State Estimation for Linear Discrete-Time Systems with State-Dependent Noise

By

Hajime AKASHI*, Kazuo NOSE* and Nobuo KOIZUMI*

(Received December 28, 1974)

Abstract

In this paper, the state estimation problem is considered for a class of linear systems with state-dependent noise. The optimal nonlinear estimator in the mean square sense is first derived on the basis of the Bayesian approach. Then a sub-optimal estimator is proposed, in which the estimate is still nonlinear in the observation data, and the covariances are obtained recursively using the observation data. The case where the state-dependent noise is white is treated specifically. Some simulations for this case are made in order to examine the practicability of the proposed sub-optimal estimator, and the result is compared with that of the linear estimator by McLane.

1. Introduction

A recursive minimum variance state estimation procedure for linear stochastic systems was first introduced by Kalman and Bucy[1,2]. Since then, there have been numerous papers written on alternative ways of deriving the conditions for the optimal estimator and on extensions of the original works.

The present paper considers the estimator which can be applied to the system with state-dependent noise. This kind of system can model either a system in which the parameters vary randomly, or a system in which an additive stochastic disturbance depends linearly on the state variables. The former is found in process control systems while the latter is seen in aerospace systems.

In these systems, the *a posteriori* probability density function of the state based on observation data becomes non-Gaussian, even though the dynamics and the observation mechanism are linear. Therefore, the optimal estimator in the mean square sense requires a nonlinear estimation procedure. An approximate

* Department of Precision Mechanics.

method for the state estimation of these systems was discussed by McLane[3]. He derived a sub-optimal estimator by solving the Wiener-Hopf equation, which is the optimal linear operation on the observation data. This approach is applicable only when the influence of the state-dependent noise is relatively small, and the state variable is nearly Gaussian. The present paper proposes a sub-optimal estimator by an approach different from the above.

In Section 2, the optimal state estimation problem is formulated with the specification of a system. The Bayesian approach[4] is then used to derive the optimal estimator in Section 3. By treating the *a posteriori* probability density function as a mixture density[5], we show that the best estimate is obtained as the weighted mean of estimates, each accompanied by a specified state-dependent noise sequence. This procedure, while it is optimal, is not a realistic procedure, since it requires evergrowing computational labor. Therefore, we propose a sub-optimal estimator based on a certain assumption, which will be given in Section 4. In Section 5, the sub-optimal estimator is derived for the case where the state-dependent noise is white. In Section 6, some simulations are made to show the practicability of the sub-optimal estimator. Concluding remarks are given in Section 7.

2. Problem Statement

The system is characterized by the equations

$$x(t+1)=A(t)x(t)+B[t,\eta(t)]x(t)+u(t) \tag{2.1}$$

$$y(t)=C(t)x(t)+D[t,\xi(t)]x(t)+v(t) \qquad t=0,1,2,\cdots \tag{2.2}$$

where $x(t)$ is an n-dimensional state vector; $y(t)$ is a p-dimensional observation vector; $A(t)$ is a known $n\times n$ matrix; $C(t)$ is a known $p\times n$ matrix; and $u(t)$ and $v(t)$ are additive white Gaussian noises with zero means and covariances $Q(t)$ and $R(t)$, respectively. The state-dependent noises are denoted by $\eta(t)$ and $\xi(t)$, and $\eta(t)$ influences the dynamics through an $n\times n$ matrix,

