Cramer–Rao Bounds for Discrete-Time Nonlinear Filtering Problems

Peter C. Doerschuk

***Abstract*— In this note, a Cramer–Rao bound for the mean squared error that can be achieved with nonlinear observations of a nonlinear pth 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.**

I. Introduction

This note concerns lower bounds on the mean squared error (MSE) in nonlinear filtering problems. Specifically, Cramer–Rao bounds (CRB's) are derived for dynamical systems that are more general than those used previously [1], [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

$$x_{k+1} = a(x_k, k) + b(x_k, k)w_k$$
$$y_k = c(x_k, k) + d(x_k, k)v_k$$

where w and v are white Gaussian noise sequences. Let $\hat{x}_k$, a function of $y_0, y_1, \cdots, y_k$, 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 [3]. Those bounds appropriate for nonlinear filtering problems were also recently reviewed [4]–[6]. Additional works not cited in [4] can be found in [7] and [8].

In this paragraph we summarize the notation used in this note. The function $\mathcal{I}(e)$ is one if e is true and zero otherwise. The real numbers are $\mathcal{R}$. The Gaussian probability density function (pdf) with mean m and covariance P is $\mathcal{N}(m, P)$. The $n \times n$ 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 $(z_l', z_{l+1}', \cdots, z_m')'$. $\nabla o = \partial o_i / \partial x_j$ $(o : \mathcal{R}^{n\times 1} \to \mathcal{R}^{m \times 1})$. Occasionally we write a matrix by giving its general entry as a function of the indexes α and β.

The remainder of the note is organized in the following fashion. The new CRB is described in Section II, an example is given in Section III, and conclusions follow in Section IV.

Manuscript received May 24, 1994. This work was supported by National Science Foundation Grant MIP-9110919, a Whirlpool Faculty Fellowship, and the School of Electrical Engineering, Purdue University.

The author is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907–1285 USA.

IEEE Log Number 9412523.

II. Cramer–Rao Bounds

The model is a nonlinear pth order autoregressive (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

$$\begin{aligned} x_{k+1} &= f_k(x_k, \cdots, x_{k-p+1}) \\ &\quad + [q_k(x_k, \cdots, x_{k-p+1}), d_k(x_k, \cdots, x_{k-p+1})]\xi_k \quad (1)\\ y_k &= h_k(x_k, \cdots, x_{k-p+1}) \\ &\quad + [e_k(x_k, \cdots, x_{k-p+1}), r_k(x_k, \cdots, x_{k-p+1})]\xi_k \quad (2)\\ y_k &= h_k(x_k, \cdots, x_{\max(k-p+1, 1-p)}) \\ &\quad + r_k(x_k, \cdots, x_{\max(k-p+1, 1-p)})v_k \quad (3)\end{aligned}$$

where (1) and (2) hold for $k = 0, \cdots, K-1$; (3) holds for $k = 1-p, \cdots, -1$ and $k = K$; $x_k, y_k \in \mathcal{R}^n$; 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. $\mathcal{N}(0, I_{2n})$; v_k is i.i.d. $\mathcal{N}(0, I_n)$; $(x_0', \cdots, x_{1-p}')'$ is $p_0(x_0, \cdots, x_{1-p})$ which is never zero; ξ, v, and $(x_0', \cdots, x_{1-p}')'$ are independent; the covariance Σ_k defined by

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

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

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

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 [1] is restricted to the case

1) $p = 1$,
2) $q_k(x_k, \cdots, x_{k-p+1}) = q_k$ independent of x and full rank,
3) $r_k(x_k, \cdots, x_{k-p+1}) = r_k$ independent of x and full rank,
4) $d_k(x_k, \cdots, x_{k-p+1}) = e_k(x_k, \cdots, x_{k-p+1}) = 0$.

