l=1}^{N} \delta(x_{0:k+1} - x_{0:k+1}^{l}) p(x_{k+1} | x_{k+1}^{l}) \] (53)

**Proof:** The joint pdf up to time \( k+1 \) can be factorized as
\[
p(x_{0:k+1}, z_{1:k+1}) = p(x_{0:k+1} | z_{1:k+1}) p(z_{k+1} | z_{1:k}) p(z_{1:k}) \] (54)

If the transition density of the state \( p(x_{k+1} | x_{k}) \) is chosen as the importance density function [7], then the weights at time \( k+1 \) are given by
\[
\omega_{k+1}^{l} \propto \omega_{k}^{l} p(z_{k+1} | x_{k+1}^{l}) \] (55)

However, suppose the resampling is taken at time \( k \), we will have \( \omega_{k}^{l} = 1/N, \forall l \). This yields the normalized weights
\[
\omega_{k+1}^{l} = \frac{p(z_{k+1} | x_{k+1}^{l})}{\sum_{l=1}^{N} p(z_{k+1} | x_{k+1}^{l})} \] (56)

Then the posterior pdf at time \( k+1 \) can be approximated by
\[
p(x_{0:k+1} | z_{1:k+1}) \approx \frac{1}{N} \sum_{l=1}^{N} \omega_{k+1}^{l} \delta(x_{0:k+1} - x_{0:k+1}^{l}) \] (57)

The second pdf in (54) involves an integral
\[
p(z_{k+1} | x_{1:k}) = \int p(z_{k+1} | x_{k+1}) p(x_{k+1} | z_{1:k}) dx_{k+1} \] (58)
We assume that the additive Gaussian noise dynamic system has the following state and measurement equations:

\[ x_{k+1} = f_k(x_k) + u_k \]  

(62)

\[ z_k = h_k(x_k) + v_k \]  

(63)

where \( f_k(\cdot) \) and \( h_k(\cdot) \) are nonlinear state transition and measurement functions respectively, \( u_k \) is the white Gaussian state process noise with zero mean and covariance matrix \( Q_k \), and \( v_k \) is the white Gaussian measurement noise with zero mean and covariance matrix \( R_k \). The sequences \{\( u_k \}\} and \{\( v_k \}\} are mutually independent. With these assumptions and notations, the transition prior of the state can be written as follows

\[
p(x_{k+1}|x_k) = \frac{1}{(2\pi)^{n_x/2}|Q_k|^{1/2}} \exp\left\{ -\frac{1}{2} \|x_{k+1} - f_k(x_k)\|^2 Q_k^{-1} \|x_{k+1} - f_k(x_k)\| \right\}
\]  

(64)

Taking the logarithm of the above pdf, we have

\[
-\log p(x_{k+1}|x_k) = -\Delta x_k^k \log p(x_{k+1}|x_k)
\]  

(65)

where \( -\Delta x_k^k \) denotes a constant independent of \( x_k \) and \( x_{k+1} \). Then the first and second-order partial derivatives of \( \log p(x_{k+1}|x_k) \) with respect to \( x_k \) can be derived respectively as

\[
\nabla x_k \log p(x_{k+1}|x_k) = \nabla x_k f_k(x_k) Q_k^{-1} (x_{k+1} - f_k(x_k))
\]  

and

\[
-\Delta x_k^k \log p(x_{k+1}|x_k) = -\nabla x_k \{ (x_{k+1} - f_k(x_k))^T Q_k^{-1} \nabla x_k f_k(x_k) \}
\]  

(66)

\[
= [\nabla x_k f_k(x_k) Q_k^{-1} [\nabla^T x_k f_k(x_k)] - [\Delta x_k^k f_k(x_k)] \Sigma^{-1} \Sigma_{u_k}^{-1} \]
\]  

(67)

where

\[
\Sigma_{u_k}^{-1} = \begin{bmatrix}
Q_k^{-1} & 0 & \cdots & 0 \\
0 & Q_k^{-1} & 0 & \vdots \\
\vdots & 0 & \ddots & 0 \\
0 & \cdots & 0 & Q_k^{-1}
\end{bmatrix}_{n_x^2 \times n_x^2}
\]  

