

Finite State Machines in State Estimation for Dynamic Systems with an Nth Order Memory and Nonlinear Interference†

by KERİM DEMİRBAŞ

Department of Electrical Engineering and Computer Science (M/C 154), University of Illinois at Chicago, P.O. Box 4348, Chicago, IL 60680, U.S.A.

ABSTRACT: *The Viterbi decoding algorithm is used for the state estimation of dynamic systems having Nth order memory and nonlinear interference. The state model is a nonlinear function of the disturbance noise and discrete values of the state. The observation model is a nonlinear function of the observation noise, random interference and discrete values of the state. Some simulation results are also presented.*

1. Introduction

The original work of Kalman (1) and Kalman and Bucy (2) initiated extensive research on recursive state estimation of dynamic systems with first-order memory. As a result, many estimation schemes have been proposed. These are optimum estimation schemes for linear models with Gaussian noise, and suboptimum estimation schemes for nonlinear dynamic systems (1–7). These schemes have been also applied for practical systems (8). One of the major restrictions of these estimation schemes is that dynamic models must be linear functions of the disturbance noise and (additive) white noise (1–7).

After the work of Nahi (9), researchers also considered state estimation of dynamic models having a Markov chain (which may be considered as a special case of interference), as well as the disturbance noise and observation noise (9, 10). Further, state estimation of general dynamic models with nonlinear random interference has been recently considered by Demirbaş (11) and Demirbaş and Leondes (12). Considered dynamic models are nonlinear functions of the state, disturbance noise, interference and observation noise. Proposed estimation schemes are based upon decoding techniques of Information Theory. These schemes are superior to the classical estimation schemes, such as the Kalman filter, since they can handle state estimation of dynamic systems with nonlinear disturbance noise, interference and observation noise, whereas the classical estimation schemes, in general, cannot (11).

The estimation schemes cited above have been developed for dynamic models with first-order memory. State estimation of dynamic models with higher-order memory can be accomplished by first representing them by higher dimensional dynamic systems with first-order memory and then using the estimation schemes cited above for these higher dimensional dynamic systems. However, the increase

† This work was carried out while the author was visiting Bilkent University, Ankara, Turkey.

in the dimension increases the complexity of the implementation of state estimation.

In this paper, an estimation approach which does not require any higher dimensional model representation is presented for dynamic systems with higher-order memory and nonlinear random interference. This scheme results in memory reduction for state estimation of dynamic systems with higher-order memory and nonlinear random interference.

II. Problem Statement

Consider the discrete dynamic systems with N th order memory and interference, which are defined by

$$x(j+1) = f(j, x(j), X(j), w(j)) \quad \text{the state model,} \quad (1)$$

$$z(j) = g(j, x(j), X(j), I(j), v(j)) \quad \text{the observed model,} \quad (2)$$

with

$$X(j) \triangleq \{x(j-1), x(j-2), \dots, x(j-N+1)\}$$

where j denotes a discrete moment in time; $v(j)$ is an $l \times 1$ observation noise vector at time j with zero mean and known statistics; $I(j)$ is an $r \times 1$ random interference vector at time j with known statistics; $z(j)$ is a $p \times 1$ observation vector at time j ; $w(j)$ is an $m \times 1$ disturbance noise vector at time j with zero mean and known statistics; $x(0)$ is an $n \times 1$ random initial state vector with known statistics; $x(j)$, $j > 0$, is an $n \times 1$ state vector at time j ; $f(j, x(j), X(j), w(j))$ and $g(j, x(j), X(j), I(j), v(j))$ are either linear or nonlinear functions which define the state at time $j+1$ and observation at time j in terms of the disturbance noise, observation noise, interference and present and $N-1$ past discrete values of the state at time j . Moreover, $x(0)$, $I(p)$, $I(q)$, $w(j)$, $w(k)$, $v(l)$ and $v(m)$ are assumed to be independent for all p, q, j, k, l and m .

Our interest is to determine an estimate of the state sequence $X^K \triangleq \{x(0), x(1), \dots, x(K)\}$ by using the observation sequence $Z^K \triangleq \{z(1), z(2), \dots, z(K)\}$.

III. Estimation Approach

The state is quantized so that the state model is represented by a finite state model (or machine) and the observation model by an approximate observation model. Further, the finite state model is represented by a trellis diagram. Then, the concept of multiple composite hypothesis testing (13) and the Viterbi decoding algorithm (14, 15) are used to estimate the state.

The state and observation models of (1) and (2) are approximated by

$$x_q(j+1) = Q(f(j, x_q(j), \hat{X}(j|j), w_d(j))) \quad \text{finite state model,} \quad (3)$$

$$z(j) = g(j, x_q(j), \hat{X}(j|j), I_d(j), v(j)) \quad \text{approximate observation model} \quad (4)$$

with

$$\hat{X}(j|j) \triangleq \{\hat{x}(j-1|j), \hat{x}(j-2|j), \dots, \hat{x}(j-N+1|j)\}$$