The generalization described here is worthwhile for two reasons: 1) state equations derived from system identification procedures typically have $p > 1$ and 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 processes 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 one and hence can be written $Q = qq'$ where $q \in \mathcal{R}^{n\times 1}$. If (F, q) is controllable, then a similarity transformation exists which transforms the system to canonical controllable form [9, Section 1.9f] which is exactly the form of a nth 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 [1] and [2]. Define the trajectory error covariance $\Lambda \in \mathcal{R}^{(K+p)n\times(K+p)n}$ with blocks $\Lambda_{l,k} \in \mathcal{R}^{n\times n}$ given by $\Lambda_{l,k} = E\{(x_l - \hat{x}_l^*)(x_k -$

$\hat{x}_k^{*\prime}\}$. Define the Fisher information matrix J by, under appropriate regularity assumptions, $J = E\{\nabla_{x_{1-p}^K}\nabla_{x_{1-p}^K} \ln p(y_{1-p}^K, x_{1-p}^K)\}$ 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 semidefinite. The estimation error at time K is $\Lambda_{K,K}$ so the desired CRB is

$$\Lambda_{K,K} - [0 \quad I_n]J^{-1}\begin{bmatrix} 0 \\ I_n \end{bmatrix} \geq 0. \tag{4}$$

There are two difficulties in applying (4). The first difficulty is that for a long trajectory the matrix J is large, and it is difficult to compute J^{-1}. This problem is circumvented in this note and [1] and [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 pth order AR process and an observation equation where the observation at time k depends on the AR process at times $k, k-1, \cdots, k-p+1$. This system can be transformed by state augmentation into a linear Gaussian state-variable system. In the state-variable system $\Lambda_{K,K}^{\text{state}}$ 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 $\Lambda_{K,K}^{\text{state}}$ is the desired bound on the performance of any estimator for the nonlinear system. The second difficulty is that the computation of J typically requires numerical computation of expectations. While the approach of this note and [1] and [2] still require such computations, the approach organizes the computations so that they can be done 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

$$f_k(x_{k-p+1}^k) = \sum_{i=0}^{p-1} A_{k,i}x_{k-i}$$

$$h_k(x_{\max(k-p+1,1-p)}^k) = \sum_{i=0}^{\min(p-1,k+p-1)} C_{k,i}x_{k-i}$$

$$p_0(x_{1-p}^0) = \mathcal{N}(0, P_0)(x_{1-p}^0)$$

$$\Sigma_k(x_{k-p+1}^k) = \check{\Sigma}_k \quad \text{for } k = 0, \cdots, K-1$$

$$R_k(x_{\max(k-p+1,1-p)}^k) = \check{R}_k \quad \text{for } k = 1-p, \cdots, -1 \text{ and } k = K$$

where $\check{\Sigma}_k$ and $\check{R}_k$ are independent of x, P_0 (P_0^{-1}) has $n \times n$ blocks $P_{i,j}$ ($P^{i,j}$) for i and j in $0, \cdots, 1-p$, and

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

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

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

For later convenience, define

$$\Delta_{i,k,l}^{so} = \frac{1}{2}E\left\{\operatorname{tr}\left[\Sigma_i \frac{\partial^2 \Sigma_i^{-1}}{\partial x_{l,\beta}\partial x_{k,\alpha}}\right] + \nabla_{x_l}\nabla_{x_k} \ln\det\Sigma_i\right\} + E\left\{[\nabla_{x_l}f_i', \nabla_{x_l}h_i']\Sigma_i^{-1}\begin{bmatrix} \nabla_{x_k}f_i \\ \nabla_{x_k}h_i \end{bmatrix}\right\} \tag{5}$$

$$\Delta_{i,k,l}^{o} = \frac{1}{2}E\left\{\operatorname{tr}\left[R_i \frac{\partial^2 R_i^{-1}}{\partial x_{l,\beta}\partial x_{k,\alpha}}\right] + \nabla_{x_l}\nabla_{x_k} \ln\det R_i\right\} + E\{\nabla_{x_l}h_i'R_i^{-1}\nabla_{x_k}h_i\} \tag{6}$$

$$\Lambda_{k,l} = -E\left\{[\nabla_{x_l}f_k', \nabla_{x_l}h_k']\Sigma_k^{-1}\begin{bmatrix} I_n \\ 0_n \end{bmatrix}\right\} = -E\{\nabla_{x_l}f_k'W_k + \nabla_{x_l}h_k'U_k'\} \tag{7}$$

$$\Gamma_k = E\left\{[I_n, 0_n]\Sigma_k^{-1}\begin{bmatrix} I_n \\ 0_n \end{bmatrix}\right\} = E\{W_k\}. \tag{8}$$

