Purdue University
Purdue e-Pubs

ECE Technical Reports

Electrical and Computer Engineering

5-17-1994

Cramer-Rao Bounds for Discrete-Time Nonlinear Filtering Problems

Peter C. Doerschuk
Purdue University School of Electrical Engineering

Cramer-Rao Bounds for Discrete-Time Nonlinear Filtering Problems

Peter C. Doerschuk

TR-EE 94-20
May 1994

School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907-1285

Cramer-Rao Bounds for Discrete-Time Nonlinear Filtering Problems

Peter C. Doerschuk[1]

School of Electrical Engineering, Purdue University

May 17, 1994

[1]Supported by U. S. National Science Foundation grant MIP-9110919, a Whirlpool Faculty Fellowship, and the School of Electrical Engineering, Purdue University. Mailing address: Peter C. Doerschuk. 1285 Electrical Engineering Building, Purdue University, West Lafayette, IN 47907-1285; Telephone: (317) 494-1742; Fax: (317) 494-6440; Internet: doerschu@ecn.purdue.edu.

Abstract

A Cramer-Rao bound for the mean squared error that can be achieved with nonlinear observations of a nonlinear p-th order autoregressive (AR) process where both the process and observation noise covariances can be state dependent is presented. The major limitation is that the AR process must be driven by an additive white Gaussian noise process that has a full-rank covariance. A numerical example demonstrating the tightness of the bound for a particular problem is included.

Contents

1 Introduction

This note concerns lower bounds on the mean squared error (MSE) in nonlinear filtering problems. Specifically, Cramer-Rao bounds (CRBs) are derived for dynamical systems that are more general than those used previously [5, 2]. Such bounds give an indication of whether accuracy requirements are realistic before a design effort is undertaken and, during a design, aid in determining whether further design effort may not be fruitful. This note concerns only discrete time problems which, while much less discussed than continuous time problems, are of great practical importance.

The nonlinear filtering problem is to causally estimate the n-dimensional state x_k of a source or message model described by a nonlinear stochastic difference equation given m-dimensional measurements y_k that are a stochastic nonlinear transformation of x_k:

$$\begin{aligned} x_{k+1} &= a(x_k, k) + b(x_k, k)w_k \\ y_k &= c(x_k, k) + d(x_k, k)v_k \end{aligned}$$

where w and v are white Gaussian noise sequences. Let $\hat{x}_k$, a function of $y_0, y_1, \ldots, y_k$, be the estimate of x_k. If the estimator is chosen to minimize the MSE $\epsilon_k = E[(x_k - \hat{x}_k)'(x_k - \hat{x}_k)]$ then the optimal estimator (denoted $\hat{x}_k^*$) is the conditional mean $\hat{x}_k^* = E[x_k|y_m, m \leq k]$ and the resulting MSE is $\epsilon_k^* = E[(x_k - \hat{x}_k^*)'(x_k - \hat{x}_k^*)]$. Fix some time M. The CRB is a lower bound on ϵ_M^*. A wide class of lower bounds on the MSE in parameter estimation problems were recently reviewed and unified [1]. Those hounds appropriate for nonlinear filtering problems were also recently reviewed [7, 3, 8]. Additional work not cited in Ref. [7] includes Refs. [11, 4].

In this paragraph we summarize the notation used in this note. The function $\mathcal{I}(e)$ is 1 if e is true and 0 otherwise. The real numbers are 'R. The Gaussian probability density function (pdf) with mean m and covariance P is $\mathcal{N}(m, \mathbf{P})$. The n x n identity (zero) matrix is I_n (0_n). E denotes expectation. The abbreviation i.i.d. means independent and identically distributed. Prime (i.e., $'$) denotes transpose. If z is a sequence indexed by the integers then

z_l^m is the vector $(z_l', z_{l+1}', ..., z_m')'$. $\nabla\phi = \partial\phi_i/\partial x_j$ ($\phi : \mathcal{R}^{n\times 1} \to \mathrm{R}^{m\times 1}$). Occasionally we write a matrix by giving its general entry as a function of the indices a and β.

The remainder of the note is organized in the following fashion. The new CRB is described in Section 2, an example is given in Section **3,** and conclusions follow in Section 4.

2 Cramer-Rao Bounds

The model is a nonlinear p-th order AR process driven by additive Gaussian noise with state-dependent gain of which nonlinear observations are made in the presence of additive Gaussian noise with state-dependent gain:

$$\begin{aligned} x_{k+1} &= f_k(x_k, \ldots, x_{k-p+1}) + [q_k(x_k, \ldots, x_{k-p+1}), d_k(x_k, \ldots, x_{k-p+1})]\xi_k, \quad k = 0, \ldots, K-1 \\ y_k &= h_k(x_k, \ldots, x_{k-p+1}) + [e_k(x_k, \ldots, x_{k-p+1}), r_k(x_k, \ldots, x_{k-p+1})]\xi_k, \quad k = 0, \ldots, K-1 \\ y_k &= h_k(x_k, \ldots, x_{\max(k-p+1,1-p)}) + r_k(x_k, \ldots, x_{\max(k-p+1,1-p)})v_k, \quad k = 1-p, \ldots, -1 \text{ and } k = K \end{aligned}$$

where $x_k, y_k \in \mathcal{R}^n$; the range of f_k and h_k is $\mathcal{R}^n$; the range of q_k, d_k, r_k, and e_k is $\mathcal{R}^{n\times n}$; ξ_k is i.i.d. $N(0, I_{2n})$; v_k is i.i.d. $N(0, \mathrm{I},)$; $(x_0', \ldots, x_{1-p}')'$ is $p_0(x_0, \ldots, x_{1-p})$ which is never zero: ξ, v and $(x_0', \ldots, x_{1-p}')'$ are independent; the covariance Σ_k defined by

$$\Sigma_k(x_k, \ldots, x_{k-p+1}) = \begin{bmatrix} q_k & d_k \\ e_k & r_k \end{bmatrix} \begin{bmatrix} q_k & d_k \\ e_k & r_k \end{bmatrix}'$$

is full rank (i.e., 2n) for k $= 0, \ldots K-1$; and the covariance R_k defined by $R_k(x_k, \ldots, x_{\max(k-p+1,1-p)}) = r_k r_k'$ is full rank (i.e., IZ) for k $= 1-p, \ldots, -1$ and k $= K$. Define the n x n block components of Σ_k and Σ_k^{-1}:

$$\Sigma_k = \begin{bmatrix} Q_k & S_k \\ S_k' & R_k \end{bmatrix}, \quad \Sigma_k^{-1} = \begin{bmatrix} W_k & U_k \\ U_k' & V_k \end{bmatrix}.$$

The assumption that x_k and y_k have the same dimension does not entail a loss of generality since f or h can be used to compensate. The major assumption is that Σ is full rank. By state augmentation the AR model can be replaced by a nonlinear state equation model but the additive noise will then have a covariance that is not full rank.

