Performance of Bayesian Parameter Estimators for Linear Signal Models

R. M. HAWKES, STUDENT MEMBER, IEEE, AND J. B. MOORE, MEMBER, IEEE

***Abstract*—In this short paper the Bayesian estimation of parameters of discrete time, linear, finite-dimensional stochastic systems is discussed. Upper bounds for the estimator mean-square error are obtained under the assumption of a finite parameter set. Necessary and sufficient conditions are established for exponential convergence of the Bayesian estimate to the true parameter values in the mean-square error sense for systems with measurements which are stationary Gaussian random processes. The conditions for convergence are given in terms of a finite set of signal model Markov parameters. The performance results for parameter estimation are shown to yield bounds on the performance of the nonlinear state estimators for the class of signal models under discussion.**

Manuscript received August 18, 1975; revised February 17, 1976. Paper recommended by D. G. Lainiotis, Past Chairman of the IEEE S-CS Adaptive and Learning Systems, Pattern Recognition Committee. This work was supported by the Australian Research Grants Committee and the Radio Research Board.

The authors are with the Department of Electrical Engineering, University of Newcastle, N.S.W., Australia.

I. Introduction

Bayesian conditional-mean parameter estimators are useful for the estimation of unknown system or noise parameters of linear finite-dimensional dynamic signal models [1]. Exact performance calculation for this type of estimator, however, appears to be difficult—a common feature of nonlinear estimators [2]. In this short paper we derive an upper bound for the estimator mean-square error. This bound is then manipulated to reveal conditions which guarantee an exponential convergence rate of the estimator.

Quite a lot of work has been done on the convergence properties of maximum-likelihood estimators (MLE). For example, Wald [3] proved the strong consistency of the MLE in the case of independent identically distributed random variables. Recently, Caines and Rissanen [4] have proved the strong consistency of the MLE of the parameters of Gaussian random processes possessing linear autoregressive moving average or state-space representations.

In [5] Liporace discusses the performance of Bayes' conditional-mean estimators on a finite parameter set using the assumption of independently and identically distributed measurement records. Specifically, the mean-square error in the Bayes' estimate is found to be bounded by a quantity that diminishes exponentially as the number of observations increases. Corresponding results are obtained when the true parameter is not a member of the parameter set.

Lainiotis *et al.* [1], [10]–[13] have also investigated the performance of the class of estimator studied in this short paper. In [1] and [10] expressions for the mean-square error of adaptive state estimators are given in a form convenient for on-line implementation. Evaluating these expressions is tantamount to simulating the system under study. In [11]–[13] probability of error bounds for certain classes of multihypothesis detection problems are given. We also note that Tse [14] is currently investigating bounds for identification error.

In this short paper we study the mean-square convergence of Bayesian estimates of unknown system and noise parameters of linear, stochastic discrete time signal models where the parameter space is assumed to be finite. For systems with stationary output sequences it is shown that certain identifiability conditions guarantee that the Bayesian parameter estimator will converge exponentially to the true parameter value. The identifiability conditions are characterized by a finite set of Markov parameters. Signal models with the same output statistics (or equivalently Markov parameters) are not identifiable. Adaptive state estimation algorithms given in [1] can be analyzed by a straightforward extension to the results for parameter estimation. Under certain stationarity conditions on the signal models the mean-square error for adaptive estimation tends at an exponential rate to the optimal mean-square error for linear filtering given complete knowledge of the signal model.

The significance of the convergence results obtained in this short paper is that for the first time, we believe, exponential convergence in the mean-square error sense is established for nonindependently and identically distributed data sequences.

II. Problem Formulation

Signal model: Suppose that a sequence of measurements are obtained that are a realization of the m-vector random process $z(k)$ with the time invariant innovations representation (IR) [7]

$$z(k) = H_i x(k) + w(k), \qquad k = 1, 2, \cdots \tag{2.1}$$

$$x(k+1) = \phi_i x(k) + L_i w(k), \qquad k = 0, 1, 2, \cdots \tag{2.2}$$

where $x(k)$ is the state n-vector and $x(0) \sim N[0, \Sigma_i]$. Without loss of generality it can be assumed that all signal models are of the same dimension, n. The sequence $\{w(k)\}$ is a zero mean, white, Gaussian m-vector sequence with covariance matrix $d_i = (d_i)^T > 0$ and independent of $x(0)$. The model of this system is completely specified except for an unknown, time-invariant parameter which is assumed to be a random vector γ taking values from finite set $\Theta=\{\theta_i\}_{i=1}^{N}\subset R^p$ with *a priori* probabilities $p(\theta_i)$. The initial state covariance Σ_i is the solution of