$$B[t,\eta(t)]=\sum_{i=1}^{m}B_i(t)\eta_i(t), \tag{2.3}$$

and $\xi(t)$ influences the observation mechanism through a $p\times n$ matrix,

$$D[t,\xi(t)]=\sum_{i=1}^{l}D_i(t)\xi_i(t), \tag{2.4}$$

where m and l are dimensions of $\eta(t)$ and $\xi(t)$ respectively; $\eta_i(t)$ and $\xi_i(t)$ are the i-th components of $\eta(t)$ and $\xi(t)$ respectively; $B_i(t)$ $(i=1,2,\cdots,m)$ is a known $n\times n$ matrix; and $D_i(t)$ $(i=1,2,\cdots,l)$ is a known $p\times n$ matrix. These state-dependent noises are Markov chains with transition probability density functions $p(\eta(t)|\eta(t-1))$ and $p(\xi(t)|\xi(t-1))$, and with initial probability density functions $p(\eta(0))$ and $p(\xi(0))$. The noises $\eta(\cdot)$ and $\xi(\cdot)$ may be distributed in any form, but their covariances must be bounded. The initial state $x(0)$ is a Gaussian random vector with the mean $\bar{x}(0)$ and the covariance $\bar{P}(0)$. The noise sequences $\{u(t)\}$, $\{v(t)\}$, $\{\eta(t)\}$, $\{\xi(t)\}$, and the initial state $x(0)$ are assumed to be mutually independent.

The problem to be considered in this paper is the derivation of the estimate $\hat{x}(t)$ which minimizes the conditional expectation,

$$E\{[x(t)-\hat{x}(t)]'M[x(t)-\hat{x}(t)]|Y(t)\} \tag{2.5}$$

based on the accumulated observation data,

$$Y(t) \triangleq \{y(0), \cdots, y(t)\}. \tag{2.6}$$

Here M is a positive definite symmetric matrix, $E\{\cdot|Y(t)\}$ denotes the conditional expectation given $Y(t)$, and the prime denotes the transpose of a vector or a matrix.

3. The Optimal Estimator

It is well known that the best estimate which minimizes the conditional expectation (2.5) is given by

$$\hat{x}(t)=\int x(t)p(x(t)|Y(t))dx(t). \tag{3.1}$$

Define the sequence of the state-dependent noise by

$$S(t) \triangleq \{s(0), \cdots, s(t)\} \tag{3.2}$$

where $s(t)$ is the index of the distribution, and $\eta(t-1)$ and $\xi(t)$ are its samples, i.e.,

$$s(t) \triangleq \{\eta(t-1), \xi(t)\}.$$

Using definition (3.2), the probability density function in eq. (3.1) can be described as the mixture density[5],

$$p(x(t)|Y(t))=\int p(S(t)|Y(t))p(x(t)|S(t), Y(t))dS(t). \tag{3.3}$$

According to Bayes' estimation theory, we have

$$p(x(t)|S(t), Y(t)) \sim N\{\hat{x}(t|S(t)), \hat{P}(t|S(t))\}^{*} \tag{3.4}$$

* $p(x) \sim N\{m, R\}$ denotes the Gaussian distribution

$$p(x) = (2\pi)^{-(n/2)}|R|^{-(1/2)}\exp\left\{-\frac{1}{2}(x-m)'R^{-1}(x-m)\right\}$$

where we define

$$\hat{x}(t|S(t)) \triangleq E\{x(t)|S(t), Y(t)\}$$
$$\hat{P}(t|S(t)) \triangleq \mathrm{cov}\{x(t)|S(t), Y(t)\}.$$

Since the state-dependent noise sequence $S(t)$ is fixed, the system described by eqs. (2.1) and (2.2) becomes an ordinary linear system with additive noises and its estimation formulae are exactly the same as those of a Kalman filter[1]:

$$\hat{x}(t|S(t)) = \bar{x}(t|S(t)) + \Gamma(t|S(t))[y(t) - \{C(t) + D[t, \xi(t)]\}\bar{x}(t|S(t))] \qquad (3.5)$$

$$\hat{P}(t|S(t)) = \bar{P}(t|S(t)) - \Gamma(t|S(t))\{C(t) + D[t, \xi(t)]\}\bar{P}(t|S(t)) \qquad (3.6)$$

$$\Gamma(t|S(t)) = \bar{P}(t|S(t))\{C(t) + D[t, \xi(t)]\}' \times [\{C(t) + D[t, \xi(t)]\}\bar{P}(t|S(t))\{C(t) + D[t, \xi(t)]\}' + R(t)]^{-1} \qquad (3.7)$$