These formulas simplify when, for example, $\Sigma_k(x_{k-p+1}^k)$ is independent of x. To determine the linear Gaussian system the user must give values for $\Delta_{i,k,l}^{so}$, $\Delta_{i,k,l}^{o}$, $\Lambda_{k,l}$, 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 (16).

We now derive the equations which must be satisfied by A, C, P_0, $\check{\Sigma}$, and $\check{R}$. Define

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

Then, the natural logarithm of the joint pdf for the x and y trajectories is

$$\begin{aligned}
\ln p = K_1 &- \frac{1}{2}\sum_{i=0}^{K-1} z_i(x_{i-p+1}^{i+1})'\Sigma_i^{-1}(x_{i-p+1}^i)z_i(x_{i-p+1}^{i+1}) \\
&- \frac{1}{2}\sum_{i=0}^{K-1} \ln\det\Sigma_i(x_{i-p+1}^i) \\
&- \frac{1}{2}[y_K - h_K(x_{\max(K-p+1,1-p)}^K)]'R_K^{-1}(x_{\max(K-p+1,1-p)}^K) \\
&\times [y_K - h_K(x_{\max(K-p+1,1-p)}^K)] \\
&- \frac{1}{2}\ln\det R_K(x_{\max(K-p+1,1-p)}^K) \\
&- \frac{1}{2}\sum_{i=1-p}^{-1} [y_i - h_i(x_{\max(i-p+1,1-p)}^i)]'R_i^{-1}(x_{\max(i-p+1,1-p)}^i) \\
&\times [y_i - h_i(x_{\max(i-p+1,1-p)}^i)] \\
&- \frac{1}{2}\sum_{i=1-p}^{-1} \ln\det R_i(x_{\max(i-p+1,1-p)}^i) + \ln p_0(x_{1-p}^0)
\end{aligned}$$

where K_1 is a constant. By equating $E\{\nabla_{x_l}\nabla_{x_k} \ln p\}$ for the nonlinear and the linear-Gaussian systems we derive a system of equations that A, C, P_0, $\check{\Sigma}$, and $\check{R}$ must satisfy. It is only necessary to consider $l \leq k$ because $\nabla_{x_l}\nabla_{x_k} \ln p = \nabla_{x_k}\nabla_{x_l} \ln p'$. There are two sets of equations. The first set, for $k = 2-p, \cdots, K$ and $l = \max(k-p, 1-p), \cdots, k-1$, are

$$\begin{aligned}
&- \sum_{i=\max(k,0)}^{\min(l+p-1,K-1)} \Delta_{i,k,l}^{so} - \Lambda_{k-1,l}\mathcal{I}(k \geq 1) \\
&- \Delta_{K,k,l}^{o}\mathcal{I}(K-p+1 \leq l) - \sum_{i=k}^{-1} \Delta_{i,k,l}^{o} \\
&+ E\{\nabla_{x_l}\nabla_{x_k} \ln p_0\}\mathcal{I}(k \leq 0) \\
&= - \sum_{i=\max(k,0)}^{\min(l+p-1,K-1)} J_{i,i-l}'\check{\Sigma}_i^{-1}J_{i,i-k} \\
&+ \left(A_{k-1,k-1-l}'\check{W}_{k-1} + C_{k-1,k-1-l}'\check{U}_{k-1}'\right)\mathcal{I}(k \geq 1) \\
&- C_{K,K-l}'\check{R}_K^{-1}C_{K,K-k}\mathcal{I}(K-p+1 \leq l) \\
&- \sum_{i=k}^{-1} C_{i,i-l}'\check{R}_i^{-1}C_{i,i-k} - P^{l,k}\mathcal{I}(k \leq 0)
\end{aligned} \tag{9}$$

