Performance Bounds for Adaptive Estimation

RICHARD M. HAWKES AND JOHN B. MOORE, MEMBER, IEEE

Abstract—The performance of adaptive state estimators for linear dynamic systems is investigated. The adaptive state estimates are formed under the assumption that the unknown system parameter belongs to a finite set and is thus readily implementable. It is shown that, for the true parameter value in a prescribed region in the parameter space, the corresponding *a posteriori* probability (or weighting coefficient in the adaptive estimator) converges exponentially in the νth mean ($\nu > 1$) and almost surely to unity. The analysis is based on the asymptotic per sample formula for the Kullback information function, which is derived in this paper. The significance of the analysis for applications is also examined.

I. INTRODUCTION

ATTENTION IS centered in this paper on the performance analysis of a class of on-line near-optimal nonlinear estimators. These are employed for state and parameter estimation of discrete-time linear Gaussian dynamic signal models with unknown structure and parameter values.

In order to apply linear quadratic Gaussian optimization theory [1] in a given application, knowledge of the signal model structure and its parameters is required. Frequently the determination of the signal model is carried out off line, but it is clearly preferable in some circumstances to use on-line techniques if these can be devised. In fact, for signal models with parameters belonging to a finite parameter set optimal mean-square-error sense state estimators can be implemented on line. In [2] such adaptive estimators are shown to be de-

composable into two parts: a linear nonadaptive part, consisting of a bank of Kalman filters, and a nonlinear part that incorporates the adaptive nature of the estimator. There exists a large class of physical problems for which the above formulation is suitable.

The purpose of this paper can be summarized under the following four points.

1) We survey several existing performance bounds for parameter estimation which are relevant to adaptive estimation.

2) For the class of nonlinear estimator under discussion here, performance bounds are developed for the case when the true parameter value turns out to lie outside the assumed finite set. This work is also reported in [3]-[7]. Besides being of some theoretical interest, the analysis of the adaptive estimators yields valuable insights into the application of such estimators operating under suboptimal conditions.

3) The question of the behavior and tightness of the bounds is addressed, comparisons are made with earlier work, and the computational requirements are discussed.

4) Several applications for the performance analysis of this paper are also suggested.

The organization of the paper is as follows. In Section II, early work on the performance analysis of parameter estimators is surveyed and its relationship with the results of this paper are discussed. Section III introduces the concept of adaptive estimation. In Section IV, there is discussion of performance bounds and techniques used to develop these. The information function is introduced in Section V and decision regions are defined in Section VI. Section VII contains the main results of the paper. Computational aspects and tightness of the bounds are addressed in Section VIII. Various applications are considered in Section IX and Section X is the concluding discussion.

II. EARLY WORK ON PERFORMANCE ANALYSIS OF PARAMETER ESTIMATORS

Since there is a high computational burden to obtain performance results for nonlinear estimators, it is desirable to develop an analysis of such estimators by deriving performance bounds. Such bounds can then be used to examine the convergence properties of an estimator. This is the approach taken in this paper. In this section, we survey the work of others on the performance analysis of parameter estimators relevant to the performance analysis of adaptive estimators.

Quite a lot of work has been done on the asymptotic properties of maximum likelihood estimators. An excellent discussion of these properties, which summarizes the pioneering work of Fisher, Wald, and others, can be found in [8]. Much of the early work applies only to the case of independently and identically distributed data. Clearly, where dynamic models are of interest such analysis is inadequate. Recently Caines and Rissanen [9] and also Astrom and Bohlin [10] have investigated the asymptotic properties of maximum likelihood estimators of the parameters of an autoregressive moving-average process. However, due to the difficulties encountered in implementing suitable practical algorithms [11], the significance of the theoretical properties of maximum likelihood estimators such as consistency and asymptotic efficiency for practice is open to question.

The survey paper of Seidman [12] on the performance limitations and error calculations for parameter estimation is also worth mentioning here. Although a class of signal models different from the one considered in this paper is discussed in [12], many properties of estimators carry over to different

classes of signal models. In itself, [12] is a good comparative study and it also provides a standard against which new bounding techniques can be judged.

In [13], linear update model reference estimators are studied for application to dynamic signal models with system functions which are uniformly Lipschitz in the state and the control. Upper bounds on the mean-square error of such estimators are obtained where the tightness of the bounds is dependent upon the tightness of the cone bounds on the signal model nonlinearities. Bound optimal filters can be designed which become optimal linear filters as the cone bounds collapse. In fact, the theory appears to be of use only in the case of signal models which are almost linear.

Along different lines, probability-of-error bounds for large classes of multihypotheses detection problems are given by Lainiotis [15], [16] and Lainiotis and Park [14]. We also note that Tse [17] is currently investigating bounds for identification error.

Having discussed an assortment of estimation techniques and associated performance analyses, which could conceivably be used to estimate unknown parameters of linear dynamic signal models, we turn our attention now to some work which forms the starting point for the analysis of this paper.

In [18], the performance of Bayes' conditional-mean parameter estimators is studied on a finite parameter set. Such estimates exhibit a mean-square error that diminishes exponentially (to zero) with the number of observations, the observations being assumed independently and identically distributed. Two situations are discussed: the true parameter included in the parameter set and the true parameter not included in the parameter set. In the latter case, the existence of an information function must be invoked to demonstrate the exponential convergence rate (though the error in this case does not tend to zero).

In this paper, analogous results to those in [18] are presented for the case of linear dynamic stochastic signal models to which the independently and identically distributed data assumption of [18] does not apply. These new results are employed to analyze the performance of adaptive estimators.

Adaptive-estimator performance has also been investigated by Lainiotis et al. [2], [19]. Specifically, Lainiotis [2] obtained the exact expression for the mean-square error of continuous adaptive state estimators in a partitioned realization form that requires a minimum of additional computations and is convenient for on-line implementation. This exact error-covariance expression is the only exact and explicit one obtained in adaptive estimation to date. Similarly, Lainiotis and Sims [19] have obtained the analogous exact expression for the state-error-covariance matrix of the discrete-data adaptive estimator. However, evaluating these expressions is tantamount to simulating the system under study. Also, Magill [20] and Hilborn and Lainiotis [21a], [21b] discuss sufficient conditions for the convergence of the adaptive weighting coefficients. These various results do not overlap with the results given herein.