Previous work [5] is restricted to the case (1) $p = 1$, (2) $q_k(x_k, \ldots, x_{k-p+1}) = q_k$ independent of x and full rank, (3) $r_k(x_k, \ldots, x_{k-p+1}) = r_k$ independent of x and full rank, and (4) $d_k(x_k, \ldots, x_{k-p+1}) = e_k(x_k, \ldots, x_{k-p+1}) = 0$. The generalization described here is worthwhile for two reasons: (1) State equations derived from system identification procedures typically have $p > 1$. (2) Discretization of second-order differential equations of mathematical physics, when driven by a random process, naturally have $p = 2$, and sometimes require q_k to be a function of x. Allowing r_k to be a function of x and including nonzero d_k and e_k can then be done with very little additional complexity.

Some, but not all, state equations with Q_k less than full rank can be rewritten as AR process s with Q_k^{AR} full rank. For instance, consider a linear time-invariant state equation ($f_k(x_k) = Fx_k$) with n components where $Q_k(x_{k-p+1}^k) = Q$ is rank 1 and hence can be written $Q = qq'$ where $q \in \mathcal{R}^{n\times 1}$. If (F,q) is controllable then a siinilarity transformation exists which transforms the system to canonical controllable form [9, Section 1.9] which is exactly the form of a n-th order scalar AR process with $Q_k^{\mathrm{AR}} = 1$.

Application of the standard CRB to the entire trajectory x_{1-p}^K of the nonlinear AR process gives the basic bound used in this note and Refs. [5, 2]. Define the trajectory error covariance $\Lambda \in \mathcal{R}^{(K+p)n\times(K+p)n}$ with blocks $\Lambda_{l,k} \in \mathrm{R}^{n\times n}$ given by $\Lambda_{l,k} = E\{(x_l-\hat{x}_l^*)(x_k-\hat{x}_k^*)'\}$. Define the Fisher information matrix $\mathbf{J}$ by, under appropriate regularity assumptions, $\mathbf{J} = E\{\nabla_{x_1^K}\nabla_{x_1^K} \ln p(y_{1-p}^K, x_{1-p}^K)\}$ where p is the joint pdf on the x and y trajectories. Then, since $\hat{x}_k^*$ is unbiased, the standard multivariate CRB is $\Lambda - J^{-1} \geq 0$ where $\geq$ means positive semi-definite. The estimation error at time K is $\Lambda_{K,K}$ so the desired CRB is

$$\Lambda_{K,K} - \begin{bmatrix} 0 & I_n \end{bmatrix} J^{-1} \begin{bmatrix} 0 \\ I_n \end{bmatrix} \geq 0. \tag{1}$$

There are two difficulties in applying Eq. 1. The first difficulty is that for a long trajectory the matrix $\mathbf{J}$ is large and it is difficult to compute J^{-1}. This problem is circumvented in this note and Refs. [5, 2] by finding a linear Gaussian system that has the same Fisher information matrix as the nonlinear system of interest. The linear Gaussian system used in this note will be a p-th order AR process and an observation equation where the observation at time k depends on the AR process at times $k, k-1, \ldots, k-p+1$. This system can be transformed by state augmentation into a linear Gaussian state-variable system. In the state-variable system $\Lambda^{\text{state}}_{K,K}$ can be computed exactly and without excessive computation by the Kalman filter and furthermore the CRB is satisfied with equality so the known value of $\Lambda^{\text{state}}_{K,K}$ is the desired bound on the performance of any estimator for the nonlinear system. The second difficulty is that the computation of $\mathbf{J}$ typically requires numerical computation of expectations. While the approach of this note and Refs. [5, 2] still requires such computations, the approach organizes the computations so that they can be clone by simulation of the nonlinear AR process alone (i.e., not also the observation equation).

The linear Gaussian system used in this note is a special case of the nonlinear system with

$$
\begin{aligned}
f_k(x^k_{k-p+1}) &= \sum_{i=0}^{p-1} A_{k,i} x_{k-i} \\
h_k(x^k_{\max(k-p+1,1-p)}) &= \sum_{i=0}^{\min(p-1,k+p-1)} C_{k,i} x_{k-i} \\
p_0(x^0_{1-p}) &= \mathcal{N}(0, P_0)(x^0_{1-p}) \\
\Sigma_k(x^k_{k-p+1}) &= \check{\Sigma}_k \quad \text{for } k = 0, \ldots, K-1 \text{ independent of } x \\
R_k(x^k_{\max(k-p+1,1-p)}) &= \check{R}_k \quad \text{for } k = 1-p, \ldots, -1 \text{ and } k = K \text{ independent of } x
\end{aligned}
$$

where P_0 (P_0^{-1}) has $n \times n$ blocks $P_{i,j}$ ($P^{i,j}$) for i and j in $0, \ldots, 1-p$ and

$$
\check{\Sigma}_k = \begin{bmatrix} \check{Q}_k & \check{S}_k \\ \check{S}'_k & \check{R}_k \end{bmatrix}, \qquad \check{\Sigma}_k^{-1} = \begin{bmatrix} \check{W}_k & \check{U}_k \\ \check{U}'_k & \check{V}_k \end{bmatrix}
$$