and the second set, for $k = 1-p, \cdots, K$, are

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

$$
\begin{aligned}
&= -\sum_{i=\max(k,0)}^{\min(k+p-1,K-1)} J'_{i,i-k}\check{\Sigma}_i^{-1}J_{i,i-k} - \check{W}_{k-1}\mathcal{I}(k \geq 1)\\
&\quad - C'_{K,K-k}\check{R}_K^{-1}C_{K,K-k}\mathcal{I}(K-p+1 \leq k)\\
&\quad - \sum_{i=k}^{-1} C'_{i,i-k}\check{R}_i^{-1}C_{i,i-k} - P^{k,k}\mathcal{I}(k \leq 0). \qquad (10)
\end{aligned}
$$

Define in sequence the following quantities:

i)

$$P^{l,k} \doteq -E\{\nabla_{x_l}\nabla_{x_k}\ln p_0\} \qquad (11)$$

for $k = 1-p,\cdots,0$ and $l = 1-p,\cdots,k$,

ii)

$$\check{W}_k \doteq \Gamma_k \qquad (12)$$

for $k = 0,\cdots,K-1$,

iii)

$$A_{k,k-l} \doteq -\check{W}_k^{-1}\Lambda'_{k,l} \qquad (13)$$

for $k = 0,\cdots,K-1$ and $l = k-p+1,\cdots,k$,

iv)

$$\check{U}_k \doteq 0 \qquad (14)$$

for $k = 0,\cdots,K-1$,

v)

$$\check{R}_k \doteq \text{arbitrary positive definite matrix} \qquad (15)$$

for $k = 1-p,\cdots,K$, and

vi)

$$
\begin{aligned}
D_{k,l} \doteq &\sum_{i=\max(k,0)}^{\min(l+p-1,K-1)} \left(\Delta^{so}_{i,k,l} - A'_{i,i-l}\check{W}_i A_{i,i-k}\right)\\
&+ \Delta^{o}_{K,k,l}\mathcal{I}(K-p+1 \leq l) + \sum_{i=k}^{-1}\Delta^{o}_{i,k,l} \qquad (16)
\end{aligned}
$$

for $k = 1-p,\cdots,K$ and $l = \max(k-p,1-p),\cdots,k$.
Use these definitions in (9) and (10) to get

$$\sum_{i=k}^{\min(l+p-1,K)} C'_{i,i-l}\check{R}_i^{-1}C_{i,i-k} = D_{k,l} \qquad (17)$$

for $k = 2-p,\cdots,K$ and $l = \max(k-p,1-p),\cdots,k-1$ and

$$\sum_{i=k}^{\min(k+p-1,K)} C'_{i,i-k}\check{R}_i^{-1}C_{i,i-k} = D_{k,k} \qquad (18)$$

for $k = 1-p,\cdots,K$. In spite of being quadratic in $C_{i,j}$, (17) and (18) can be solved recursively for $C_{i,j}$. The procedure is given in Fig. 1. Once A, C, P_0, $\check{\Sigma}$, and $\check{R}$ have been determined by this procedure, then the state augmentation and Kalman filter computations needed to determine the bound 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 [1]. (There is a typographical error in [1, 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 C's 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