The important features of the analysis of this paper are the handling of nonindependently and identically distributed data, drawing the distinction between the true parameter being included and excluded in the finite parameter set, and demonstrating the exponentially fast convergence rates.

III. ADAPTIVE ESTIMATION

The concept of adaptive estimation has been considered by many authors [41]. In particular, it has been considered by

Fralick [42] for the case of the unsupervised estimation of time-invariant parameters under the assumption of independent measurements. Magill [20] extended these results to dependent measurements generated by state-variable models. However, unlike Fralick [42], Magill's results were restricted to the case of measurement certainty, scalar observations, and discrete unknown parameters. Moreover, Magill's results were not in a recursive form. Subsequently, Lainiotis and Sims [19], [21] extended Magill's results to vector measurements and arbitrary continuous or discrete parameters. More importantly they obtained the optimal adaptive estimator in a recursive form and also derived the exact error-covariance expression. Further, Lainiotis [2a], [2b] and Lainiotis and Park [2c] extended the results of Fralick and Magill to the unsupervised estimation of unknown arbitrary parameters for dependent measurements given by state-variable models. The more difficult corresponding continuous-data adaptive-estimation problem was first considered by Lainiotis [2], [2d], [2e], who essentially obtained the continuous equivalent of all of the above results. It is finally noted that the work of Lo [22], Cameron [23], as well as of Lainiotis and Park [43], and the more general results of Hilborn and Lainiotis [21a] are also related.

We now consider in turn, a signal model, an optimal adaptive-estimation scheme, a finite-dimensional estimator, and a discussion on the adaptive-estimation performance.

A. Signal Model

Consider the scalar stationary Gaussian stochastic measurement process $y(k)$ with innovations representation (IR) [24]-[27]

$$x(k + 1) = \phi_A x(k) + L_A w(k) \qquad (3.1)$$

$$y(k) = H_A x(k) + w(k). \qquad (3.2)$$

Here, $x(k)$ is the state n-vector with stochastic initial condition $x(0) \sim N[0, \Sigma_A]$. The random input sequence $w(k) \sim N[0, d_A]$, $d_A > 0$ is assumed to be white and independent of $x(0)$. The system matrices $\phi_., L_., H_.,$ and $d_.$ are functions of the unknown parameter A taken to be a sample value of the random vector a distributed on $\mathcal{R}_a \subset R^P$ with probability density $p_a(A)$. We assume that $p_a(A) = 0$ in the complement of $\mathcal{R}_a$ denoted by $\overline{\mathcal{R}}_a$. The initial state covariance matrix Σ_A is the solution of

$$\Sigma_A = \phi_A \Sigma_A \phi_A' + L_A d_A L_A' \qquad (3.3)$$

which can be expressed as

$$\Sigma_A = \sum_{r=0}^{\infty} \phi_A^r L_A d_A L_A' [\phi_A']^r. \qquad (3.4)$$

Denoting the sequence of measurements up to the kth discrete time instant by the vector $Y_k' = [y(1), y(2), \cdots, y(k)]$, the probability density of Y_k given $a = A$ is

$$p(Y_k | A) = N[0, T_A(k)] \qquad (3.5)$$

where the output covariance matrix of the system is Toeplitz, viz.,

$$T_A(k) = \begin{bmatrix} t_A(0) & t_A(1) & \cdots t_A(k-1) \\ t_A(1) & t_A(0) & t_A(k-2) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ t_A(k-1) & t_A(k-2) \cdots t_A(0) \end{bmatrix} \qquad (3.6)$$

The entries of $T_A(k)$ are given by

$$t_A(s) = \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} u_A(j) u_A(l) d_A \quad \delta(s + l - j) \qquad (3.7)$$

where

$$\begin{aligned} u_A(0) &= 1 \\ u_A(s) &= H_A \phi_A^{s-1} L_A, \qquad s = 1, 2, \cdots \\ \delta(s) &= \begin{cases} 1, & s = 0 \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \qquad (3.8)$$

The pulse transfer function of the system (3.1), (3.2) is introduced:

$$U_A(z) = \sum_{r=0}^{\infty} z^{-r} u_A(r) \qquad (3.9)$$

$$= 1 + H_A [zI - \phi_A]^{-1} L_A. \qquad (3.10)$$

The spectral density of the observed stochastic process $y(k)$ is given by [28]

$$f_A(\omega) = \sum_{s=-\infty}^{\infty} t_A(s) e^{is\omega} \qquad (3.11)$$

$$= U_A(e^{-i\omega}) U_A(e^{i\omega}) d_A. \qquad (3.12)$$

We denote the least upper bound and greatest lower bound of $f_A(\cdot)$ on $[0, 2\pi]$ as M_A and m_A, respectively.

Certain subsets of the vector space R^P are considered in the following definitions.

Definition 3.1: $\mathcal{R}_s \subset R^P$ is the set of all $A \in R^P$ for which the system (3.1), (3.2) is asymptotically stable and minimum phase (i.e., the zeros and poles of $U_A(z)$ lie within the unit circle).

For $A \in \mathcal{R}_s$, $t_A(k)$ is absolutely summable over k [29, ch. 4] and, hence, $f_A(\cdot)$ is bounded and Riemann-integrable in $[0, 2\pi]$ [30]. Also, for $A \in \mathcal{R}_s$, $m_A > 0$.

Denoting $Z_A^0 = d_A$ and Markov parameters $Z_A^s = H_A \phi_A^{s-1} L_A$, $s = 1, 2, \cdots$, we define a set $Z_A = \{Z_A^0, Z_A^1, \cdots, Z_A^{2n}\}$, where n is the signal-model dimension.

Definition 3.2: An identifiable parameter set $\mathcal{R}_i \subset R^P$ has the property that $Z_{A_1} = Z_{A_2}$ if and only if $A_1 = A_2$ for all $A_1, A_2 \in \mathcal{R}_i$.