For $k = 0, \ldots, K-1$ and $l = 0, \ldots, p-1$ define

$$
J_{k,l} = \begin{bmatrix} A_{k,l} \\ C_{k,l} \end{bmatrix}
$$

For later convenience, define

$$\Delta^{so}_{i,k,l} = E\left\{ [\nabla_{x_l} f_i', \nabla_{x_l} h_i'] \, \Sigma_i^{-1} \begin{bmatrix} \nabla_{x_k} f_i \\ \nabla_{x_k} h_i \end{bmatrix} \right\} + \frac{1}{2} E\{\mathrm{tr} \left[\Sigma_i \frac{\partial^2 \Sigma_i^{-1}}{\partial x_{l,\beta} \partial x_{k,\alpha}} \right] + \nabla_{x_l} \nabla_{x_k} \ln \det \Sigma_i \right\} \tag{2}$$

$$\Delta^{o}_{i,k,l} = E\left\{ \nabla_{x_l} h_i' R_i^{-1} \nabla_{x_k} h_i \right\} + \frac{1}{2} E\left\{ \mathrm{tr} \left[R_i \frac{\partial^2 R_i^{-1}}{\partial x_{l,\beta} \partial x_{k,\alpha}} \right] + \nabla_{x_l} \nabla_{x_k} \ln \det R_i \right\} \tag{3}$$

$$\Lambda_{k,l} = -E\left\{ [\nabla_{x_l} f_k', \nabla_{x_l} h_k'] \Sigma_k^{-1} \begin{bmatrix} I_n \\ 0_n \end{bmatrix} \right\} = -E\left\{ \nabla_{x_l} f_k' W_k + \nabla_{x_l} h_k' U_k' \right\} \tag{4}$$

$$\Gamma_k = E\left\{ [I_n, 0_n] \Sigma_k^{-1} \begin{bmatrix} I_n \\ 0_n \end{bmatrix} \right\} = E\left\{ W_k \right\} \tag{5}$$

These formulae simplify when, for example, $\Sigma_k(x^k_{k-p+1})$ is independent of x. In order to determine the linear Gaussian system the user must give values for $\Delta^{so}_{i,k,l}$, $\Delta^{o}_{i,k,l}$, $\Lambda_{k,l}$, Γ_k, and $E\{\nabla_{x_l} \nabla_{x_k} \ln p_0\}$ in the nonlinear system. Computation of the first four will likely require Monte Carlo simulation unless f is linear and p_0 is Gaussian or. for Γ_k, the gain for the process noise is state independent. The exact range of i, k, and l required is determined by Eq. 13.

We now derive the equations which must be satisfied by $A, C, P_0, \check{\Sigma}$, and R. The natural logarithm of the joint pdf for the x and y trajectories is

$$\begin{aligned}
\ln p &= K_1 - \frac{1}{2} \sum_{i=0}^{K-1} \left\{ z_i(x^{i+1}_{i-p+1})' \Sigma_i^{-1}(x^i_{i-p+1}) z_i(x^{i+1}_{i-p+1}) + \ln \det \Sigma_i(x^i_{i-p+1}) \right\} \\
&\quad - \frac{1}{2} \Big\{ [y_K - h_K(x^K_{\max(K-p+1,1-p)})]' R_K^{-1}(x^K_{\max(K-p+1,1-p)}) [y_K - h_K(x^K_{\max(K-p+1,1-p)})] \\
&\qquad + \ln \det R_K(x^K_{\max(K-p+1,1-p)}) \Big\} \\
&\quad - \frac{1}{2} \sum_{i=1-p}^{-1} \Big\{ [y_i - h_i(x^i_{\max(i-p+1,1-p)})]' R_i^{-1}(x^i_{\max(i-p+1,1-p)}) [y_i - h_i(x^i_{\max(i-p+1,1-p)})] \\
&\qquad + \ln \det R_i(x^i_{\max(i-p+1,1-p)}) \Big\} \\
&\quad + \ln p_0(x^0_{1-p})
\end{aligned}$$

where K_1 is a constant. By equating $\mathrm{E}\{\nabla_{x_l} \nabla_{x_k} \ln p\}$ for the nonlinear and the linear-Gaussian systems we derive a system of equations that A, C, P_0, %. and R must satisfy. It is only necessary to consider $l \leq$ k because $\nabla_{x_l}\nabla_{x_k} \ln p = \nabla_{x_k}\nabla_{x_l} \ln p'$. The equations, derived in Appendix A, are

$$
\begin{aligned}
& k = 2-p, \ldots, K; l = \max(k-p, 1-p), \ldots, k-1: \\
& - \sum_{i=\max(k,0)}^{\min(l+p-1,K-1)} \Delta^{so}_{i,k,l} - \Lambda_{k-1,l}\mathcal{I}(k \geq 1) - \Delta^{o}_{K,k,l}\mathcal{I}(K-p+1 \leq l) \\
& - \sum_{i=k}^{-1} \Delta^{o}_{i,k,l} + E\{\nabla_{x_l}\nabla_{x_k} \ln p_0\}\mathcal{I}(k \leq 0) \\
= & - \sum_{i=\max(k,0)}^{\min(l+p-1,K-1)} J'_{i,i-l}\check{\Sigma}_i^{-1}J_{i,i-k} + \left(A'_{k-1,k-1-l}\check{W}_{k-1} + C'_{k-1,k-1-l}\check{U}'_{k-1}\right)\mathcal{I}(k \geq 1) \\
& - C'_{K,K-l}\check{R}_K^{-1}C_{K,K-k}\mathcal{I}(K-p+1 \leq l) - \sum_{i=k}^{-1} C'_{i,i-l}\check{R}_i^{-1}C_{i,i-k} - P^{l,k}\mathcal{I}(k \leq 0)
\end{aligned} \tag{6}
$$

$$
\begin{aligned}
& k = 1-p, \ldots, K: \\
& - \sum_{i=\max(k,0)}^{\min(k+p-1,K-1)} \Delta^{so}_{i,k,k} - \Gamma_{k-1}\mathcal{I}(k \geq 1) - \Delta^{o}_{K,k,k}\mathcal{I}(K-p+1 \leq k) \\
& - \sum_{i=k}^{-1} \Delta^{o}_{i,k,k} + E\left\{\nabla_{x_k}\nabla_{x_k} \ln p_0\right\}\mathcal{I}(k \leq 0) \\
= & - \sum_{i=\max(k,0)}^{\min(k+p-1,K-1)} J'_{i,i-k}\check{\Sigma}_i^{-1}J_{i,i-k} - \check{W}_{k-1}\mathcal{I}(k \geq 1) \\
& - C'_{K,K-k}\check{R}_K^{-1}C_{K,K-k}\mathcal{I}(K-p+1 \leq k) - \sum_{i=k}^{-1} C'_{i,i-k}\check{R}_i^{-1}C_{i,i-k} - P^{k,k}\mathcal{I}(k \leq 0)
\end{aligned} \tag{7}
$$