$$\bar{x}(t+1|S(t+1)) = \{A(t) + B[t, \eta(t)]\}\hat{x}(t|S(t)) \qquad (3.8)$$

$$\bar{P}(t+1|S(t+1)) = \{A(t) + B[t, \eta(t)]\}\hat{P}(t|S(t))\{A(t) + B[t, \eta(t)]\}' + Q(t) \qquad (3.9)$$

where $\bar{x}(t|S(t))$ and $\bar{P}(t|S(t))$ are defined by

$$\bar{x}(t|S(t)) \triangleq E\{x(t)|S(t), Y(t-1)\}$$

$$\bar{P}(t|S(t)) \triangleq \mathrm{cov}\{x(t)|S(t), Y(t-1)\}.$$

Using Bayes' rule, the probability density function $p(S(t)|Y(t))$ is transformed to

$$p(S(t)|Y(t)) = \frac{p(S(t)|Y(t-1))p(y(t)|S(t), Y(t-1))}{\int p(S(t)|Y(t-1))p(y(t)|S(t), Y(t-1))dS(t)}. \qquad (3.10)$$

From eq. (2.2), $p(y(t)|S(t), Y(t-1))$ is Gaussian of the form,

$$p(y(t)|S(t), Y(t-1)) \sim N\{\{C(t) + D[t, \xi(t)]\}\bar{x}(t|S(t)), \{C(t) + D[t, \xi(t)]\}\bar{P}(t|S(t))\{C(t) + D[t, \xi(t)]\}' + R(t)\}. \qquad (3.11)$$

The Markov property of $\{s(t)\}$ yields

$$p(S(t)|Y(t-1)) = p(s(t)|s(t-1))p(S(t-1)|Y(t-1)). \qquad (3.12)$$

Therefore eqs. (3.10) and (3.12) give a recursive relation for $p(S(t)|Y(t))$.

By treating the *a posteriori* probability density function $p(x(t)|Y(t))$ as the mixture density given by eq. (3.3), the best estimate is obtained as the weighted mean of estimates, each containing a specified state-dependent noise sequence. It is obvious from eqs. (3.3) and (3.10), however, that these equations require evergrowing computational labor for the numerical integration. Therefore, the algorithm derived in this section, while it is optimal, is not a realistic procedure.

4. Derivation of a Sub-optimal Estimator

Avoiding the computational difficulty, a sub-optimal estimator is proposed. It is based on the assumption that the probability density function of $x(t)$ given $Y(t-1)$ and $s(t)$ is Gaussian, i.e.,

$$p(x(t)|s(t), Y(t-1)) \sim N\{\bar{x}(t|s(t)), \bar{P}(t|s(t))\} \tag{4.1}$$

where we define

$$\bar{x}(t|s(t)) \triangleq E\{x(t)|s(t), Y(t-1)\}$$

$$\bar{P}(t|s(t)) \triangleq \text{cov}\{x(t)|s(t), Y(t-1)\}.$$

Using assumption (4.1), the amount of computation will be constant with respect to time t. The probability density function $p(x(t)|Y(t))$ can be transformed into,

$$p(x(t)|Y(t)) = \int p(s(t)|Y(t))p(x(t)|s(t), Y(t))ds(t). \tag{4.2}$$

Bayes' formula provides

$$p(x(t)|s(t), Y(t)) = \frac{p(y(t)|s(t), x(t))p(x(t)|s(t), Y(t-1))}{p(y(t)|s(t), Y(t-1))}. \tag{4.3}$$