The theorems in later sections require the assumption $A \in \mathcal{R}_a \subset \mathcal{R}_i \cap \mathcal{R}_s$.

B. Optimal Adaptive Estimation

The optimal mean-square state estimate $\hat{x}_{k+1|k}$ of $x(k + 1)$ is given by [2], [31]

$$\hat{x}_{k+1|k} = \int \hat{x}_{k+1|k,A} \, p(A | Y_k) \, dA \qquad (3.13)$$

where

$$\hat{x}_{k+1|k,A} = \phi_A \hat{x}_{k|k-1,A} + L_{k|A} \nu_{k|A} \qquad (3.14)$$

$$\hat{x}_{1|0,A} = 0 \qquad (3.15)$$

$$\nu_{k|A} = y(k) - H_A \hat{x}_{k|k-1,A} \qquad (3.16)$$

$$L_{k|A} = [K_{k|A} + L_A d_A] d_{k|A}^{-1} \qquad (3.17)$$

$$d_{k|A} = H_A \Sigma_{k|k-1,A} H_A' + d_A \qquad (3.18)$$

$$K_{k|A} = \phi_A \Sigma_{k|k-1,A} H_A' \qquad (3.19)$$

$$\Sigma_{k+1|k,A} = [\phi_A - L_{k|A} H_A] \Sigma_{k|k-1,A} [\phi_A - L_{k|A} H_A]'$$
$$+ [L_A - L_{k|A}] d_A [L_A - L_{k|A}] \qquad (3.20)$$

$$\Sigma_{1|0,A} = \Sigma_A \qquad (3.21)$$

$$p(A \mid Y_k) = \frac{p(Y_k \mid A) p_a(A)}{\int p(Y_k \mid A) p_a(A) \, dA} \qquad (3.22)$$

$$p(Y_k \mid A) = \left[(2\pi)^k \prod_{s=1}^{k} d_{s|A} \right]^{-1/2} \exp\left[-\frac{1}{2} \sum_{s=1}^{k} \nu_{s|A}^2 d_{s|A}^{-1} \right].$$
$$\qquad (3.23)$$

Since the IR and Kalman filter are inverse systems, and by virtue of our asymptotic stability and minimum phase assumptions, as $k \to \infty$, $\Sigma_{k+1|k,A} \to 0$, $L_{k|A} \to L_A$, and $d_{k|A} \to d_A$.

Usually adaptive estimators in the form given above cannot be readily implemented and we are led to consider suboptimal estimators as in the next subsection.

C. Finite Dimensional Estimators

The adaptive state estimator designed under the assumption of a finite parameter set is now described. The situation where the true value of the parameter may not belong to this finite set will be included in the performance analysis to follow.

We introduce a finite set $\mathcal{A} = \{A_1, A_2, \cdots, A_N\} \subset \mathcal{R}_a$ and assign a set of "a priori probabilities" p_r, $r = 1, 2, \cdots, N$, for the finite set on an ad hoc basis in proportion to the values $p_a(A_r)$, $r = 1, 2, \cdots, N$. Without loss of generality, we impose the restriction $p_a(A_r) \neq 0$, $A_r \in \mathcal{A}$.

The adaptive state estimate on $\mathcal{A}$ is defined as

$$\hat{x}_{k+1|k} = \sum_{r=1}^{N} \hat{x}_{k+1|k,A_r} p(A_r \mid Y_k) \qquad (3.24)$$

where the a posteriori probabilities are

$$p(A_l \mid Y_k) = \frac{p(Y_k \mid A_l) p_l}{\sum_{r=1}^{N} p(Y_k \mid A_r) p_r} = \frac{p(Y_k \mid A_l) p_a(A_l)}{\sum_{r=1}^{N} p(Y_k \mid A_r) p_a(A_r)} \qquad (3.25)$$

Notice that the normalizing constant for the assigned p_r is not required in calculating $p(A_l \mid Y_k)$. The quantities $\hat{x}_{k+1|k,A_r}$ and $p(Y_k \mid A_r)$ can be calculated recursively via equations (3.14)–(3.21) and (3.23) with $A = A_r$.

Thus the finite-dimensional adaptive estimator on $\mathcal{A}$ consists of a bank of N Kalman filters plus certain nonlinear operations on the data processed by the Kalman filters.

D. Adaptive Estimator-Performance

In our performance analysis to follow, the convergence behavior of the finite set of a posteriori probabilities $p(A_r \mid Y_k)$, $r = 1, 2, \cdots, N$, is examined. These are the weighting coefficients in the adaptive estimate given by (3.24) and their behavior determines the behavior of the adaptive estimator.

The main result to be presented here is that the weighting coefficients will converge almost surely [32] to unity for the coefficient corresponding to the signal model (defined in the sequel) nearest the true signal model and to zero for the others. Thus the adaptive state estimate converges to the state estimate produced by the Kalman filter that best matches the true signal model according to a criterion to be introduced.

IV. Performance Bounds

In this section, some useful bounding techniques are discussed and the reader is introduced to the problem of analyzing the bounds so obtained in order to determine the convergence properties of the adaptive-estimation algorithm. The set $\mathcal{R}_a$ is assumed, in this section, to be the finite set $\{A_1, A_2, \cdots, A_{N'}\}$ with a priori probabilities $p_a(A_r)$, $r = 1, 2, \cdots, N'$. We further assume that $\mathcal{A} = \mathcal{R}_a$ and $N = N'$ so that the finite dimensional estimator on $\mathcal{A}$ of Section III is optimal for this case. The analysis of this simpler case forms a useful background for the more general theory to follow.

The key quantity which we seek to bound is the expectation

$$E_{Y_k|a} \{p(A_r \mid Y_k) \mid A_j\}, \quad A_r, A_j \in \mathcal{A}. \qquad (4.1)$$

Before explaining the importance of this quantity, the apparent paradox of taking the expectation of the probability of one sample value of a random variable conditioned on another sample value is resolved. Referring to Section III, the a posteriori probabilities $p(A_r \mid Y_k)$ are seen to be functions only of the received data Y_k and other deterministic quantities. Thus the expectation is taken, given that the data comes from the true model with parameter value A_j.

