ADAPTIVE ESTIMATION USING PARALLEL PROCESSING TECHNIQUES*

by

Peter K.S.Tam†

John B. Moore†

Technical Report EE7411

ISBN 0 7259 0122

Revised August 1974

* Work supported by the Australian Research Grants Committee.

† Department of Electrical Engineering,
University of Newcastle
New South Wales, 2308
Australia

ABSTRACT

A parallel processing technique for adaptive estimation is investigated. In particular, for the case of unknown system or statistical parameters denoted by the vector θ belonging to a finite set $\{\theta_1,\theta_2,\ldots,\theta_N\}$, the maximum likelihood θ is determined and denoted $\hat{\theta}$, and the minimum mean square error state estimate conditioned on this $\hat{\theta}$, namely $\hat{x}(t|t,\hat{\theta})$ is taken to be the state estimate. Using this approach new estimators are derived which require less computational effort and have less limitations than previous adaptive estimators using parallel processing techniques described in the literature. Results for the case of time-varying unknown parameters are also derived.

An example is included of state estimation for a known signal model but with unknown noise statistics. The filter banks are constrained to be time-invariant and so only appropimate maximum likelihood parameter estimation is achieved.

1. Introduction

The Kalman-Bucy filter [1,2] for the estimation of the states of a linear dynamical system requires an exact knowledge of the system parameters and noise covariances. We consider the adaptive estimation problem of estimating the states when the dynamical and/or statistical model is specified up to a set of unknown parameters, denoted by the vector θ.

Parallel processing techniques have been applied by a number of authors [3-5] to the adaptive estimation problem and in fact adaptive estimators requiring one hundred or so Kalman filters can be implemented using mini-computers. In essence, the standard Bayesian approach to the adaptive estimation problem is as follows [5]. Assuming that the unknown parameter vector θ is discrete or suitably quantized to a finite number of grid points $\{\theta_1,\ldots,\theta_N\}$, with known or assumed a priori probability for each θ_i, the conditional mean estimator includes a bank of N Kalman filters where the i^{th} filter is a standard Kalman filter designed on the assumption that $\theta=\theta_i$. The filter bank is driven by the noisy signal measurements. The conditional mean state estimate is given by a weighted sum of the states of the Kalman filters. The weighting coefficient of the state of the i^{th} Kalman filter is the a posteriori probability of θ_i, which can be updated recursively using the noisy signal measurements and the state of the i^{th} Kalman filter.

Unfortunately, for systems with continuous-time measurements, the above Bayesian approach has the drawback that, first, the measurement noise covariance R has to be known, and second, the unknown parameter vector θ has to be time-invariant.

For the special case of systems with discrete measurements, parallel processing estimation techniques have been developed when the measurement noise covariance is unknown, at the expense of either complexity or loss of optimality [4], [6]. In the practically important direction of reducing the complexity of the adaptive estimator, Alspach and Abiri [6,7] obtain time-invariant state estimators for time-invariant systems with unknown noise covariances by considering a grid of possible time-invariant Kalman gains directly rather than a more involved grid of possible noise covariances in the unknown parameter space. The results can be extended to cases when the noise covariances are time-varying quantities [8]. However, for the estimator of [6] to be close to optimal, it is required that the Kalman filters have reached steady-state and that the number of measurements received be large. Thus, during the transient periods, the estimator operates suboptimally.

Some parallel processing estimation techniques are also available for discrete-time systems when θ is time-varying. Using a combination of digital and analogue techniques, optimal adaptive estimators for the case when the unknown parameter is a scalar Markov sequence of known statistics have been developed [9]. By approximating the a posteriori density of the state vector with a Gaussian probability density, Ackerson and Fu [10] derived a suboptimal estimator for the particular time-varying unknown parameter case in which the input noise or the measurement noise comes from a group of white Gaussian noise sources, which act one at a time, with the transition from one noise source to another being described by a discrete Markov transition matrix.

It is clear from the above survey of existing adaptive estimators using parallel processing techniques that it would be

worthwhile to investigate any adaptive estimation approach using parallel processing techniques suitable for both continuous-time and discrete-time problems, and which can handle unknown time-varying parameters and unknown measurement noise parameters with but a small addition in complexity over the case of time-invariant unknown parameters and known measurement noise covariance. In this paper, such an approach is investigated. This approach can be briefly described as follows.

An unknown parameter vector θ is defined in such a way that there is one and only one Kalman filter corresponding to a particular grid point θ_i in the θ space. The states of the Kalman filters are denoted $\hat{x}(t|t,\theta_i)$. By comparing the relative magnitudes of the likelihood functions of the θ_i's evaluated using the measurement data and $\hat{x}(t|t,\theta_i)$ for all i, the most likely θ_i at time t, denoted $\hat{\theta}(t)$, is determined. The state $\hat{x}[t|t,\hat{\theta}(t)]$ is taken as the estimate of the signal model state $x(t)$. Note that since the likelihood functions of the unknown parameters are used instead of their exact a posteriori probabilities, it is not necessary to assign a priori probabilities to the different θ_i's

The advantages of using the approach just outlined can be summarised as follows. First, for continuous-time systems, in contrast to the standard Bayesian approach, our approach does not require that the measurement noise covariance be known exactly. Second, for the discrete-time case, our estimators require less computational effort than alternative known estimators using parallel processing techniques. Third, for both the continuous-time case and the discrete-time case, the various results can be extended in a simple manner to give useful adaptive estimators for the case of time-varying θ.

The above advantages are of course useless unless the estimators using the above approach perform satisfactorily. Monte Carlo simulations show that they work very well in minimizing the mean square estimation error when compared to the more complex adaptive estimator for discrete-time problems described in [8].

An outline of the subsequent sections of this paper is as follows. In section II, we present the discrete-time results for time-invariant unknown parameters. In section III, the results of section II are extended to the case of time-varying unknown parameters. In section IV, the results of sections II, III are extended to systems with continuous-time measurements.

II. DISCRETE-TIME RESULTS FOR TIME INVARIANT UNKNOWN PARAMETERS

In this section, we first review appropriate discrete-time optimal adaptive estimator results from [3]-[5] and then we apply these results to achieve an alternative adaptive estimator which in most applications is simpler and thus more attractive. A simulation example is given to demonstrate the performance characteristics of the alternative estimator.

Consider the system

$$x(t_{k+1}) = \phi(t_{k+1},t_k|\psi)x(t_k) + G(t_k|\psi)\, u(t_k) \tag{1}$$

$$y(t_{k+1}) = H(t_{k+1}|\psi)\, x(t_{k+1}) \tag{2}$$

$$z(t_{k+1}) = y(t_{k+1}) + v(t_{k+1}) \tag{3}$$

where $u(\cdot)$, $v(\cdot)$ are independent zero-mean Gaussian white noise sequences with covariance matrices I and R respectively. The positive definite matrix R will be assumed known in some instances, and not known in others. The initial state $x(t_0)$ is a Gaussian random vector with mean $x_0(\psi)$ and variance $P(t_0|t_0,\psi)$ and is independent of $u(\cdot)$ and $v(\cdot)$. The entities $\phi(t_{k+1},t_k|\psi)$, $G(t_k|\psi)$, $H(t_{k+1}|\psi)$, $x_0(\psi)$, $P(t_0|t_0,\psi)$ are completely specified by the parameter vector ψ, i.e. they are known if ψ is known, possibly unknown if ψ is not known. This formulation effectively allows an unknown input noise covariance (although we have assumed $E\{u(t_k)u'(t_k)\} = I$) through the intervention of $G(t_k|\psi)$ in (1).