where $w_d(j)$ is a discrete disturbance noise vector which approximates the disturbance noise $w(j)$, and the possible values of this discrete disturbance noise vector are denoted by $w_{d1}(j), w_{d2}(j), \dots$ and $w_{dm_{nj}}(j)$ (11); $I_d(j)$ is a discrete interference random vector which approximates the interference $I(j)$, and the possible values of this discrete random vector are indicated by $I_{d1}(j), I_{d2}(j), \dots$ and $I_{dr_j}(j)$; $Q(\cdot)$ is the quantizer defined in (11), which divides the n -dimensional Euclidean space into generalized rectangles (referred to as *gates*) of equal size (called *gate size*) and which then assigns to each rectangle its center; $x_q(0)$ is an initial discrete random vector approximating the initial state vector $x(0)$, and the possible values of this discrete random vector are called the initial quantization levels (or quantization levels of the state at time zero), and they are denoted by $x_{q1}(0), x_{q2}(0), \dots$, and $x_{qn_0}(0)$; $x_q(j)$ is the quantized state vector at time j whose quantization levels are denoted by $x_{q1}(j), x_{q2}(j), \dots$, and $x_{qn_j}(j)$, where the subscript n_j is the number of quantization levels at time j ; $\hat{X}(j|j)$ is the estimate of $X(j)$ given the observation sequence from time one to time j , that is Z^j . In other words, $\hat{x}(k|j)$ is the estimate of the state at time k given the observation sequence from time one to time j , except for $\hat{x}(k|j)$ which is by definition, the mean value of the initial state vector $x(0)$ for $k < 0$ or $k = j = 0$.

The gate size and the numbers of possible values of the discrete random vectors in (3) and (4) are preselected, depending upon the desired estimation accuracy with available memory for state estimation. The finite state model of (3) and the approximate observation model of (4) are better approximations of the state model of (1) and the observation model of (2) for smaller gate sizes and greater numbers of possible values of the discrete random vectors since a random vector is better approximated by a discrete random vector having a greater number of possible values, and a smaller gate size results in smaller quantization errors. However, the complexity of the proposed estimation scheme increases with smaller gate sizes or greater numbers of possible values of the discrete random vectors. Thus, one should make a compromise between the complexity of the scheme and the desired estimation accuracy with available memory when selecting the gate size and the numbers of possible values of the discrete random vectors in (3) and (4).

Denoting the quantization levels of the state at each time by nodes at a column and transitions between quantization levels by directed lines allows one to represent the finite state model by a trellis diagram (Fig. 1), which is said to be the trellis diagram of the state. The trellis diagram from time zero to time K has many paths. The quantization levels along only one of these paths can be taken by the state from time zero to time K . Thus, state estimation is to find a path through the trellis diagram by using the observation sequence from time one to time K . The quantization levels along this path will be the estimates of the state from time zero to time K . The path along which the quantization levels will become the estimates of the state is found by treating state estimation as a multiple composite hypothesis testing since the observation model contains a random interference. It can be shown (11) that the optimum testing rule which minimizes the overall error probability is to choose the path with the greatest metric, namely

$$\text{choose } P^m \text{ if } M(P^m) > M(P^j) \quad \text{for all } j \neq m \quad (5)$$

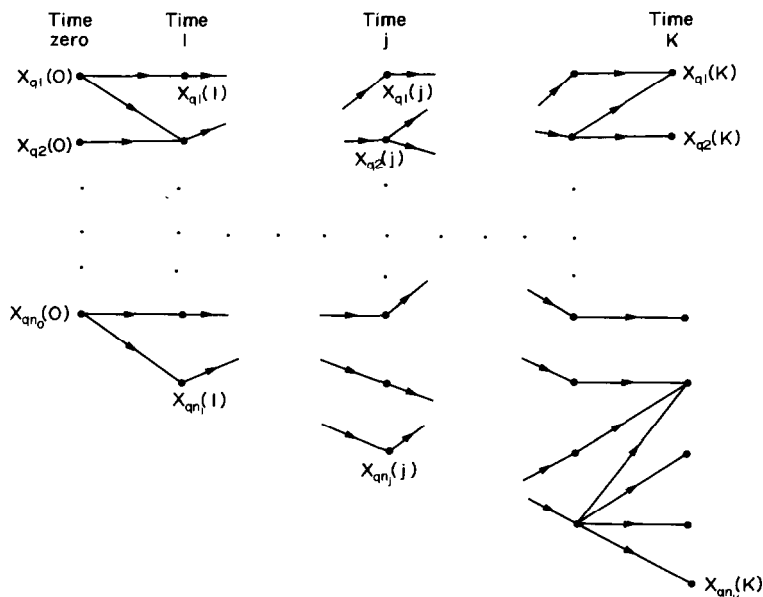


FIG. 1. Trellis diagram of state.

where the equality is resolved at random, among the paths satisfying the equality. In (5) P^m denotes the m th path through the trellis diagram, in which the superscript m denotes the path label, and $M(P^m)$ is, by definition, the metric of the m th path through the trellis diagram, which is defined by

$$M(P^m) \triangleq \sum_{k=0}^K M(x_q^m(k)) + \sum_{k=1}^K M(x_q^m(k-1) \rightarrow x_q^m(k))$$

where $x_q^m(k)$ is the quantization level (or node) at time k along the m th path (P^m) of the trellis diagram;

$$M(x_q^m(k)) \triangleq \begin{cases} \ln \{\text{Prob} \{x_q(k) = x_q^m(k)\}\} & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$$

which is, by definition, the metric of the node (or quantization level) $x_q^m(k)$, where \ln indicates the natural logarithm;