The following lemma gives a condition for almost sure convergence of a sequence of random variables X_m to zero.

Lemma 4.1: If there exists an n such that for $m > n$ $E\{|X_m|^\nu\} \leqslant \rho^m$, $0 < \rho < 1$ and $\nu > 0$, then $\{X_m\}$ converges almost surely to zero.

Proof: See [3]. The proof follows from the Markov inequality [32] and the Borel–Cantelli lemma [32]. The discussion of [32, sec. 6, ch. 1] is helpful here and specifically proposition 6.4 should be noted.   Q.E.D.

The following inequality is easily verified for $l \geqslant 1$:

$$E_{Y_k|a} \{|p(A_r \mid Y_k)|^l \mid A_j\} \leqslant E_{Y_k|a} \{p(A_r \mid Y_k) \mid A_j\}. \quad (4.2)$$

In light of Lemma 4.1 and inequality (4.2), the importance of bounding the expectation of (4.1) should be obvious. To make use of Lemma 4.1, however, the bound must display an exponential convergence rate to zero.

It should also be noted that if the bound for the expectation of (4.1) converges to zero then $p(A_r \mid Y_k)$ also converges to zero in νth mean ($\nu > 1$) [32] in view of (4.2).

The following lemma introduces a useful bound for the expectation of (4.1).

Lemma 4.2: For the a posteriori probabilities $p(A_r \mid Y_k)$, $r = 1, 2, \cdots, N$, of the finite dimensional estimator on $\mathcal{A}$ of Section III, $r \neq j$, the following bound applies:

$$E_{Y_k|a} \{p(A_r \mid Y_k) \mid A_j\} \leqslant I_{rj}^k C_{rj} \qquad (4.3)$$

$$I_{rj}^k = 2^{k/2} |T_{A_r}(k)|^{1/4} |T_{A_j}(k)|^{1/4} |T_{A_r}(k) + T_{A_j}(k)|^{-1/2} \qquad (4.4)$$

$$C_{rj} = p_a^{1/2}(A_r) / p_a^{1/2}(A_j). \qquad (4.5)$$

Proof: See [3] and [5]. The bounding technique is analogous to that used in [18].   Q.E.D.

Remark: In a temporary digression, observe that the probability-of-error bounds of [14] can be related to the bound of Lemma 4.2. Consider the decision scheme for deciding which is the true parameter value out of the set $\mathcal{A}$ by choosing that member of $\mathcal{A}$ corresponding to the maximum a posteriori probability. The probability of error (i.e., of making a decision that the true parameter value is A_r when in fact it is A_j) can be bounded above, using the Markov inequality [32], by the quantity $I_{rj}^k C_{rj}$ of Lemma 4.2 [3]. This bound is in agreement with the work of Lainiotis [14].

The exponential convergence rate required by Lemma 4.1 for the bound of Lemma 4.2 remains to be established. To do this the bound must be analyzed and two possible techniques are available here.

The first method is reported in [3] and [5] and involves taking an LDU factorization of the output covariance matrix of the system $T_{A_r}(k)$. Bounds are obtained that require knowledge of only a finite part of the system memory to be calculated. These so-called finite memory bounds are useful in establishing that the bound of Lemma 4.2 displays an exponential convergence rate to zero. The identifiability of $\mathfrak{R}_a$ (definition 3.2) is necessary and sufficient for exponential convergence.

A second method, which is used in later sections, is to invoke some theorems relating to the asymptotic behavior of the eigenvalue distribution of Toeplitz matrices [30], [33], [34]. The main result is now stated.

Lemma 4.3: (Grenander and Szego [33].) For the $(k \times k)$ Toeplitz output covariance matrix $T_A(k)$, $A \in \mathfrak{R}_a$, of equation (3.5) with eigenvalues $\tau_{k,s}$, $s = 0, 1, \cdots, k - 1$, and for any function $F(x)$ continuous on $[m_A, M_A]$ where m_A and M_A are the bounds on the output spectral density, as defined in Section III,

$$\lim_{k \to \infty} k^{-1} \sum_{s=0}^{k-1} F(\tau_{k,s}) = (2\pi)^{-1} \int_0^{2\pi} F[f_A(\omega)] \, d\omega. \quad (4.6)$$

Using several results which are given in [30] and follow from Lemma 4.3, the required exponential convergence rate can again be established.

The significance of Lemma 4.3 is that it allows a transfer between the time domain and the frequency domain to illuminate the asymptotic behavior of the bound of Lemma 4.2.

In [3], the two approaches to analyzing the bound of Lemma 4.2 are reconciled, but more work could be done on this aspect.

In the remainder of the paper, we dispense with the somewhat restrictive assumption that $\mathfrak{R}_a$ be a finite set and investigate the behavior of the finite-dimensional adaptive state estimators of Section III in a more interesting environment.

V. THE INFORMATION FUNCTION

Consider the binary decision problem of choosing between hypotheses H_s and H_r on the basis of data Y_k, where H_s: Y_k is from the statistical population with probability density $p(Y_k | A_s)$, H_r is defined likewise, and $A_s, A_r \in \mathfrak{R}_a$. The information function $J_k(A_s, A_r)$ is conceived [35, ch. 1] as a measure for the discrimination in favor of H_s against H_r given Y_k under H_s.