$$\Sigma_i=\phi_i\Sigma_i\phi_i^T+L_id_iL_i^T \tag{2.3}$$

and can be expressed as an infinite series

$$\Sigma_i=\sum_{r=0}^{\infty}\phi_i^rL_id_iL_i^T(\phi_i^T)^r. \tag{2.4}$$

(The homogeneous system $x(k+1)=\phi_ix(k)$ is assumed asymptotically stable in this short paper.)

Denoting the sequence of measurements up to the kth discrete time instant by the vector $Z_k^T=[z^T(1),\cdots,z^T(k)]$, the probability densities, $\{p(Z_k|\theta_i):\theta_i\in\Theta\}$ are a family of normal densities with

$$p(Z_k|\theta_i)=N[0,P_i(k)] \tag{2.5}$$

where $P_i(k)$, is the output covariance matrix of the system of (2.1) and (2.2).

The output covariance matrix $P_i(k)$ is block Toeplitz with the $(rs)^{th}$ $(m\times m)$ block denoted by

$$[P_i(k)]_{rs}=P_i^{|r-s|},\qquad r,s=1,2,\cdots,k. \tag{2.6}$$

An equivalent description of signal model (2.1), (2.2) is

$$Z_k=A(k|\theta_i)U_k \tag{2.7}$$

where

$$U_k^T=[u^T(1),\cdots,u^T(k)] \tag{2.8}$$

and $A(k|\theta_i)$ is of block lower triangular form with $(m\times m)$ block elements, $a_{r,s}^i$, r, $s=1,2,\cdots,k$, given by

$$a_{r,s}^i=0,\qquad s>r \tag{2.9a}$$

$$a_{r,s}^i=I_m,\qquad r=s \tag{2.9b}$$

$$a_{r,r-1}^i=H_iL_i(r-1) \tag{2.9c}$$

$$a_{r,s}^i=H_i\phi_i^{r-s-1}L_i(s),\qquad r>s+2. \tag{2.9d}$$

The quantity $L_i(s)$ can be related to the Kalman filter associated with signal model (2.1), (2.2) via

$$L_i(s)=\phi_iK_i(s) \tag{2.10}$$

where $K_i(s)$ is the Kalman gain sequence. The random sequence $u(s)$, $s=1,2,\cdots,k$ is a zero mean, white, Gaussian sequence with covariance matrix d_i^k equal to that of the innovations sequence of the Kalman filter associated with the ith model (2.1), (2.2). We comment that it is well known that under the standing assumptions of this short paper

$$\lim_{k\to\infty}d_i^k=d_i\quad\text{and}\quad\lim_{k\to\infty}L_i(k)=L_i. \tag{2.11}$$

With $A(k|\theta_i)$ as defined in (2.9), we introduce the definitions

$$P_i(k)=A(k|\theta_i)D_i^kA^T(k|\theta_i),\qquad D_i^k=\text{diag}\{d_i^1,\cdots,d_i^k\} \tag{2.12}$$

$$Y_i^0=d_i,\qquad Y_i^s=H_i\phi_i^{s-1}L_i,\qquad s=1,2,\cdots. \tag{2.13}$$

The output covariance matrix $P_i(k)$ can be expressed in terms of the Markov parameters Y_i^s via (2.6) and

$$P_i^0=Y_i^0+\sum_{r=1}^{\infty}Y_i^rY_i^0[Y_i^r]^T$$

$$P_i^s=Y_i^sY_i^0+\sum_{r=1}^{\infty}Y_i^{r+s}Y_i^0[Y_i^r]^T,\qquad s=1,2,\cdots. \tag{2.14}$$

It is well known [9] that $Y_i^0\cdots Y_i^{2n}$ specify Y_i^s for all s.

The following identifiability condition is imposed on the set of N possible signal models.

Identifiability condition (C1): No two signal models have the same sets $Y_i^0\cdots Y_i^{2n}$ (or equivalently $Y_i^0\cdots Y_i^n$, $\alpha_i^0\cdots\alpha_i^{n-1}$ where α_i^s are the coefficients of the characteristic equation of ϕ_i and n is the signal model dimension).