$$M(x_q^m(k-1) \rightarrow x_q^m(k)) \triangleq \ln \{\pi_k^m p(z(k) | x_q^m(k), \hat{X}(k | k))\}$$

which is, by definition, the metric of the branch connecting the node $x_q^m(k-1)$ to the node $x_q^m(k)$, where π_k^m is the transition probability from the quantization level $x_q^m(k-1)$ to the quantization level $x_q^m(k)$ which is, by definition, the probability that the quantized state takes $x_q^m(k)$ at time k when it was equal to $x_q^m(k-1)$ at time $k-1$, that is

$$\begin{aligned} \pi_k^m &\triangleq \text{Prob} \{x_q(k) = x_q^m(k) | x_q(k-1) = x_q^m(k-1)\} \\ &= \sum_i \text{Prob} \{w_d(k-1) = w_{di}(k-1)\} \end{aligned}$$

where the summation is taken over all l such that

$$Q(f(k-1), x_q^m(k-1), \hat{X}(k-1|k-1), w_{dl}(k-1))) = x_q^m(k),$$

and the conditional density function $p(z(k)|x_q^m(k), \hat{X}(k|k))$ is the density function of the observation at time k given that the quantized state $x_q(k)$ is equal to $x_q^m(k)$ and $X(k|k) = \hat{X}(k|k)$. This conditional density is given by

$$p(z(k)|x_q^m(k), \hat{X}(k|k)) = \sum_{l=1}^{r_k} p(z(k)|x_q^m(k), \hat{X}(k|k), I_{dl}(k)) \text{Prob} \{I_d(k) = I_{dl}(k)\},$$

where $p(z(k)|x_q^m(k), \hat{X}(k|k), I_{dl}(k))$ is the conditional density of the observation at time k given that $x_q(k) = x_q^m(k)$, $X(k) = \hat{X}(k|k)$, and $I_d(k) = I_{dl}(k)$.

It follows from the previous definitions that the metric of a path is the sum of the metrics of the nodes and branches along the path. The quantization levels along the path with the greatest metric (from time zero to time K) are the estimates of the state from time zero to time K . The path with the greatest metric is easily found by using the Viterbi decoding algorithm recursively (11). The recursive estimation steps are as follows:

Step j ($j = 1, 2, \dots, K$): Represent the state model by a trellis diagram from time zero to time j . Then, find the path with the greatest metric from time zero to time j through this trellis diagram by using the Viterbi decoding algorithm. The quantization levels along this path determine the set $\hat{X}(j|j)$ in (3) and (4). This set is used to find the trellis diagram from time zero to time $j+1$; and repeat this process. Finally, when j becomes K , stop and decide that the quantization levels along the path with the greatest metric from time zero to time K are the estimates of the state from time zero to time K .

IV. Performance

The performance of the proposed estimation approach is based upon the performance of the Viterbi decoding algorithm (VDA). The performance of the VDA can be quantified by a Gallager-type ensemble upper bound (15, 16), since the evaluation of the exact error probability or error probability bound for choosing the correct path is complex. It can be shown (11) that such an ensemble bound is given by

$$\bar{P}_E \leq R(\mu) \sum_{j=1}^K \left\{ \int_{z(j)} \left[\sum_{x \in X^e} q(x) p(z(j)|x, \hat{X}(j|j))^{1/(1+\mu)} \right]^{1+\mu} dz(j) \right\}$$

for any $\mu \in [0, 1]$, (6)

with

$$R(\mu) \triangleq (M-1)^\mu \left[\prod_{j=0}^K \left(\frac{\pi_j^{\max}}{\pi_j^{\min}} \right)^{\mu/(1+\mu)} \right]$$

where \bar{P}_E is the ensemble averaged overall error probability for the state estimation; X^e is the set of all possible quantization levels of the state from time one to time

$K; q(\cdot)$ is an arbitrary probability density function on X^e ; $p(z(j)|x, \hat{X}(j|j))$ is the conditional density function of the observation, given that the state is equal to x and $X(j) = \hat{X}(j|j)$; M is the number of possible paths through the trellis diagram of the state; π_0^{\min} and π_0^{\max} are the minimum and maximum values of the occurrence probabilities of possible values of the discrete initial state $x_q(0)$; and π_j^{\min} and π_j^{\max} , $j > 0$, are the minimum and maximum values of the transition probabilities from time $j-1$ to time j , respectively.

As the performance measure of the proposed approach, the uniformly weighted ensemble bound with $\mu = 1$ is used since it is the easiest bound to evaluate (11); "uniformly weighted" is, by definition, that $q(x) = 1/N^e$ for all x , where N^e is the number of elements in X^e . Consider, as an example, the models which are given by

$$x(j+1) = f(j, x(j), X(j), w(j)) \quad \text{the state model,} \quad (7)$$

$$z(j) = g(j, x(j), X(j), I(j)) + h(j, x(j), X(j), I(j))v(j) \quad \text{the observation model} \quad (8)$$

where $x(0)$ and $v(j)$ are Gaussian noises with means $m_0, 0$; and variances R_0 and $R_v(j)$, respectively. $h(j, x(j), X(j), I(j))$ is a given (linear or nonlinear) function such that

$$[h(j, x_{qi}(j), \hat{X}(j|j), I_{di}(j))]R_v(j)[h(j, x_{qi}(j), \hat{X}(j|j), I_{di}(j))]^T \neq 0$$