If ψ and R are specified, the conditional-mean state estimate $\hat{x}(t_k|t_k,\psi,R)$ is of course given from the familiar Kalman filter equations:

$$\hat{x}(t_{k+1}|t_{k+1},\psi,R) = \hat{x}(t_{k+1}|t_k,\psi,R) + K(t_{k+1}|\psi,R)\,\tilde{z}(t_{k+1}|\psi,R) \quad (4)$$

$$\hat{x}(t_{k+1}|t_k,\psi,R) = \phi(t_{k+1}|t_k,\psi)\,\hat{x}(t_k|t_k,\psi,R),\ \hat{x}(t_0|t_0,\psi,R) = \hat{x}_0(\psi) \quad (5)$$

$$\tilde{z}(t_{k+1}|\psi,R) = z(t_{k+1}) - H(t_{k+1}|\psi)\,\hat{x}(t_{k+1}|t_k,\psi,R) \quad (6)$$

$$K(t_{k+1}|\psi,R) = P(t_{k+1}|t_k,\psi,R)\,H'(t_{k+1}|\psi,R)\,P_z^{-1}(t_{k+1}|t_k,\psi,R) \quad (7)$$

$$P_z(t_{k+1}|t_k,\psi,R) = H(t_{k+1}|\psi)P(t_{k+1}|t_k,\psi,R)H'(t_{k+1}|\psi) + R \quad (8)$$

$$P(t_{k+1}|t_k,\psi,R) = \phi(t_{k+1},t_k|\psi)P(t_k|t_k,\psi,R)\phi'(t_{k+1},t_k|\psi) + G(t_k|\psi)G'(t_k|\psi) \quad (9)$$

$$P(t_{k+1}|t_{k+1},\psi,R) = P(t_{k+1}|t_k,\psi,R) - K(t_{k+1}|\psi,R)P_z(t_{k+1}|t_k,\psi,R)K'(t_{k+1}|\psi,R) \quad (10)$$

Next we consider that R is known but ψ is unknown. Assuming that Ψ, the space of admissible values of ψ, is discrete or suitably quantized to a finite number of grid points $\{\psi_1,\ldots,\psi_p\}$, with known or assumed a priori probability $p(\psi_i|t_0,R)$ for each ψ_i, the conditional mean state estimate $\hat{x}(t_k|t_k,R)$ is given by [3]-[5]:

$$\hat{x}(t_k|t_k,R) = \sum_{i=1}^{p} \hat{x}(t_k|t_k,\psi_i,R)p(\psi_i|t_k,R) \quad (11)$$

$$p(\psi_i|t_k,R) = c|P_z(t_k|t_{k-1},\psi_i,R)|^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}||\tilde{z}(t_k|\psi_i,R)||^2_{P_z^{-1}(t_k|t_{k-1},\psi_i,R)}\}\,p(\psi_i|t_{k-1},R) \quad (12)$$

where c is a normalizing constant independent of ψ_i so that $\sum_{i=1}^{p} p(\psi_i|t_k,R) = 1$.

Now consider the more general case where R is unknown with a continuous range $\mathcal{R}$ of admissible values. An obvious approach for extending the above results is to first approximate $\mathcal{R}$ with a suitable finite set of quantized points $\{R_1,\ldots,R_M\}$.

Then with known or assumed a priori probability $p(\psi_i, R_j | t_0)$ for each (ψ_i, R_j), the conditional mean state estimate would be given from [4]

$$\hat{x}(t_k|t_k) = \sum_{i=1}^{P} \sum_{j=1}^{M} \hat{x}(t_k|t_k,\psi_i,R_j)\, p(\psi_i,R_j|t_k) \tag{13}$$

$$p(\psi_i,R_j|t_k) = c' P_z(t_k|t_{k-1},\psi_i,R_j)\,|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}||\tilde{z}(t_k|\psi_i,R_j)||^2 P_z^{-1}(t_k|t_{k-1},\psi_i,R_j)\}$$
$$p(\psi_i,R_j|t_{k-1}) \tag{14}$$

where c' is a normalizing constant independent of ψ_i and R_j.

To implement equation (13), $P \times M$ Kalman filters are required compared to the requirement of P Kalman filters for the implementation of equation (11). This represents a large increase in the number of parallel processing units when R is unknown.

As a first step towards reducing the filter complexity, we replace the above discretization of the Cartesian product $\Psi \times R$ with the discretization of an alternative set which we now describe. Through equations (7) - (10), the product space $\Psi \times R$ defines a (continuous) space Θ of hypotheses of possible Kalman filter configurations, specified by the quadruples $\{K(\cdot|\psi,R), \phi(\cdot|\psi), \hat{x}_0(\psi), H(\cdot|\psi)\}$. (In many cases, the various time-varying Kalman gains have to be approximated by time-invariant gains to simplify calculations). For each θ in Θ, we have the following equations in lieu of equations (4)-(6).

$$\hat{x}(t_{k+1}|t_{k+1},\theta) = \hat{x}(t_{k+1}|t_k,\theta) + K(t_{k+1}|\theta)\, \tilde{z}(t_{k+1}|\theta) \tag{4}'$$

$$\hat{x}(t_{k+1}|t_k,\theta) = \phi(t_{k+1}, t_k|\theta)\hat{x}(t_k|t_k,\theta) \quad \hat{x}(t_0|t_0,\theta) = \hat{x}_0(\theta) \tag{5}'$$

$$\tilde{z}(t_{k+1}|\theta) = z(t_{k+1}) - H(t_{k+1}|\theta)\, \hat{x}(t_{k+1}|t_k,\theta) \tag{6}'$$

Note that while equations (4)-(6) can be re-written in terms of θ as above, it is not possible to do so with equations (7)-(10). Thus the mapping $\Psi \times R \to \Theta$ incurs some loss of information as the specification of θ alone determines only the Kalman filter configuration but not the error covariance of the state estimate associated with that Kalman filter.

It turns out that if Θ is to be approximated by a set of N grid points $\{\theta_1,\ldots,\theta_N\}$, then in many cases N can be much smaller than $P \times M$. As an illustration of the above statement, we point out that in [6], a numerical example for state estimation of a linear time-invariant system with unknown noise covariances is given whereby 2500 grid points in the space of unknown noise covariances can be adequately replaced by 100 grid points in the space of possible steady-state Kalman gains.

Using the set of grid points $\{\theta_1,\ldots,\theta_N\}$, the conditional mean state estimate is now given from

$$\hat{x}(t_k|t_k) = \sum_{i=1}^{N} \hat{x}(t_k|t_k,\theta_i)p(\theta_i|t_k) \tag{15}$$

where $p(\theta_i|t_k)$ is the a posteriori probability that $\theta = \theta_i$.

A crude approach for calculating $p(\theta_i|t_k)$ is to first observe that the joint a posteriori probability that $\theta = \theta_i$ and $R = R_j$ can be updated recursively using the following equation

$$p(\theta_i,R_j|t_k) = c''|P_z(t_k|t_{k-1},\theta_i,R_j)|^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}||\tilde{z}(t_k|\theta_i)||^2_{P_z^{-1}(t_k|t_{k-1},\theta_i,R_j)}\} \; p(\theta_i,R_j|t_{k-1}) \tag{16}$$

where c'' is a normalizing constant independent of θ_i,R_j; and $P_z(t_k|t_{k-1},\theta_i,R_j) = E\{\tilde{z}(t_k|\theta_i)\tilde{z}'(t_k|\theta_i)|R_j\}$ is related to $[K(t_k|\theta_i),R_j]$ through the following relation (c.f. equations (7),(8)), which holds when

$p(\theta_i, R_j | t_0) \neq 0$

$$P_z(t_k | t_{k-1}, \theta_i, R_j) = [I - H(t_k | \theta_i) K(t_k | \theta_i)] P_{k-1, j} \tag{17}$$

Then by approximating the Cartesian product space $\Theta \times R$ with a set of grid points $\{(\theta_i, R_j) | i=1,\ldots,N;\ j=1,\ldots,M\}$ and applying the theorem of total probability, we can express $p(\theta_i | t_k)$ as a marginal probability, viz,

$$p(\theta_i | t_k) = \sum_{j=1}^{M} p(\theta_i, R_j | t_k) \tag{18}$$

This approach of using equations (16)-(18) however requires considerable on-line calculations as we have to store and update M separate quantities to obtain a single $p(\theta_i | t_k)$. Moreover, a lot of off-line calculations may be required to obtain the a priori probabilities $p(\theta_i, R_j | t_0)$.

An alternative approach is to find an approximate expression for $p(\theta_i | t_k)$. For example, using the techniques described in [6], it can be shown that for scalar measurements and assuming that for each θ_i the range of admissible values of $P_z(t_k | t_{k-1}, \theta_i, R)$ is the interval $(0, \infty)$, then we have the following approximate expression

$$p(\theta_i | t_k) = c\left[\prod_{j=1}^{k} z^2(t_j | \theta_i)\right]^{\frac{2-k}{2}} \tag{19}$$

where c is a normalizing constant independent of θ_i and k is a large even number (greater than about 1000).

We now propose a simple estimation scheme to yield a suboptimal minimum mean square error estimate for the case when R is unknown with a

continuous range of admissible values. Our approach is based on the observation (shown below) that under a simplifying assumption on the range of admissible values of R, then given the measurements up to time t_k, it is relatively easy to find $[\hat{\theta}(t_k), \hat{R}(t_k)]$, which we shall denote simply as $(\hat{\theta},\hat{R})$, that maximizes the likelihood function $p(Z_k|\theta,R)$ of (θ,R), where $Z_k = \{z(t_1),\ldots,z(t_k)\}$. Using the estimate $\hat{\theta}$ so obtained, our proposed simple estimator is one yielding the estimates $\hat{x}(t_k|t_k,\hat{\theta})$. (Note that $\hat{R}$ is not required other than as a step to find $\hat{\theta}$).

