Posterior Cramér–Rao Bounds for Discrete-Time Nonlinear Filtering

Petr Tichavský, *Member, IEEE*, Carlos H. Muravchik, *Member, IEEE*, and Arye Nehorai, *Fellow, IEEE*

***Abstract*—A mean-square error lower bound for the discrete-time nonlinear filtering problem is derived based on the Van Trees (posterior) version of the Cramér–Rao inequality. This lower bound is applicable to multidimensional nonlinear, possibly non-Gaussian, dynamical systems and is more general than the previous bounds in the literature. The case of singular conditional distribution of the one-step-ahead state vector given the present state is considered. The bound is evaluated for three important examples: the recursive estimation of slowly varying parameters of an autoregressive process, tracking a slowly varying frequency of a single cisoid in noise, and tracking parameters of a sinusoidal frequency with sinusoidal phase modulation.**

***Index Terms*—Adaptive estimation, Kalman filtering, nonlinear filters, time-varying systems, tracking filters.**

I. Introduction

DISCRETE-TIME nonlinear filtering or the associated problem of adaptive system identification arise in various applications such as adaptive control, analysis, and prediction of nonstationary time series. As is well known, the optimal estimator for this problem cannot be built in general, and it is necessary to turn to one of the large number of existing suboptimal filtering techniques [1]. Assessing the achievable performance may be difficult, and we have to resort to simulations and comparing proximity to lower bounds corresponding to optimum performance. Lower bounds give an indication of performance limitations, and consequently, they can also be used to determine whether imposed performance requirements are realistic or not.

In time-invariant statistical models, a commonly used lower bound is the Cramér–Rao bound (CRB), given by the inverse of the Fisher information matrix. In the time-varying systems context we deal with here, the estimated parameter vector has to be considered random since it corresponds to an underlying nonlinear, randomly driven model. A lower bound that is analogous to the CRB for random parameters was derived in [11]; this bound is usually referred to as the Van Trees version of the CRB, or posterior CRB (PCRB) [16]. Some properties of the PCRB are summarized in Section II.

Several lower bounds for nonlinear dynamical systems have appeared in the literature; see the overview in [6]. However, the continuous-time case has received heavy emphasis but not the discrete-time case, which is of greater practical importance. Bobrovsky and Zakai [2] were the first to apply the Cramér–Rao theory to scalar discrete-time systems. The bound was later improved and generalized to the multidimensional case by Galdos [3]. Both of these bounds were obtained by comparing the information matrix of the original system with an information matrix of a suitable Gaussian system. The bound in [3] is already quite general, but it still has some limitations (see the discussion in [6]), i.e., the assumption that the dimension of the system and measurements are identical. Recently, the approach by Galdos has been generalized for nonlinear pth-order autoregressive processes driven by additive Gaussian noise with state-dependent gain [4].

In Section III of this paper, a novel and simple derivation of the posterior CRB for the discrete-time multidimensional nonlinear filtering problem that avoids any Gaussian assumptions is presented. The derivation is obtained from first principles and differs from other approaches that instead consider comparison of the original nonlinear system with an appropriate linear Gaussian system. We present an example of a linear Gaussian system (which is different from those in [2] and [3]) that has the same associated information matrix as the original system. In Section IV, the lower bound is extended for a frequently occurring case of nonlinear filtering, where the conditional distribution of the state one step ahead, given the current state, is singular. Note that a special case of a similar extension was proposed in [3]. Section V illustrates an application of the bound for three important examples:

- recursive estimation of slowly varying parameters of an autoregressive process;
- tracking of a slowly varying frequency of a single cisoid in noise (a new alternate derivation of the lower bound in [16]);
- tracking parameters of a varying frequency that is modulated by a sinusoid [17].

Conclusions are drawn in Section VI.

II. Properties of the PCRB (Review)

Let x represent a vector of measured data, let θ be an r-dimensional estimated random parameter, let $p_{x,\theta}(X, \Theta)$ be

Manuscript received December 3, 1996; revised September 12, 1997. A short version of this paper was presented at the First European Conference on Signal Analysis and Prediction ECSAP-97, June 1997, Prague, Czech Republic. This work was supported in part by Grant 102/97/0466 of the Grant Agency of the Czech Republic, CICPBA and UNLP, the Air Force Office of Scientific Research under Grant F49620-97-1-0481, the National Science Foundation under Grant MIP-9615590, and the Office of Naval Research under Grant N00014-96-1-1078.

The associate editor coordinating the review of this paper and approving it for publication was Prof. James A. Bucklew.

P. Tichavský is with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic (e-mail: tichavsk@utia.cas.cz).

C. Muravchik is with LEICI, Departimento Electrotecnia, Universidad Nacional de La Plata, La Plata, Argentina.

A. Nehorai is with the Department of Electrical Engineering and Computer Sciences, University of Illinois at Chicago, Chicago, IL 60607-7053 USA.

Publisher Item Identifier S 1053-587X(98)03256-5.

the joint probability density of the pair (x, θ), and let $g(x)$ be a function of x, which is an estimate of θ. The PCRB on the estimation error has the form

$$P \triangleq \mathrm{E}\{[g(x)-\theta][g(x)-\theta]^T\} \geq J^{-1} \tag{1}$$

where J is the $r \times r$ (Fisher) information matrix with the elements

$$J_{ij} = \mathrm{E}\left[-\frac{\partial^2 \log p_{x,\theta}(X,\Theta)}{\partial\Theta_i\,\partial\Theta_j}\right] \qquad i, j = 1, \cdots, r \tag{2}$$

provided that the derivatives and expectations in (1) and (2) exist. The superscript "T" in (1) denotes the transpose of a matrix, and the inequality in (1) means that the difference $P - J^{-1}$ is a positive semidefinite matrix. The proof given in [10] or [11] holds under the additional condition of

$$\lim_{\Theta_i\to\infty} B(\Theta)p_\theta(\Theta) = 0, \; \lim_{\Theta_i\to-\infty} B(\Theta)p_\theta(\Theta) = 0 \qquad i = 1, \cdots, r \tag{3}$$

where $B(\Theta)$ is the estimation bias conditioned by $\theta = \Theta$, and

$$B(\Theta) = \int_{-\infty}^{\infty} [g(X) - \Theta]\, p_{x|\theta}(X|\Theta)\, dX. \tag{4}$$

Let ∇ and δ be operators of the first and second-order partial derivatives, respectively

$$\nabla_\Theta = \left[\frac{\partial}{\partial\Theta_1}, \cdots, \frac{\partial}{\partial\Theta_r}\right]^T \tag{5}$$

$$\Delta_\Psi^\Theta = \nabla_\Psi \nabla_\Theta^T. \tag{6}$$

Using this notation, (2) can be written as

$$J = \mathrm{E}[-\Delta_\Theta^\Theta \log p_{x,\theta}(X, \Theta)]. \tag{7}$$

Since $p_{x,\theta}(X, \Theta) = p_{x|\theta}(X|\Theta) \cdot p_\theta(\Theta)$, it can easily be seen that J can be decomposed into two additive parts:

$$J = J_D + J_P \tag{8}$$

where J_D represents the information obtained from the data, and J_P represents the *a priori* information

$$J_D = \mathrm{E}\{-\Delta_\Theta^\Theta \log p_{x|\theta}(X|\Theta)\} \qquad (r \times r) \tag{9}$$

$$J_P = \mathrm{E}\{-\Delta_\Theta^\Theta \log p_\theta(\Theta)\} \qquad (r \times r) \tag{10}$$

provided that the expectations in (9) and (10) exist. Note that J_D is an expectation of the standard Fisher information matrix over the *a priori* distribution of Θ.

An alternative expression for the information matrix can be derived from the equality $p_{x,\theta}(X, \Theta) = p_{\theta|x}(\Theta|X) \cdot p_x(X)$. Since $p_x(X)$ is an integral of $p_{x,\theta}(X, \Theta)$ over Θ, it does not depend any longer on Θ; therefore, we have

$$J = \mathrm{E}\{-\Delta_\Theta^\Theta \log p_{\theta|x}(\Theta|X)\}. \tag{11}$$

For example, if the posterior distribution of θ conditioned on the data vector x is Gaussian with mean $\overline{\theta}_x$ and a (regular) covariance matrix Σ_x

$$-\log p_{\theta|x}(\Theta|X) = c_0 + \tfrac{1}{2}(\Theta - \overline{\theta}_x)^T \Sigma_x^{-1} (\Theta - \overline{\theta}_x) \tag{12}$$

holds, where c_0 denotes a constant independent of Θ. Then, the information matrix in (11) reads

$$J = \mathrm{E}\{\Sigma_x^{-1}\}. \tag{13}$$

If θ is estimated by $g(x) = \mathrm{E}(\theta|x)$, then (1) is satisfied with equality. This is exactly the case for the Kalman filter when performing the task of linear filtering.