Define in sequence the following quantities:

$$P^{l,k} \doteq -E\{\nabla_{x_l}\nabla_{x_k} \ln p_0\} \quad \text{for } k = 1-p, \ldots, 0;\ l = 1-p, \ldots, k. \tag{8}$$

$$\check{W}_k \doteq \Gamma_k \quad \text{for } k = 0, \ldots, K-1. \tag{9}$$

$$A_{k,k-l} \doteq -\check{W}_k^{-1}\Lambda'_{k,l} \quad \text{for } k = 0, \ldots, K-1;\ l = k-p+1, \ldots, k. \tag{10}$$

$$\check{U}_k \doteq 0 \quad \text{for } k = 0, \ldots, K-1. \tag{11}$$

$$R_k \doteq \text{arbitrary positive definite matrix} \quad \text{for } k = 1-p, \ldots, K. \tag{12}$$

$$
\begin{aligned}
D_{k,l} \doteq & \sum_{i=\max(k,0)}^{\min(l+p-1,K-1)} \left(\Delta^{so}_{i,k,l} - A'_{i,i-l}\check{W}_iA_{i,i-k}\right) + \Delta^{o}_{K,k,l}\mathcal{I}(K-p+1 \leq l) + \sum_{i=k}^{-1} \Delta^{o}_{i,k,l} \\
& \text{for } k = 1-p, \ldots, K;\ l = \max(k-p, 1-p), \ldots, k.
\end{aligned} \tag{13}
$$

Use these definitions in Eqs. 6 and 7 to get

$$\sum_{i=k}^{\min(l+p-1,K)} C'_{i,i-l}\check{R}_i^{-1}C_{i,i-k} = D_{k,l} \text{ for } \begin{array}{l} k = 2-p,\ldots,K; \\ l = \max(k-p, 1-p),\ldots,k-1 \end{array} \tag{14}$$

$$\sum_{i=k}^{\min(k+p-1,K)} C'_{i,i-k}\check{R}_i^{-1}C_{i,i-k} = D_{k,k} \text{ for } k = 1-p,\ldots,K. \tag{15}$$

In spite of being quadratic in $C_{i,j}$, Eqs. 14 and 15 can be solved recursively for $C_{i,j}$. The procedure, with control structures written in the C programming language, is:

$$\begin{array}{l}
\text{for}(k = K; k >= 1-p; k--)\{ \\
\quad C_{k,0} = \check{R}_k^{1/2}\left(D_{k,k} - \sum_{i=k+1}^{\min(k+p-1,K)} C'_{i,i-k}\check{R}_i^{-1}C_{i,i-k}\right)^{T/2} \\
\quad \text{for}(l = k-1; l >= \max(k-p+1, 1-p); l--)\{ \\
\qquad C_{k,k-l} = \left[\left(D_{k,l} - \sum_{i=k+1}^{\min(l+p-1,K)} C'_{i,i-l}\check{R}_i^{-1}C_{i,i-k}\right)\left(\check{R}_k^{-1}C_{k,0}\right)^{-1}\right]' \\
\quad \} \\
\}
\end{array}$$

where $^{1/2}$ are matrix square roots (R $= R^{1/2}\left(R^{1/2}\right)'$, $R^{T/2} \doteq \left(R^{1/2}\right)'$). Once A, C. P_0, Σ. and R have been determined by the procedure described above then the state augmentation and Kalman filter computations needed in order to determine the hound are standard.

Evaluation of these equations for $p = 1$ with $d = e = 0$, $Q_k(x^k_{k-p+1})$ and $R_k(x^k_{\max(k-p+1,1-p)})$ independent of x, and $\check{R}_k = R_k$ recovers the results of Ref. [5]. (There is a typographical error in [5, Eq. 7f] which reads $P_0 = \nabla_{x_0}\nabla_{x_0}p_{x_0}(x_0)$ but should read $-P_0^{-1} = E\{\nabla_{x_0}\nabla_{x_0}\ln p_{x_0}(x_0)\}$). More generally, for any $p = 1$ problem, the equations for $k \neq l$ are satisfied by the choice of A so the inner loop of the algorithm vanishes. Furthermore, in the outer loop, the sums over quadratic forms in the Cs are empty and therefore the outer loop can be executed in any order. Therefore each $C_{k,0}$ can be chosen independently of all other $C_{k',0}$ for $k' \neq k$. The generalization from $p = 1$ to $p > 1$ is the main contribution of this note. Another interesting special case is the case of d = e = 0 for which the definition of $D_{k,l}$ simplifies to

$$D_{k,l} = \sum_{i=\max(k,0)}^{\min(l+p-1,K-1)} \left(\Delta^{s}_{i,k,l} - A'_{i,i-l}\check{W}_i A_{i,i-k}\right) + \sum_{i=k}^{\min(l+p-1,K)} \Delta^{o}_{i,k,l}$$

where

$$\Delta^{s}_{i,k,l} = E\left\{\nabla_{x_l} f_i' Q_i^{-1} \nabla_{x_k} f_i\right\} + \frac{1}{2} E\left\{\operatorname{tr}\left[Q_i \frac{\partial^2 Q_i^{-1}}{\partial x_{l,\beta} \partial x_{k,\alpha}}\right] + \nabla_{x_l}\nabla_{x_k} \ln\det Q_i\right\}.$$