Using the fact that $E\{\tilde{z}(t_j|\theta_i)\tilde{z}'(t_j|\theta_i)|R\} = [I-H(t_j|\theta_i)K(t_j|\theta_i)]^{-1}R$, the likelihood function $p(Z_k|\theta_i,R)$, which is the conditional probability density of Z_k as a function of θ_i and R, can be established as

$$p(Z_k|\theta_i,R) = \prod_{j=1}^{k} \{|2\pi R|^{-\frac{1}{2}}|I-H(t_j|\theta_i)K(t_j|\theta_i)|^{\frac{1}{2}}\exp\{-\tfrac{1}{2}\sum_{j=1}^{k}||\tilde{z}(t_j|\theta_i)||^2 R^{-1}[I-H(t_j|\theta_i)K(t_j|\theta_i)]\} \tag{20}$$

Now we make the simplifying assumption (quite similar to that made in [6] when the measurements are scalars) that for each θ_i, R can take any value provided it is a positive definite symmetric matrix. For each $\theta = \theta_i$ let $\hat{R}(\theta_i)$ denote the value of R which maximizes $p(Z_k|\theta_i,R)$. If $\hat{R}(\theta_i)$ exists, it can be obtained by setting $\frac{\partial p(Z_k|\theta_i,R)}{\partial R^{-1}} = 0$, which gives for $k > 1$,

$$\hat{R}(\theta_i) = \tfrac{1}{2}(\Omega+\Omega') \tag{21}$$

where

$$\Omega = \frac{1}{k}\sum_{j=1}^{k} \tilde{z}(t_j|\theta_i)\tilde{z}'(t_j|\theta_i)[I-H(t_j|\theta_i)K(t_j|\theta_i)]' \tag{22}$$

Substituting the above value of $\hat{R}(\theta_i)$ into the expression for $p(Z_k|\theta_i,R)$ gives

$$p[Z_k|\theta_i,\hat{R}(\theta_i)] = c \prod_{j=1}^{k} \{|\hat{R}(\theta_i)|^{-1} |I-H(t_j|\theta_i)K(t_j|\theta_i)|\}^{\frac{1}{2}} \qquad (23)$$

where c is a constant independent of $[\theta_i,\hat{R}(\theta_i)]$.

Equations (21) and (22), though useful in a number of cases, have to be used with caution because they apply only when $\hat{R}(\theta_i)$ so obtained is positive definite. (Hence they are not applicable for $k=1$). In order to ensure a positive definite $\hat{R}(\theta_i)$, we need further assumption on the range of R for each $\theta = \theta_i$. Let us assume that for $\theta = \theta_i$, R is of the form $R = \alpha_i R_i$ where R_i is a known positive definite matrix and α_i is an unknown scalar which can take any value in $(0,\infty)$. Then setting $\frac{\partial p(Z_k|\theta_i,R)}{\partial \alpha_i} = 0$ gives

$$\hat{R}(\theta_i) = \frac{R_i}{k} \{\sum_{j=1}^{k} ||z(t_j|\theta_i)||^2 R_i^{-1}[I - H(t_j|\theta_i)K(t_j|\theta_i)]\} \qquad (24)$$

This equation (24) gives a positive definite $\hat{R}(\theta_i)$ for all $k=1,2,3,\ldots$ Substituting this $\hat{R}(\theta_i)$ into the expression for $p(Z_k|\theta_i,R)$ again gives equation (23).

Note that $\hat{R}(\theta_i)$ and $p[Z_k|\theta_i,\hat{R}(\theta_i)]$ can be updated recursively, and that $\max_i \{p[Z_k|\theta_i,\hat{R}(\theta_i)]\}$ is in fact the global maximum of $p(Z_k|\theta_i,R)$ with respect to i and R.

Our proposed estimator consists of a bank of N Kalman Filters. The conditional innovation processes of the Kalman filters are used to update recursively $p[Z_k|\theta_i,\hat{R}(\theta_i)]$, by means of equation (23). Then $\hat{\theta}$ is chosen to maximize $p[Z_k|\theta_i,\hat{R}(\theta_i)]$, and $\hat{x}(t_k|t_k,\hat{\theta})$ is chosen as the state estimate.

Such an estimator is sketched diagrammatically in Figure 1.

Under certain circumstances, the computations can be further simplified. For example, if the measurements are scalars, then as maybe seen from equation (23), the problem of maximizing $p(Z_k|\theta,R)$ with respect to (θ,R) reduces to the simpler problem of minimizing $\{\sum_{j=1}^{k} z^2(t_j|\theta_i)|i=1,\ldots N\}$

Application to state estimators in unknown noise environment.

It is interesting to compare the performance of our proposed estimator with the estimator proposed by Alspach and Abiri in [6] for state estimation of a linear time-invariant system when both the plant driving noise covariance Q and the measurement noise covariance R are unknown. In [6], to reduce the complexity of the filter structure, the $[Q,R]$ space is mapped to the space of Kalman gains (assumed to be time invariant). The simplifying assumption that the Kalman gains are time invariant leads to sub-optimal performance during the transient period but is still a very useful assumption to make. In order to provide conditions for comparison purposes we will also make this assumption for our proposed estimator. Thus we consider the grid $\{\theta_1,\ldots,\theta_N\}$ to be the grid of possible time-invariant (sub-optimal) Kalman gains.

We now consider the state estimation of a first-order system where the noise covariances are unknown. The system equations are

$$x(t_{k+1}) = 0.8x(t_k) + G\,u(t_k) \tag{25}$$

$$z(t_k) = x(t_k) + J\,v(t_k) \tag{26}$$

where $u(\cdot)$, $v(\cdot)$ are zero-mean independent white Gaussian sequences with unity covariances and G, J are 'unknown' quantities.

The unknown Kalman gain for the above system can take a value between 0 and 1. We consider the following simple Kalman gain space consisting of

N grid points $\{\theta_1,\ldots,\theta_N\}$ as the unknown parameter space. The ith Kalman filter (that is the filter conditioned on $\theta = \theta_i$) has Kalman gain $K(\theta_i) = \theta_i$ where $\theta_i = (i-.5)/N$.

For the simulation results, the signal generating system is first run until steady-state can be assumed. Then measurements $z(k)$ are processed for estimation of $x(t_k)$ for $k=0,1,\ldots,21$. Typical performances (the mean-square-error of the state estimates at each t_k) are shown in figures 2,3 using 300 sample paths and 10 grid points in the Kalman gain space. The results show that our estimator actually performs slightly better than that of [6].

To conclude this section we comment that not only does our scheme give better performance than that of [6], at least for the somewhat random examples chosen, but it is certainly considerably less complicated to implement.

III DISCRETE-TIME RESULTS FOR TIME-VARYING UNKNOWN PARAMETERS

In this section, we show that the estimation scheme described in the last section can be extended in a simple but suboptimal manner to handle unknown time-varying (but not slowly varying) parameters and unknown measurement noise parameters. The addition in complexity, which involves feedback from the estimator output to the individual Kalman filters, is very small relative to that required for an optimal minimum square error estimation scheme.

For our problem, equations (4)′ - (6)′ generalize to

$$\hat{x}[t_{k+1}|t_{k+1},\Theta(t_{k+1})] = \hat{x}[t_{k+1}|t_k,\Theta(t_k)] + K[t_{k+1},\Theta(t_{k+1})]\tilde{z}[t_{k+1}|\Theta(t_k)] \quad (27)$$

$$\hat{x}[t_{k+1}|t_k,\Theta(t_k)] = \Phi[t_{k+1},t_k|\theta(t_k)]\hat{x}[t_k|t_k,\Theta(t_k)], \quad \hat{x}[t_0|t_0,\Theta(t_0)] = \hat{x}[\Theta(t_0)] \quad (28)$$

$$\hat{z}[t_{k+1}|\Theta(t_{k+1})] = z(t_{k+1}) - H[t_{k+1},\Theta(t_{k+1})]\hat{x}[t_{k+1}|t_k,\Theta(t_k)] \quad (29)$$

where $\Theta(t_k) = \{\theta(t_0),\ldots,\theta(t_k)\}$.