Now, the probability density functions in the right hand side of eq. (4.3) will be examined. First, $p(x(t)|s(t), Y(t-1))$ is Gaussian according to eq. (4.1). From observation equation (2.2), $p(y(t)|s(t), x(t))$ becomes Gaussian of the form,

$$p(y(t)|s(t), x(t)) \sim N\{\{C(t)+D[t, \xi(t)]\}x(t), R(t)\}. \tag{4.4}$$

Furthermore, from eqs. (2.2) and (4.1), $p(y(t)|s(t), Y(t-1))$ becomes Gaussian with the mean,

$$\{C(t)+D[t, \xi(t)]\}\bar{x}(t|s(t)), \tag{4.5a}$$

and the covariance,

$$\{C(t)+D[t, \xi(t)]\}\bar{P}(t|s(t))\{C(t)+D[t, \xi(t)]\}'+R(t). \tag{4.5b}$$

Let us define

$$\hat{x}(t|s(t)) \triangleq E\{x(t)|s(t), Y(t)\}$$

$$\hat{P}(t|s(t)) \triangleq \text{cov}\{x(t)|s(t), Y(t)\}.$$

Substituting eqs. (4.4) and (4.5) into eq. (4.3), and using the assumption (4.1), $p(x(t)|s(t), Y(t))$ can be shown to be Gaussian of the form,

$$p(x(t)|s(t), Y(t)) \sim N\{\hat{x}(t|s(t)), \hat{P}(t|s(t))\}$$

where

$$\hat{x}(t|s(t))=\bar{x}(t|s(t))+\Gamma(t|s(t))[y(t)-\{C(t)+D[t,\xi(t)]\}\bar{x}(t|s(t))], \tag{4.6}$$

$$\hat{P}(t|s(t))=\bar{P}(t|s(t))-\Gamma(t|s(t))\{C(t)+D[t,\xi(t)]\}\bar{P}(t|s(t)), \tag{4.7}$$

and where

$$\Gamma(t|s(t))=\bar{P}(t|s(t))\{C(t)+D[t,\xi(t)]\}' \times[\{C(t)+D[t,\xi(t)]\}\bar{P}(t|s(t))\{C(t)+D[t,\xi(t)]\}'+R(t)]^{-1}. \tag{4.8}$$

Using Bayes' rule, the probability density function $p(s(t)|Y(t))$ in eq. (4.2) becomes

$$p(s(t)|Y(t))=\frac{p(y(t)|s(t),Y(t-1))p(s(t)|Y(t-1))}{\int p(y(t)|s(t),Y(t-1))p(s(t)|Y(t-1))ds(t)}. \tag{4.9}$$

where $p(y(t)|s(t),Y(t-1))$ is given by eq. (4.5) ; and from the Markov property of $\{s(t)\}$, $p(s(t)|Y(t-1))$ can be described as

$$p(s(t)|Y(t-1))=\int p(s(t)|s(t-1))p(s(t-1)|Y(t-1))ds(t-1). \tag{4.10}$$

Substituting eq. (4.10) into eq. (4.9), it becomes

$$p(s(t)|Y(t)) =\frac{p(y(t)|s(t),Y(t-1))\int p(s(t)|s(t-1))p(s(t-1)|Y(t-1))ds(t-1)}{\int p(y(t)|s(t),Y(t-1))\int p(s(t)|s(t-1))p(s(t-1)|Y(t-1))ds(t-1)ds(t)}. \tag{4.11}$$

Eq. (4.11) gives the recursive formula for the calculation of $p(s(t)|Y(t))$.

In order to make use of eqs. (4.6), (4.7) and (4.8) as recursive formulae to obtain $\hat{x}(t|s(t))$ and $\hat{P}(t|s(t))$, it is necessary to have equations for the one-step-ahead predicted mean and covariance. Let us consider the probability density function $p(x(t+1)|s(t+1),Y(t))$. From the system equation (2.1), we have

$$\bar{x}(t+1|s(t+1))=\{A(t)+B[t,\eta(t)]\}\hat{x}(t|s(t+1)) \tag{4.12}$$

$$\bar{P}(t+1|s(t+1))=\{A(t)+B[t,\eta(t)]\} \times\hat{P}(t|s(t+1))\{A(t)+B[t,\eta(t)]\}'+Q(t), \tag{4.13}$$