3 Example

In this section we consider an example estimation problem in order to demonstrate the computations involved and that the bound is tight at least for some estimation problems. (Please contact the author for copies of the software). The physical system is a damped pendulum driven by a random torque where noisy measurements are made of the horizontal component of the pendulum bob's location and the goal is to estimate the angular location of the pendulum (measured from the negative-going vertical). This system demonstrates $p = 2$ and nonlinear state and observation equations. After discretization the equations are

$$\begin{aligned} \phi_{n+1} &= \phi_n \left(2 - \frac{\gamma T}{l^2 m}\right) + \phi_{n-1}\left(\frac{\gamma T}{l^2 m} - 1\right) - \frac{g}{l} T^2 \sin\phi_{n-1} + \sigma_w \frac{T^2}{l^2 m} \frac{\tau((n-1)T)}{\sigma_w} \\ & \qquad \text{for } n = 0, \ldots, K - 1 \\ y_n &= l \sin\phi_n + v(nT) \qquad \text{for } n = 0, \ldots, K \end{aligned}$$

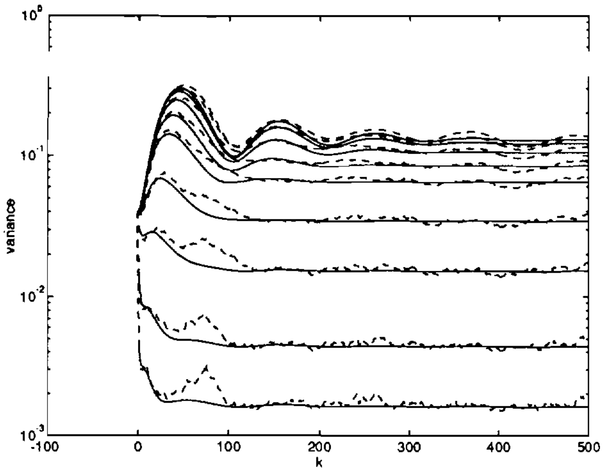
where l is the length of the pendulum, m is the mass of the bob, γ is the coefficient of friction, g is the acceleration due to gravity, T is the sampling interval, τ is the random applied torque ($\mathcal{N}(0, \sigma_w^2)$), and v is the observation noise ($\mathcal{N}(0, a:)$). The MSE of the optimal filter was underbounded by the CRB developed in this note and overbounded by the performance of the Extended Kalman Filter (EKF). When computing expectations for the CRB and the EKF by Monte Carlo methods, the state trajectories used were identical. Monte Carlo evaluation of the performance of the EKF must, however, also include realizations of the observation trajectory. We consider three cases and for each case compute the CRB and the

MSE for the EKF at each point in a $K = 500$ point trajectory for a variety of a,. Notice that only one set of expectations need be computed for the CRB in order to determine the bound for any a,. The parameters are $l = 1$, $m = 1$, $\gamma = 1$, $g = 10.$ and $T = .01$.

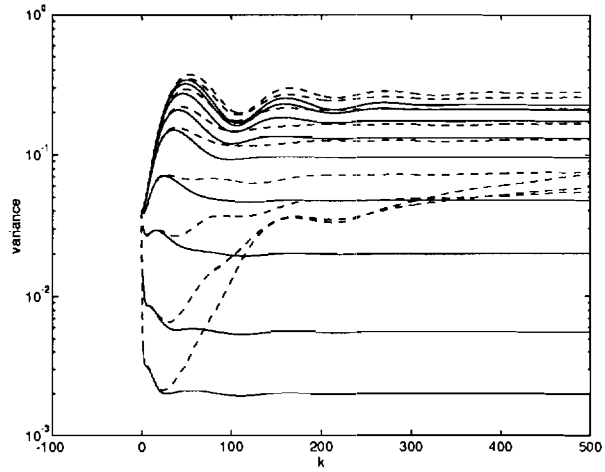
In the first and second cases, the pdf on the initial condition is jointly Gaussian: $p(\phi_{-1}, \phi_0) = \mathcal{N}(0, \sigma_\phi^2)(\phi_{-1})\mathcal{N}(0, \sigma_{\dot\phi}^2)((\phi_0 - \phi_{-1})/T)$ with $\sigma_\phi = .2$ and $\sigma_{\dot\phi} = 2$. In the first case (Figure la), where the bound is tight, a, $= 15$ and the expectations were computed by summing over 10^3 state trajectories. This case is weakly nonlinear since the sample mean and standard deviation of $\max_{k=-1,\ldots,500} |\phi_k|$ is 0.758271 ± 0.271391. In the second case (Figure Ib), where the bound is loose, a, $= 20$ and the expectations were computed by summing over 10^5 state trajectories. This case is moderately nonlinear since the sample mean and standard deviation of $\max_{k=-1,\ldots,500} |\phi_k|$ is 0.97351 ± 0.363154. With these parameters the EKF makes rare but large errors analogous to cycle slips in a phase-locked loop so a larger number of trajectories were used to evaluate its performance than in the first case. Finally, in the third case? the system is very nonlinear and the SNR is poor. This problem, modeled after acquisition in a phase locked loop, has a nearly uniformly distributed initial condition on the angle and a large variance process noise. Specifically, a, $= 25$ and the pdf on the initial condition is $p(\phi_{-1}, \phi_0) = [U(\pi) * \mathcal{N}(0, \sigma_p^2)](\phi_{-1})\mathcal{N}(0, \sigma_{\dot\phi}^2)((\phi_0 - \phi_{-1})/T)$ where $U(x)$ is the uniform distribution on the interval [$-$x,$+x$], $*$ is convolution, and a, $= .5$ and $\sigma_{\dot\phi} = 2$. This case is strongly nonlinear since the sample mean and standard deviation of $\max_{k=-1,\ldots,500} |\phi_k|$ is 2.55681 ± 2.21283. Though the EKF performance is poor and is not shown, the CRB can still be computed without difficulty, as shown in Figure Ic based on 10^3 state trajectories.