Bayesian parameter estimation: The parameter estimate $\hat\theta(k)$ minimizing Bayes' risk for a quadratic loss function is given by [6]

$$\hat\theta(k)=\sum_{i=1}^{N}\theta_ip(\theta_i|Z_k) \tag{2.15}$$

where the *a posteriori* probability, $p(\theta_i|Z_k)$, is expressed as

$$p(\theta_i|Z_k)=\frac{p(Z_k|\theta_i)p(\theta_i)}{\sum_{r=1}^{N}p(Z_k|\theta_r)p(\theta_r)}. \tag{2.16}$$

For on-line computation of the probabilities $p(\theta_i|Z_k)$, (2.16) is usually manipulated into an equivalent recursive form as in [1], [8]. The estimator in recursive form consists of a bank of parallel processors or Kalman filters yielding the pseudo-innovations processes $\tilde z(k|k-1,\theta_i)$. The *a posteriori* probabilities $p(\theta_i|Z_k)$ are calculated recursively in terms of $p(\theta_i|Z_{k-1})$ and $\tilde z(k|k-1,\theta_i)$, $i=1,\cdots,N$. The recursive algorithms are efficient in terms of memory and computational requirements compared with nonrecursive methods.

III. Bounds on the Mean-Square Error

In this section bounds are derived for $\sigma^2(k)$, the mean-square error of the Bayesian parameter estimator, defined as

$$\sigma^2(k)=E\{[\hat\theta(k)-\gamma]^T[\hat\theta(k)-\gamma]\}. \tag{3.1}$$

The bounds for $\sigma^2(k)$ are obtained using similar techniques to those used in [5]. The differences between our analysis and that of [5] are as follows

1) We do not restrict attention to independently and identically distributed measurement sequences.

2) We restrict attention to measurement sequences obtained as outputs of linear, dynamical, finite-dimensional, stochastic systems.

3) An explicit expression for the bound is derived.

Lemma (3.1): The mean-square error of the Bayesian parameter estimator is bounded as

$$\sigma^2(k)<\sum_{j=1}^{N}p(\theta_j)\sum_{\substack{r=1\\r,s\neq j}}^{N}\sum_{s=1}^{N}(\theta_r-\theta_j)^T(\theta_s-\theta_j)c_{rsj}\left[I_{rj}^kI_{sj}^k\right]^{1/2} \tag{3.2}$$

where

$$c_{rsj}=p^{1/4}(\theta_r)p^{1/4}(\theta_s)/p^{1/2}(\theta_j) \tag{3.3}$$

$$I_{rj}^k=2km/2|P_r(k)|^{1/4}|P_j(k)|^{1/4}|P_r(k)+P_j(k)|^{-1/2}. \tag{3.4}$$

Proof (See [15]): A serious drawback with the bounds for $\sigma^2(k)$ given in Lemma 3.1, at least for the usual case when the signal model has infinite memory, is that the number of computations required to calculate the bound increases as the *cube* of the number of measurements k. For the special case when the signal model has finite memory, the growth in computations is linear in k. Taking cognizance of this last fact, we now seek *relaxed bounds* for $\sigma^2(k)$ which can be 1) characterized by a computational effort *linear* in k, and 2) made arbitrarily close to the bounds of the previous section.

First the following *definitions* are introduced:

$$B_{rj}^k=A^{-1}(k|\theta_j)A(k|\theta_r)=\left[\begin{array}{c|c}B_{rj}^{k-1}&0\\\hline(b_{rj}^{k-1})^T&I_m\end{array}\right] \tag{3.5}$$

$$V_{rj}^{k-1}=\left[D_{rj}^{k-1}\right]^{-1}+\left[B_{rj}^{k-1}\right]^T\left[D_j^{k-1}\right]^{-1}\left[B_{rj}^{k-1}\right]>0. \tag{3.6}$$

In the next theorem the following *partitioning* of the matrices $A(k|\theta_i)$, D_i^k defined in Section II and B_{rj}^{k-1} and V_{rj}^{k-1} defined in (3.5)–(3.6) is assumed.