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

where

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

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

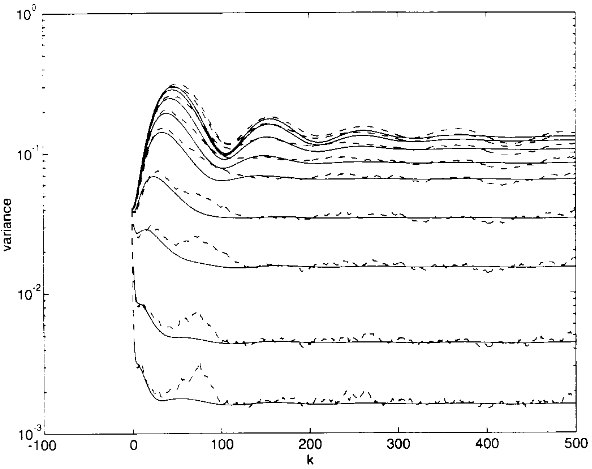
Fig. 1. Algorithm for the solution of (17) and (18). The control structures are written in the C programming language. The notation $^{1/2}$ indicates matrix square roots ($R = R^{1/2}(R^{1/2})'$, $R^{T/2} \doteq (R^{1/2})'$).

III. EXAMPLE

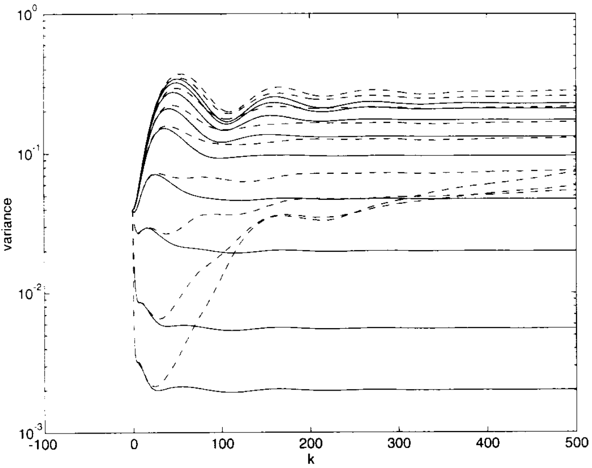
In this section we consider an example estimation problem to demonstrate the computations involved and that the bound is tight at least for some estimation problems. (Please contact the author for copies of the software.) The physical system is a damped pendulum driven by a random torque where noisy 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

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

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 ($\mathcal{N}(0,\sigma_v^2)$). 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 σ_v. Notice that only one set of expectations need be computed for the CRB to determine the bound for any σ_v. The parameters are $l = 1$, $m = 1$, $\gamma = 1$, $g = 10$, and $T = .01$.



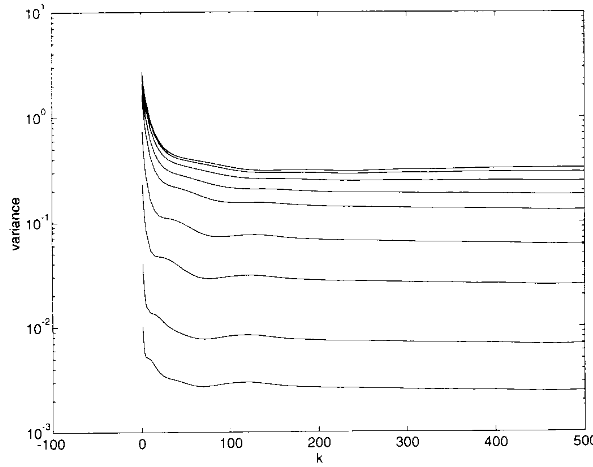
(a)



(b)

Fig. 2. Comparison of CRB (solid line) and EKF estimation variance (dotted line) for a range of measurement noise variance: $\sigma_v \in \{.1, .2, .5, 1, 2, 3, 5, 10, 100\}$. (a) Gaussian initial condition, $\sigma_w = 15$. (b) Gaussian initial condition, $\sigma_w = 20$.

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 (Fig. 2(a)), where the bound is tight, $\sigma_w = 15$ 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 $\max_{k=-1,\cdots,500} |\phi_k|$ is 0.758271 ± 0.271391. In the second case (Fig. 2(b)), where the bound is loose, $\sigma_w = 20$ 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 $\max_{k=-1,\cdots,500} |\phi_k|$ 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 than in the first case were used to evaluate its performance. 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, $\sigma_w = 25$ and the pdf on the initial condition



(c)

Fig. 2. Continued. (c) Nearly uniform initial condition, $\sigma_w = 25$.