Assume now that the parameter θ is decomposed into two parts as $\theta = [\theta_\alpha^T, \theta_\beta^T]^T$, and the information matrix J is correspondingly decomposed into blocks

$$J = \begin{bmatrix} J_{\alpha\alpha} & J_{\alpha\beta} \\ J_{\beta\alpha} & J_{\beta\beta} \end{bmatrix}. \tag{14}$$

It can easily be shown that the covariance of estimation of θ_β is lower bounded by the right-lower block of J^{-1}, i.e.,

$$P_\beta \triangleq \mathrm{E}\{[g_\beta(x) - \theta_\beta][g_\beta(x) - \theta_\beta]^T\} \geq [J_{\beta\beta} - J_{\beta\alpha} J_{\alpha\alpha}^{-1} J_{\alpha\beta}]^{-1} \tag{15}$$

assuming that $J_{\alpha\alpha}^{-1}$ exists. In the following, the matrix $J_{\beta\beta} - J_{\beta\alpha} J_{\alpha\alpha}^{-1} J_{\alpha\beta}$ will be called the *information submatrix* for parameter θ_β.

III. A Lower Bound for the Nonlinear Filtering Problem

Consider the nonlinear filtering problem

$$x_{n+1} = f_n(x_n, w_n) \tag{16}$$

$$z_n = h_n(x_n, v_n) \tag{17}$$

where

x_n	system state at time n;
$\{z_n\}$	measurement process;
$\{w_n\}$ and $\{v_n\}$	independent white processes (i.e., sequences of mutually independent random variables or vectors);
f_n and h_n	(in general) nonlinear functions.

The functions f_n and h_n may depend on time n. Further assume that the initial state x_0 has a known probability density function $p(x_0)$. Let the dimension of the states $\{x_n\}$ be r.

Equations (16) and (17) together with $p(x_0)$ determine unambiguously the joint probability distribution of $X_n = (x_0, \cdots, x_n)$ and $Z_n = (z_0, \cdots, z_n)$ for an arbitrary n [2]

$$p(X_n, Z_n) = p(x_0) \prod_{j=1}^{n} p(z_j|x_j) \prod_{k=1}^{n} p(x_k|x_{k-1}). \tag{18}$$

In (18) as well as in the sequel, $p(\cdot)$'s refer to (unconditional and conditional) probability densities of the variables depicted in the argument of p's. The conditional probability densities $p(x_k|x_{k-1})$ and $p(z_k|x_k)$ follow from (16) and (17), respectively, under suitable hypotheses.

Let $J(X_n)$ be the $(nr \times nr)$ information matrix of X_n derived from the above joint distribution. The problem that we wish to solve in this section is the computation of the information submatrix for estimating x_n, which is denoted J_n, which is given as the inverse of the $(r \times r)$ right-lower block of $[J(X_n)]^{-1}$. The matrix J_n^{-1} will provide a lower bound on the mean square error of estimating x_n. In the sequel, $p(X_n, Z_n)$ is denoted by p_n for brevity.

Decompose X_n as $X_n = [X_{n-1}^T, x_n^T]^T$ and $J(X_n)$ correspondingly as

$$J(X_n) = \begin{bmatrix} A_n & B_n \\ B_n^T & C_n \end{bmatrix} \triangleq \begin{bmatrix} \mathrm{E}\{-\Delta_{X_{n-1}}^{X_{n-1}} \log p_n\} & \mathrm{E}\{-\Delta_{X_{n-1}}^{x_n} \log p_n\} \\ \mathrm{E}\{-\Delta_{x_n}^{X_{n-1}} \log p_n\} & \mathrm{E}\{-\Delta_{x_n}^{x_n} \log p_n\} \end{bmatrix} \quad (19)$$

provided that the derivatives and the expectations exist. Comparison of (16) and (20) gives

$$J_n = C_n - B_n^T A_n^{-1} B_n. \quad (20)$$

Thus, computation of the $(r \times r)$ matrix J_n involves either calculation of the inverse of $[(n-1)r \times (n-1)r]$ matrix A_n or inverse of the full $(nr \times nr)$ matrix $J(X_n)$.

The following proposition gives a recipe for computing J_n recursively without manipulating large matrices such as A_n or $J(X_n)$. In particular, an efficient method for computing the limit of J_n for $n \to \infty$ follows from the recursion.

Proposition 1: The sequence $\{J_n\}$ of posterior information submatrices for estimating state vectors $\{x_n\}$ obeys the recursion

$$J_{n+1} = D_n^{22} - D_n^{21}(J_n + D_n^{11})^{-1} D_n^{12} \quad (21)$$

where

$$D_n^{11} = \mathrm{E}\{-\Delta_{x_n}^{x_n} \log p(x_{n+1}|x_n)\} \qquad (r \times r) \quad (22)$$
$$D_n^{12} = \mathrm{E}\{-\Delta_{x_n}^{x_{n+1}} \log p(x_{n+1}|x_n)\} \qquad (r \times r) \quad (23)$$
$$D_n^{21} = \mathrm{E}\{-\Delta_{x_{n+1}}^{x_n} \log p(x_{n+1}|x_n)\} = [D_n^{12}]^T \quad (24)$$
$$D_n^{22} = \mathrm{E}\{-\Delta_{x_{n+1}}^{x_{n+1}} \log p(x_{n+1}|x_n)\} + \mathrm{E}\{-\Delta_{x_{n+1}}^{x_{n+1}} \log p(z_{n+1}|x_{n+1})\} \qquad (r \times r). \quad (25)$$

Proof: The joint probability function of X_{n+1} and Z_{n+1} can be written as

$$\begin{aligned} p_{n+1} &\triangleq p(X_{n+1}, Z_{n+1}) \\ &= p(X_n, Z_n) \cdot p(x_{n+1}|X_n, Z_n) \cdot p(z_{n+1}|x_{n+1}, X_n, Z_n) \\ &= p_n \cdot p(x_{n+1}|x_n) \cdot p(z_{n+1}|x_{n+1}). \end{aligned} \quad (26)$$

Using (26) and the notations in (19) and (22)–(25), the posterior information matrix for X_{n+1} can be written in block form as

$$J(X_{n+1}) = \begin{bmatrix} A_n & B_n & 0 \\ B_n^T & C_n + D_n^{11} & D_n^{12} \\ 0 & D_n^{21} & D_n^{22} \end{bmatrix} \quad (27)$$

where 0's stand for zero blocks of appropriate dimensions.

The information submatrix J_{n+1} can be found as an inverse of the right-lower $(r \times r)$ submatrix of $[J(X_{n+1})]^{-1}$

$$\begin{aligned} J_{n+1} &= D_n^{22} - [0 \quad D_n^{21}] \begin{bmatrix} A_n & B_n \\ B_n^T & C_n + D_n^{11} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ D_n^{12} \end{bmatrix} \\ &= D_n^{22} - D_n^{21}[C_n + D_n^{11} - B_n^T A_n^{-1} B_n]^{-1} D_n^{12}. \end{aligned} \quad (28)$$

Using the definition of J_n in (20), we obtain the desired formula (21). ■

Note that the recursion in (21) involves computations with matrices of dimension $(r \times r)$. The initial information submatrix J_0 can be calculated from the *a priori* probability function $p(x_0)$

$$J_0 = \mathrm{E}\{-\Delta_{x_0}^{x_0} \log p(x_0)\}. \quad (29)$$

A few remarks follow to elucidate special cases.

A. Additive Gaussian Noise

Assume that the nonlinear filtering problem in (16) and (17) has the form

$$x_{n+1} = f_n(x_n) + w_n \quad (30)$$
$$z_n = h_n(x_n) + v_n \quad (31)$$

and that the noises $\{w_n\}$ and $\{v_n\}$ are Gaussian with zero mean and invertible covariance matrices Q_n and R_n, respectively. From these assumptions, it follows that

$$-\log p(x_{n+1}|x_n) = c_1 + \tfrac{1}{2}[x_{n+1} - f_n(x_n)]^T Q_n^{-1}[x_{n+1} - f_n(x_n)] \quad (32)$$
$$-\log p(z_{n+1}|x_{n+1}) = c_2 + \tfrac{1}{2}[z_{n+1} - h_{n+1}(x_{n+1})]^T \cdot R_{n+1}^{-1}[z_{n+1} - h_{n+1}(x_{n+1})] \quad (33)$$

where c_1 and c_2 are constants, and

$$D_n^{11} = \mathrm{E}\{[\nabla_{x_n} f_n^T(x_n)] Q_n^{-1} [\nabla_{x_n} f_n^T(x_n)]^T\} \quad (34)$$
$$D_n^{12} = -\mathrm{E}\{\nabla_{x_n} f_n^T(x_n)\} Q_n^{-1} \quad (35)$$
$$D_n^{22} = Q_n^{-1} + \mathrm{E}\{[\nabla_{x_{n+1}} h_{n+1}^T(x_{n+1})] \cdot R_{n+1}^{-1}[\nabla_{x_{n+1}} h_{n+1}^T(x_{n+1})]^T\}. \quad (36)$$

The well-known solution of the problem in the linear case [with linear functions f_n and g_n in (30) and (31)] is the Kalman filter. This is an algorithm that computes parameters of the conditional distribution of the state x_n given the data Z_n. The distribution is Gaussian, and its mean and covariance matrix are usually denoted $\hat{x}_{n|n}$ and $\Sigma_{n|n}$, respectively. It can easily be shown that the recursion (21) for J_n is identical to those that are usually derived for $\Sigma_{n|n}^{-1}$ from the Kalman filter equations [1].