Here we have to assume that for $\theta=\theta_i$ the measurement noise covariance $R(t_k)$ for $v(t_k)$ is of the form $R(t_k) = \alpha(t_k)R_i(t_k)$ where $\alpha(t_k)$ is an unknown scalar and $R_i(t_k)$ is a known matrix. We assume that $\alpha(t_k)$, $\theta(t_k)$, $\theta(t_{k-1})$ are independent of each other and that $\alpha(t_k)$ can take any value in $(0,\infty)$. The extimator structure that we are going to propose can be slightly modified to include the case where there is a particular range of admissible values of $\theta(t_{k-1})$ for each $\theta(t_k)$. However, simulation results so far indicate that very little improvement in performance can be gained by imposing such restrictions.

At $t=t_2$ $\max_{\theta(t_2),\alpha(t_2)} [p[Z_2|\theta(t_2),\theta(t_1),\alpha(t_2)]$ for each $\theta(t_2)$ can be found in two stages as follows. First, note that we can express $p[Z_2|\theta(t_2),\theta(t_1),\alpha(t_2)]$ in an equation similar to that of equation (20) for $p[Z_k|\theta_j,R]$. Therefore, for each of the grid points $\{[\theta_i(t_2),\theta_j(t_1)]$ $i,j=1,\ldots,N\}$, $p[Z_2|\theta(t_2),\theta(t_1),\alpha(t_2)] = \max_{\alpha(t_2)} p[Z_2|\theta(t_2),\theta(t_1),\alpha(t_2)]$ can be found via a pair of equations similar to that of equations (23) and (24). Second, $\hat{\theta}(t_0)$ is chosen such that $p[Z_2|\theta(t_2),\hat{\theta}(t_1),\hat{\alpha}(t_2)]$ $= \max_{\theta(t_1)} \{p[Z_2|\theta(t_2),\theta(t_1),\hat{\alpha}(t_2)]\}$ where $\hat{\theta}(t_0)$ can in general be different for different $\theta(t_1)$. Having thus found $\hat{\theta}(t_0)$ for each $\theta(t_1)$, the one step ahead prediction state estimate given by the Kalman filter conditioned on each $\theta(t_1)$ is then reset to $\hat{x}[t_1|t_0,\theta(t_1)]$, where again $\hat{x}[t_1|t_0,\theta(t_1)]$ can be different for different $\theta(t_1)$.

To recapitulate, based on the 'maximum likelihood' approach as described in the last section and under some simplifying assumptions, we reject, for each $\theta(t_1)$ at $t = t_1$, all the combinations of $[\theta(t_2),\alpha(t_2)]$ except the pair $[\hat{\theta}(t_0),\hat{\alpha}(t_1)]$ which maximizes the likelihood function $p[Z_1|\theta(t_1),\theta(t_0),\alpha(t_1)]$.

In general, at $t = t_k$, $\max_{\theta(t_{k-1}),\alpha(t_k)} p[Z_k|\theta(t_k),\theta(t_{k-1}),\theta(t_{k-2}),\theta(t_k),A(t_{k-1})]$ for each $\theta(t_k)$ can be found in two stages similar to those described above for $t = t_1$. Here $\hat{\Theta}(t_{k-2}) = \{\hat{\theta}(t_0),\theta(t_1),\ldots,\theta(t_{k-2})\}$ and $\hat{A}(t_{k-1}) = \{\hat{\alpha}(t_1),\ldots,\hat{\alpha}(t_{k-1})\}$ and these two sequences can in general be different for different $\theta(t_{k-1})$. Having thus obtained $\hat{\theta}(t_{k-1})$ for each $\theta(t_k)$, the one step ahead prediction state estimate given by the Kalman filter conditioned on each $\theta(t_k)$ is then reset to $\hat{x}[t_k|t_{k-1},\hat{\theta}(t_{k-1})]$.

Resetting the states of the Kalman filters as described above allows us to limit the number of Kalman filters to a finite number N although the number of grid points for $\Theta(t_k)$ grows exponentially with k.

As an illustration of the above procedure, we shall describe the derivation of the various guantities in more details for the case of scalar measurements (i.e. $R(t_k) = \alpha(t_k)$) and $H[t_k|\theta(t_k)] = H(t_k)$ independent of $\theta(t_k)$, which gives particularly simple results.

At $t = t_1$, we have

$$p[Z_1|\theta(t_1),\theta(t_0),\alpha(t_1)] = \{2\pi\alpha(t_1)\}^{-\frac{1}{2}}\{1-H(t_1)k[t_1|\theta(t_1)]\}^{\frac{1}{2}}\exp\{-\frac{\ell[t_1|\theta(t_1)]}{2\alpha(t_1)}\}$$

where (30)

$$\ell[t_1|\Theta(t_1)] = \{1-H(t_1)K[t_1|\theta(t_1)]\}\tilde{z}^2[t_1|\Theta(t_0)] \tag{31}$$

$$\tilde{z}[t_1|\upsilon(t_0)] = z(t_1) - H(t_1)\hat{x}[t_1|t_0,\theta(t_0)] \tag{32}$$

Observe that, for each $[\theta(t_1),\theta(t_0)]$, $p = p[Z_1|\theta(t_1),\theta(t_0),\alpha(t_1)]$ is a unimodel function of $\alpha(t_1)$. Setting $\frac{\partial p}{\partial\alpha(t_1)} = 0$ gives $\hat{\alpha}(t_1) = \ell[t_1|\Theta(t_1)]$ and $p[Z_1|\theta(t_1),\theta(t_0),\hat{\alpha}(t_1)] = c|\tilde{z}[t_1|\theta(t_0)]|$. Choosing $\hat{\theta}(t_0)$ to maximize $\{|\tilde{z}[t_1|\theta(t_0)]\}$, the one step ahead prediction state estimate given by the N Kalman filters arethen reset to $\hat{x}[t_1|t_0,\theta(t_0)]$. (Note that in this case $\hat{\theta}(t_0)$ is independent of $\theta(t_1)$ so that the calculations are much simplified.)

By induction, we easily obtain for each $[\theta(t_k),\theta(t_{k-1})]$

$$\hat{\alpha}(t_k) = \ell[t_k|\theta(t_k),\theta(t_{k-1}),\Theta(t_{k-2})] \tag{33}$$

$$p[Z_k|\theta(t_k),\theta(t_{k-1}),\hat{\theta}(t_{k-2}),\hat{A}(t_k)] = c'|\tilde{z}[t_k|\theta(t_{k-1}),\hat{\theta}(t_{k-2})]| \qquad (34)$$

where $\tilde{z}[t_k|\theta(t_k),\theta(t_{k-1}),\hat{\theta}(t_{k-2})] = \{1-H(t_k)K[t_k|\theta(t_k)]\}\tilde{z}^2[t_k|\theta(t_{k-1}),\theta(t_{k-2})]$ $\tilde{z}[t_k|\theta(t_{k-1}),\hat{\theta}(t_{k-2})] = z(t_k) - H(t_k)\hat{x}[t_k|t_{k-1},\theta(t_{k-1}),\hat{\theta}(t_{k-2})]$. The normalising constant c' and the sequence $\hat{\theta}(t_{k-2})$ are independent of $[\theta(t_k),\theta(t_{k-1})]$.

From equation (34), $p[Z_k|\theta(t_k),\theta(t_{k-1}),\hat{\theta}(t_{k-2}),\hat{A}(t_k)]$ is independent of $\theta(t_k)$. Thus, the problem of finding the global maximum of $p[Z_k|\theta(t_k),\theta(t_{k-1}),\hat{\theta}(t_{k-2}),\alpha(t_k),\hat{A}(t_{k-1})]$ with respect to $[\theta(t_{k-1}),\alpha(t_k)]$ reduces to the problem of finding $\min_i\{\tilde{z}^2[t_k|\theta_i(t_{k-1}),\hat{\theta}(t_{k-2})]\}$. The one step ahead prediction state estimate given by the N Kalman filters are then reset to $\hat{x}[t_k|t_{k-1},\hat{\theta}(t_{k-1})]$. The estimator structure for this problem is illustrated diagramatically in figure 4.

Example

A modification of the first-order system given by equations (25) and (26) is used to compare the performances of the estimators described in the last section and the estimator described in this section. Here we assume G,J to be uniformly distributed white noise sequences. Figure 5 shows a typical set of simulation results which demonstrate that the estimator described in this section gives superior performance when the unknown parameters are varying rapidly with time.

It might be thought that the scheme proposed in this section could be better than that of the previous section for the example of the previous section since the unknown parameter is really time-varying although slowly varying in this example. Simulations showed that when the parameter variations are slowly varying the non-resetting scheme of the previous section works best.

IV CONTINUOUS-TIME RESULTS.

The 'maximum likelihood' technique described in this paper has the advantage that even for continuous-time measurements, an estimator can be obtained which does not require the exact knowledge of the measurement noise covariance.