$$D_i^k = \left[\begin{array}{c|c|c} * & 0 & 0 \\ \hline 0 & D_i & 0 \\ \hline 0 & 0 & d_i^k \end{array}\right], \quad A(k|\theta_i) = \left[\begin{array}{c|c} * & 0 \\ \hline * & A_i \end{array}\right] \tag{3.7}$$

$$B_{rj}^{k-1} = \left[\begin{array}{c|c} * & 0 \\ \hline * & B_{rj} \end{array}\right], \quad V_{rj}^{k-1} = \left[\begin{array}{c|c} * & * \\ \hline * & V_{rj} \end{array}\right] \tag{3.8}$$

$$\left(b_{rj}^{k-1}\right)^T = \left[\begin{array}{c|c} 0 & b_{rj}^T \end{array}\right]. \tag{3.9}$$

The dimension of $A_i, D_i, d_i^k, B_{rj}, b_{rj}$, and V_{rj} is $mM \times mM$, $(mM-m)\times(mM-m)$, $m\times m$, $(mM-m)\times(mM-m)$, $m\times(mM-m)$, and $(mM-m)\times(mM-m)$, respectively, for all k. Here M is the extent of the assumed finite memory. The asterisks * denote terms that do not enter our calculations. The following relationships are readily established:

$$\left[\begin{array}{c|c} B_{rj} & 0 \\ \hline b_{rj}^T & I_m \end{array}\right] = A_j^{-1}A_r, \qquad V_{rj} = D_r^{-1} + B_{rj}^T D_j^{-1} B_{rj} > 0. \tag{3.10}$$

Theorem 3.1: 1) The term I_{rj}^k of Lemma 3.1 can be calculated recursively via

$$I_{rj}^k = I_{rj}^{k-1}\rho_{rj}^k; \qquad I_{rj}^0 = 1 \tag{3.11}$$

$$\rho_{rj}^k = |d_r^k d_j^k|^{1/4} |1/2 d_r^k + 1/2 d_j^k + \epsilon_{rj}^k|^{-1/2} \tag{3.12}$$

$$\epsilon_{rj}^k = 1/2\left[b_{rj}^{k-1}\right]^T\left[V_{rj}^{k-1}\right]^{-1}\left[b_{rj}^{k-1}\right] \tag{3.13}$$

with d_r^k defined in Section II and b_{rj}^{k-1} and $V_{rj}^{k-1} > 0$ defined in (3.5).

2) The term ρ_{rj}^k can be bounded as

$$\rho_{rj}^k \leqslant \bar{\rho}_{rj}^k(M) \tag{3.14}$$

$$\bar{\rho}_{rj}^k(M) = |d_r^k d_j^k|^{1/4} |1/2 d_r^k + 1/2 d_j^k + \bar{\epsilon}_{rj}^k(M)|^{-1/2} \tag{3.15}$$

$$\bar{\epsilon}_{rj}^k(M) = 1/2 b_{rj}^T V_{rj}^{-1} b_{rj}, \qquad \text{with } V_{rj} > 0. \tag{3.16}$$

The bound $\bar{\rho}_{rj}^k(M)$ is a monotonically decreasing function of M and depends only on $A_r, A_j, D_r, D_j, d_r^k$, and d_j^k in the partitioning of (3.7).

Proof: 1) Using (3.4), (2.12), and the fact $|A(k|\theta_i)| = 1$ we have that

$$|P_i(k)| = |D_i^k| = |P_i(k-1)||d_i^k| \tag{3.17}$$

$$|P_r(k) + P_j(k)| = |\left[B_{rj}^k\right] D_r^k \left[B_{rj}^k\right]^T + D_j^k| \tag{3.18}$$

where B_{rj}^k is defined in (3.5).

Now Result (A1) of the Appendix can be applied to (3.18) to yield

$$|P_r(k) + P_j(k)| = |P_r(k-1) + P_j(k-1)||d_r^k + d_j^k + 2\epsilon_{rj}^k| \tag{3.19}$$

where the matrix ϵ_{rj}^k is defined in (3.13). Substituting (3.17) and (3.19) into (3.4) establishes (3.11) and (3.12). Finally, observe $\epsilon_{rj}^k \geqslant 0$.

2) To establish that $\bar{\rho}_{rj}^k(M)$ is a monotonically decreasing function of M bounded below by ρ_{rj}^k it suffices to show that $\bar{\epsilon}_{rj}^k(M)$ is a monotonically increasing function of M bounded above by ϵ_{rj}^k. This result follows immediately from Result (A2) of the Appendix. ∇∇∇

Asymptotic Results: We now *define*

$$\rho_{rj} = \lim_{k\to\infty} \rho_{rj}^k, \qquad \bar{\rho}_{rj}(M) = \lim_{k\to\infty} \bar{\rho}_{rj}^k(M). \tag{3.20}$$

Theorem 3.2: For the identification of signal models of the form (2.1), (2.2) the Bayesian parameter estimator converges exponentially to the true parameter value if and only if the identifiability condition (C1) is satisfied.