4 Conclusions

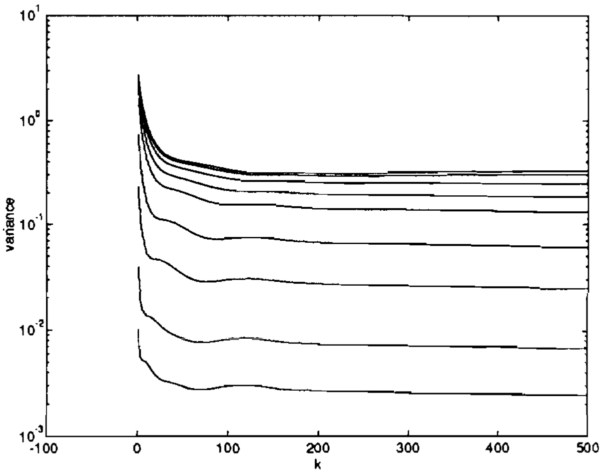
A Cramer-Rao bound for the mean squared error that can be achieved with nonlinear observations of a nonlinear p-th order AR process where both the process and observation noise covariances can be state dependent is presented. The major limitation is that the AR



(a)



(b)



(c)

Figure 1: Comparison of CRB (solid line) and EKF estimation variance (dotted line) for a range of measurement noise variance: $\sigma_v \in \{.1, .2, .5, 1, 2, 3, 5, 10, 100)$. *(it)* Gaussian initial condition, $a,$ = 15. (b) Gaussian initial condition, $a_{,,,}$ = 20. (c) Neatly uniform initial condition, $a,$ = 25.

process must be driven by an additive white Gaussian noise process that has a full-rank covariance. The bound is a generalization of the results of Refs. [5, 2] to the case p > 1 and state-dependent noises. In addition, its computation requires different methods, specifically the solution of a system of quadratic equations for which a recursive method is described.

Relaxation of the full-rank condition on the process noise covariance will probably require constrained CRB tools [10, 6]. The merger of the constrained CRR tools with the dynamical system approach of this note and Refs. [5,2] does not appear to be straightforward.

A Derivation of Eqs. 6 and 7

A part of the derivation of Eqs. 6 and 7 concerns expectations of mixed second partial derivatives of quadratic forms. Let $\Sigma : \mathcal{R}^{n\times 1}$ x $\mathcal{R}^{n\times 1} \rightarrow \mathcal{R}^{m\times m}$ with $\Sigma' = \Sigma$, ϕ . $\mathcal{R}^{n\times 1}$ x $\mathcal{R}^{n\times 1} \rightarrow \mathcal{R}^{m\times 1}$, $\psi : \mathcal{R}^{n\times 1}$ x $\mathcal{R}^{n\times 1} \rightarrow \mathcal{R}$, and $\psi = \frac{1}{2}\phi'\Sigma^{-1}\phi$. Let $x_k \in \mathcal{R}^{n\times 1}$ with components $x_{k,\alpha}$ and likewise for x_l. It is straightforward to show that the α, β element of $\nabla_{x_l}\nabla_{x_k}\psi$ is

$$\begin{aligned}\frac{\partial^2\psi}{\partial x_{l,\beta}\partial x_{k,\alpha}} &= \left(\nabla_{x_l}\phi'\Sigma^{-1}\nabla_{x_k}\phi\right)_{\beta,\alpha} + \left(\phi'\frac{\partial\Sigma^{-1}}{\partial x_{l,\beta}}\nabla_{x_k}\phi\right)_{1,\alpha} + \left(\phi'\frac{\partial\Sigma^{-1}}{\partial x_{k,\alpha}}\nabla_{x_l}\phi\right)_{1,\beta} \\ &+ \frac{1}{2}\phi'\frac{\partial^2\Sigma^{-1}}{\partial x_{l,\beta}\partial x_{k,\alpha}}\phi + \sum_j \left(\phi'\Sigma^{-1}\right)_{1,j}\frac{\partial^2\phi_j}{\partial x_{l,\beta}\partial x_{k,\alpha}}. \end{aligned} \tag{16}$$

Consider k = 1 − p, . . . , −1 and k = K, define

$$\psi(x^i_{\max(i-p+1,1-p)}) = \frac{1}{2}[y_i - h_i(x^i_{\max(i-p+1,1-p)})]'R_i^{-1}(x^i_{\max(i-p+1,1-p)})[y_i - h_i(x^i_{\max(i-p+1,1-p)})]$$

where i is such that k and l are in $\{\max(i-p+1, 1-p), \ldots, i\}$, and compute $E\{\nabla_{x_l}\nabla_{x_k}\psi\}$ using Eq. 16. The second, third, and fifth terms of Eq. 16 are each the product of (1) a random vector $\mu_i = y_i - h_i(x^i_{\max(i-p+1,1-p)})$ which, conditional on $x^i_{\max(i-p+1,1-p)}$, has mean zero and covariance $R_i(x^i_{\max(i-p+1,1-p)})$ and (2) a function; call it Ψ, of $x^i_{\max(i-p+1,1-p)}$. Therefore, the expectation of these terms is zero, for example,

$$E\left\{\mu'\Psi(x^i_{\max(i-p+1,1-p)})\right\} = E\left\{E^{x^i_{\max(i-p+1,1-p)}}\left\{\mu'\Psi(x^i_{\max(i-p+1,1-p)})\right\}\right\} = E\left\{0\right\} = 0.$$

(E^e denotes conditional expectation). The expectation of the fourth term need not vanish:

$$\begin{aligned}\frac{1}{2}E\left\{\mu_i'\frac{\partial^2 R_i^{-1}(x^i_{\max(i-p+1,1-p)})}{\partial x_{l,\beta}\partial x_{k,\alpha}}\mu_i\right\} &= \frac{1}{2}E\left\{E^{x^i_{\max(i-p+1,1-p)}}\left\{\mu_i'\frac{\partial^2 R_i^{-1}(x^i_{\max(i-p+1,1-p)})}{\partial x_{l,\beta}\partial x_{k,\alpha}}\mu_i\right\}\right\}\\ &= \frac{1}{2}E\left\{\mathrm{tr}\left[R_i(x^i_{\max(i-p+1,1-p)})\frac{\partial^2 R_i^{-1}(x^i_{\max(i-p+1,1-p)})}{\partial x_{l,\beta}\partial x_{k,\alpha}}\right]\right\}\end{aligned}$$

In this and other instances,

$$\mathrm{tr}\left[M\frac{\partial^2 M^{-1}}{\partial x_{l,\beta}\partial x_{k,\alpha}}\right] = \mathrm{tr}\left[\frac{\partial M}{\partial x_{l,\beta}}M^{-1}\frac{\partial M}{\partial x_{k,\alpha}}M^{-1} - \frac{1}{2}\frac{\partial^2 M}{\partial x_{l,\beta}\partial x_{k,\alpha}}M^{-1}\right].$$