The sequence \(\{u_k\}\} is independent of the sequence \(\{v_k\}\}. The following identity is used.

\[
E_{p(x_{k+1}|x_k)} \Upsilon^1_{k} = 0
\]  

(68)

From (65), we have

\[
-\Delta x_k^k \log p(x_{k+1}|x_k)
\]  

(69)

By substituting (52) and (69) into (42), we have

\[
B_{k+1}^{11} = E_{p(x_{k+1}|x_k)} \{ -\Delta x_k^k \log p(x_{k+1}|x_k) \}
\]  

\[
\approx \frac{1}{N} \sum_{l=1}^{N} \delta(x_{0:k+1} - x_{0:k+1}) p(x_{k+1}|x_k) \times
\]  

(70)

Now for \( B_{k+1}^{22,\alpha} \), it is evident that

\[
B_{k+1}^{22,\alpha} = Q_k^{-1}
\]  

(71)

And the first and second-order partial derivatives of \( \log p(z_k|x_k) \) can be derived respectively as follows

\[
\nabla x_k \log p(z_k|x_k) = \nabla x_k h_k(x_k) R_k^{-1} (z_k - h_k(x_k))
\]  

(72)

\[
-\Delta x_k^k \log p(z_k|x_k) = \nabla x_k h_k(x_k) R_k^{-1} [\nabla^T x_k h_k(x_k)] - \Delta x_k^k h_k(x_k) \Sigma_{v_k}^{-1} \Upsilon^2_{k}
\]  

(73)
Now substituting (61) and (73) into the above equation, we have

$$B_{22}^{b,k} = E_{p(x_{k+1}|x_k)} \{ -\Delta x_{k+1} \log p(z_{k+1}|x_{k+1}) \}$$

$$\approx \int \frac{1}{N} \sum_{l=1}^{N} \delta(x_{0:k+1} - x_{l:k+1}) p(z_{k+1}|x_{k+1})$$

$$(-\Delta x_{k+1} \log p(z_{k+1}|x_{k+1})) dx_{k+1} dz_{k+1}$$

$$= \frac{1}{N} \sum_{l=1}^{N} \nabla_{x_{k+1}} h(x_{k+1}) R_{k+1}^{-1} \nabla_{x_{k+1}} h(x_{k+1})|_{x_{k+1} = x_{l:k+1}}$$

with the following identity being used

$$E_{p(x_{k+1}|x_k)} \tilde{Y}_{k+1}^{22,b} = 0$$

(74)

The approximations for $S_{k+1}^{11}$, $S_{k+1}^{21}$, and $S_{k+1}^{22}$ can be derived similarly. Using (51), we have

$$S_{k+1}^{11} \approx \frac{1}{N} \sum_{l=1}^{N} g^{\hat{\gamma}}(x_{l:k-1}^{a}, x_{k})$$

(75)

where

$$g^{\hat{\gamma}}(x_{k-1}, x_{k}) =$$

$$[\nabla_{x_{k-1}} f_{k-1}(x_{k-1})] Q_{k-1}^{-1} [\nabla_{x_{k-1}} f_{k-1}(x_{k-1})] - [\Delta x_{k-1} f_{k-1}(x_{k-1})] \bar{\Sigma}_{x_{k-1}}^{-1} \tilde{Y}_{k-1}^{11} (x_{k-1}, x_{k})$$

$$S_{k}^{22,1} \approx -\frac{1}{N} \sum_{l=1}^{N} \left[ Q_{k-1}^{-1} \nabla_{x_{k-1}} f_{k-1}(x_{k-1}) \right]|_{x_{k-1} = x_{l:k-1}}^{\ast}$$

$$S_{k}^{22,0} = Q_{k-1}^{-1}$$

(77)

$$S_{k}^{22,0} \approx \frac{1}{N} \sum_{l=1}^{N} \left[ (\nabla_{x_{k}} h_{k}(x_{k})) R_{k}^{-1} (\nabla_{x_{k}} h_{k}(x_{k})) \right]$$

$$- [\Delta x_{k} h_{k}(x_{k})] \bar{\Sigma}_{x_{k}}^{-1} \tilde{Y}_{k}^{22,b}|_{x_{k} = x_{k}}^{\ast}$$