Proof—Sufficiency: Assume that k is sufficiently large so that A_r, A_j, D_r, D_j, d_r^k, and d_j^k have attained steady-state values for $M \leqslant 2n+1$. Thus, $\bar{\rho}_{rj}^k(M)$ which can be calculated in terms of these quantities has assumed its steady-state value $\bar{\rho}_{rj}(M)$ for $r, j = 1, \cdots, N$. Now if $Y_r^0 \neq Y_j^0$ (or equivalently $d_r \neq d_j$) then it can be shown that $|d_r d_j|^{1/2}|1/2 d_r + 1/2 d_j|^{-1} < 1$. Theorem 3.1-1) then implies that $\rho_{rj} < 1$.

If $Y_r^0 = Y_j^0$ then define $\bar{s}$ as the maximum s for $s \in \{1, \cdots, 2n\}$ such that $Y_r^t = Y_j^t$ for all $t < \bar{s}$. By assumption then $Y_r^{\bar{s}} \neq Y_j^{\bar{s}}$. Let $M = \bar{s}$ in Theorem 3.1. Referring to (3.7) we have

$$A_i = \begin{bmatrix} I_m & 0 & - & - & 0 \\ Y_i^1 & I_m & - & - & 0 \\ - & - & & & - \\ Y_i^{\bar{s}} & Y^{\bar{s}-1} & - & - & I_m \end{bmatrix} = \left[\begin{array}{c|c} L_i & 0 \\ \hline Y_i^{\bar{s}} Y_i^{\bar{s}-1} \text{—} Y_i^1 & I_m \end{array}\right]. \tag{3.21}$$

Now, in (3.10)

$$\left[\begin{array}{c|c} B_{rj} & 0 \\ \hline b_{rj}^T & I_m \end{array}\right] = A_j^{-1}A_r = \left[\begin{array}{c|c} L_j^{-1} & 0 \\ \hline -\left[Y_j^{\bar{s}} \cdots Y_j^1\right] L_j^{-1} & I_m \end{array}\right] \left[\begin{array}{c|c} L_j & 0 \\ \hline Y_r^{\bar{s}} \cdots Y_r^1 & I_m \end{array}\right] \tag{3.22}$$

$$= \left[\begin{array}{c|c} I_{m(\bar{s}-1)} & 0 \\ \hline \left(Y_r^{\bar{s}} - Y_j^{\bar{s}}\right) 0 \cdots 0 & I_m \end{array}\right]. \tag{3.23}$$

Thus, in (3.15)

$$\bar{\rho}_{rj}(\bar{s}) = |d_j|^{1/2} |d_j + \bar{\epsilon}_{rj}(\bar{s})|^{-1/2}, \bar{\epsilon}_{rj}(\bar{s}) = \left[Y_r^{\bar{s}} - Y_j^{\bar{s}}\right] d_j \left[Y_r^{\bar{s}} - Y_j^{\bar{s}}\right]^T, \tag{3.24}$$

since from (3.10) $V_{rj} = 2D_j^{-1}$. Now $\bar{\epsilon}_{rj}(\bar{s})$ is a positive definite matrix. Thus, using the matrix inequality $|A+B| > |A|$ for A, $B > 0$, $\rho_{rj} \leqslant \bar{\rho}_{rj}(\bar{s}) < 1$.

Necessity: If $Y_j^s = Y_r^s$ for $s = 0, \cdots, 2n$ and for all $r, j = 1, \cdots, N$, then by virtue of (2.6) and (2.14) $P_r(k) = P_j(k)$ for all $r, j = 1, \cdots, N$ and for all k since $Y_j^s = Y_r^s$ for all s. Consequently, $p(Z_k|\theta_r) = p(Z_k|\theta_j)$ for a given realization of the data Z_k and thus the estimator does not converge. ∇∇∇

IV. Adaptive State Estimation

For signal models as described in Section II an optimal mean-square error sense (MSE) state estimate $\hat{x}(k|k)$ can be obtained in the presence of unknown system parameters via [1]

$$\hat{x}(k|k) = \sum_{r=1}^{N} \hat{x}(k|k,\theta_r) p(\theta_r|Z_k) \tag{4.1}$$

where $\hat{x}(k|k,\theta_r)$ is the MSE state estimate conditioned on $\gamma = \theta_r$. These estimates, $\hat{x}(k|k,\theta_r)$, are obtained using a Kalman filter conditioned on θ_r, and the *a posteriori* probabilities $p(\theta_r|Z_k)$ are calculated using the innovations of the Kalman filter. The convergence results of previous sections are applicable to the nonlinear estimation scheme given by (4.1).