Definition 5.1: The information function is defined for $A_s, A_r \in \mathfrak{R}_a$ by

$$J_k(A_s, A_r) = E_{Y_k|a} \left\{ \ln \frac{p(Y_k | A_s)}{p(Y_k | A_r)} \middle| A_s \right\}. \quad (5.1)$$

As we are concerned with zero-mean multivariate normal populations, we have [35, ch. 9]

$$J_k(A_s, A_r) = \tfrac{1}{2} \{ \ln [|T_{A_r}(k)| \, |T_{A_s}(k)|^{-1}]$$

$$+ \operatorname{tr} [T_{A_s}(k) \, T_{A_r}^{-1}(k)] - k \}. \quad (5.2)$$

The asymptotic per sample information function $\bar{J}(A_s, A_r)$, for $A_s, A_r \in \mathfrak{R}_a$, is defined by

$$\bar{J}(A_s, A_r) = \lim_{k \to \infty} k^{-1} J_k(A_s, A_r). \quad (5.3)$$

Theorem 5.1: The asymptotic per sample information function $\bar{J}(A_s, A_r)$ can be expressed as

$$\bar{J}(A_s, A_r) = \tfrac{1}{2} \left\{ (2\pi)^{-1} \int_0^{2\pi} \ln [f_{A_r}(\omega) f_{A_s}^{-1}(\omega)] \, d\omega \right.$$

$$\left. + (2\pi)^{-1} \int_0^{2\pi} f_{A_s}(\omega) f_{A_r}^{-1}(\omega) \, d\omega - 1 \right\} \quad (5.4)$$

or as the alternate expression

$$\bar{J}(A_s, A_r) = \tfrac{1}{2} \left\{ \ln d_{A_r} - \ln d_{A_s} + (2\pi i)^{-1} \, d_{A_s} d_{A_r}^{-1} \right.$$

$$\left. \cdot \oint U_{A_s}(z) \, U_{A_r}^{-1}(z) \, U_{A_s}(z^{-1}) \, U_{A_r}^{-1}(z^{-1}) z^{-1} \, dz - 1 \right\} \quad (5.5)$$

where $\oint$ denotes integration around the unit circle.

Proof: The proof is given in [3] and follows from the asymptotic behavior of the eigenvalue distribution of Toeplitz matrices as discussed in [30], [33], and [34]. Q.E.D.

This asymptotic per sample formula for Kullback's information measure does not appear to be in the literature. An asymptotic expression is given in [36] but is much more complicated than (5.4) and (5.5). The properties of the integral in (5.5) are discussed at length in [11] and the efficient machine calculation of such integrals is discussed in [28]. For low-order cases, hand calculations are possible using Cauchy's residue theorem.

In [35, p. 26], Fisher's information measure is related to Kullback's information measure. In [3], we relate the asymptotic per sample formulas (5.4), (5.5) to the asymptotic per sample formula for Fisher's information measure $\bar{F}_A$, which is discussed in recent literature concerning the frequency-domain synthesis of optimal inputs for system identification [37].

The following lemma is related to [11, theorem 1] and [35, theorem 3.1].

Lemma 5.1: For parameters A_1, $A_2 \in \mathfrak{R}_a \subset \mathfrak{R}_i \cap \mathfrak{R}_s$, if $A_1 \neq A_2$ then $\bar{J}(A_1, A_2) > 0$, and $\bar{J}(A_1, A_2) = 0$ if and only if $A_1 = A_2$ or, equivalently, $Z_{A_1} = Z_{A_2}$.

Proof: See [3].

Remarks:

1) An important point about the information function, which is implied by the relationships in [35, p. 28], is that, for sufficiently small ΔA, $\bar{J}(A, A + \Delta A) = \bar{J}(A + \Delta A, A) = \tfrac{1}{2} \Delta A' \bar{F}_A \Delta A$. For $\mathfrak{R}_a \subset \mathfrak{R}_i \cap \mathfrak{R}_s$, we have $\bar{F}_A > 0$. For $A_1, A_2 \in R^p$, $\sqrt{(A_1 - A_2)' C (A_1 - A_2)}$, $C > 0$, enjoys the usual metric properties of topology. Thus $\sqrt{\bar{J}(\cdot, \cdot)}$ can be regarded having the properties of a metric locally (but not globally).

2) Functions related to the information function, viz., the divergence and Bhattacharyya distance, have been used in detection, pattern recognition, and signal design as "distance" measures between probability distributions [38], [15]. This type of function is discussed in [18] to investigate the convergence of Bayes' estimates for independently and identically distributed samples and in [17] to develop bounds on identification error.

VI. DECISION REGIONS IN THE PARAMETER SPACE $\mathfrak{R}_a$

The following definition introduces an obviously desirable restriction on the finite estimator set $\mathfrak{A}$.

Assumption (A1): With $A_r, A_s \in \mathfrak{A}$, $\bar{J}(A_s, A_r) > 0$, for $r \neq s$.

With the aid of the information function, it is possible to define the member of $\mathfrak{A}$ nearest A for each $A \in \mathfrak{R}_a$. Also de-

cision regions R_j can be defined as consisting of all points $A \in \mathfrak{C}$ such that A_j is the member of $\mathfrak{C}$ nearest to A for $j = 1, 2, \cdots, N$.

Definition 6.1: $A_j \in \mathfrak{C}$ is nearest to $A \in \mathfrak{R}_a$ if and only if $\bar{J}(A, A_j) < \bar{J}(A, A_r), r = 1, 2, \cdots, N, r \neq j$.

Definition 6.2: The decision region $R_j \subset \mathfrak{R}_a$ is the set of all points $A \in \mathfrak{R}_a$ for which $A_j \in \mathfrak{C}$ is nearest to A.

Of course, the decision regions have boundaries, and it is clear that the nearest $A_j \in \mathfrak{C}$ to $A \in \mathfrak{R}_a$ is not unique when A is on the boundary of more than one decision region.

VII. PERFORMANCE BOUNDS II

In this section, the performance of finite dimensional adaptive estimators are investigated for the case where the true parameter lies outside the finite set $\mathfrak{C}$. The results in Section IV are a special case of those in this section.

The following lemma is a necessary preliminary.

Lemma 7.1: For $\lambda \in [0, 1]$ and either

a) $M_{A_r}^{-1} - m_{A_j}^{-1} \geqslant 0$ or b) $\lambda < -M_A^{-1} [M_{A_r}^{-1} - m_{A_j}^{-1}]^{-1}$

$$T_{rj}^\lambda(k) = T_{rj}^\lambda = T_A^{-1} - \lambda T_{A_j}^{-1} + \lambda T_{A_r}^{-1} > 0 \qquad (7.1)$$

for all k where T is the $(k \times k)$ output covariance matrix of (3.5) and m and M are the bounds on the output spectral density as defined in Section III.