where

$$\hat{x}(t|s(t+1))\triangleq E\{x(t)|s(t+1),Y(t)\}$$

$$\hat{P}(t|s(t+1))\triangleq \mathrm{cov}\{x(t)|s(t+1),Y(t)\}.$$

Now, the formulae for the calculation of $\hat{x}(t|s(t+1))$ and $\hat{P}(t|s(t+1))$ should be obtained. Applying Bayes' rule, $\hat{x}(t|s(t+1))$ can be obtained* from $\hat{x}(t|s(t))$ by

$$\hat{x}(t|s(t+1))=\int \hat{x}(t|s(t))\frac{p(s(t)|Y(t))p(s(t+1)|s(t))}{\int p(s(t)|Y(t))p(s(t+1)|s(t))ds(t)}ds(t). \tag{4.14}$$

The conditional covariance $\hat{P}(t|s(t+1))$ can be obtained* from $\hat{P}(t|s(t))$ by

$$\hat{P}(t|s(t+1))=\int \{\hat{P}(t|s(t))+[\hat{x}(t|s(t))-\hat{x}(t|s(t+1))] \times[\hat{x}(t|s(t))-\hat{x}(t|s(t+1))]'\} p(s(t)|s(t+1),Y(t))ds(t) \tag{4.15}$$

where

$$p(s(t)|s(t+1),Y(t))=\frac{p(s(t)|Y(t))p(s(t+1)|s(t))}{\int p(s(t)|Y(t))p(s(t+1)|s(t))ds(t)}. \tag{4.16}$$

Eqs. (4.6)~(4.8) and (4.12)~(4.16) provide the formula for deriving $\hat{x}(t|s(t))$. The probability density function $p(s(t)|Y(t))$ can be obtained from eqs. (4.5) and (4.11). Therefore, with the initial conditions,

$$\bar{x}(0|s(0))=\bar{x}(0), \qquad \bar{P}(0|s(0))=\bar{P}(0), \qquad p(s(0)|Y(-1))=p(s(0)),$$

and given the transition probability density function $p(s(t)|s(t-1))$, the sub-

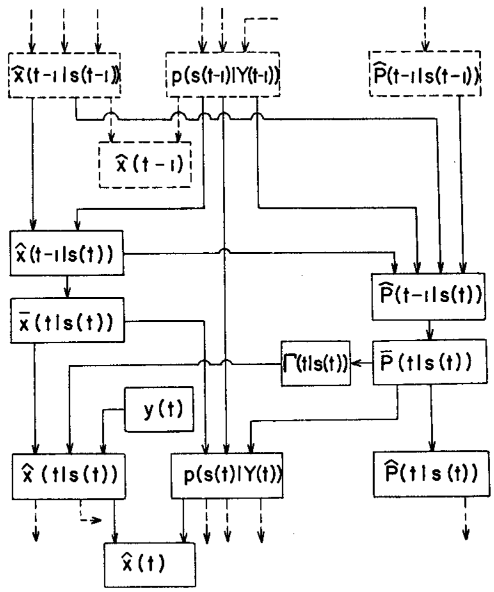

Fig. 1. The sequential structure of the estimator proposed in Section 4.

* For derivation, see Appendix.

optimal estimate can be obtained using the following equation,

$$\hat{x}(t)=\int\hat{x}(t|s(t))p(s(t)|Y(t))ds(t). \tag{4.17}$$

The sequential structure of the sub-optimal estimator is illustrated in Fig. 1.

5. The Sub-optimal Estimator when the State-dependent Noise is White

When the state-dependent noise is white, the sub-optimal estimation algorithm can be simplified. Eqs. (4.6), (4.7) and (4.8) are same, but the remaining formulae in the previous section can be simplified.