In order to compare the result (21) with the PCRB computations in [2] and [3], we find matrices $\tilde{F}_n$, $\tilde{H}_n$, $\tilde{Q}_n$, and $\tilde{R}_n$ such that the linear system

$$\tilde{x}_{n+1} = \tilde{F}_n \tilde{x}_n + \tilde{w}_n \quad (37)$$
$$\tilde{z}_n = \tilde{H}_n \tilde{x}_n + \tilde{v}_n \quad (38)$$

has the same information matrix as the original nonlinear system; in (37) and (38), $\{\tilde{w}_n\}$ and $\{\tilde{v}_n\}$ are independent white Gaussian noises with zero means and covariance matrices $\tilde{Q}_n$ and $\tilde{R}_n$, respectively. The matrices $\tilde{F}_n$, $\tilde{H}_n$, $\tilde{Q}_n$, and $\tilde{R}_n$ can be determined by comparing the matrices D_n^{11}, D_n^{12}, and D_n^{22} of the original system, which are obtained from (34)–(37) to those of the linear system in (37) and (38), yielding

$$\tilde{D}_n^{11} = \tilde{F}_n^T \tilde{Q}_n^{-1} \tilde{F}_n \quad (39)$$
$$\tilde{D}_n^{12} = -\tilde{F}_n^T \tilde{Q}_n^{-1} \quad (40)$$
$$\tilde{D}_n^{22} = \tilde{Q}_n^{-1} + \tilde{H}_{n+1}^T \tilde{R}_{n+1}^{-1} \tilde{H}_{n+1}. \quad (41)$$

One possible solution of the above system of equations is

$$\tilde{F}_n = -(D_n^{12})^{-1} D_n^{11} \tag{42}$$
$$\tilde{Q}_n = D_n^{21} (D_n^{11})^{-1} D_n^{12} \tag{43}$$
$$\tilde{R}_n = R_n \tag{44}$$
$$\tilde{H}_n = R_n^{1/2} [D_{n-1}^{22} - (D_{n-1}^{12})^{-1} D_{n-1}^{11} (D_{n-1}^{12})^{-1}]^{1/2} \tag{45}$$

where $A^{1/2}$ denotes the square root of a positive semidefinite matrix A, assuming that the requisite inverses in (42)–(45) exist. Note that the above linear filter is different from those proposed in [2] and [3].

B. A Generalization

Consider the generalization of the nonlinear system in (16) and (17) as

$$x_{n+1} = f_n(x_n, w_n) \tag{46}$$
$$z_n = h_n(x_n, v_n, z_{n-1}, \cdots, z_{n-m}) \tag{47}$$

where m is an integer. It can easily be seen that for the generalized system, the whole derivation of (21) can be repeated en masse, with only two small differences: First, in the initialization, it has to be assumed that $z_{-1}, \cdots, z_{-m}$ are known constants and second that $p(z_{n+1}|x_{n+1}, X_n, Z_n)$ in (26) cannot be reduced to $p(z_{n+1}|x_{n+1})$ but merely to $p(z_{n+1}|x_{n+1}, z_n, \cdots, z_{n-m+1})$. The latter term will also replace the former one in (25).

C. Time-Invariant Solutions

Now, assume that the functions $f_n(\cdot, \cdot)$ and $h_n(\cdot)$ are time invariant (independent of n). It can easily be seen that the matrices $D_n^{11}, \cdots, D_n^{22}$ also do not depend on n. It can be shown that for $n \to \infty$, the matrix J_n converges to a matrix J_∞, which is given as a solution to the equation

$$J_\infty = D_n^{22} - D_n^{21} (J_\infty + D_n^{11})^{-1} D_n^{12}. \tag{48}$$

Note that (48) is a discrete-time algebraic Riccati equation. A more common form of the Riccati equation is obtained if the recursion (21) is equivalently written as

$$\begin{aligned} J_{n+1} = & D_n^{21} (D_n^{11})^{-1} J_n (D_n^{11})^{-1} D_n^{12} - D_n^{21} (D_n^{11})^{-1} \\ & \cdot J_n (J_n + D_n^{11})^{-1} J_n (D_n^{11})^{-1} D_n^{12} \\ & + D_n^{22} - D_n^{21} (D_n^{11})^{-1} D_n^{12} \end{aligned} \tag{49}$$

which can be easily proved by simple algebraic manipulations. Then, put $J_{n+1} = J_n = J_\infty$.

Two popular methods for solving the Riccati equation are derived in [5] and [8], respectively; for a more comprehensive survey, see [7]. In addition, note that there is an available software for solving the equation in Matlab, namely, a function DARESOLV or an older function DLQR.

IV. A Frequent Singular Case

Computation of the information submatrix J_n, as described in the previous section, fails if the conditional distribution of x_{n+1}, given x_n is singular, and therefore, the probability density $p(x_{n+1}|x_n)$ is not defined. In the case of the additive Gaussian noise considered in the previous section, this happens when the matrix Q_n is singular. In order to deal with these cases, consider the following modification of the original problem.

Assume that the state vector x_n can be written in block form as

$$x_n = \begin{bmatrix} x_n^{(1)} \\ x_n^{(2)} \end{bmatrix} \tag{50}$$

where $x_n^{(j)}$ has the length r_j, $j = 1, 2$, with $r_1 + r_2 = r$. The filtering is described by the set of equations

$$x_{n+1}^{(1)} = f_n(x_n, w_n) \tag{51}$$
$$x_{n+1}^{(2)} = g_n(x_n, x_{n+1}^{(1)}) \tag{52}$$
$$z_n = h_n(x_n, v_n) \tag{53}$$

where f_n, g_n, and h_n are (in general) nonlinear functions. Again, the task is to calculate the information submatrix J_n for x_n. The partitioning restriction (51)–(53) of the problem is somewhat general and includes, among others, the case $x_{n+1}^{(2)} = x_n^{(2)}$, which means that the second part of the state vector is constant in time, and it can be considered for use when there are unknown constant parameters in the model. Note that in [3], the case was considered when g_n is only a function of $x_n^{(2)}$.

In this section, we present first an explicit solution—a recursive equation for J_n—for a special case of the system (51)–(53) with a linear function g_n and then a conceptual solution for general g_n.

Case 1—Linear g_n:

Proposition 2: Consider the linear filtering in (51)–(53), and assume that the function g_n is linear so that (52) can be written as

$$x_{n+1}^{(2)} = G_n^{(1)} x_n^{(1)} + G_n^{(2)} x_n^{(2)} + G_n^{(3)} x_{n+1}^{(1)}. \tag{54}$$

In addition, assume that $G_n^{(2)}$ is invertible for all n. Put

$$X_n^{(1)} = \begin{bmatrix} x_1^{(1)} \\ \vdots \\ x_n^{(1)} \end{bmatrix}. \tag{55}$$

Let $J(X_n^{(1)}, x_n^{(2)})$ be an information matrix derived from the joint probability density $p(X_n^{(1)}, x_n^{(2)}, Z_n)$, and let S_n and J_n be the information submatrices for $[x_{n-1}^{(1)}, x_n]$ and for x_n, respectively. Then, S_n and J_n obey the recursions

$$\begin{aligned} S_{n+1} \triangleq & \begin{bmatrix} S_{n+1}^{11} & S_{n+1}^{12} & S_{n+1}^{13} \\ S_{n+1}^{21} & S_{n+1}^{22} & S_{n+1}^{23} \\ S_{n+1}^{31} & S_{n+1}^{32} & S_{n+1}^{33} \end{bmatrix} \\ = & M_n^{-T} \begin{bmatrix} J_n^{11} + H_n^{11} & J_n^{12} + H_n^{12} & H_n^{13} \\ (J_n^{12} + H_n^{12})^T & J_n^{22} + H_n^{22} & H_n^{23} \\ (H_n^{13})^T & (H_n^{23})^T & H_n^{33} \end{bmatrix} M_n^{-1} \end{aligned} \tag{56}$$

$$J_{n+1} = \begin{bmatrix} S_{n+1}^{22} & S_{n+1}^{23} \\ S_{n+1}^{32} & S_{n+1}^{33} \end{bmatrix} - \begin{bmatrix} S_{n+1}^{21} \\ S_{n+1}^{31} \end{bmatrix} [S_{n+1}^{11}]^{-1} [S_{n+1}^{12} \quad S_{n+1}^{13}] \tag{57}$$

where

$$J_n = \begin{bmatrix} J_n^{11} & J_n^{12} \\ J_n^{21} & J_n^{22} \end{bmatrix} \qquad (r \times r) \tag{58}$$