Therefore,

$$E\left\{\nabla_{x_l}\nabla_{x_k}\frac{1}{2}[y_i-h_i]'R_i^{-1}[y_i-h_i]\right\} = E\left\{\nabla_{x_l}h_i'R_i^{-1}\nabla_{x_k}h_i\right\} + \frac{1}{2}E\left\{\mathrm{tr}\left[R_i\frac{\partial^2 R_i^{-1}}{\partial x_{l,\beta}\partial x_{k,\alpha}}\right]\right\}. \quad (17)$$

Now consider $k = 0,\ldots,K-1$, define

$$z_k(x^{k+1}_{k-p+1}) = \begin{bmatrix} x_{k+1} - f_k(x^k_{k-p+1}) \\ y_k - h_k(x^k_{\max(k-p+1,1-p)}) \end{bmatrix},$$

and redefine ψ:

$$\psi(x^{i+1}_{i-p+1}) = \frac{1}{2}z_i(x^{i+1}_{i-p+1})'\Sigma_i^{-1}(x^i_{i-p+1})z_i(x^{i+1}_{i-p+1})$$

where i is such that k and l are in $\{i-p+1,\ldots,i+1\}$. The calculation of $E\{\nabla_{x_l}\nabla_{x_k}\psi\}$, which is omitted, is similar to the calculation leading to Eq. 17 with one complication: if $k = i+1$ then $\nabla_{x_{i+1}}[x_{i+1} - f_i(x^i_{i-p+1})] = \nabla_{x_{i+1}}x_{i+1} = I_n$, $\nabla_{x_{i+1}}[y_i - h_i(x^i_{\max(i-p+1,1-p)})] = 0$, and all second derivatives $\partial^2\Sigma_i^{-1}(x^i_{i-p+1})/\partial x_{l,\beta}\partial x_{k,\beta}$ and $\partial^2 z_i(x^{i+1}_{i-p+1})_j/\partial x_{l,\beta}\partial x_{k,\beta}$ are zero (and likewise for $l = i+1$).

We now use these expectation formulae to derive Eqs. 6 and 7. The natural logarithm of the joint pdf for the x and y trajectories is

$$\begin{aligned}\ln p &= K_1 - \frac{1}{2}\sum_{i=0}^{K-1}\left\{z_i(x^{i+1}_{i-p+1})'\Sigma_i^{-1}(x^i_{i-p+1})z_i(x^{i+1}_{i-p+1}) + \ln\det\Sigma_i(x^i_{i-p+1})\right\}\\ &\quad - \frac{1}{2}\Big\{[y_K - h_K(x^K_{\max(K-p+1,1-p)})]'R_K^{-1}(x^K_{\max(K-p+1,1-p)})[y_K - h_K(x^K_{\max(K-p+1,1-p)})]\\ &\qquad + \ln\det R_K(x^K_{\max(K-p+1,1-p)})\Big\}\end{aligned}$$

$$-\frac{1}{2}\sum_{i=1-p}^{-1}\left\{[y_i - h_i(x^i_{\max(i-p+1,1-p)})]'R_i^{-1}(x^i_{\max(i-p+1,1-p)})[y_i - h_i(x^i_{\max(i-p+1,1-p)})]\right.$$
$$\left. + \ln\det R_i(x^i_{\max(i-p+1,1-p)})\right\}$$
$$+ \ln p_0(x^0_{1-p})$$

where K_1 is a constant. The part of lnp that depends on x_k for $k = 1-p,\ldots,K$ is

$$\begin{aligned}
\ln p|_{x_k} &= -\frac{1}{2}\sum_{i=\max(k,0)}^{\min(k+p-1,K-1)}\left\{z_i(x^{i+1}_{i-p+1})'\Sigma_i^{-1}(x^i_{i-p+1})z_i(x^{i+1}_{i-p+1}) + \ln\det\Sigma_i(x^i_{i-p+1})\right\}\\
&\quad -\frac{1}{2}\left\{z_{k-1}(x^k_{k-p})'\Sigma_{k-1}^{-1}(x^{k-1}_{k-p})z_{k-1}(x^k_{k-p}) + \ln\det\Sigma_{k-1}(x^{k-1}_{k-p})\right\}\mathcal{I}(k\geq 1)\\
&\quad -\frac{1}{2}\left\{[y_K - h_K(x^K_{\max(K-p+1,1-p)})]'R_K^{-1}(x^K_{\max(K-p+1,1-p)})[y_K - h_K(x^K_{\max(K-p+1,1-p)})]\right.\\
&\qquad\left. + \ln\det R_K(x^K_{\max(K-p+1,1-p)})\right\}\mathcal{I}(K-p+1\leq k)\\
&\quad -\frac{1}{2}\sum_{i=k}^{-1}\left\{[y_i - h_i(x^i_{\max(i-p+1,1-p)})]'R_i^{-1}(x^i_{\max(i-p+1,1-p)})[y_i - h_i(x^i_{\max(i-p+1,1-p)})]\right.\\
&\qquad\left. + \ln\det R_i(x^i_{\max(i-p+1,1-p)})\right\}\\
&\quad + \left[\ln p_0(x^0_{1-p})\right]\mathcal{I}(k\leq 0)
\end{aligned}$$

where $\mathcal{I}(e)$ is 1 if e is true and 0 otherwise. Using Eq. 17 (and analogous results) and Eqs. 2, 3, and 5 we find that

$$\begin{aligned}
E\left\{\nabla_{x_k}\nabla_{x_k}\ln p\right\} &= -\sum_{i=\max(k,0)}^{\min(k+p-1,K-1)}\Delta^{so}_{i,k,k} - \Gamma_{k-1}\mathcal{I}(k\geq 1) - \Delta^{o}_{K,k,k}\mathcal{I}(K-p+1\leq k)\\
&\quad -\sum_{i=k}^{-1}\Delta^{o}_{i,k,k} + E\left\{\nabla_{x_k}\nabla_{x_k}\ln p_0\right\}\mathcal{I}(k\leq 0). \qquad (18)
\end{aligned}$$

where the fact that $\Sigma_{k-1}(x^{k-1}_{k-p})$ does not depend on x_k was used in the Γ_{k-1} term.