Proof: See [3]. The proof follows from the fact that the eigenvalues of output covariance T_A lie between m_A and M_A [30]. Q.E.D.

The subset of $\lambda \in [0, 1]$ for which $T_{rj}^\lambda > 0$ is denoted Π'.

Lemma 7.2: For the *a posteriori* probabilities $p(A_r|Y_k)$, $r = 1, 2, \cdots, N$, of the finite dimensional estimator on $\mathfrak{C}$ of Section III the following bound applies for $\lambda \in \Pi'$ and $A \in R_j$, $r \neq j$:

$$E_{Y_k|a} \{p(A_r| Y_k)|A\} \leqslant I_{rj}^\lambda(k) c_{rj}^\lambda \qquad (7.2)$$

$$I_{rj}^\lambda(k) = |T_{A_r}|^{-\lambda/2} |T_{A_j}|^{\lambda/2} |T_A|^{-1/2} |T_{rj}^\lambda|^{-1/2} \qquad (7.3)$$

$$c_{rj}^\lambda = p_a^\lambda(A_r)/p_a^\lambda(A_j). \qquad (7.4)$$

Proof: See [3]. The bounding technique is analogous to that used in [18]. Q.E.D.

Theorem 7.1: An asymptotic expression for the bound of Lemma 7.2 as the number of data points increase is given by

$$\bar{I}_{rj}^\lambda \triangleq \lim_{k \to \infty} [I_{rj}^\lambda(k)]^{1/k}$$

$$= \exp\left[-\frac{\lambda}{2}(2\pi)^{-1} \int_0^{2\pi} \ln f_{A_r}(\omega) d\omega + \frac{\lambda}{2}(2\pi)^{-1}\right.$$

$$\cdot \int_0^{2\pi} \ln f_{A_j}(\omega) d\omega - \frac{1}{2}(2\pi)^{-1} \int_0^{2\pi} \ln f_A(\omega) d\omega$$

$$\left. - \frac{1}{2}(2\pi)^{-1} \int_0^{2\pi} \ln \{f_A^{-1}(\omega) - \lambda f_{A_j}^{-1}(\omega) + \lambda f_{A_r}^{-1}(\omega)\} d\omega\right].$$

$$(7.5)$$

Proof: See [3]. The proof exploits the asymptotic behavior of the eigenvalue distribution of Toeplitz matrices as discussed in [30], [33], and [34]. Q.E.D.

Theorem 7.2: For $A \in R_j$, where $\bar{J}(A, A_r) > \bar{J}(A, A_j), r \neq j$, and some nonzero $\lambda \in \Pi', \bar{I}_{rj} < 1$. Thus $p(A_r| Y_k), r \neq j$, converges almost surely to zero.

Proof: See [3]. The argument used to show that $\bar{I}_{rj}^\lambda < 1$ is similar to one used in [18]. The fact that $\bar{I}_{rj}^\lambda < 1$ implies, by the definition of limit, that there exists a Q such that, for all $q > Q$,

$$I_{rj}^\lambda(q) < [\rho_{rj}^\lambda]^q$$

where $\rho_{rj}^\lambda < 1$. Using Lemma 4.1 gives the desired result. Q.E.D.

Remarks

1) When the true parameter value A lies on a decision boundary, the analysis techniques of this paper cannot completely resolve the convergence question. This question is discussed further in Section VIII. The ambiguity problem occurs only on decision boundaries, which are a set of measure zero in $\mathfrak{R}_a$. On such boundaries more than one member of A can be regarded as nearest the true parameter value in the sense of definition 6.1. In this case, it can be shown that all other *a posteriori* probabilities of $\mathfrak{C}$ converge almost surely to zero.

2) If a decision scheme such as that outlined in the remark following Lemma 4.2 is considered for the case where A lies outside $\mathfrak{C}$, then probability-of-error bounds can be obtained from Lemma 7.2 using the Markov inequality [32].

3) Solving for the λ that minimizes $\bar{I}_{rj}^\lambda$ subject to the constraints of Lemma 7.1 appears too difficult without the aid of a computer.

4) It may be possible to weaken the conditions of Lemma 7.1 so the matrix T_{rj}^λ is positive definite.

5) It can be seen that calculating integrals of the form

$$I = (2\pi)^{-1} \int_0^{2\pi} \ln \{f_1(\omega) + f_2(\omega)\} d\omega$$

$$= (2\pi i)^{-1} \oint \ln [U_1(z^{-1}) U_1(z) d_1 + U_2(z^{-1}) U_2(z) d_2] \frac{dz}{z}$$

$$(7.6)$$

is required to evaluate the asymptotic formulas for the bounds of this paper. To evaluate I, the first step is to perform a spectral factorization of $U_1(z^{-1}) U_1(z) d_1 + U_2(z^{-1}) U_2(z) d_2$, using standard techniques such as polynomial zero solving algorithms [39], to yield

$$I = (2\pi i)^{-1} \oint \ln \left[g \frac{p(z^{-1}) p(z)}{q(z^{-1}) q(z)}\right] \frac{dz}{z} \qquad (7.7)$$

where $p(z)$ and $q(z)$ are monic polynomials with zeros inside the unit circle. Then the integral can be evaluated as

$$I = \ln g. \qquad (7.8)$$

Alternatively, numerical methods could be used for the calculation of such integrals [40].

VIII. COMPUTATIONAL ASPECTS AND TIGHTNESS OF THE BOUNDS

Bounds are given in Section VII which have several uses in parameter estimation and adaptive estimation for dynamic systems. Methods of analyzing and calculating the bounds are also discussed in Sections VII and IV. However, to carry out a complete analysis of the bounds for all possible true parameter values A would, in most practical problems, be an enormous task since integration of the bounds over the parameter space

would be necessary in order to obtain a single performance index. Fortunately, this colossal task is not really necessary. Once several simple examples are studied, it is easier to appreciate the fact that the behavior of the adaptive estimator can be understood from knowledge of the Kullback information measure and that design studies can be done using this information function. Working with the Kullback information measure, given in Section V, is much simpler than working with the bounds themselves.