$$M_n = \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & I \\ G_n^{(1)} & G_n^{(2)} & G_n^{(3)} \end{bmatrix} \qquad [(r + r_1) \times (r + r_1)] \tag{59}$$

$$H_n^{11} = \mathrm{E}\{-\Delta_{x_n^{(1)}}^{x_n^{(1)}} \log \overline{p}_n\} \qquad (r_1 \times r_1) \tag{60}$$

$$H_n^{12} = \mathrm{E}\{-\Delta_{x_n^{(1)}}^{x_n^{(2)}} \log \overline{p}_n\} \qquad (r_1 \times r_2) \tag{61}$$

$$H_n^{13} = \mathrm{E}\{-\Delta_{x_n^{(1)}}^{x_{n+1}^{(1)}} \log \overline{p}_n\} \qquad (r_1 \times r_1) \tag{62}$$

$$H_n^{22} = \mathrm{E}\{-\Delta_{x_n^{(2)}}^{x_n^{(2)}} \log \overline{p}_n\} \qquad (r_2 \times r_2) \tag{63}$$

$$H_n^{23} = \mathrm{E}\{-\Delta_{x_n^{(2)}}^{x_{n+1}^{(1)}} \log \overline{p}_n\} \qquad (r_2 \times r_1) \tag{64}$$

$$H_n^{33} = \mathrm{E}\{-\Delta_{x_{n+1}^{(1)}}^{x_{n+1}^{(1)}} \log \overline{p}_n\} \qquad (r_1 \times r_1) \tag{65}$$

and

$$\overline{p}_n = p[x_{n+1}^{(1)} | x_n] \cdot p[z_{n+1} | x_n, x_{n+1}^{(1)}] \tag{66}$$

provided that the above derivatives and expectations exist. In (59), the I's and 0's stand for identity and zero matrices of appropriate dimensions.

Proof: See the Appendix.

Note that the conditional probability function $p[z_{n+1}|x_n, x_{n+1}^{(1)}]$ in (66) is obtained from $p(z_{n+1}|x_{n+1})$ by substituting for $x_{n+1}^{(2)}$ from (54).

A stationary solution for J_n would be obtained by inserting J_∞ for J_n and J_{n+1}. Note that the resulting equation no longer has the form of a Riccati equation, unlike (48) in the previous section.

For example, consider the above-mentioned case when $x_n^{(2)}$ is a constant unknown parameter. Comparing the equation $x_{n+1}^{(2)} = x_n^{(2)}$ with (54), we have $G_n^{(1)} = 0$, $G_n^{(2)} = I$, and $G_n^{(3)} = 0$, where 0's and I's stand again for zero and identity matrices of appropriate dimensions. Utilizing the special form of the matrix M_n in (59), from (56) and (57), the recursions

$$J_{n+1}^{11} = H_n^{33} - (H_n^{13})^T [J_n^{11} + H_n^{11}]^{-1} H_n^{13} \tag{67}$$

$$\begin{aligned} J_{n+1}^{12} &= (J_{n+1}^{21})^T \\ &= (H_n^{23})^T - (H_n^{13})^T [J_n^{11} + H_n^{11}]^{-1} (J_n^{12} + H_n^{12}) \end{aligned} \tag{68}$$

$$\begin{aligned} J_{n+1}^{22} &= J_n^{22} + H_n^{22} - (J_n^{12} + H_n^{12})^T \\ &\quad \cdot [J_n^{11} + H_n^{11}]^{-1} (J_n^{12} + H_n^{12}). \end{aligned} \tag{69}$$

can be derived. Note that in the stationary case, where $H_n^{11}, \cdots, H_n^{33}$ do not depend on n, the matrix sequence $\{J_n^{11}\}$ converges for $n \to \infty$ to the solution of the Riccati-type of equation

$$J_\infty^{11} = H_n^{33} - (H_n^{13})^T [J_\infty^{11} + H_n^{11}]^{-1} H_n^{13}. \tag{70}$$

The sequence $\{J_n^{12}\}$ either converges to a constant matrix

$$\begin{aligned} J_\infty^{12} &= [I + (H_n^{13})^T (J_\infty^{11} + H_n^{11})^{-1}]^{-1} \\ &\quad \cdot [(H_n^{23})^T - (H_n^{13})^T (J_\infty^{11} + H_n^{11})^{-1} H_n^{12}] \end{aligned} \tag{71}$$

or diverges to infinity when at least one of the eigenvalues of $(H_n^{13})^T (J_\infty^{11} + H_n^{11})^{-1}$ has magnitude larger or equal to one.

The matrices J_n^{22} in (69) grow without any bound in general. If this happens, then the limit PCRB for estimating $x_n^{(1)}$ for $n \to \infty$ is the same as if $x_n^{(2)}$ were known. Indeed, these results can be expected because if the data bear any information about the parameter $x_n^{(2)}$, this information is accumulated as the time n goes to infinity.

Another example of application of Proposition 2 is given in Example 2 in the next section.

Case 2—Nonlinear g_n: The main idea for handling the singular case of the nonlinear filter in (51)–(53) is to "regularize" the filter, e.g., to replace (52) by a perturbed equation

$$x_{n+1}^{(2)} = g(x_n, x_{n+1}^{(1)}) + w_n^{(2)} \tag{72}$$

where $\{w_n^{(2)}\}$ is a sequence of pairwise independent *Gaussian* random vectors with zero mean and covariance matrix εI, independent of $\{w_n\}$ and $\{v_n\}$, with ε close to 0. For the modified system, it is possible to apply the result (21) from Section III.

Let $p_\varepsilon(\cdot)$'s and E_ε denote probability densities and the expectation operator induced by the perturbed system (51), (53), and (72). Note that

$$p_\varepsilon(x_{n+1}|x_n) = p(x_{n+1}^{(1)}|x_n) \cdot p_\varepsilon(x_{n+1}^{(2)}|x_n, x_{n+1}^{(1)}) \tag{73}$$

where $p(x_{n+1}^{(1)}|x_n)$ is determined by (51), and

$$\begin{aligned} &-\log p_\varepsilon(x_{n+1}^{(2)}|x_n, x_{n+1}^{(1)}) \\ &\quad = c_3 + \frac{1}{2\varepsilon} ||x_{n+1}^{(2)} - g[x_n, x_{n+1}^{(1)}]||^2 \end{aligned} \tag{74}$$

where c_3 is a constant. The matrices $D_{\varepsilon,n}^{11}, \cdots, D_{\varepsilon,n}^{22}$ for the regularized system can be written as

$$D_{\varepsilon,n}^{ij} = \overline{D}_{\varepsilon,n}^{ij} + \frac{1}{\varepsilon} K_{\varepsilon,n}^{ij} \qquad i, j = 1, 2 \tag{75}$$

where $\overline{D}_{\varepsilon,n}^{ij}$ is given as an E_ε-expectation of the second-order derivative of $-\log p(x_{n+1}^{(1)}|x_n)$ w.r.t. x_n and x_{n+1}, as in (22)–(25), $\overline{D}_{\varepsilon,n}^{22}$ contains, in addition, an E_ε-expectation of the second-order derivative of $-\log p(z_{n+1}|x_{n+1})$ w.r.t. x_{n+1}, and $K_{\varepsilon,n}^{ij}$, $i, j = 1, 2$ are given as an E_ε-expectation of the same derivatives of $\frac{1}{2}||x_{n+1}^{(2)} - g(x_n, x_{n+1}^{(1)})||^2$. In particular

$$K_{\varepsilon,n}^{11} = \mathrm{E}_\varepsilon\{[\nabla_{x_n} g^T][\nabla_{x_n} g^T]^T\} \tag{76}$$

$$K_{\varepsilon,n}^{12} = [\mathrm{E}_\varepsilon\{[\nabla_{x_n} g^T][\nabla_{x_{n+1}^{(1)}} g^T]^T\} \quad \mathrm{E}_\varepsilon\{\nabla_{x_n} g^T\}] \tag{77}$$

$$K_{\varepsilon,n}^{22} = \begin{bmatrix} \mathrm{E}_\varepsilon\{[\nabla_{x_{n+1}^{(1)}} g^T][\nabla_{x_{n+1}^{(1)}} g^T]^T\} & -\mathrm{E}_\varepsilon\{\nabla_{x_{n+1}^{(1)}} g^T\} \\ -\mathrm{E}_\varepsilon\{[\nabla_{x_{n+1}^{(1)}} g^T]^T\} & I \end{bmatrix} \tag{78}$$

where the arguments of g are omitted for brevity. The information submatrix for the original system will be obtained from the result (21) in the limit $\varepsilon \to 0$

$$J_{n+1} = \lim_{\varepsilon \to 0} \left[\overline{D}^{22}_{\varepsilon,n} + \frac{1}{\varepsilon} K^{22}_{\varepsilon,n} - \left(\overline{D}^{21}_{\varepsilon,n} + \frac{1}{\varepsilon} K^{21}_{\varepsilon,n} \right) \cdot \left(J_n + \overline{D}^{11}_{\varepsilon,n} + \frac{1}{\varepsilon} K^{11}_{\varepsilon,n} \right)^{-1} \left(\overline{D}^{12}_{\varepsilon,n} + \frac{1}{\varepsilon} K^{12}_{\varepsilon,n} \right) \right]. \tag{79}$$

An example of application of (79) is given in Example 3 in the following section.