for all j, l and i , where the superscript T stands for the transpose. Substituting $p(z(j)|x(j) = x, X(j) = \hat{X}(j|j))$, $\mu = 1$, and $q(\cdot) = (1/N^e)$ into the bound in (6), and further using the inequality

$$\left[\sum_k c_k \right]^\eta \leq \sum_k [c_k]^\eta \quad \text{for any } c_k \geq 0 \text{ and } \eta \in [0, 1],$$

we can obtain the bound

$$\bar{P}_E \leq R(1) \left(\frac{1}{N^e} \right)^{2K} \prod_{j=1}^K \left\{ \sum_{x_1 \in X^e} \sum_{x_2 \in X^e} \sum_{l=1}^{r_l} \sum_{i=1}^{r_l} [p(I_{di}(j))p(I_{di}(j))]^{1/2} \right. \\ \left. \times F(j, x_1, I_{di}(j), x_2, I_{di}(j)) \right\}, \quad (9)$$

where

$$F(j, x_1, I_{di}(j), x_2, I_{di}(j)) \triangleq C \exp \left\{ \frac{D}{4} \right\} \\ C \triangleq \frac{\{\det [2(R_l^{-1} + R_l^{-1})^{-1}]\}^{1/2}}{(\det R_l)^{1/4}(\det R_l)^{1/4}}, \\ D \triangleq b_l^T(R_l^{-1} + R_l^{-1})^{-1}b_l - D_1 - D_2, \\ D_1 \triangleq [g(j, x_1, \hat{X}(j|j), I_l(j))]^T R_l^{-1} [g(j, x_1, \hat{X}(j|j), I_l(j))], \\ D_2 \triangleq [g(j, x_2, \hat{X}(j|j), I_l(j))]^T R_l^{-1} [g(j, x_2, \hat{X}(j|j), I_l(j))],$$

$$b_{ii} \triangleq R_i^{-1}[g(j, x_1, \hat{X}(j|j), I_i(j))] + R_i^{-1}[g(j, x_2, \hat{X}(j|j), I_i(j))],$$

$$R_i \triangleq [h(j, x_1, \hat{X}(j|j), I_i(j))]R_i(j)[h(j, x_1, \hat{X}(j|j), I_i(j))]^T,$$

$$R_i \triangleq [h(j, x_2, \hat{X}(j|j), I_i(j))]R_i(j)[h(j, x_2, \hat{X}(j|j), I_i(j))]^T,$$

$$P(I_{di}(j)) \triangleq \text{Prob} \{I_{di}(j) = I_{di}(j)\}.$$

The bound of (9) is the one used as the performance measure of the proposed approach for the models of (7) and (8).