Several examples have been worked in [3], [4], and [5] to illustrate the tightness of the bounds and the role of the information measure. In [5], simulation results are given for an example involving second-order dynamic models. The results in [4] are for a different class of signal model and allow a comparison of our techniques with earlier work summarized in [12]. The simplicity of the model in [4] also facilitates understanding of the theory. In [3], the asymptotic per sample information measure is plotted for several simple dynamic linear models, one of which displays multiple minima in the information measure. The information measure for the pulse frequency modulation example considered in [4] also has multiple minima.

The experimental work cited above suggests that the bounds presented here are quite tight in cases of practical interest.

When the true parameter value is on the boundary between two decision regions, the finite dimensional estimator cannot decide between the two members of the finite set concerned. We stress that this ambiguity problem is a limitation of the estimator itself and not of the bounds.

IX. APPLICATIONS

The implications of the preceding analysis results for applications are discussed under two subheadings.

A. Approximating High-Order Systems by Low-Order Systems

In order that construction of the adaptive state estimator of Section III be feasible, it is necessary to assume that the unknown parameter vector of the system considered comes from a finite set of *a priori* values. Fortunately, many practical problems may be represented or adequately approximated by such a model. For example, in a control problem such as in antiaircraft systems, it may be necessary to estimate certain states of the target such as position and velocity from a data record. In forming such estimates, the dynamic behavior of the target could be taken into account. The finite set of *a priori* parameter values could represent different targets performing different maneuvers. There may be two approximations which have entered the model at this stage. The true targets may be very high-order systems which have been approximated by low-order systems. Also, not all the known targets may have been included in the finite set of possibilities used to design the adaptive estimator. The advantage with the analysis of this paper is that the nature and implications for the learning system of such approximations can be precisely investigated. In our analysis, it is assumed for convenience that all signal models are of the same dimension n. However, nonminimal representations are permissible and thus systems of all orders up to n are effectively included. As discussed in the previous section, it can be seen in light of the analysis of this paper that for the purposes of investigating a particular adaptive-estimator design the Kullback information measure, defined in Section V, should be used. It is much easier to work with than the bounds themselves, and yet it defines the

various decision regions and gives a guide to the convergence rate expected.

B. Two-Step Parameter Estimation

It is proposed in [3] that parameter estimation for a wide range of stochastic signal models can be achieved using decision methods in a first stage to yield an approximate estimate and estimation techniques in a second stage to refine this estimate. The decision method involves examining the finite set of *a posteriori* probabilities, introduced in Section III, in order to decide in which of several possible regions of the parameter space the true parameter lies. In the second stage, estimation techniques such as gradient search methods could be used.

One motivation for suggesting this two-step scheme is the link between the ideas of the maximum likelihood (ML) method for identification and the decision method suggested above. In the ML approach [11], an estimate $\hat{A}_{ml}$ is obtained by minimizing a certain function $\hat{V}(A_{ml})$. In the limit, as the number of data points increase, $V(\hat{A}_{ml})$, given $a = A$, converges almost surely to our function $J(\hat{A}_{ml}, A)$ under the given assumptions of this paper [11]. As pointed out in [11], if the likelihood function is to be maximized using an algorithm based on gradients there is always the possibility that the algorithm may converge to a local extremum. In [3], [6], decision methods for parameter estimation that are aimed at overcoming this difficulty are proposed. The details of the two-step scheme are not given here. In a somewhat different context, the two-step idea is investigated for pulse frequency modulation systems in [3].

X. DISCUSSION

For discrete-time linear dynamic stochastic systems, an adaptive state estimate can be calculated by an appropriately weighted summation of conditional estimates, which are formed by a bank of Kalman filters. The weighting coefficients are determined by relatively simple nonlinear operations on the observed data.

In this paper, the performance of such adaptive estimators is studied to provide some insight into the design of these estimators. The Kullback information measure is a useful tool for such design considerations.

Areas which have not been covered in this paper are the case of time-varying systems and the vector-measurements case. In the time-varying case, the finite memory bounds of [3] and [5] are useful. The extension to vector measurements is straightforward.

REFERENCES

[1] J. M. Mendel and D. L. Gieseking, "Bibliography on the linear-quadratic-Gaussian problem," *IEEE Trans. Automat. Contr.*, vol. AC-16, Dec. 1971.

[2] a) D. G. Lainiotis, "Optimal adaptive estimation: Structure and parameter adaption," *IEEE Trans. Automat. Contr.*, vol. AC-16, Apr. 1971.
b) ——, "Joint detection, estimation, and system identification," *Inform. Control J.*, vol. 19, no. 8, pp. 75–92, Aug. 1971.
c) ——, "Adaptive pattern recognition: A state variable approach," in *Advances in Pattern Recognition*, M. Watanabe, Ed. New York: Academic Press, May 1972.
d) D. G. Lainiotis and S. K. Park, "On joint detection, estimation and system identification: Discrete data case," *Int. J. Control*, vol. 17, no. 3, pp. 609–633, Mar. 1973.
e) D. G. Lainiotis, "Partitioned estimation algorithms, I: Nonlinear estimation," *J. Inform. Sciences*, vol. 7, no. 3, pp. 203–255, 1974.
f) ——, "Optimal adaptive estimation: Structure and parameter adaptation," in *Proc. IEEE Symp. Adaptive Processes*, Nov. 1969; also Electronics Research Center, Univ. of Texas, Austin, TX,

Tech. Rep. no. 74, Sept. 1969.

[3] R. M. Hawkes, "Performance analysis of Bayesian parameter estimators," Ph.D. dissertation, Univ. of Newcastle, Newcastle, New South Wales, Dec. 1975.

[4] R. M. Hawkes and J. B. Moore, "Upper bound on mean-square error for Bayesian parameter estimators," Univ. of Newcastle, Newcastle, New South Wales, Tech. Rep. EE7514, May 1975. Also to appear in IEEE Trans. Inform. Theory.