V. Examples

Example 1—AR Process with Time-Varying Parameters: Consider a scalar-valued random process $\{z_n\}$ and introduce the notation

$$\underline{z}_n = [z_n, z_{n-1}, \cdots, z_{n-r+1}]^T. \tag{80}$$

Let z_n obey the recursion

$$z_{n+1} = x^T_{n+1} \underline{z}_n + v_n \tag{81}$$

where x_{n+1} is a vector of instantaneous autoregressive coefficients at time instant n, and $\{v_n\}$ is a Gaussian white noise with zero mean and variance σ^2. Further, assume that x_n has Gaussian random increments

$$x_{n+1} = x_n + w_n \tag{82}$$

where $\{w_n\}$ is white, independent of $\{v_n\}$, zero mean, and has covariance matrices $\{Q_n\}$.

The system (81) and (82) has the form of (46) and (47). The information submatrix J_n can be obtained by a straightforward application of (21) and (34)–(37). The result is

$$D^{11}_n = -D^{12}_n = Q^{-1}_n \tag{83}$$

$$D^{22}_n = Q^{-1}_n + Z_n \tag{84}$$

where

$$Z_n = \sigma^{-2} \operatorname{E}\{\underline{z}_n \underline{z}^T_n\} \tag{85}$$

so that

$$J_{n+1} = Q^{-1}_n + Z_n - Q^{-1}_n [J_n + Q^{-1}_n]^{-1} Q^{-1}_n. \tag{86}$$

Note that the optimum estimate of x_n from the data Z_n in the mean-of-square sense is the Kalman filter; the conditional distribution of x_n given Z_n is Gaussian. Let $\hat{x}_{n|n}$ and $\Sigma_{n|n}$ denote parameters of this distribution, namely, the mean and the covariance matrix. As mentioned in the introduction, the PCRB is tight in this case, and J_n is equal to the expected value of $\Sigma^{-1}_{n|n}$. Note that $\Sigma^{-1}_{n|n}$ in the Kalman filter obeys the same recursion as J_n with the exception that Z_n in (85) is replaced by $\sigma^{-2}\, \underline{z}_n \underline{z}^T_n$ without the expectation operator.

In order to achieve practical conclusions from the above theory, assume that drift of the autoregressive parameter is slow, i.e., that the trace of Q_n is much lower than 1, and that x_n fluctuates around a mean value $\overline{x}$ for a considerably long period of time. Then, the covariance function of $\{z_n\}$ is approximately equal to the covariance function of an AR process with parameter $\overline{x}$. The matrix Z_n in (85) can be replaced by a covariance matrix $\overline{Z}$ of the above process, which is a function of $\overline{x}$. Note that $\overline{Z}$ is independent of the variance of innovations σ^2. Some methods for calculating the covariance matrix of an AR process are presented, e.g., in [14]. For example, for the first-order autoregressive process [abbreviated as AR(1) in the sequel]

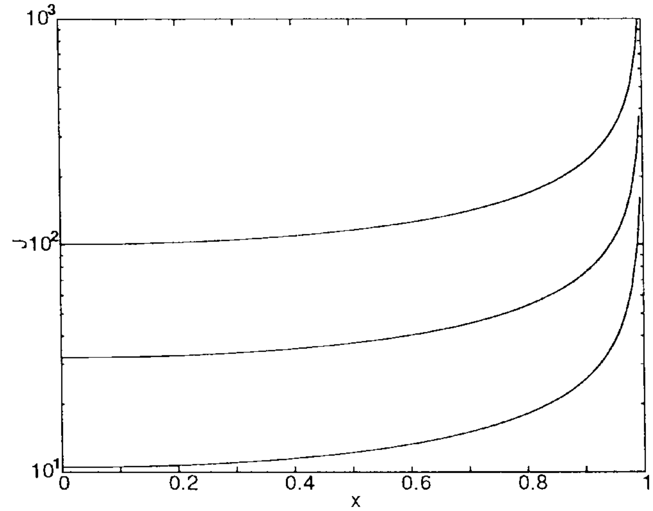

Fig. 1. Fisher information for slowly varying parameter of an AR(1) process as a function of this parameter for $Q = 10^{-2}$, 10^{-3}, and 10^{-4} (from the bottom up), respectively.

$$\overline{Z} = \frac{1}{1 - \overline{x}^2} \tag{87}$$

holds. Here, $\overline{x}$ is restricted to the interval $(-1, 1)$ to assure stability of the model. If, in addition, the matrix sequence $\{Q_n\}$ is constant, $Q_n \equiv Q$, and it is possible to calculate the limit information matrix (which is a scalar, in the case of $r = 1$) from the equation

$$J_\infty = Q^{-1} + \overline{Z} - Q^{-1}[J_\infty + Q^{-1}]^{-1} Q^{-1}. \tag{88}$$

In particular, for the AR(1) process, we obtain the solution

$$J_\infty = \frac{1}{2(1 - \overline{x}^2)} \left[1 + \sqrt{1 + 4Q^{-1}(1 - \overline{x}^2)} \right]. \tag{89}$$

Numerical values of (89) for $Q = 10^{-2}$, 10^{-3}, and 10^{-4} are plotted in Fig. 1. It is shown that the information about the parameter increases rapidly if the pole approaches unity. For the pole well separated from ± 1, i.e., $\overline{x} \approx 0$, it holds that $J_\infty \approx Q^{-1/2}$.

The matrix J_∞ in (88) [or the corresponding scalar in (89) in the special case] describes the information content that the AR process bears about the fluctuating AR parameter. This information content depends on the actual value of the estimated parameter. If it happens that J_∞ is small and, consequently, that the limit PCRB J^{-1}_∞ is large, it indicates that the assumed data model might not be appropriate.

Example 2—Sinusoidal Frequency Estimation: In this subsection, the developed methodology is applied to computation of the posterior CRB for tracking parameters of a single noisy cisoid with slowly varying frequency. This computation is easier than those recently presented in [16]. Second, as a special case of a single time-invariant frequency, the well known Cramér–Rao bound by Rife and Boorstyn [9] is derived.

The signal is assumed to have the form

$$z_n = m_0\, e^{i\varphi_n} + v_n \qquad n = 0, 1, 2, \cdots \tag{90}$$

where

- m_0 magnitude;
- φ_n instantaneous phase of cisoid at time instant n;
- $\{v_n\}$ noise.

The instantaneous frequency (denoted ω_n) is defined as the one-step increment of φ_n. Thus, the signal with randomly varying frequency can be described by the state vector

$$x_n = (\omega_n, \varphi_n)^T \tag{91}$$

and time update of x_n is given by the pair of the equations

$$\omega_{n+1} = \omega_n + e_n \tag{92}$$

$$\varphi_{n+1} = \varphi_n + \omega_{n+1} = \varphi_n + \omega_n + e_n. \tag{93}$$

It is assumed that $\{v_n\}$ and $\{e_n\}$ are independent sequences of independent random variables with zero mean values and variances σ^2 and γ^2, respectively; $\{e_n\}$ is Gaussian, and $\{v_n\}$ is complex circular Gaussian (i.e., the real and imaginary parts of $\{v_n\}$ are independent normally distributed with zero means and equal variances $\sigma^2/2$). Next, assume that the probability distribution of the initial instantaneous phase and frequency is known.

Obviously, in the standard formulation, the covariance matrix of the system noise $w_n = (e_n, e_n)$ is not invertible, and the conditional probability $p(x_{n+1}|x_n)$ is singular. The calculation of the information submatrix as in Section II fails, but it is possible to apply the approach developed in Section III with $x_n^{(1)} = \omega_n$ and $x_n^{(2)} = \varphi_n$. Comparing (93) with (54), we get $G_n^{(1)} = 0$, and $G_n^{(2)} = G_n^{(3)} = 1$. The assumed probability distributions of the noise $\{e_n\}$ and $\{v_n\}$ imply that

$$\begin{aligned} -\log p(x_{n+1}^{(1)}|x_n) &= -\log p(\omega_{n+1}|\omega_n) \\ &= c_4 + \frac{1}{2\gamma^2}(\omega_n - \omega_{n+1})^2 \end{aligned} \tag{94}$$

$$\begin{aligned} -\log p(z_{n+1}|x_n, x_{n+1}^{(1)}) &= -\log p(z_{n+1}|\varphi_n, \omega_{n+1}) \\ &= c_5 + \frac{1}{\sigma^2}\left|m_0 e^{i(\varphi_n+\omega_{n+1})} - z_{n+1}\right|^2 \end{aligned} \tag{95}$$

where c_4 and c_5 are normalization constants. A straightforward calculation of (60)–(65) gives

$$H_n^{11} = \frac{1}{\gamma^2},\ H_n^{12} = 0,\ H_n^{13} = -\frac{1}{\gamma^2} \tag{96}$$

$$H_n^{22} = H_n^{23} = \frac{2m_0^2}{\sigma^2},\ H_n^{33} = \frac{2m_0^2}{\sigma^2} + \frac{1}{\gamma^2}. \tag{97}$$

Inserting the above relations into (56) and (57) and we get, after some simplifications, (98), shown at the bottom of the page, where

$$d_n = \det J_n = J_n^{\omega\omega} J_n^{\varphi\varphi} - [J_n^{\omega\varphi}]^2. \tag{99}$$

In (98) and (99), $J_n^{\omega\omega}$, $J_n^{\omega\varphi}$, and $J_n^{\varphi\varphi}$ denote elements of the matrix J_n.