Time-invariant unknown parameters

Consider the system

$$\dot{x}(t) = F(t|\psi)x(t) + G(t|\psi)u(t) \qquad t \geq 0 \tag{35}$$

$$y(t) = H(t|\psi)x(t) \tag{36}$$

$$z(t) = y(t) + v(t) \tag{37}$$

where $\{u(t)\}$, $\{v(t)\}$ are independent zero mean Gaussian white noise processes with covariance matrices I, $R(t)$ respectively. The initial state $x(t_0)$ is a Gaussian random vector with mean $\hat{x}(t_0|t_0,\psi)$ and variance $P(t_0|t_0,\psi)$ and is independent of $\{u(t)\}$ and $\{v(t)\}$. The measurement noise covariance $R(t)$ is assumed to be of the form $R(t) = \alpha R_0(t)$ where $R_0(t)$ is known but α is an unknown scalar. The entities $F(t|\psi), G(t|\psi), H(t|\psi), \hat{x}(t_0|t_0,\psi), P(t_0|t_0,\psi)$ are known if the parameter vector ψ is known, possibly unknown if ψ is not known.

If ψ and α are specified, the conditional mean state estimate $\hat{x}(t|t,\psi,\alpha)$ is given from the Kalman-Bucy filter equations:

$$\dot{\hat{x}}(t|t,\psi,\alpha) = F(t|\psi)\hat{x}(t|t,\psi,\alpha) + K(t|\psi,\alpha)[z(t) - \hat{y}(t|t,\psi,\alpha)] \tag{38}$$

$$\hat{y}(t|t,\psi,\alpha) = H(t|\psi)\hat{x}(t|t,\psi,\alpha) \tag{39}$$

$$K(t|\psi,\alpha) = P(t|t,\psi,\alpha)H'(t|\psi)[\alpha R_0(t)]^{-1} \tag{40}$$

$$\dot{P}(t|t,\psi,\alpha) = F(t|\psi)P(t|t,\psi,\alpha) + P(t|t,\psi,\alpha)F'(t|\psi) - \alpha K(t|\psi,\alpha)$$
$$R_0(t)K'(t|\psi,\alpha) + G(t|\psi)G'(t|\psi) \tag{41}$$

If α is specified but ψ is unknown, then assuming Ψ, the space of admissible values of ψ, is discrete or suitably quantized to a finite number of grid points $\{\psi_1,\ldots,\psi_p\}$ with known or assumed a priori probability $p(\psi_i|t,\alpha)$ for each ψ_i, the conditional mean state estimate $\hat{x}(t|t,\alpha)$ is given by [5]:

$$\hat{x}(t|t,\alpha) = \sum_{i=1}^{p} \hat{x}(t|t,\psi_i,\alpha)p(\psi_i|t,\alpha) \tag{42}$$

$$p(\psi_i|t,\alpha) = c\exp\{\ell(\psi_i|t,\alpha)\}p(\psi_i|t_0,\alpha) \tag{43}$$

$$\ell(\psi_i|t,\alpha) = \int_{t_0}^{t} \hat{y}'(\tau|\tau,\psi_i,\alpha)R_0^{-1}(\tau)z(\tau)d\tau - \frac{1}{2}\int_{t_0}^{t} \|\hat{y}(\tau|\tau,\psi_i,\alpha)\|^2_{R_0^{-1}}d\tau$$

(44)

where c is a normalizing constant independent of ψ_i and $\ell(\psi_i|t,\alpha)$ is the log likelihood function of ψ_i given α and the measurements $\{z(\tau)|\tau\in(t_0,t)\}$.

Again, allowing for more parallel processing effort, it might appear to the reader that a natural extension of the above results for the case when α is unspecified is 'readily achieved' using the following approach. Assume A, the range of admissible values of α, can be approximated by the finite set $\{\alpha_1,\ldots,\alpha_M\}$. Then, with known or assumed a priori probability

$p(\psi_i,\alpha_j|t_0)$ for each (ψ_i,α_j), the conditional mean state estimate would be given from

$$\hat{x}(t|t) = \sum_{i=1}^{P}\sum_{j=1}^{M} \hat{x}(t|t,\psi_i,\alpha_j)p(\psi_i,\alpha_j|t) \tag{45}$$

However, further difficulty arises because there is no direct and simple technique to recursively update $p(\psi_i,\alpha_j|t)$.

We now propose a simple estimation scheme to yield a suboptimal minimum mean square error estimate for the case when α is unknown. Following the approach described in section II we shall now consider directly the space Θ of hypotheses of possible Kalman filter configurations. For each θ in Θ, the conditional mean state estimate $\hat{x}(t|t,\theta)$ is given by the following equations:

$$\dot{\hat{x}}(t|t,\theta) = F(t|\theta)\hat{x}(t|t,\theta) + K(t|\theta)[z(t)-\hat{y}(t|t,\theta)] \tag{46}$$

$$\hat{y}(t|t,\theta) = H(t|\theta)\hat{x}(t|t,\theta) \tag{47}$$

Again, we approximate Θ with a finite set of grid points $\{\theta_1,\ldots,\theta_N\}$, and we make the simplifying assumption that each θ_i allows the same range of admissible values of α. Then, for any admissible value of α, the log likelihood function $\ell(\theta_i|t,\alpha)$ of θ_i given the measurements $\{z(\tau)|\tau\in(t_0,t)\}$ can be expressed as

$$\ell(\theta_i|t,\alpha) = L(\theta_i|t)/_{\alpha} \tag{48}$$

where

$$L(\theta_i|t) = \int_t^t \hat{y}'(\tau|\tau,\theta_i)R_0^{-1}(\tau)z(\tau)d\tau - \frac{1}{2}\int_{t_0}^{t} ||\hat{y}(\tau|\tau,\theta_i)||^2_{R_0^{-1}}d\tau \qquad (49)$$

Obviously, given the measurements $\{z(\tau)|\tau \in (t_0,t)\}$, the most likely θ_i, denoted $\hat{\theta}$, is the θ_i which maximizes $L(\theta_i|t)$, which is independent of α. This observation suggests that in the event that α is not known, a very simple state estimator is one yielding the estimate $\hat{x}(t|t,\hat{\theta})$. Such an estimator is sketched diagramatically in Figure 5. Simulation results have not been obtained for this estimator since it is not anticipated that they can extablish any more than those for the discrete-time results.

Time-Varying Unknown Parameters

Here we consider obtaining estimators when the unknown parameter vector $\theta(t)$ and the measurement noise intensity α are modelled as piecewise constant functions of time with $\theta(t) = \theta(t_k)$, $\alpha(t) = \alpha(t_k)$ for $t_k \leq t < t_{k+1}$. We further assume that $\theta(t_k)$ can take one of the values $\{\theta_1,\ldots,\theta_N\}$.

The estimation scheme that we propose is as follows. At $t=t_1$, $\hat{\theta}(t_0)$ is found such that $L[\hat{\theta}(t_0)|t_1] = \max_i\{L[\theta_i(t_0)|t_1]\}$. The states of all the N Kalman filters are then reset to $\hat{x}[t_1|t_1,\hat{\theta}(t_0)]$, viz, the state of the Kalman filter conditioned on $\hat{\theta}(t_0)$. In general, at $t=t_{k+1}$, $\hat{\theta}(t_k)$ is found such that $L[\hat{\theta}(t_k)|t_{k+1},\hat{\theta}(t_{k-1})] = \max_i\{L[\theta_i(t_k)|t_{k+1},\hat{\theta}(t_{k-1})]\}$, where $\hat{\theta}(t_{k-1}) = \{\hat{\theta}(t_0),\ldots,\hat{\theta}(t_{k-1})\}$ and

$$L[\theta_i(t_k)|t,\hat{\theta}(t_{k-1})] = \int_{t_k}^{t} \hat{y}'[\tau|\tau,\hat{\theta}(t_{k-1}),\theta(t_k)]R_0^{-1}(\tau)z(\tau)d\tau - \frac{1}{2}\int_{t_k}^{t} ||\hat{y}(\tau|\tau,$$

$$\hat{\theta}(t_{k-1}),\theta(t_k))||^2_{R_0^{-1}}d\tau$$

Simulations have not been carried out for this scheme since as for the previous not too much additional information can be gained from this. However, it could be stressed that the important point here is that the ideas investigated in this paper, in contrast to the standard Bayesian approach, can in fact be applied to continuous-time systems in a straightforward manner.

REFERENCES

[1] R.E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", Trans. ASME, J. Basic Engng., Vol. 82, March 1960, pp. 34-45.

[2] R.E. Kalman and R.S. Bucy, "New Results in Linear Filtering and Prediction Theory", Trans. ASME, J. Basic Engng., sec. D Vol. 83, December 1961, pp. 95-107.