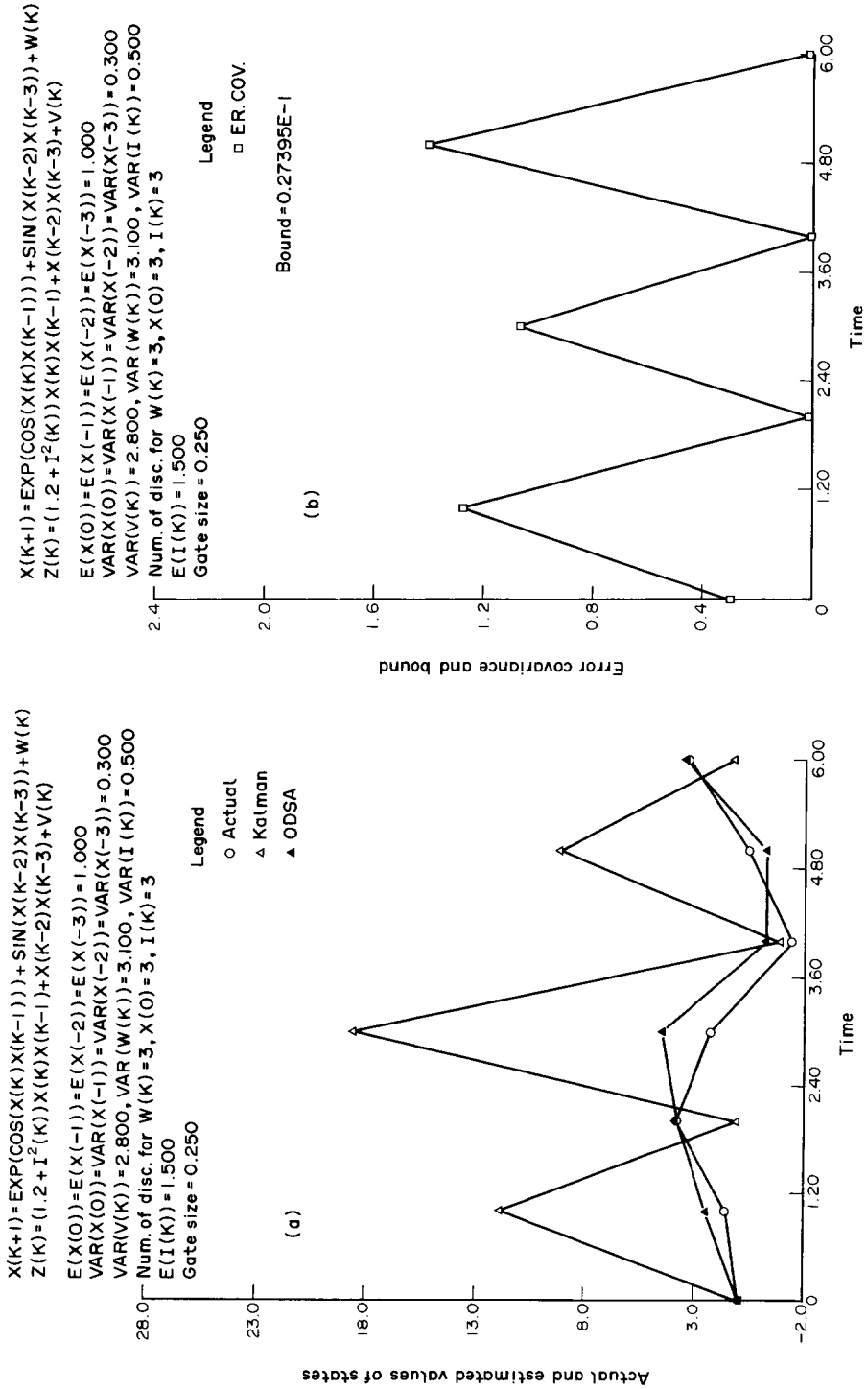
V. Simulations

Many examples were simulated on the IBM 3081K main frame computer. The aim of simulations was to observe the performance of the proposed approach and the divergence of the Kalman filter estimates assuming zero interference and $X(j) = \hat{X}(j|j)$. In simulations, the Gaussian random variables were used and approximated by the discrete random variables which are presented in (11). Simulations were performed up to time 8 since the implementation of the proposed approach requires an exponentially increasing memory with time. The performance bound, in (9), of the proposed approach was also simulated. However one should realize that this bound is an ensemble bound and it can sometimes yield a number being greater than one (which is useless) since in the derivations of the bound, some inequalities are used. Thus, this bound does not yield the exact performance of the proposed approach (11).

Simulation results of three examples are presented in Figs 2–4. At the top left hand corner of each figure, the simulated models and statistics of random variables used are presented. AAEOP and AAEK indicate the time-averaged absolute errors for the estimates by the proposed estimation approach and the Kalman filter; ER, COV, and Bound show the error variances of the Kalman estimates and the performance bound, given by (9), of the proposed estimation approach; $E(G(\cdot))$ and $\text{VAR}(G(\cdot))$ represent the mean value and variance of the random variable $G(\cdot)$; Actual, ODSA and Kalman denote the actual values, and their estimates by the proposed estimation approach and Kalman filter, respectively; and Num. of disc. for $G(\cdot)$ indicates the number of possible values of the discrete random variable used to approximate the random variable $G(\cdot)$.

Simulations were performed for different parameters (which are the gate size and the numbers of possible values of the discrete random variables). As in Figs 2–4, the discrete random variables with three possible values yield a good estimation performance.

Figures 2–4 present simulation results of three nonlinear dynamic models with a 4th order memory and nonlinear interference. As one knows, the Kalman filter cannot, in general, be used for state estimation of dynamic models with nonlinear interference. If it were used by assuming zero interference and $X(j) = \hat{X}(j|j)$, the Kalman estimates could diverge from the actual state values. This is observed in Figs 2–4. However, the estimates obtained by the proposed approach closely follow the actual values of the states.



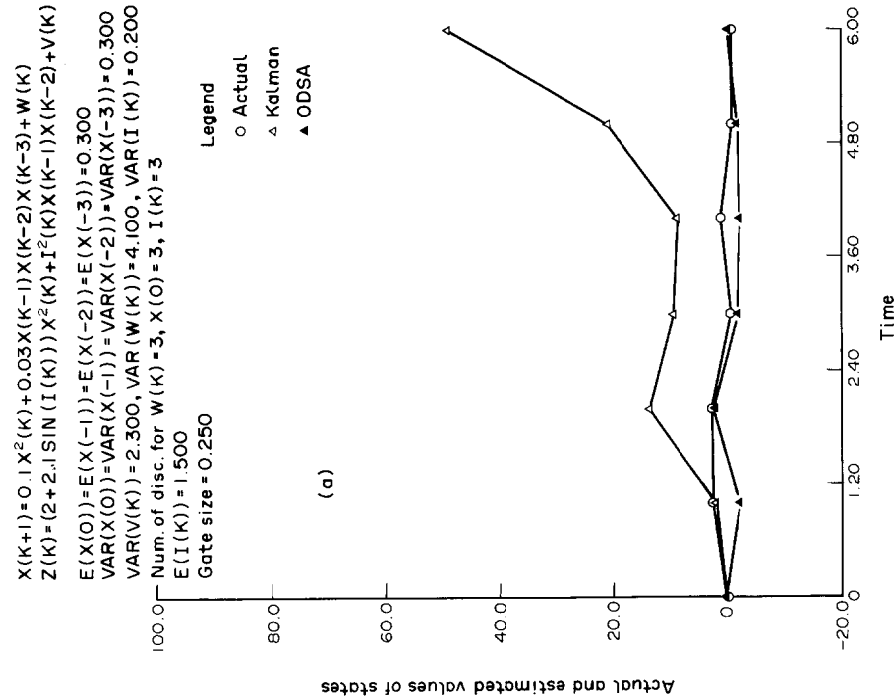


Fig. 3(a). Actual and estimated values of states.