The stationary solution of (98) can be found by putting $J_n = J_{n+1} = J_\infty$. After excluding the terms $J_\infty^{\omega\varphi}$ and $J_\infty^{\varphi\varphi}$, a fourth-order polynomial equation for $J_\infty^{\omega\omega}$ is obtained. This equation can be shown to have only one positive real-valued root. The final result is

$$J_\infty = \frac{1}{2\gamma^2}\begin{bmatrix} h & -\dfrac{h^2}{4+h} \\ -\dfrac{h^2}{4+h} & \dfrac{2h^3}{(4+h)^2} \end{bmatrix} \tag{100}$$

where

$$h = w + \sqrt{w^2 + 4w} \tag{101}$$

$$w = \Gamma + \sqrt{\Gamma^2 + 8\Gamma} \tag{102}$$

$$\Gamma = \gamma^2\,\frac{m_0^2}{\sigma^2}. \tag{103}$$

The limit PCRB on the instantaneous frequency is equal to the left-upper corner element of J_∞^{-1}, i.e.,

$$\begin{aligned} \mathrm{LPCRB}(\hat{\omega}_n) &= [J_\infty^{\omega\omega} - (J_\infty^{\omega\varphi})^2/J_\infty^{\varphi\varphi}]^{-1} = \gamma^2\,\frac{4}{h} \\ &= \frac{\gamma^2}{w}\left(-w + \sqrt{w^2 + 4w}\right) \end{aligned} \tag{104}$$

which coincides with the result derived in [16].

Finally, let us consider estimation of stationary frequency, i.e., put $\gamma^2 = 0$. Then, (98) is reduced to

$$\begin{aligned} J_{n+1} &= \begin{bmatrix} J_{n+1}^{\omega\omega} & J_{n+1}^{\omega\varphi} \\ J_{n+1}^{\omega\varphi} & J_{n+1}^{\varphi\varphi} \end{bmatrix} \\ &= \begin{bmatrix} J_n^{\omega\omega} - 2J_n^{\omega\varphi} + J_n^{\varphi\varphi} & J_n^{\omega\varphi} - J_n^{\varphi\varphi} \\ J_n^{\omega\varphi} - J_n^{\varphi\varphi} & J_n^{\varphi\varphi} + \dfrac{2m_0^2}{\sigma^2} \end{bmatrix}. \end{aligned} \tag{105}$$

For $J_0 = 0$ (no *a priori* information about the frequency and phase), the recursion (105) has a solution

$$J_n^{\varphi\varphi} = n\,\frac{2m_0^2}{\sigma^2} \tag{106}$$

$$J_n^{\omega\varphi} = -\sum_{k=1}^{n-1} J_k^{\varphi\varphi} = -n(n-1)\frac{m_0^2}{\sigma^2} \tag{107}$$

$$J_n^{\omega\omega} = \sum_{k=1}^{n-1}(J_k^{\varphi\varphi} - 2J_k^{\omega\varphi}) = n(n-1)(2n-1)\,\frac{m_0^2}{3\sigma^2}. \tag{108}$$

$$J_{n+1} = \frac{1}{1+\gamma^2 J_n^{\omega\omega}} \cdot \begin{bmatrix} \gamma^2 d_n + J_n^{\omega\omega} - 2J_n^{\omega\varphi} + J_n^{\varphi\varphi} & -\gamma^2 d_n + J_n^{\omega\varphi} - J_n^{\varphi\varphi} \\ -\gamma^2 d_n + J_n^{\omega\varphi} - J_n^{\varphi\varphi} & \gamma^2 d_n + J_n^{\varphi\varphi} + \dfrac{2m_0^2}{\sigma^2}(1+\gamma^2 J_n^{\omega\omega}) \end{bmatrix} \tag{98}$$

The PCRB on the frequency is equal to the left-upper corner element of J_n^{-1}, i.e.,

$$\mathrm{PCRB}(\hat{\omega}_n) = [J_n^{\omega\omega} - (J_n^{\omega\varphi})^2/J_n^{\varphi\varphi}]^{-1} = \frac{6}{n(n^2-1)}\,\frac{\sigma^2}{m_0^2} \tag{109}$$

which coincides with the CRB for the problem [9].

Example 3—Sinusoidal Signal with Sinusoidal Phase Modulation: Consider a sinusoidal signal as in (90), define the instantaneous frequency ω_n of the carrier as a one-step backward difference of the instantaneous phase φ_n as in (93), and assume that the frequency evolves in time like a sinusoid within the range $(-\pi, \pi)$. We refer to this sinusoid as a message and assume that the frequency of the message evolves like a random walk. Note that an algorithm for tracking parameters of signals of this kind was proposed in [17].

At each time instant, the signal can be characterized by a state vector with three components

$$x_n = [\nu_n,\, \phi_n,\, \varphi_n]^T \tag{110}$$

where

- φ_n instantaneous phase of the carrier;
- ϕ_n instantaneous phase of the message;
- ν_n frequency of the message.

Assume that the instantaneous frequency of the carrier equals

$$\omega_n = \omega_c + \eta\cos\phi_n \tag{111}$$

where ω_c is the central frequency of the carrier, and η is the maximum deviation of the carrier frequency from ω_c.

The time update of the state vector is given by the set of equations

$$\nu_{n+1} = \nu_n + e_n \tag{112}$$
$$\phi_{n+1} = \phi_n + \nu_{n+1} \tag{113}$$
$$\varphi_{n+1} = \varphi_n + \omega_c + \eta\cos\phi_{n+1} = \varphi_n + \omega_c + \eta\cos(\phi_n + \nu_{n+1}). \tag{114}$$

As in the previous subsection, assume that $\{e_n\}$ is a Gaussian white noise with variance γ^2. The filtering in (112)–(114) and (90) is an example of the singular case from Section IV with nonlinear function g_n and

$$x_n^{(1)} = \nu_n \tag{115}$$
$$x_n^{(2)} = [\phi_n,\, \varphi_n]^T \tag{116}$$
$$-\log p(x_{n+1}^{(1)}|x_n) = c_6 + \frac{1}{2\gamma^2}(\nu_{n+1}-\nu_n)^2 \tag{117}$$

and

$$\tfrac{1}{2}||x_{n+1}^{(2)} - g(x_n, x_{n+1}^{(1)})||^2 = \tfrac{1}{2}(\phi_{n+1} - \phi_n - \nu_{n+1})^2 + \tfrac{1}{2}[\varphi_{n+1} - \varphi_n - \omega_c - \eta\cos(\phi_n + \nu_{n+1})]^2. \tag{118}$$

A straightforward calculation gives

$$\overline{D}_{\varepsilon,n}^{11} = -\overline{D}_{\varepsilon,n}^{12} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{119}$$

$$\overline{D}_{\varepsilon,n}^{22} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{2m_0^2}{\sigma^2} \end{bmatrix} \tag{120}$$

$$K_{\varepsilon,n}^{11} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & a_{\varepsilon,n} & b_{\varepsilon,n} \\ 0 & b_{\varepsilon,n} & 1 \end{bmatrix} \tag{121}$$

$$K_{\varepsilon,n}^{12} = \begin{bmatrix} 0 & 0 & 0 \\ a_{\varepsilon,n} & -1 & -b_{\varepsilon,n} \\ b_{\varepsilon,n} & 0 & -1 \end{bmatrix} \tag{122}$$

$$K_{\varepsilon,n}^{22} = \begin{bmatrix} a_{\varepsilon,n} & -1 & -b_{\varepsilon,n} \\ -1 & 1 & 0 \\ -b_{\varepsilon,n} & 0 & 1 \end{bmatrix} \tag{123}$$

where

$$a_{\varepsilon,n} = 1 + \eta^2 \mathrm{E}_\varepsilon\{\sin^2(\phi_n + \nu_{n+1})\} \tag{124}$$
$$b_{\varepsilon,n} = -\eta\mathrm{E}_\varepsilon\{\sin(\phi_n + \nu_{n+1})\}. \tag{125}$$

An available but tedious method of computing an approximate value of J_n is to choose a small fixed ε, do a number of independent simulations of the data according to the "regularized" model, and replace the expectations in (124) and (125) by corresponding sample averages. Then, evaluate J_n as in (79).