[5] ——, "Performance of Bayesian parameter estimators for linear signal models," Univ. of Newcastle, Newcastle, New South Wales, Tech. Rep. EE7410, July 1974. Also to appear in IEEE Trans. Automat. Contr.

[6] ——, "Decision methods in dynamic system identification," Invited Paper, in 14th Symp. Adaptive Processes, Houston, TX, Dec. 1975.

[7] ——, "Performance analysis of adaptive estimators," in Proc. Sixth Symp. Nonlinear Estimation and its Applications, San Diego, CA, Sept. 1975.

[8] S. Zacks, The Theory of Statistical Inference. New York: Wiley, 1971.

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

[10] K. J. Astrom and T. Bohlin, "Numerical identification of linear dynamical systems from normal operating records," presented at the IFAC Symp. Theory of Self-Adaptive Systems, Teddington, England, 1965; also in Theory of Self-Adaptive Control Systems, P. H. Hammond, Ed. New York: Plenum, 1966.

[11] K. J. Astrom and T. Soderstrom, "Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model," IEEE Trans. Automat. Contr., vol. AC-19, Dec. 1974.

[12] L. P. Seidman, "Performance limitations and error calculations for parameter estimation," Proc. IEEE, vol. 58, May 1970.

[13] A. S. Gilman and I. B. Rhodes, "Cone-bounded nonlinearities, mean-square bounds," IEEE Trans. Automat. Contr., vol. AC-18, June 1973.

[14] D. G. Lainiotis and S. K. Park, "Probability of error bounds," IEEE Trans. Syst., Man, Cybern., Apr. 1971.

[15] D. G. Lainiotis, "On general relationship between estimation, detection, and the Bhattacharyya coefficient," IEEE Trans. Inform. Theory, vol. IT-15, July 1969.

[16] ——, "A class of upper bounds on probability of error for multihypotheses pattern recognition," IEEE Trans. Inform. Theory, vol. IT-15, Nov. 1969.

[17] E. Tse, "Bounds for identification error and a quantitative measure of identifiability," in IEEE Conf. Decision and Control, Nov. 1974.

[18] L. A. Liporace, "Variance of Bayes estimates," IEEE Trans. Inform. Theory, vol. IT-17, Nov. 1971.

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

[20] D. T. Magill, "Optimal adaptive estimation of sampled stochastic processes," IEEE Trans. Automat. Contr., vol. AC-10, Oct. 1965.

[21] a) 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, Apr. 1969.
b) C. G. Hilborn and D. G. Lainiotis, "Optimal adaptive filter realization for sampled stochastic processes with an unknown pa-

rameter," IEEE Trans. Automat. Contr., vol. AC-14, Dec. 1969.

[22] J. T. H. Lo, "Finite-dimensional senser orbits and optimal nonlinear filtering," IEEE Trans. Inform. Theory, vol. IT-18, Sept. 1972.

[23] A. V. Cameron, "Control and estimation of linear systems with non-Gaussian a priori distributions," in Proc. 2nd Annu. Allerton Conf. System Science, Oct. 1968.

[24] T. Kailath, "The innovations approach to detection and estimation," Proc. IEEE, vol. 58, 1970.

[25] B. D. O. Anderson and T. Kailath, "The choice of signal-process models in Kalman-Bucy filtering," J. Math. Anal. Appl., vol. 35, 1971.

[26] D. Q. Mayne, "A canonical model for identification of multivariable linear systems," IEEE Trans. Automat. Contr., vol. AC-17, Oct. 1972.

[27] L. H. Son and B. D. O. Anderson, "Design of Kalman filters using signal model output statistics," Proc. Inst. Elec. Eng., vol. 120, no. 2, Feb. 1973.

[28] K. J. Astrom, Introduction to Stochastic Control Theory. New York: Academic Press, 1970.

[29] S. Barnett and C. Storey, Matrix Methods in Stability Theory. New York: Barnes and Noble, Inc., 1970.

[30] R. M. Gray, "Toeplitz and circulant matrices: A review," Stanford Electronics Lab., Stanford Univ., Stanford, CA, Tech. Rep. 6502-1, June 1971.

[31] A. H. Jazwinski, Stochastic Processes and Filtering Theory. New York: Academic Press, 1970.

[32] E. Wong, Stochastic Processes in Information and Dynamical Systems. New York: McGraw-Hill, 1971.

[33] U. Grenander and G. Szegö, Toeplitz Forms and Their Applications. Berkeley, CA.: Univ. of California Press, 1958.

[34] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," IEEE Trans. Inform. Theory, Vol. IT-18, Nov. 1972.

[35] S. Kullback, Information Theory and Statistics. New York: Wiley, 1959.

[36] ——, "An application of information theory to multivariate analysis, II," Annals Mathematical Statistics, vol. 27, pp. 122–145, 1956.

[37] M. B. Zarrop, R. L. Payne, and G. C. Goodwin, "Experiment design for time series analysis: The multivariate case," Univ. of Newcastle, Newcastle, New South Wales, Tech. Rep. EE7506, May 1975.

[38] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," IEEE Trans. Commun. Technol., vol. COM-15, Feb. 1967.

[39] B. D. O. Anderson, K. L. Hitz, and N. D. Diem, "Recursive algorithm for spectral factorization," IEEE Trans. Circuits and Syst., vol. CAS-21, Nov. 1974.

[40] S. O. Rice, "Efficient evaluation of integrals of analytic functions by the trapezoidal rule," Bell Syst. Tech. J., vol. 52, pp. 707–722, May–June 1973.

[41] D. G. Lainiotis, "Estimation: A brief survey," J. Inform. Sciences, vol. 7, no. 3, pp. 197–202, 1974.

[42] S. Fralick, "Learning to recognize patterns without a teacher," Stanford Electronics Labs, Stanford Univ., Stanford, CA, Tech. Rep. SEL-65-011, Mar. 1965.

[43] S. K. Park and D. G. Lainiotis, "Monte-Carlo study of the optimal nonlinear estimator: Linear systems with non-gaussian initial state," Int. J. Control, vol. 16, no. 6, pp. 1029–1040, 1972.