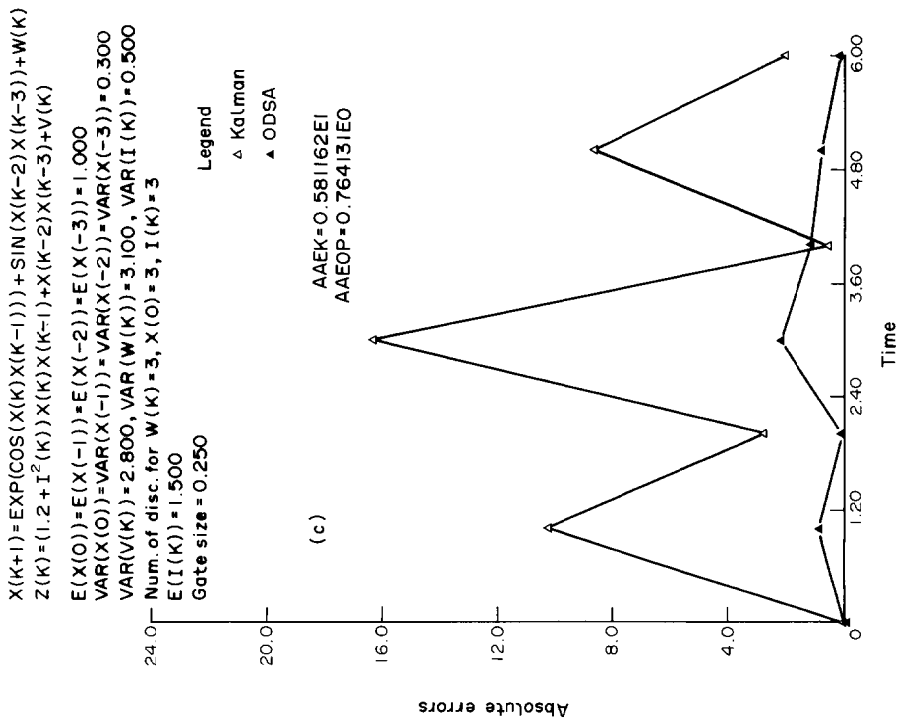


Fig. 2(c). Absolute and time-averaged absolute errors for estimates of states.