[3] D.T. Magill, "Optimal Adaptive Estimation of Sam led Stochastic Processes" IEEE Trans. Automat. Contr., Vol. AC-10. No 4 Oct 1965, pp. 215-218.

[4] 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, April 1969.

[5] D.G. Lainiotis, "Optimal Adaptive Estimation: Structure and Parameter Adaptation", IEEE Trans. Automat. Contr. Vol. AC-16, No. 2, April 1971 pp. 160-169.

[6] D.L. Alspach and A. Abiri, "Optimal Nonlinear Estimation for a Linear System with Unknown Plant and Measurement Noise Covariances", Proc. of 3rd Symposium on Nonlinear Estimation Theory and its Applications, San Diego, September 1972, pp. 11-13.

[7] D.L. Alspach, "A Parallel Processing Solution to the Adaptive Kalman Filtering Problem with Vector Measurements", Intern. J. Computers and Electrical Engng., Vol. 1 No. 1, June 1973, pp 83-94.

[8] D.L. Alspach, "A Bayesian Nonlinear Filtering Algorithm for Linear Systems with Unknown Time Varying Noise Statistics", Proc. IEEE Conference on Decision and Control incl. 12th Symposium on Adaptive Processes, December 1973, pp. 533-539.

[9] C.G. Hilborn, Jr., and D.G. Lainiotis, "Optimal Adaptive Filter Realizations for Sample Stochastic Processes with an unknown Parameter" IEEE Trans. on Automat. Contr. Vol AC-14, December 1969, pp. 767-770.

[10] G.A. Ackerson and K.S. Fu, "On State Estimation in Switching Environments" IEEE Trans. Automat. Contr. Vol AC-15, February 1970, pp. 10-16.

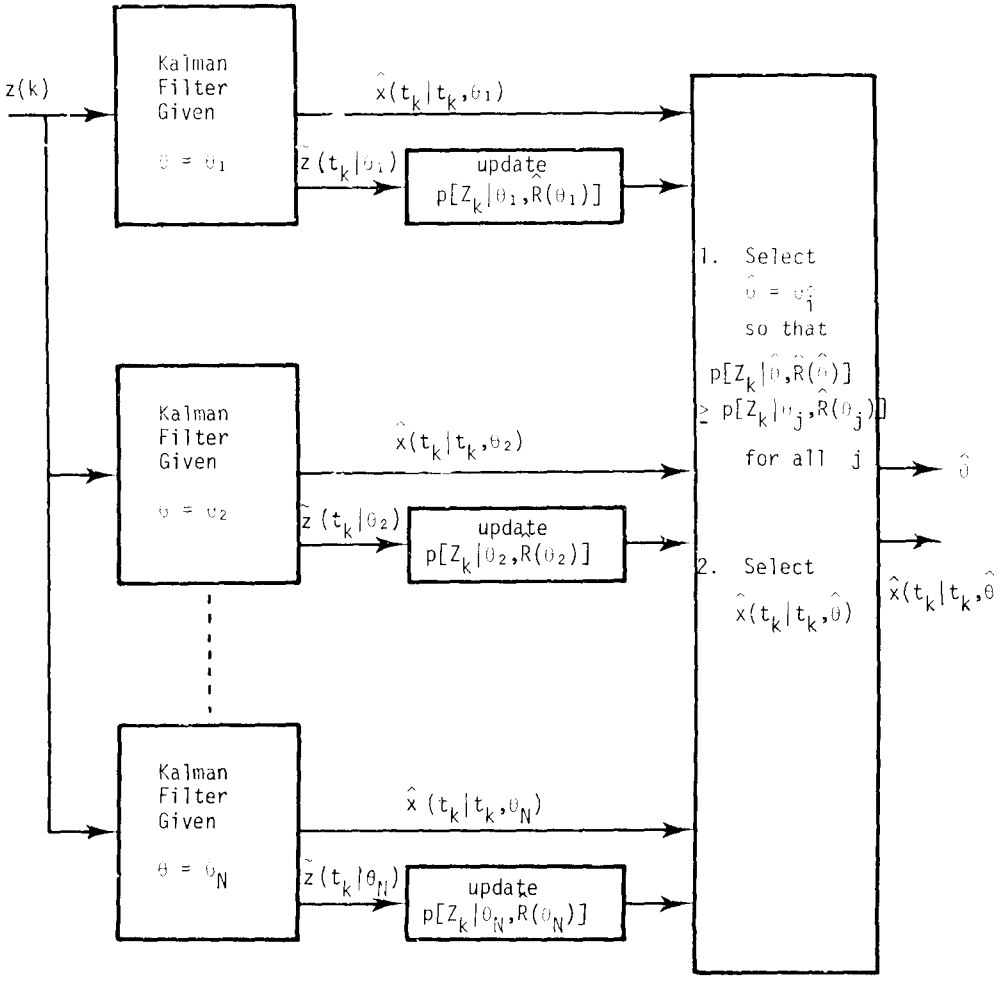

FIG. 1 DISCRETE-TIME ADAPTIVE ESTIMATOR FOR TIME-INVARIANT θ.

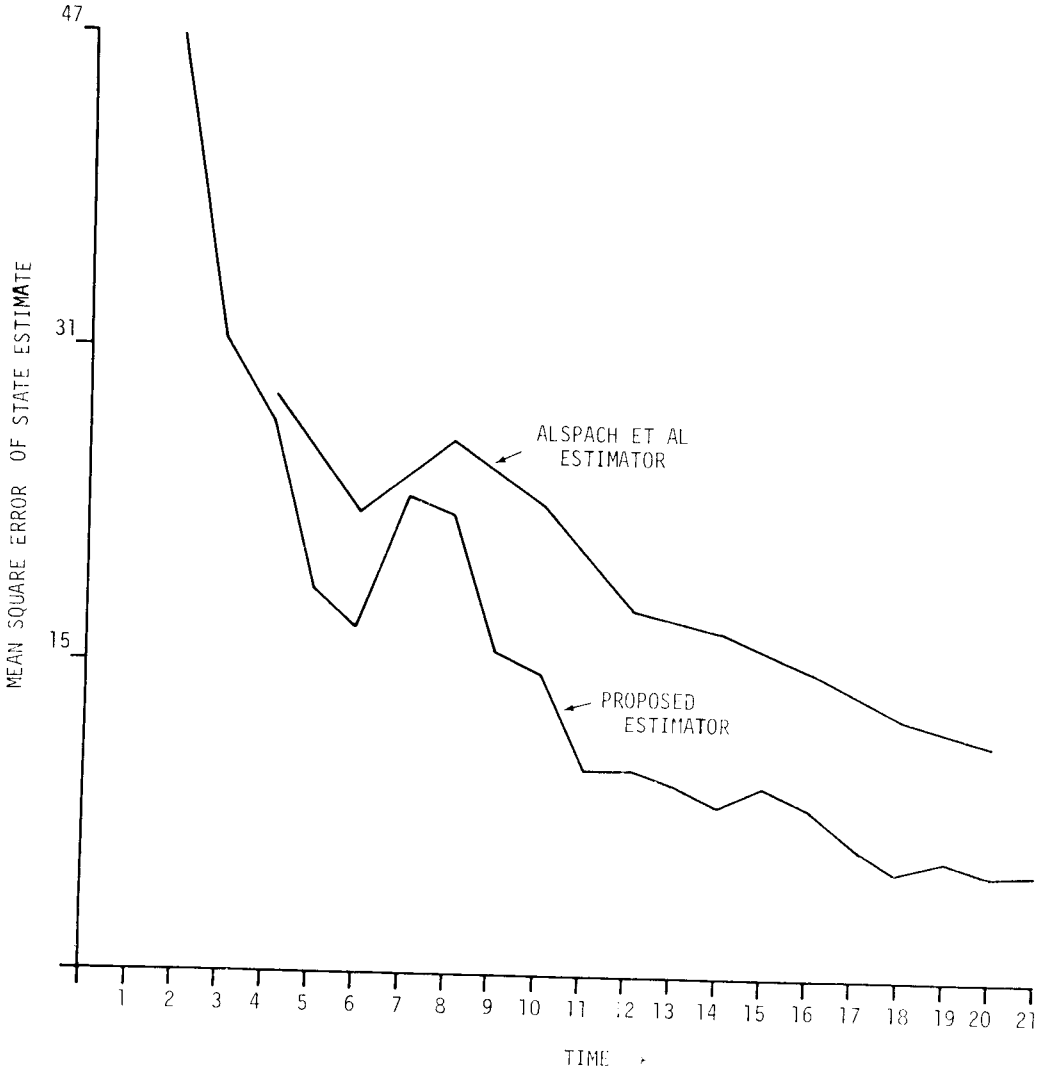

FIG. 2 MEAN SQUARE ERROR FOR STATE ESTIMATION IN HIGH NOISE (G = .6, J = 10)