(78)

where $\bar{\Sigma}_{x_{k}}$ and $\tilde{Y}_{k}^{22,b}$ are defined as

$$\bar{\Sigma}_{x_{k}}^{-1} = \begin{bmatrix} R_{k}^{-1} & 0 & \ldots & 0 \\ 0 & R_{k}^{-1} & \vdots \\ \vdots & \ddots & 0 \\ 0 & \ldots & 0 & R_{k}^{-1} \end{bmatrix}_{(n_{x} n_{z}) \times (n_{x} n_{z})}$$

and

$$\tilde{Y}_{k}^{22,b} = \begin{bmatrix} z_{k} - h_{k}(x_{k}) & \ldots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & z_{k} - h_{k}(x_{k}) \end{bmatrix}_{(n_{x} n_{z}) \times n_{z}}$$

### 4 Simulation Results

In this section, the univariate non-stationary growth model (UNGM), which is a highly nonlinear and bi-modal model, is adopted to demonstrate the accuracy of the computed bounds. The UNGM is very useful in econometrics, and has been used in [7, 9, 10]. The dynamic state space equations for UNGM are given by

$$x_{k+1} = \alpha x_{k} + \beta \frac{x_{k}}{1 + x_{k}^{2}} + \gamma \cos(1.2k) + u_{k}$$

(79)

where $u_{k}$ and $v_{k}$ are the state process noise and measurement noise respectively, and they are white Gaussian with zero means and variances $\sigma_{u}^{2}$ and $\sigma_{v}^{2}$.

We set parameters $\alpha = 1$, $\beta = 5$, $\gamma = 8$, $\sigma_{u}^{2} = 1$, $\sigma_{v}^{2} = 1$, and $\kappa = 1/20$ for UNGM. Fig. 1 shows the system states and measurements over a period of 20 discrete time steps. Due to the measurement equation of the UNGM specified in (80), there is bi-modality inherent in the filtering problem. This is very clear from Figure 1, since the observation does not follow the system state very closely. In such a case, it is very difficult to track the state using conventional methods, and the particle filter demonstrates better tracking performance than the extended Kalman filter, as illustrated in Figures 1 and 2. Fig. 3 shows the conditional PCRLB and the conditional MSE. It is clearly shown that the conditional PCRLB gives a lower bound on the conditional MSE that an estimator can achieve. It is also clear that the conditional PCRLB and the conditional MSE follow the same trend. Since the proposed bound utilizes the available measurement information, it will predict the estimator’s performance more faithfully than the unconditional PCRLB, and it can be used as a criterion for selecting the best sensors that lead to the minimal conditional PCRLB value in a sensor selection problem.

### 5 Conclusion

In this paper, we proposed the notion of conditional PCRLB, which takes the past available measurements as realizations rather than random vectors to evaluate the state in the upcoming time step. Both the mathematical closed form and the sequential Monte Carlo approximation for this bound were presented. Future work will focus on investigating the properties of the proposed bound. Theorem 1 shows that the conditional PCRLB is not only a bound on the filtering estimator $\hat{x}_{k+1}$, it also sets a bound on the smoothing estimator $\hat{x}_{0:k}$, when the new measurement $z_{k+1}$ becomes available. The conditional PCRLB derived in this paper provides an approach to recursively predict the MSE one-step ahead. It can be extended to multi-step ahead cases in the future. An important application of the proposed bound is in a variety of sensor management.
problems in sensor networks. Choosing the most informative set of sensors will improve the tracking performance, while at the same time reduce the requirement for communication bandwidth and the energy needed by sensors for sensing, local computation and communication.

References