Theorem 4.1: For the adaptive estimation scheme of (4.1) the identifiability condition (C1) is necessary and sufficient for the mean-square error in the state estimate $\hat{x}(k|k)$ to converge exponentially to its lower bound, the mean-square error for linear filtering assuming complete knowledge of the model statistics.

Proof: See [15].

V. Numerical Results

In this section two examples are given to illustrate the theory presented in earlier sections.

Example 5.1: In this example the quantity $\bar{\rho}_{rj}^k(M)$ of Theorem 3.1 is computed for the case of two signal models in time-invariant IR form. The two models are specified by

$$\phi_r = \begin{bmatrix} 0.0 & -0.7 \\ 1.0 & 0.0 \end{bmatrix} \qquad \phi_j = \begin{bmatrix} 0.6 & -0.61 \\ 1.0 & 0.0 \end{bmatrix}$$

$$L_r = \begin{bmatrix} -0.54874 \\ 0.54874 \end{bmatrix} \qquad L_j = \begin{bmatrix} 0.0039341 \\ 0.71681 \end{bmatrix}$$

$$H_r = H_j = [0.5 \quad 0.5]$$

$$d_r = 0.80241 \qquad d_j = 1.0264.$$

The Σ_r and Σ_j which give a stationary output sequence can be computed from (2.4). Denoting the roots of the characteristic polynomial of ϕ_r as

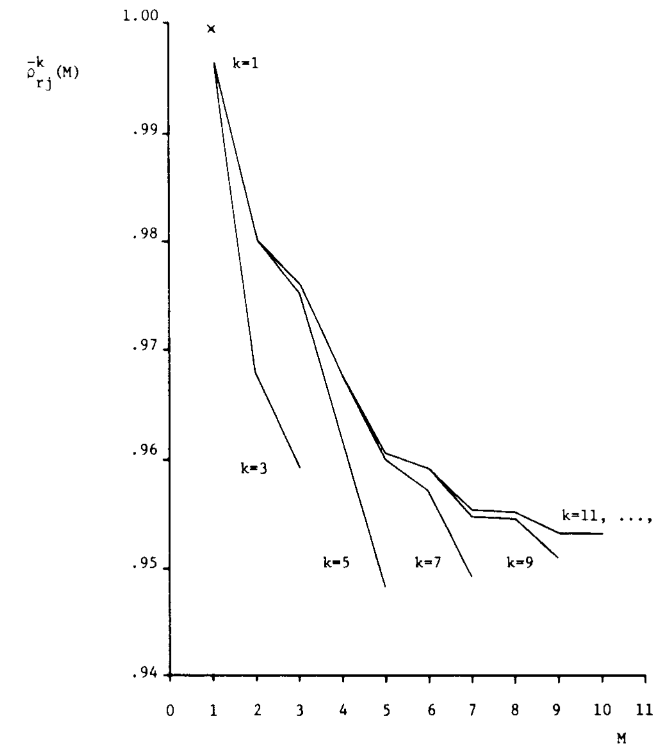

Fig. 1. Effect of memory M on upper bounds.