Another approach for computing J_n can be utilized in cases when the rate of evolution of ν_n, i.e., the variance γ^2, and the variance of the observation noise are small. Consider sequences $\{\overline{\nu}_n\}$, $\{\overline{\phi}_n\}$, $\{\overline{\varphi}_n\}$ that obey (112)–(114) with $\{e_n \equiv 0\}$ (this is called an "equilibrium state" in [15]), and assume that the probability densities of $\{\nu_n\}$, $\{\phi_n\}$, $\{\varphi_n\}$ are concentrated in neighborhoods of $\{\overline{\nu}_n\}$, $\{\overline{\phi}_n\}$, and $\{\overline{\varphi}_n\}$. Then, $a_{\varepsilon,n}$ and $b_{\varepsilon,n}$ are approximated by

$$a_{\varepsilon,n} \approx \overline{a}_n \triangleq 1 + \eta^2\sin^2(\overline{\phi}_n + \overline{\nu}_{n+1}) \tag{126}$$
$$b_{\varepsilon,n} \approx \overline{b}_n \triangleq -\eta\sin(\overline{\phi}_n + \overline{\nu}_{n+1}). \tag{127}$$

Using the above approximation the limit in (79) can be evaluated analytically. The result, which is obtained with the aid of symbolic Mathematica, is

$$J_{n+1}^{\nu\nu} = (J_n^{\nu\nu} + J_n^{\phi\phi} - 2J_n^{\nu\phi} + \gamma^2 d_1)/(1+\gamma^2 J_n^{\nu\nu}) \tag{128}$$
$$J_{n+1}^{\nu\phi} = [J_n^{\nu\phi} - J_n^{\phi\phi} - \gamma^2 d_1 + \gamma^2 d_2 \overline{b}_n + (J_n^{\phi\varphi} - J_n^{\nu\varphi})\overline{b}_n] /(1+\gamma^2 J_n^{\nu\nu}) \tag{129}$$
$$J_{n+1}^{\nu\varphi} = (J_n^{\nu\varphi} - J_n^{\phi\varphi} - \gamma^2 d_2)/(1+\gamma^2 J_n^{\nu\nu}) \tag{130}$$
$$J_{n+1}^{\phi\phi} = [J_n^{\phi\phi} + \overline{b}_n^2 J_n^{\varphi\varphi} - 2\overline{b}_n J_n^{\phi\varphi} + \gamma^2(d_1 - 2\overline{b}_n d_2 + \overline{b}_n^2 d_3)] /(1+\gamma^2 J_n^{\nu\nu}) \tag{131}$$
$$J_{n+1}^{\phi\varphi} = [J_n^{\phi\varphi} - \overline{b}_n J_n^{\varphi\varphi} + \gamma^2(d_2 - \overline{b}_n d_3)] /(1+\gamma^2 J_n^{\nu\nu}) \tag{132}$$
$$J_{n+1}^{\varphi\varphi} = (J_n^{\varphi\varphi} + \gamma^2 d_3)/(1+\gamma^2 J_n^{\nu\nu}) + \frac{2m_0^2}{\sigma^2}. \tag{133}$$

where

$$d_1 = J_n^{\nu\nu} J_n^{\phi\phi} - [J_n^{\nu\phi}]^2 \tag{134}$$
$$d_2 = J_n^{\nu\nu} J_n^{\phi\varphi} - J_n^{\nu\phi} J_n^{\nu\varphi} \tag{135}$$
$$d_3 = J_n^{\nu\nu} J_n^{\varphi\varphi} - [J_n^{\nu\varphi}]^2 \tag{136}$$

and $J_n^{\nu\nu}, \cdots, J_n^{\varphi\varphi}$ are the elements of J_n.

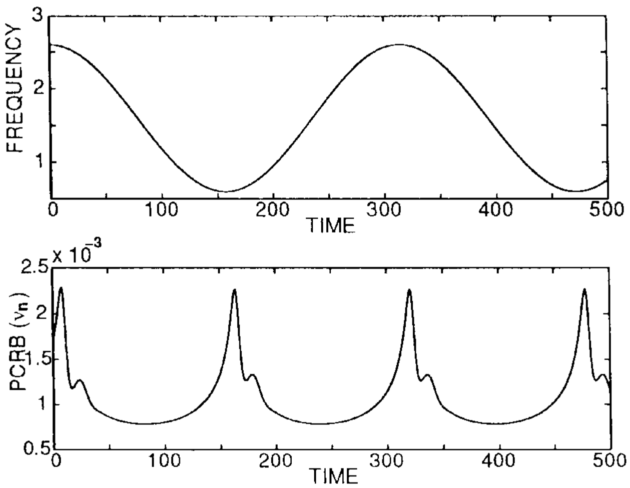

Fig. 2. Instantaneous frequency of the carrier of a sinusoidal signal and the PCRB on the signal frequency as functions of time in the model considered in Example 3.

To illustrate the above result, consider a signal of the length $N = 500$ with the following parameters: $m_0^2/\sigma^2 = 1/2$, $\overline{\nu} = 0.02$, $\omega_c = 1.6$, $\eta = 1$, $\gamma^2 = 10^{-4}$, $\phi_0 = 0$. Fig. 2 shows the posterior CRB on parameter ν_n, which was derived from J_n, as a function of time. Simultaneously, the instantaneous frequency of the carrier is plotted. Note that the nonlinear character of the signal model implies that the PCRB does not converge to any limit value for $n \to \infty$, but it is periodic in time with the frequency that is twice greater than the frequency of the message ν_n. In particular, if the frequency of the carrier is close to its minimum or maximum and its rate of change is low, the amount of information that the signal bears about the possible changes of ν_n is small, the PCRB increases, and vice versa.

VI. Conclusions

A simple and straightforward derivation of the posterior Cramér–Rao lower bound for the discrete-time nonlinear filtering problem was presented. Explicit realizations of this lower bound were calculated for three important examples.

1) tracking a slowly varying AR parameter;
2) tracking a slowly varying sinusoidal frequency;
3) tracking a slowly varying frequency that is modulated by a sinusoid.

The derived lower bound can be used for evaluating the performance of existing suboptimal methods of nonlinear filtering. It is believed that a similar bound can be derived for a more general model of nonlinear autoregressive systems as well.

Appendix
Proof of Proposition 2

The proof of Proposition 2 utilizes the following lemma.

Lemma 1: Consider the problem of estimating a random vector x from an observation vector z. Let $p(x, z)$ be the joint probability density of (x, z), and assume that information matrix

$$J(x) = \mathrm{E}\{-\Delta_x^x \log p(x, z)\}. \quad (137)$$

exists. Let $y = Mx$, where M is a constant invertible matrix. Then, the probability density $p(y, z)$ exists, and the corresponding information matrix for estimating y is given by

$$J(y) = M^{-T} J(x) M^{-1}. \quad (138)$$

Proof: The proof is based on the well-known rule for change of coordinates of the estimated parameters (see, e.g., [13]), is straightforward, and is therefore omitted here. ■

Proof of Proposition: Let p_n denote the probability density of the triplet $[X_n^{(1)}, x_n^{(2)}, Z_n]$

$$p_n \triangleq p(X_n^{(1)}, x_n^{(2)}, Z_n) = p(X_{n-1}^{(1)}, x_n^{(1)}, x_n^{(2)}, Z_n). \quad (139)$$

It will be shown by induction that p_n exists.

The information matrix that corresponds to the triplet $[X_{n-1}^{(1)}, x_n^{(1)}, x_n^{(2)}]$ can be written in block form as

$$J(X_{n-1}^{(1)}, x_n^{(1)}, x_n^{(2)}) = \begin{bmatrix} A_n & B_n & C_n \\ B_n^T & D_n & E_n \\ C_n^T & E_n^T & F_n \end{bmatrix} \quad (140)$$

where the blocks $A_n, B_n, \cdots, F_n$ are obtained as expectations of the second-order derivatives of $-\log p_n$ with respect to $X_{n-1}^{(1)}, x_n^{(1)}$, and $x_n^{(2)}$.