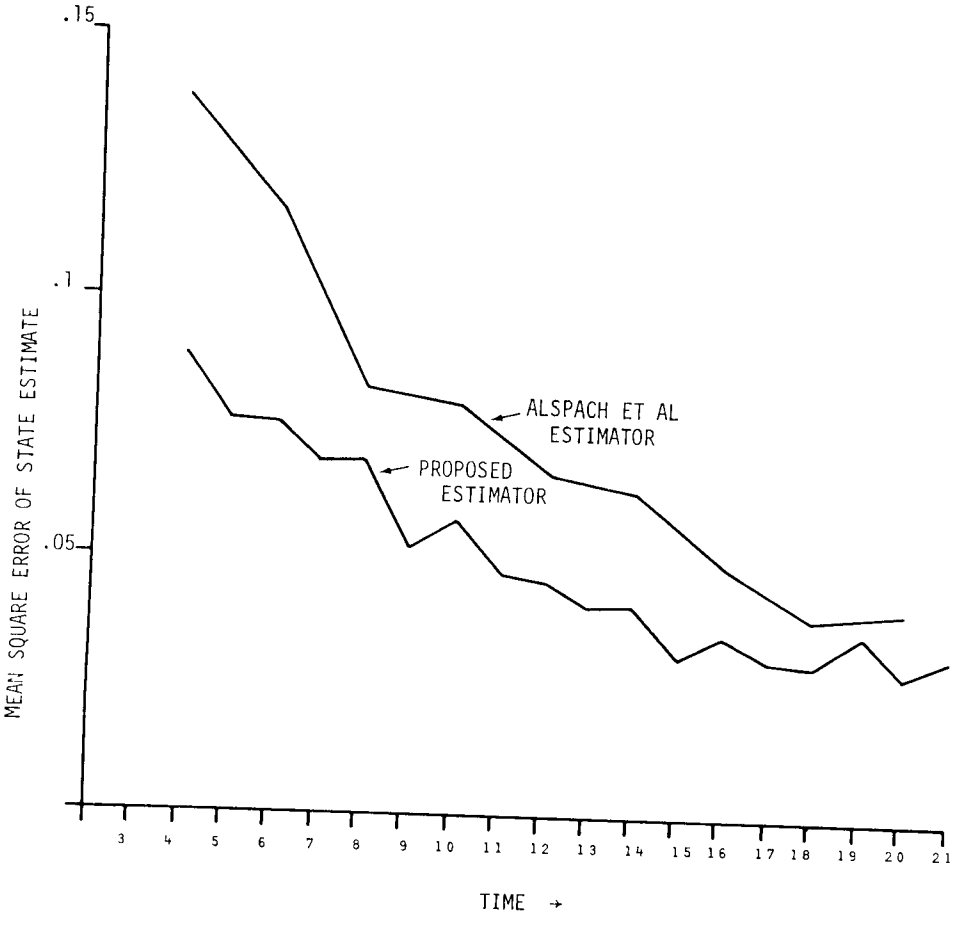

FIG. 3 MEAN SQUARE ERROR FOR STATE ESTIMATION IN LOW NOISE (G = .6, J = .1)

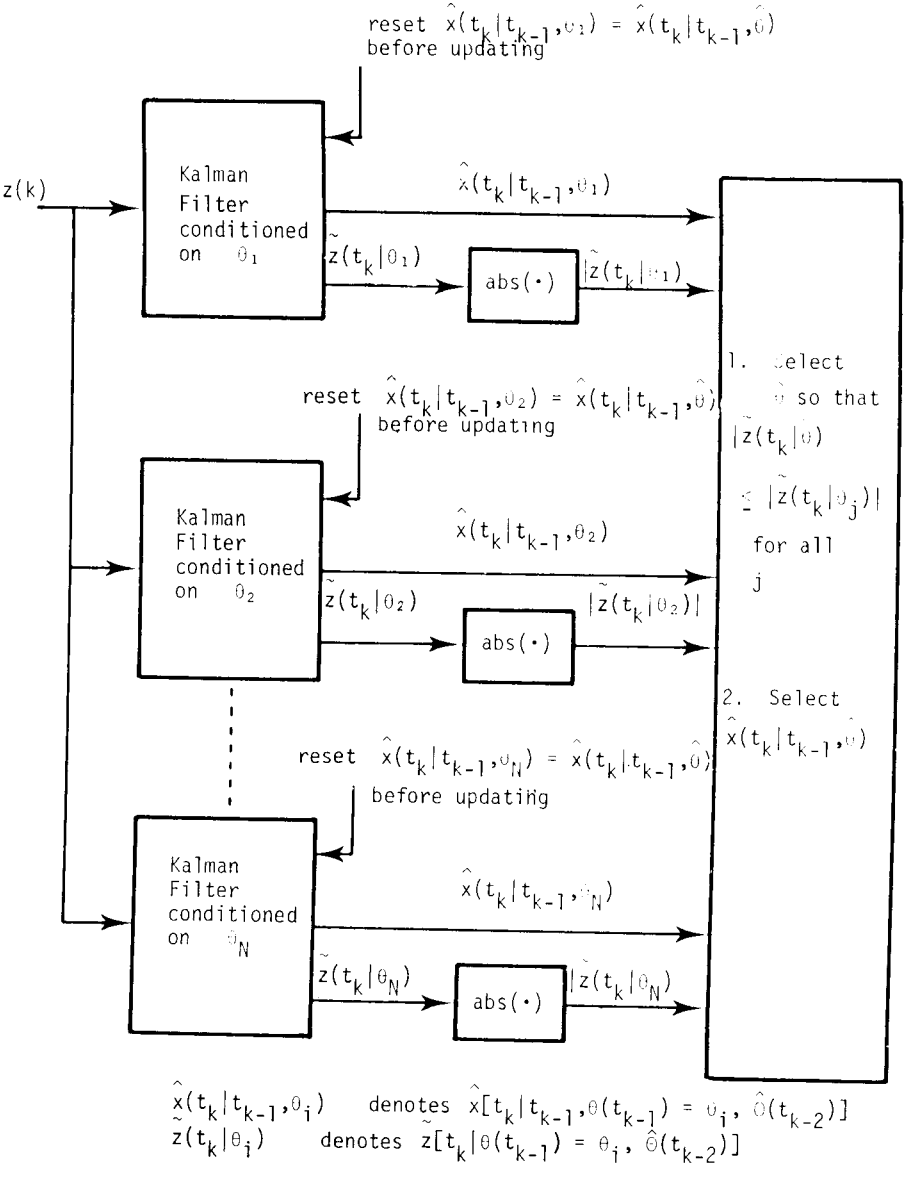

FIG. 4 DISCRETE-TIME ADAPTIVE ESTIMATOR FOR TIME-VARYING θ [ASSUMING SCALAR MEASUREMENTS AND $H[t_k|\theta(t_k)] = H(t_k)$]