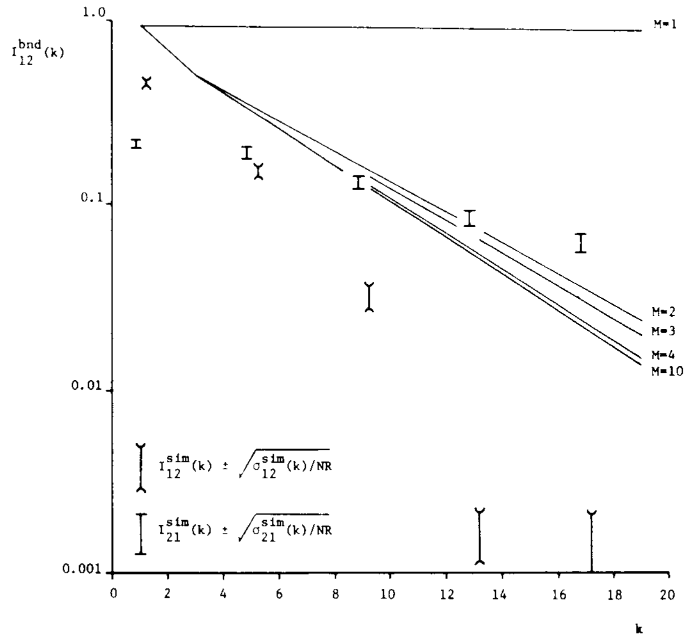

Fig. 2. Comparison of simulation results and upper bounds.

λ_r^i, $i=1, 2$, then the time constants of the system are $k_r^i=[1/\log_e|\lambda_r^i|]$ where $[\cdot]$ denotes integer part. For the models of this example $k_r^i=8$ and $k_j^i=4$, $i=1, 2$.

In Fig. 1 the effect of M on the quantity $\bar{\rho}_{rj}^k(M)$ is plotted for various time points k. For this example it would appear that a memory M of 10 time points enables a bound to be calculated which is as tight as that which could be calculated taking into account the entire memory. Choosing M of the order of the longest system time constant appears to be adequate to achieve the best bound. In this example possibly 1000 data points would be needed to distinguish between the two models. In this case there is a considerable advantage to assuming a finite memory in calculating this type of bound.

Example 5.2: In this example we compare the bounds obtained assuming a finite memory with simulation results. Which of five possible signal models is present is to be decided using a Bayesian parameter estimator. Each of the five models is assumed to be in time invariant IR form. First, details of the simulation are given, then results are presented for comparison with the bounds and discussion follows.

In the simulation the *a posteriori* probabilities $p(\theta_r|Z_k)$ are averaged over NR runs to give $I_{rj}^{\text{sim}}(k)=NR^{-1}\Sigma_{s=1}^{NR}[p(\theta_r|Z_k)]_s^j$. Here the subscript j refers to the true model present. In order that some idea of the accuracy of the simulation result can be obtained the quantity $\sigma_{rj}^{\text{sim}}(k)=(NR-1)^{-1}\Sigma_{s=1}^{NR}\{[p(\theta_r|Z_k)]_s^j-I_{rj}^{\text{sim}}(k)\}^2$ is also calculated. In the simulations $NR=1000$ and $I_{rj}^{\text{sim}}(k)$ and $\sigma_{rj}^{\text{sim}}(k)$ are calculated for time points $k=1, 5, 9, 13, 17$.

In Fig. 2 $I_{12}^{\text{sim}}(k)$ and $I_{21}^{\text{sim}}(k)$ have been plotted for time points $k=1, 5, 9, 13, 17$. The number of runs NR is 1000. The values $I_{rj}^{\text{sim}}(k)\pm\sqrt{\sigma_{rj}^{\text{sim}}(k)/NR}$ have been shown in Fig. 2. For comparison the quantity $I_{rj}^{bnd}(k)=\Pi_{s=1}^{k}\bar{\rho}_{rj}^{s}(M)$ is plotted for $k=1, 2, \cdots, 20$, and $M=1, 2, 3, 4, 10$. Although not completely obvious from (3.15) it is found that $I_{rs}^{bnd}(k)=I_{sr}^{bnd}(k)$. The two models of interest are now specified by

$$\phi_1=\begin{bmatrix}1.5 & -0.75\\ 1.0 & 0.0\end{bmatrix}\qquad \phi_2=\begin{bmatrix}-1.0 & -0.707\\ 1.0 & 0.0\end{bmatrix}$$

$$L_1=\begin{bmatrix}1.0394\\ 0.78906\end{bmatrix}\qquad L_2=\begin{bmatrix}-1.0188\\ 0.63887\end{bmatrix}$$

$$H_1=H_2=[1.0 \quad 0.5]$$

$$d_1=3.484 \qquad d_2=4.3277.$$

Observe that in Fig. 2 there is some irregularity in the convergence rate of $I_{rj}^{\text{sim}}(k)$ as k increases and that $I_{21}^{\text{sim}}(k)$ in fact is above the bound at certain time points. From our experience such irregularities are quite common in results obtained using simulations. The results in Fig. 2 do not, in our opinion, invalidate the bounds but rather are an indictment of the use of simulations to study the performance of stochastic nonlinear systems. Here the simulations do indicate that the bounds are not absurdly weak.