There are terms in $\ln p$ that depend simultaneously on x_k and x_l for $l < k$ only for the range $l = \max(k-\mathrm{p}, 1-p),\ldots,k-1$ in which case the terms are

$$\begin{aligned}
\ln p|_{x_k;x_l,l<k} &= -\frac{1}{2}\sum_{i=\max(k,0)}^{\min(l+p-1,K-1)}\left\{z_i(x^{i+1}_{i-p+1})'\Sigma_i^{-1}(x^i_{i-p+1})z_i(x^{i+1}_{i-p+1}) + \ln\det\Sigma_i(x^i_{i-p+1})\right\}\\
&\quad -\frac{1}{2}\left\{z_{k-1}(x^k_{k-p})'\Sigma_{k-1}^{-1}(x^{k-1}_{k-p})z_{k-1}(x^k_{k-p}) + \ln\det\Sigma_{k-1}(x^{k-1}_{k-p})\right\}\mathcal{I}(k\geq 1)\\
&\quad -\frac{1}{2}\left\{[y_K - h_K(x^K_{\max(K-p+1,1-p)})]'R_K^{-1}(x^K_{\max(K-p+1,1-p)})[y_K - h_K(x^K_{\max(K-p+1,1-p)})]\right.
\end{aligned}$$

$$+ \ln \det R_K(x^K_{\max(K-p+1,1-p)})\Big\}\mathcal{I}(K-p+1 \le l)$$

$$- \frac{1}{2}\sum_{i=k}^{-1}\Big\{[y_i - h_i(x^i_{\max(i-p+1,1-p)})]'R_i^{-1}(x^i_{\max(i-p+1,1-p)})[y_i - h_i(x^i_{\max(i-p+1,1-p)})]$$

$$+ \ln \det R_i(x^i_{\max(i-p+1,1-p)})\Big\}$$

$$+ \left[\ln p_0(x^0_{1-p})\right]\mathcal{I}(k \le 0).$$

Using Eq. 17 (and analogous results) and Eqs. 2, 3, and 4 we find that

$$\begin{aligned} E\{\nabla_{x_l}\nabla_{x_k}\ln p\} &= -\sum_{i=\max(k,0)}^{\min(l+p-1,K-1)} \Delta^{so}_{i,k,l} - \Lambda_{k-1,l}\mathcal{I}(k \ge 1) - \Delta^{o}_{K,k,l}\mathcal{I}(K-p+1 \le l) \\ &\quad - \sum_{i=k}^{-1}\Delta^{o}_{i,k,l} + E\{\nabla_{x_l}\nabla_{x_k}\ln p_0\}\mathcal{I}(k \le 0). \end{aligned} \tag{19}$$

where the fact that $\Sigma_{k-1}(x^{k-1}_{k-p})$ does not depend on x_k was used in the $\Lambda_{k-1,l}$ term.

Finally, Eq. 6 (Eq. 7) follows by equating Eq. 19 (Eq. 18) for the nonlinear and linear systems. Evaluation of the formulae for the linear Gaussian system uses the following results for the linear Gaussian system:

$$\begin{aligned} \nabla_{x_l} f_k(x^k_{k-p+1}) &= A_{k,k-l} \\ \nabla_{x_l} h_k(x^k_{\max(k-p+1,1-p)}) &= C_{k,k-l} \\ E\{\nabla_{x_l}\nabla_{x_k}\ln p_0(x^0_{1-p})\} &= -P^{l,k} \\ \Delta^{so}_{i,k,l} &= J'_{i,i-l}\check{\Sigma}_i^{-1}J_{i,i-k} \\ \Delta^{o}_{i,k,l} &= C'_{i,i-l}\check{R}_i^{-1}C_{i,i-k} \\ \Lambda_{k,l} &= -\left(A'_{k,k-l}\check{W}_k + C'_{k,k-l}\check{U}'_k\right) \\ \Gamma_k &= \check{W}_k. \end{aligned}$$

References

[1] J. S. Abel. A bound on mean-square-estimate error. *IEEE Trans. Info. Theory*, 39(5):1675–1680, September 1993.

[2] B. Z. Bobrovsky and M. Zakai. A lower bound on the estimation error for Markov processes. *IEEE Trans. Auto. Contr.*, 20(6):785–788, December 1975.

[3] B. Z. Bobrovsky and M. Zakai. Comments on "Status of CR-like lower bounds for nonlinear filtering". *IEEE Trans. Aero. Elect. Syst.*, 26(5):895–896, September 1990.

[4] B.-Z. Bobrovsky, M. M. Zakai, and 0 . Zeitouni. Error bounds for the nonlinear filtering of signals with small diffusion coefficients. *IEEE Trans. Info. Theory,* 34(4):710–721, July 1988.

[5] J. I. Galdos. A Cramer-Rao bound for multidimensional discrete-time dynamical systems. *IEEE Trans. Auto. Contr.*, 25(1):117–119, February 1980.

[6] J. D. Gorman and A. O. Hero. Lower bounds for parametric estimation with constraints. *IEEE Trans. Info. Theory,* 36(6):128.5–1301, November 1990.

[7] T. H. Kerr. Status of CR-like lower bounds for nonlinear filtering. *IEEE Trans. Aero. Elect. Syst.,* 25(5):590–601, September 1989.

[8] T. H. Kerr. Author's reply to comments on "Status of CR-like lower bounds for nonlinear filtering". *IEEE Trans. Aero. Elect. Syst.*, 26(5):896–898, September 1990.

[9] H. Kwakernaak and R. Sivan. *Linear Optimal Control Systems.* Wiley-Interscience, 1972.

[10] T. L. Marzetta. A simple derivation of the constrained multiple parameter Cramer-Rao bound. *IEEE Trans. Sig. Proc.*, 41(6):2247–2249, June 199.3.

[11] O. Zeitouni and A. Dembo. On the maximal achievable accuracy in nonlinear filtering problems. *IEEE Trans. Auto. Contr.*, 33(10):965–967, October 1988.