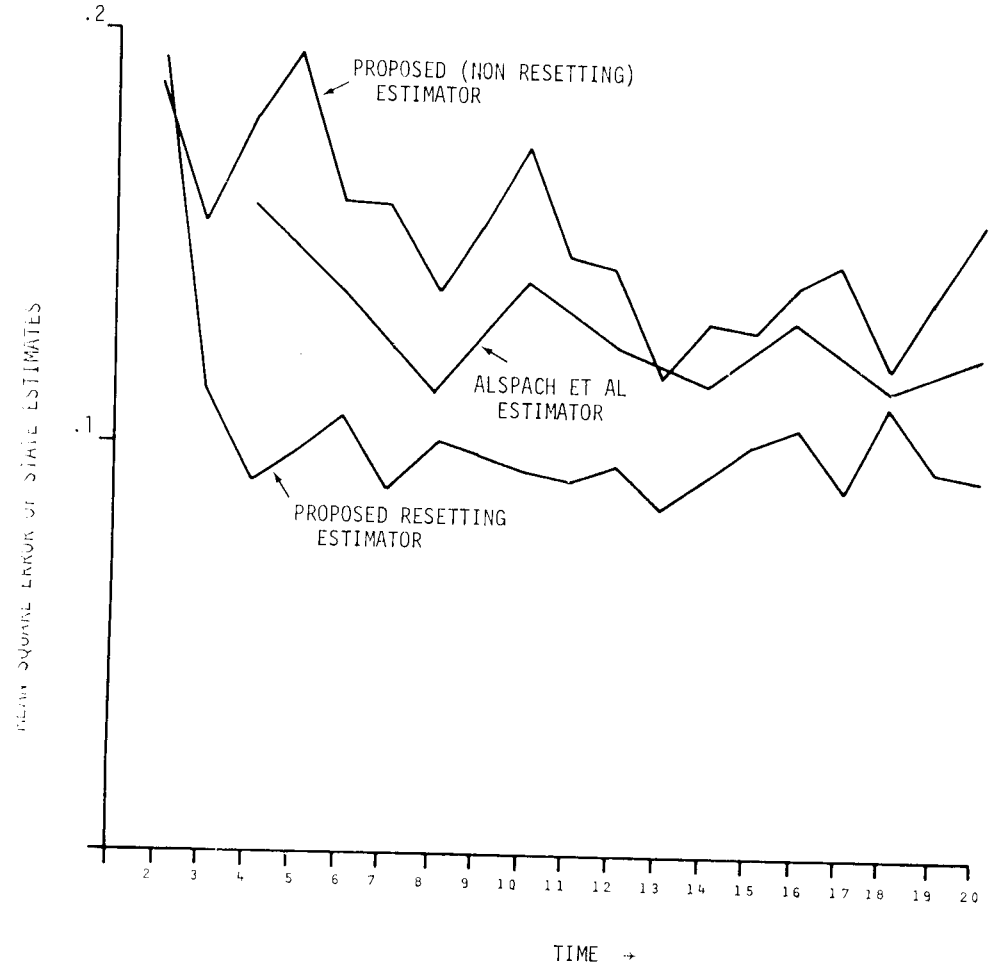

FIG. 5 MEAN SQUARE ERROR OF STATE ESTIMATES FOR TIME-VARYING NOISE COVARIANCES [G, J RANDOM VARIABLES UNIFORMLY DISTRIBUTED IN (0, 1)]

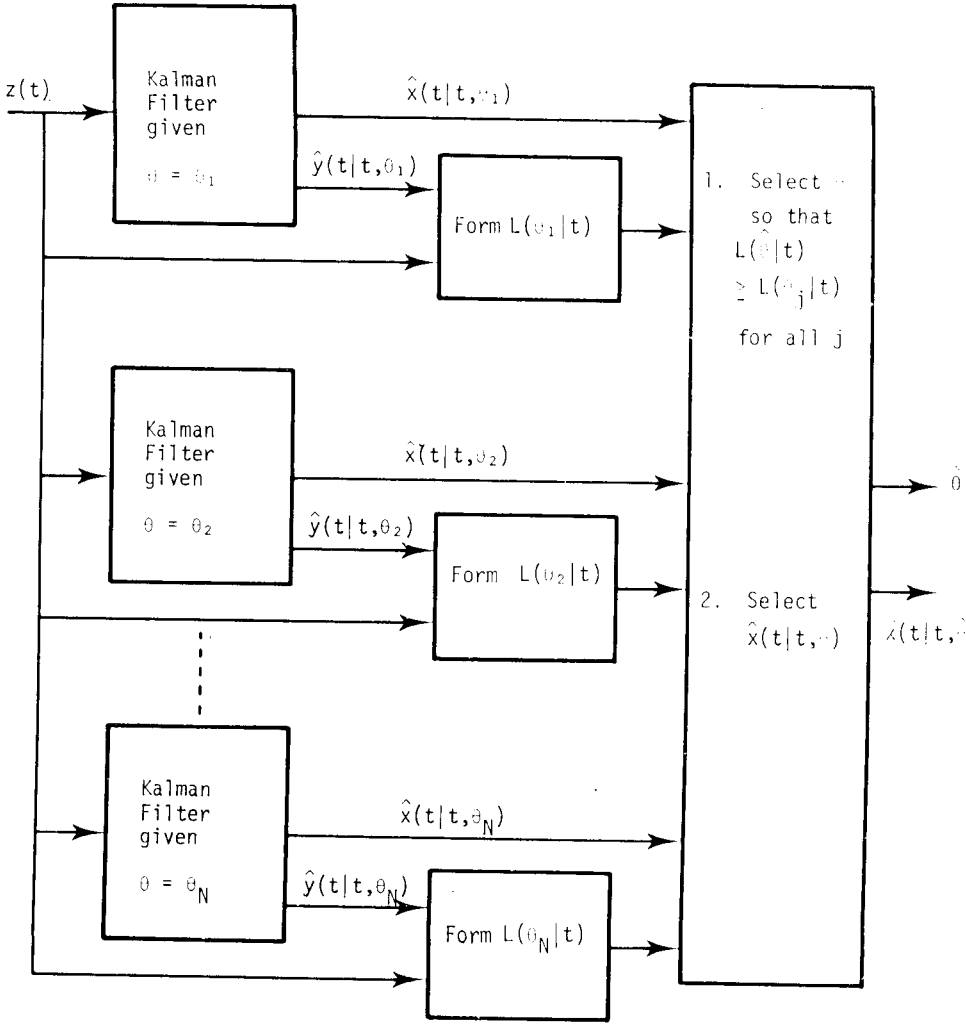

FIG. 6 CONTINUOUS-TIME ADAPTIVE ESTIMATOR FOR TIME-INVARIANT

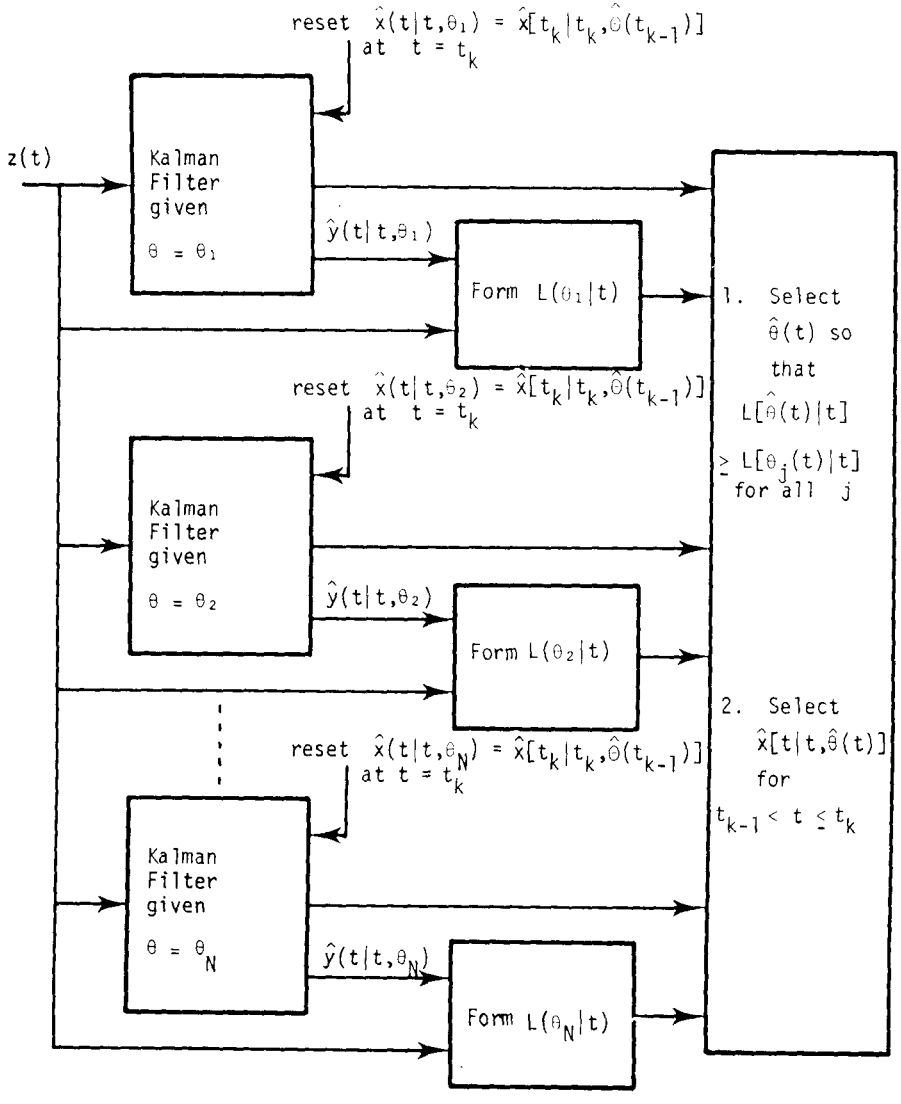

FIG. 7 CONTINUOUS-TIME ADAPTIVE ESTIMATOR FOR TIME-VARYING θ