VI. Discussion

It may be concluded that the upper bounds derived in this short paper provide a simple and effective means of determining the convergence rate of the class of parameter estimator studied here. The finite memory required to obtain bounds as tight as those calculated with no memory limitations appears to be related to the system time constants. However, the convergence rate depends upon the "distance" between the various possible signal models. This latter concept of "distance" between the models is discussed in a later paper using the *Kullback information function*.

Appendix

In this Appendix, certain not well known and nontrivial matrix results needed in this short paper are given. (These results are derived in [15].)

Result A1:

$$X=\left|\begin{bmatrix}A_1 & 0\\ A_2 & I\end{bmatrix}\begin{bmatrix}P_1^r & 0\\ 0 & P_2^r\end{bmatrix}\begin{bmatrix}A_1^T & A_2^T\\ 0 & I\end{bmatrix}+\begin{bmatrix}P_1^j & 0\\ 0 & P_2^j\end{bmatrix}\right|$$

$$=|A_1P_1^rA_1^T+P_1^j||P_2^r+P_2^j+\Sigma|$$

where

$$\Sigma=A_2\left[(P_1^r)^{-1}+A_1^T(P_1^j)^{-1}A_1\right]^{-1}A_2^T$$

and the matrices P_1^r, P_2^r, P_1^j, and P_2^j are assumed to be positive definite and symmetric.

Result A2: For the positive definite matrix $\begin{bmatrix}V_1 & V_3\\ V_2 & V_4\end{bmatrix}$

$$[A_1^T A_2^T]\begin{bmatrix} V_1 & V_3 \\ V_2 & V_4 \end{bmatrix}^{-1}\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \geqslant A_2^T V_4^{-1} A_2.$$

REFERENCES

[1] D. G. Lainiotis, "Optimal adaptive estimation: Structure and parameter adaptation," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 160–170, Apr. 1971.

[2] A. S. Gilman and I. B. Rhodes, "Cone-bounded nonlinearities and mean-square bounds—Estimation upper bound," *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 260–265, June 1973.

[3] A. Wald, "Note on the consistency of the maximum likelihood estimate," *Ann. Math. Statist.*, vol. 20, pp. 595–601.

[4] P. E. Caines and J. Rissanen, "Maximum likelihood estimation of parameters in multivariate Gaussian stochastic processes," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 102–104, Jan. 1974.

[5] L. A. Liporace, "Variance of Bayes estimates," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 665–669, Nov. 1971.

[6] E. L. Lehman, *Testing Statistical Hypothesis*. New York: Wiley, 1959, pp. 12–23.

[7] T. Kailath, "The innovations approach to detection and estimation," *Proc. IEEE*, vol. 58, pp. 680–695, 1970.

[8] F. L. Sims, D. G. Lainiotis, and D. T. Magill, "Recursive algorithm for the calculation of the adaptive Kalman filter weighting coefficients," *IEEE Trans. Automat. Contr.* (Corresp.), vol. AC-14, pp. 215–218, Apr. 1969.

[9] C. T. Chen, *Introduction to Linear System Theory*. New York: Holt, Rinehart and Winston, 1970.

[10] D. G. Lainiotis and F. L. Sims, "Performance measures for adaptive Kalman estimators," *IEEE Trans. Automat Contr.*, vol. AC-15, Apr. 1970.

[11] D. G. Lainiotis, "A class of upper bounds on probability of error for multihypothesis pattern recognition," *IEEE Trans. Inform. Theory*, vol. IT-15, Nov. 1969.

[12] ——, "On a general relationship between estimation, detection and the Bhattacharyya coefficient," *IEEE Trans. Inform. Theory*, vol. IT-15, July 1969.

[13] D. G. Lainiotis and S. K. Park, "Probability of error bounds," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-7, Apr. 1971.

[14] E. Tse, "Bounds for identification error and a quantitative measure of identifiability," presented at the IEEE Conf. Decision Contr., Nov. 1974.

[15] R. M. Hawkes and J. B. Moore, "Performance of Bayesian parameter estimators for linear signal models," Univ. Newcastle, N. S. W., Australia, Tech. Rep. EE7410, July 1974.