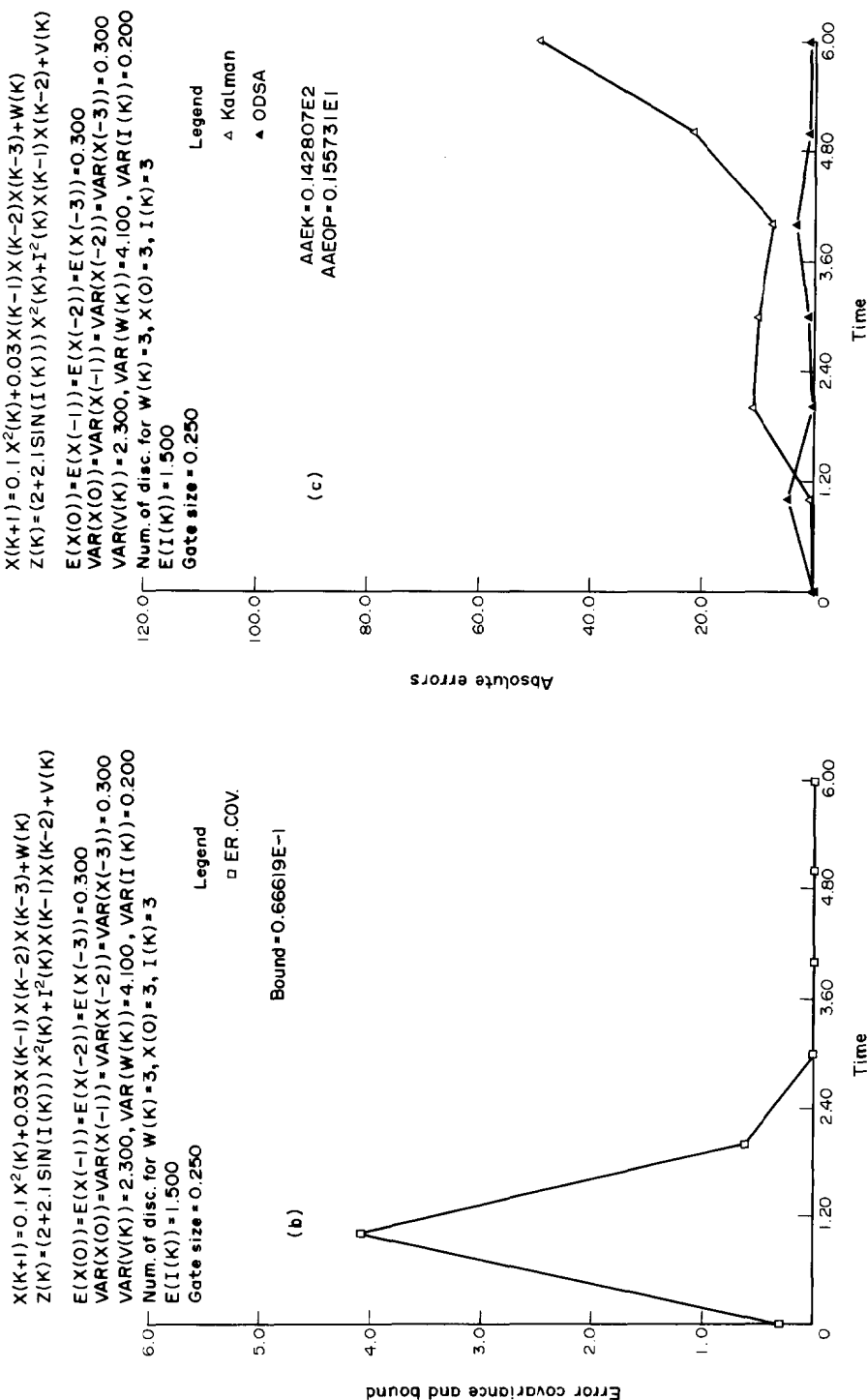


Fig. 3(b). Error variances and bound for estimates of states.

Fig. 3(c). Absolute and time-averaged absolute errors for estimates of states.

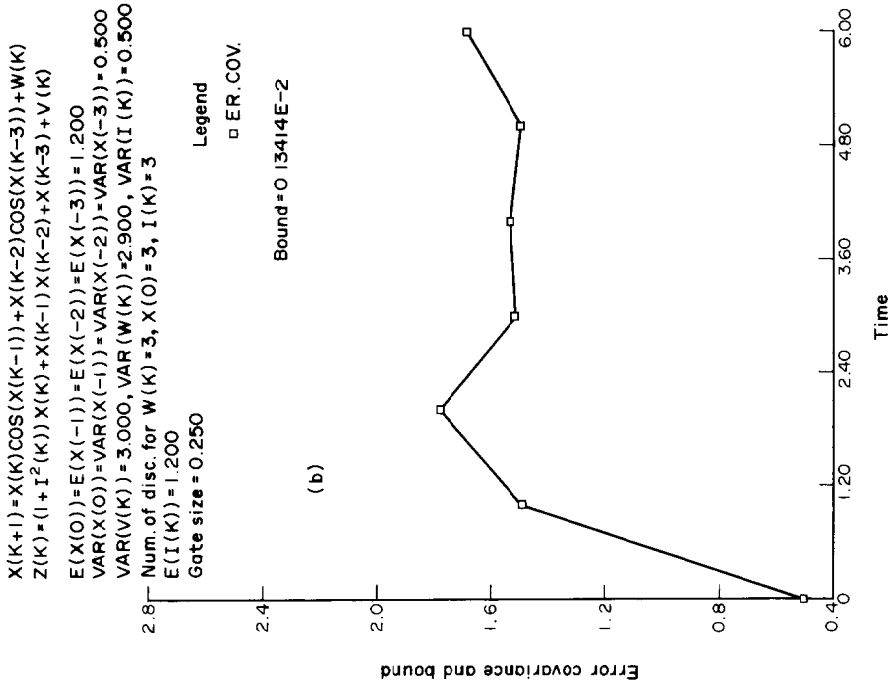


Fig. 4(b). Error variances and bound for estimates of states.

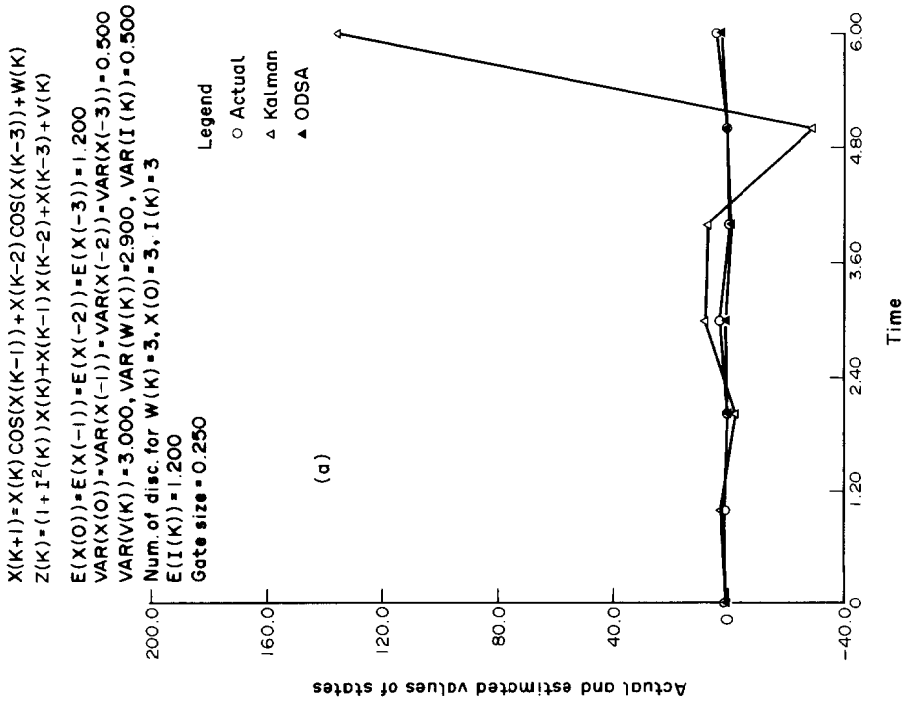


Fig. 4(a). Actual and estimated values of states.