The information submatrix for the state vector x_n can be obtained as the inverse of the right-lower submatrix of $\{J(X_{n-1}^{(1)}, x_n^{(1)}, x_n^{(2)})\}^{-1}$, i.e.,

$$\begin{aligned} J_n &\triangleq \begin{bmatrix} J_n^{11} & J_n^{12} \\ J_n^{21} & J_n^{22} \end{bmatrix} \\ &= \begin{bmatrix} D_n & E_n \\ E_n^T & F_n \end{bmatrix} - \begin{bmatrix} B_n^T \\ C_n^T \end{bmatrix} A_n^{-1} [B_n, C_n] \\ &= \begin{bmatrix} D_n - B_n^T A_n^{-1} B_n & E_n - B_n^T A_n^{-1} C_n \\ E_n^T - C_n^T A_n^{-1} B_n & F_n - C_n^T A_n^{-1} C_n \end{bmatrix}. \end{aligned} \quad (141)$$

Consider the probability density of the quartet $[X_n^{(1)}, x_n^{(2)}, x_{n+1}^{(1)}, Z_{n+1}]$, denoted by $\tilde{p}_{n+1}$. Note that two vectors

$$\begin{aligned} \begin{bmatrix} X_{n-1}^{(1)} \\ x_n^{(1)} \\ x_{n+1}^{(1)} \\ x_{n+1}^{(2)} \end{bmatrix} &= \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & I \\ 0 & G_n^{(1)} & G_n^{(2)} & G_n^{(3)} \end{bmatrix} \begin{bmatrix} X_{n-1}^{(1)} \\ x_n^{(1)} \\ x_n^{(2)} \\ x_{n+1}^{(1)} \end{bmatrix} \\ &\triangleq \tilde{M}_n \begin{bmatrix} X_{n-1}^{(1)} \\ x_n^{(1)} \\ x_n^{(2)} \\ x_{n+1}^{(1)} \end{bmatrix} \end{aligned} \quad (142)$$

obey the linear relationship. Since $G_n^{(2)}$ is assumed to be regular, it follows that $\tilde{M}_n$ is regular as well. Applying Lemma 1, it follows by induction that p_n in (139) exists for each n, and

$$J(X_n^{(1)}, x_{n+1}^{(1)}, x_{n+1}^{(2)}) = \tilde{M}_n^{-T} J(X_n^{(1)}, x_n^{(2)}, x_{n+1}^{(1)}) \tilde{M}_n^{-1}. \quad (143)$$

Using conditional densities, $\tilde{p}_{n+1}$ can be written as the product

$$\begin{aligned}\tilde{p}_{n+1} &= p_n\, p(x_{n+1}^{(1)}|X_n^{(1)}, x_n^{(2)}, Z_n)\\ &\quad \cdot p(z_{n+1}|X_n^{(1)}, x_n^{(2)}, x_{n+1}^{(1)}, Z_n)\\ &= p_n\, p(x_{n+1}^{(1)}|x_n]\, p[z_{n+1}|x_n, x_{n+1}^{(1)}). \end{aligned} \quad (144)$$

The second equality in (144) follows from the formulation of the filtering problem. From (140) and (144), it follows that

$$J(X_n^{(1)}, x_n^{(2)}, x_{n+1}^{(1)}) = \begin{bmatrix} A_n & B_n & C_n & 0 \\ B_n^T & D_n + H_n^{11} & E_n + H_n^{12} & H_n^{13} \\ C_n^T & (E_n + H_n^{12})^T & F_n + H_n^{22} & H_n^{23} \\ 0 & (H_n^{13})^T & (H_n^{23})^T & H_n^{33} \end{bmatrix} \quad (145)$$

where $H_n^{ij},\ i, j = 1, 2, 3$ were defined in (60)–(65). The information submatrix for $[x_n^{(1)}, x_n^{(2)}, x_{n+1}^{(1)}]$ then equals

$$\tilde{S}_{n+1} \triangleq \begin{bmatrix} D_n + H_n^{11} & E_n + H_n^{12} & H_n^{13} \\ (E_n + H_n^{12})^T & F_n + H_n^{22} & H_n^{23} \\ (H_n^{13})^T & (H_n^{23})^T & H_n^{33} \end{bmatrix} - \begin{bmatrix} B_n^T \\ C_n^T \\ 0 \end{bmatrix} A_n^{-1}[B_n, C_n, 0]. \quad (146)$$

This can be rewritten using (141) as

$$\tilde{S}_{n+1} = \begin{bmatrix} J_n^{11} + H_n^{11} & J_n^{12} + H_n^{12} & H_n^{13} \\ (J_n^{12} + H_n^{12})^T & J_n^{22} + H_n^{22} & H_n^{23} \\ (H_n^{13})^T & (H_n^{23})^T & H_n^{33} \end{bmatrix}. \quad (147)$$

Combining (143) and the form of $\tilde{M}_n$ in (59) and (142) implies

$$S_{n+1} = M_n^{-T} \tilde{S}_{n+1} M_n^{-1}. \quad (148)$$

The statement follows. ■

REFERENCES

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering.* Englewood Cliffs, NJ: Prentice-Hall, 1979.

[2] B. Z. Bobrovsky and M. Zakai, "A lower bound on the estimation error for Markov processes," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 785–788, Dec. 1975.

[3] J. I. Galdos, "A Cramér–Rao bound for multidimensional discrete-time dynamical systems," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 117–119, Feb. 1980.

[4] P. C. Doerschuk, "Cramer–Rao bounds for discrete-time nonlinear filtering problems," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 1465–1469, Aug. 1995.

[5] G. A. Hewer, "An iterative technique for the computation of the steady state gains for the discrete optimal regulator," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 382–384, Feb. 1971.

[6] T. H. Kerr, "Status of CR-like lower bounds for nonlinear filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, pp. 590–600, Sept. 1989.

[7] V. Kučera, "Factorization of rational spectral matrices: A survey of methods," in *Proc. IEE Int. Conf. Contr.*, Edinburgh, U.K., 1991, pp. 1074–1078.

[8] T. Pappas, A. J. Laub, and N. R. Sandel, "On the numerical solution of the discrete-time algebraic Riccati equation," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 631–641, June 1980.

[9] D. C. Rife and R. R. Boorstyn, "Multiple tone parameter estimation from discrete time observation," *Bell Syst. Tech. J.*, pp. 1389–1410, Nov. 1976.

[10] D. J. Sakrison *Notes on Analog Communication.* Princeton, NJ: Van Nostrand, 1970.

[11] H. L. van Trees, *Detection, Estimation and Modulation Theory.* New York: Wiley, 1968.

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

[13] B. Porat, *Digital Processing of Random Signals.* Englewood Cliffs, NJ: Prentice-Hall, 1993.

[14] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods.* New York: Springer-Verlag, 1987.

[15] P. Tichavský and P. Händel, "Two algorithms for adaptive retrieval of slowly varying multiple cisoids in noise," *IEEE Trans. Signal Processing*, vol. 43, pp. 1116–1127, May 1995.

[16] P. Tichavský, "Posterior Cramér–Rao bounds for adaptive harmonic retrieval," *IEEE Trans. Signal Processing*, vol. 43, pp. 1299–1302, May 1995.

[17] P. Tichavský and P. Händel, "Recursive estimation of linearly or harmonically modulated frequencies of multiple cisoids in noise," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, München, Germany, Apr. 1997, pt. III, pp. 1925–1928.

[18] P. Tichavský, C. Muravchik, and A. Nehorai, "Posterior C–R bounds for performance of adaptive parameter estimation," in *Signal Anal. Prediction I, Proc. First Euro. Conf. Signal Anal. Prediction*, A. Procházka, J. Uhlíř and P. Sovka, Eds. Prague, Czech Republic, June 24–27, 1997, pp. 153–156.

Petr Tichavský (M'98) graduated in 1987 from the Czech Technical University, Prague, Czechoslovakia. He received the Ph.D. degree in theoretical cybernetics from the Czechoslovak Academy of Sciences, Prague, in 1992.

He is now with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague. In 1994, he received the Fulbright grant for a ten-month fellowship at the Department of Electrical Engineering, Yale University, New Haven, CT. His current research interests include adaptive parameter estimation and spectrum analysis.

Carlos H. Muravchik (S'81–M'83) graduated in 1973 as an electronics engineer from the National University of La Plata, La Plata, Argentina. He received the M.Sc. degree in statistics in 1983 and the M.Sc. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1980 and 1983, respectively.

He is currently a Professor with the Department of Electrical Engineering, National University of La Plata, and a member of its Industrial Electronics, Control, and Instrumentation Laboratory. He is also a member of the Comisión de Investigaciones Científicas de la Pacifica, Buenos Aires, Argentina. He was a Visiting Professor at Yale University, New Haven, CT, in 1983 and 1994 and at the University of Illinois, Chicago, in 1996. His research interests are in the areas of statistical signal and array processing with biomedical, control, and communications applications and in nonlinear control systems.

Arye Nehorai (S'80–M'83–SM'90–F'94) received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, in 1976 and 1979, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1983.

After graduation, he was a Research Engineer at Systems Control Technology, Inc., Palo Alto, CA. From 1985 to 1995, he was with the Department of Electrical Engineering, Yale University, New Haven, CT, where he became an Associate Professor in 1989. In 1995, he joined the Department of Electrical Engineering and Computer Science, The University of Illinois, Chicago, as a Full Professor.

Dr. Nehorai has been an Associate Editor of the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING and IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION. He has also been an Associate Editor of *Circuits, Systems, and Signal Processing,* and *The Journal of the Franklin Institute.* He served as the Chairman of the Connecticut IEEE Signal Processing Chapter from 1986 to 1995 and was co-recipient of the 1989 IEEE Signal Processing Society's Senior Award for the Best Paper in Statistical Signal and Array Processing. He is a Fellow of the Royal Statistical Society.