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 $\sigma_p = .5$ and $\sigma_{\dot\phi} = 2$. This case is strongly nonlinear since the sample mean and standard deviation of $\max_{k=-1,\cdots,500} |\phi_k|$ 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 Fig. 2(c) based on 10^3 state trajectories.

IV. Conclusions

A Cramer–Rao bound for the mean squared error that can be achieved with nonlinear observations of a nonlinear pth order 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. The bound is a generalization of the results of [1] and [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], [11]. The merger of the constrained CRB tools with the dynamical system approach of this note and [1] and [2] does not appear to be straightforward.

References

[1] J. I. Galdos, "A Cramer–Rao bound for multidimensional discrete-time dynamical systems," *IEEE Trans. Automat. Contr.*, vol. AC-25, no. 1, pp. 117–119, Feb. 1980.

[2] B. Z. Bobrovsky and M. Zakai, "A lower bound on the estimation error for Markov processes," *IEEE Trans. Automat. Contr.*, vol. AC-20, no. 6, pp. 785–788, Dec. 1975.

[3] J. S. Abel, "A bound on mean-square-estimate error," *IEEE Trans. Inform. Theory*, vol. 39, no. 5, pp. 1675–1680, Sep. 1993.

[4] T. H. Kerr, "Status of CR-like lower bounds for nonlinear filtering," *IEEE Trans. Aero. Elect. Syst.*, vol. 25, no. 5, pp. 590–601, Sep. 1989.

[5] B. Z. Bobrovsky and M. Zakai, "Comments on 'Status of CR-like lower bounds for nonlinear filtering'," *IEEE Trans. Aero. Elect. Syst.*, vol. 26, no. 5, pp. 895–896, Sep. 1990.

[6] T. H. Kerr, "Author's reply to comments on 'Status of CR-like lower bounds for nonlinear filtering'," *IEEE Trans. Aero. Elect. Syst.*, vol. 26, no. 5, pp. 896–898, Sep. 1990.

[7] O. Zeitouni and A. Dembo, "On the maximal achievable accuracy in nonlinear filtering problems," *IEEE Trans. Automat. Contr.*, vol. 33, no. 10, pp. 965–967, Oct. 1988.
[8] B.-Z. Bobrovsky, M. M. Zakai, and O. Zeitouni, "Error bounds for the nonlinear filtering of signals with small diffusion coefficients," *IEEE Trans. Inform. Theory*, vol. 34, no. 4, pp. 710–721, July 1988.
[9] H. Kwakernaak and R. Sivan, *Linear Optimal Control Systems*. New York: Wiley-Interscience, 1972.
[10] T. L. Marzetta, "A simple derivation of the constrained multiple parameter Cramer–Rao bound," *IEEE Trans. Sig. Proc.*, vol. 41, no. 6, pp. 2247–2249, June 1993.
[11] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inform. Theory*, vol. 36, no. 6, pp. 1285–1301, Nov. 1990.

Adaptive Feedback Control Algorithms for Large Data Transfers in High-Speed Networks

Rauf Izmailov

Abstract— **Two linear feedback control algorithms for handling and preventing congestion in broadband asynchronous transfer mode networks are proposed and analyzed. The fluid approximation model is described with a continuous-time system of delay-differential equations. The algorithms are asymptotically stable, and the transient processes are nonoscillatory. The control parameters are locally optimal (optimality is based on the asymptotic rate of convergence). The results of numerical experiments suggest that these parameters are globally optimal as well.**

I. Introduction

Asynchronous transfer mode (ATM) transport technology is generally considered as a basis for future integrated telecommunications service. Since there would be an inevitable interaction and interference among users in the communication network, an increasing amount of research has been devoted to different control issues (see [1], [12], [15], [16], and [19] and their references). One of the basic problems arising here is the presence of propagation delays which pose a challenge for stability, as speed of data transmission in modern high-speed networks keeps increasing.