Let $s(t)$ be white. As the sequence $\{s(t)\}$ is an independent process, eq. (4.10) becomes

$$p(s(t)|Y(t-1))=p(s(t)). \tag{5.1}$$

Therefore eq. (4.9) can be described by

$$p(s(t)|Y(t))=\frac{p(s(t))p(y(t)|s(t),Y(t-1))}{\int p(s(t))p(y(t)|s(t),Y(t-1))ds(t)}. \tag{5.2}$$

From eq. (4.16),

$$p(x(t)|s(t+1),Y(t))=p(x(t)|Y(t)), \tag{5.3}$$

and

$$\hat{x}(t|s(t+1))=\hat{x}(t). \tag{5.4}$$

Therefore eq. (4.12) becomes

$$\bar{x}(t+1|s(t+1))=\{A(t)+B[t,\eta(t)]\}\hat{x}(t). \tag{5.5}$$

Also, from eqs. (5.3) and (5.4), we have

$$\hat{P}(t|s(t+1))=\hat{P}(t). \tag{5.6}$$

The covariance $\hat{P}(t)$ is obtained from

$$\hat{P}(t)=\int\{\hat{P}(t|s(t))+[\hat{x}(t|s(t))-\hat{x}(t)][\hat{x}(t|s(t))-\hat{x}(t)]'\}p(s(t)|Y(t))ds(t). \tag{5.7}$$

Finally, eq. (4.13) can be written as

$$\bar{P}(t+1|s(t+1))=\{A(t)+B[t,\eta(t)]\}\hat{P}(t)\{A(t)+B[t,\eta(t)]\}'+Q(t). \tag{5.8}$$

Therefore, eqs. (4.6)∼(4.8), (4.17), (5.2), (5.5), (5.7) and (5.8), with the initial conditions,

$$\bar{x}(0|s(0)) = \bar{x}(0), \qquad \bar{P}(0|s(0)) = \bar{P}(0),$$

and given the probability density function $p(s(t))$, complete the estimation algorithm. The sequential structure of the estimator derived in this section is illustrated in Fig. 2.

6. Simulations of the Sub-optimal Estimator

Simulation results are shown here in order to examine the practicability of the proposed sub-optimal estimation algorithm, and to compare the performance of the proposed estimator with that of the linear estimator proposed by McLane[3]. Consider the following time-invariant scalar system

$$x(t+1) = [a + b\eta(t)]x(t) + u(t) \tag{6.1}$$

$$y(t) = [c + d\xi(t)]x(t) + v(t). \tag{6.2}$$

For the comparison of the simulations with those of the linear estimator given by McLane, we treat the cases where the state-dependent noises are white Gaussian. Let $\eta(t)$ and $\xi(t)$ be white Gaussian with zero means and covariances S_1 and S_2, respectively.

Digital simulations are carried out in two cases, (1) and (2). (1) is the case when the state-dependent noise is influencing the system dynamics only, and (2) is the case when the state-dependent noise is influencing the observation mechanism only. For case (1), the following set of numerical values is chosen:

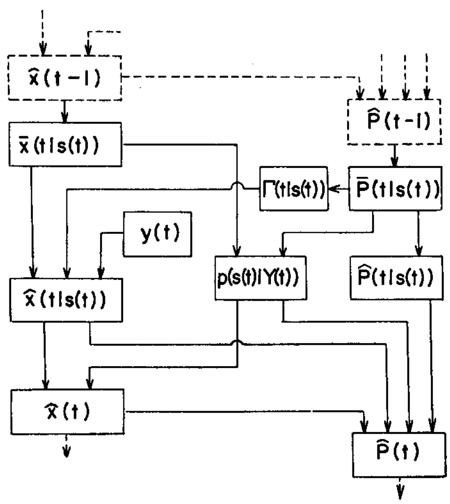

Fig. 2. The sequential structure of the estimator when the state-dependent noise is white (Section 5).