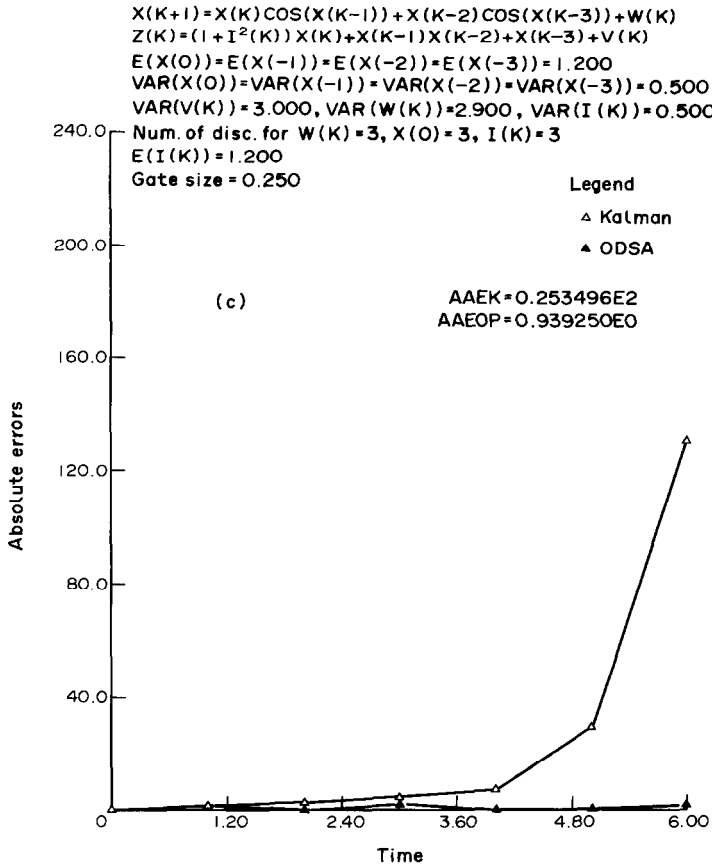


FIG. 4(c). Absolute and time-averaged absolute errors for estimates of states.

VI. Conclusions

The proposed estimation approach can be used to estimate the state of dynamic models with both an N th order memory and nonlinear random interference, whereas the classical estimation schemes, such as the Kalman filter, in general, cannot. Dynamic models of the proposed approach can be any nonlinear functions. However, the implementation of the proposed estimation approach requires an exponentially increasing memory with time.

References

- (1) R. E. Kalman, "A new approach to linear filtering and prediction problems", *Trans. ASME, J. Basic Eng., Ser. D*, Vol. 82, pp. 35-45, 1960.
- (2) R. E. Kalman and B. C. Bucy, "New results in linear filtering and prediction theory", *Trans. ASME, J. Basic Eng., Ser. D*, Vol. 83, pp. 95-108, 1961.

- (3) A. P. Sage and J. L. Melsa, "Estimation Theory with Applications to Communications and Control", McGraw-Hill, New York, 1971.
- (4) T. Kailath, "An innovation approach to least square estimation, Part I : Linear filtering in adaptive white noise", *IEEE Trans. Aut. Contr.*, Vol. AC-13, No. 6, pp. 646–655, 1968.
- (5) T. Kailath, "A view of three decades in linear filtering theory", *IEEE Trans. Inf. Theory*, Vol. IT-20, No. 2, pp. 146–181, 1974.
- (6) J. Makhoul, "Linear prediction : A tutorial review", *Proc. IEEE*, Vol. 63, pp. 561–580, 1975.
- (7) J. S. Medich, "A survey of data smoothing for linear and nonlinear dynamic systems", *Automatica*, Vol. 9, pp. 151–162, 1973.
- (8) C. E. Hutchinson, "The Kalman Filter applied to aerospace and electronic systems", *IEEE Trans. Aerospace Electron. Syst.*, Vol. AES-20, No. 4, pp. 500–504, 1984.
- (9) N. E. Nahi, "Optimal recursive estimation with uncertain observation", *IEEE Trans. Inf. Theory*, Vol. IT-15, No. 4, pp. 457–462, 1969.
- (10) R. A. Monzingo, "Discrete optimal linear smoothing for systems with uncertain observation", *IEEE Trans. Inf. Theory*, Vol. IT-21, No. 3, pp. 271–275, 1975.
- (11) K. Demirbaş, "Information theoretic smoothing algorithms for dynamic systems with or without interference", *Adv. Contr. Dyn. Syst.*, Volume XXI, pp. 175–295, Academic Press, New York, 1984.
- (12) K. Demirbaş and C. T. Leondes, "Optimum decoding based smoothing algorithm for dynamic systems with interference", *Int. J. Syst. Sci.*, Vol. 17, No. 2, pp. 251–267, 1986.
- (13) H. L. Van Trees, "Detection, Estimation and Modulation Theory", Wiley, New York, 1968.
- (14) G. D. Forney Jr, "Convolution codes II. Maximum likelihood decoding", *Inf. Contr.*, Vol. 25, pp. 222–266, 1974.
- (15) A. J. Viterbi and J. K. Omura, "Principles of Digital Communication and Coding", McGraw-Hill, New York, 1979.
- (16) R. G. Gallager, "A simple derivation of the coding theorem and some applications", *IEEE Trans. Inf. Theory*, Vol. IT-11, pp. 3–18, 1965.