In most of the proposed algorithms and models (see [4], [6], [10], [11], and [18] and their references), control decisions are based on a single number (the deviation of the state of the system from the target value) or a single bit (the sign of such deviation). Analysis and numerical simulations [4], [8], [7], [17] demonstrated that the stability of such algorithms in the presence of propagation delays has the form of bounded oscillations (occurring even in the deterministic setting).

The single number limitation on the number of control parameters appears to be the key obstacle to the elimination of these undesirable oscillations. As proved in [8], a large class of feedback algorithms based on a single number always has an unstable equilibrium. Thus it is natural to address the question of what additional parameters should be considered and how to translate them into control algorithms. This question, however, has only recently begun to be addressed [3].

Manuscript received April 5, 1994; revised August 26,1 994 and December 15, 1995.
The author is with NEC USA, Inc., C&C Research Laboratories, Princeton, NJ 08540 USA.
IEEE Log Number 9412525.

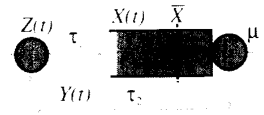


Fig. 1. The model.

In this paper we consider a single connection controlled by an access regulator whose parameters are dynamically adjusted according to a linear function of the several states of the buffer in different times. We consider two versions of the same algorithm, differing by the number of control parameters. We prove asymptotic stability of both algorithms and show that the proposed parameters guarantee a nonoscillatory asymptotic behavior. The rate of asymptotic convergence is calculated explicitly in both cases, and the parameters are locally optimal (convergence rate deteriorates with any small change of parameters). Numerical results suggest the chosen parameters are globally optimal as well.

II. Model and Control Algorithms

Consider a single connection between a source controlled by an access regulator and a distant node served with a constant transmission capacity μ (Fig. 1). The traffic source is monitored and regulated by the access regulator, and the distant bottleneck node sends back the information on its congestion status, defined as the difference between the current buffer contents and the target value (a fixed threshold).

To describe large data transfers and isolate the issue of adaptation mechanism from other considerations, assume (as in [8]) that there is an infinite amount of traffic to be sent to the remote node. To capture the small size of ATM packets as well as high rates of the network, we approximate the traffic by fluid flows. The access regulator controls the current rate $Z(t)$ basing its decisions on the buffer contents $X(t)$ of the distant node, which is continually monitored by the source. A target value $\bar{X}$ (threshold) of the remote buffer contents is fixed: if $X(t) > \bar{X}$, the node is considered congested. The delays from the source to the bottleneck and back are τ_1 and τ_2, which add up the round-trip delay $\tau = \tau_1 + \tau_2$ (we assume the propagation delay is essentially larger than the queueing delay, so τ is fixed). The control objective is to adapt $Z(t)$ to μ dynamically, while keeping $X(t)$ at an acceptable level.

The first algorithm takes into account the deviations of $X(t)$ from the target value $\bar{X}$ during two consecutive time instants, separated by the round-trip delay τ. These deviations are weighted with linear gain parameters a and b, so in a neighborhood of the threshold $\bar{X}$ the system is described by

$$X'(t) = Z(t - \tau_1) - \mu, \tag{1}$$

$$Z'(t) = -a(X(t - \tau_2) - \bar{X}) - b(X(t - \tau_2 - \tau) - \bar{X}). \tag{2}$$

We take the derivative of (2) and substitute $Y(t) = Z(t) - \mu$ to obtain the delay-differential equation in the normalized time scale $T = t/\tau$ (where $A = a\tau^2$, $B = b\tau^2$)

$$Y''(T) + AY(T-1) + BY(T-2) = 0. \tag{3}$$

Its characteristic equation $f(z) = z^2e^{2z} + Ae^z + B = 0$ has an infinite number of roots λ_i. The location of these roots on the complex plane determines [2, Theorem 6.7] the asymptotic behavior of $Y(T)$. In particular, the degree of stability $\lambda = \sup_i\{\Re\lambda_i\}$ guarantees