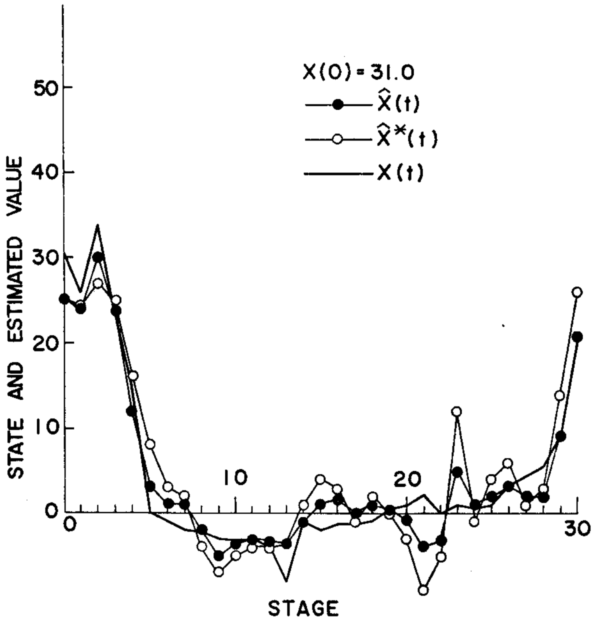

Fig. 3. Sample paths of $\hat{x}(t)$, $\hat{x}^*(t)$ and $x(t)$ for case (1).

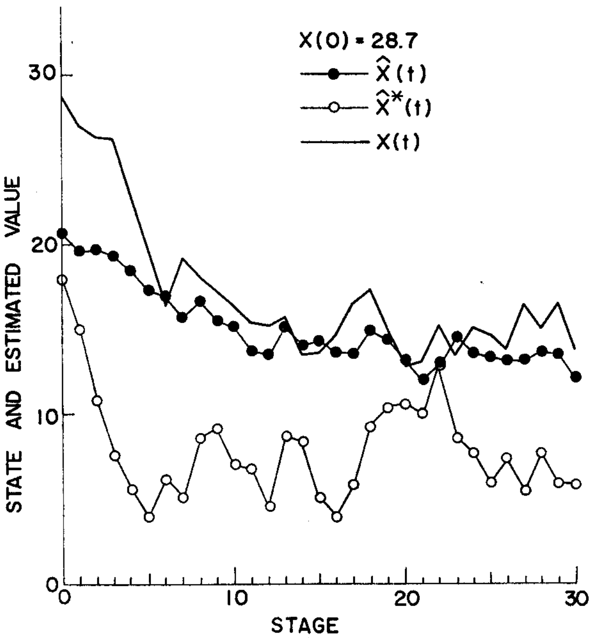

Fig. 4. Sample paths of $\hat{x}(t)$, $\hat{x}^*(t)$ and $x(t)$ for case (2).

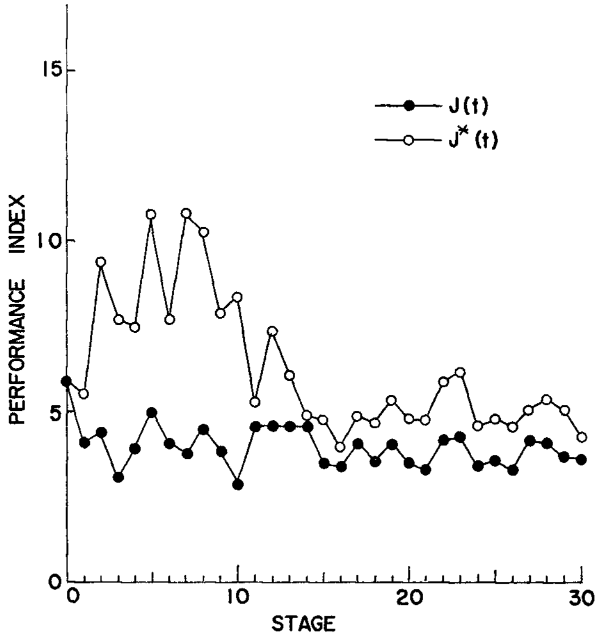

Fig. 5. Performance indices $J(t)$ and $J^*(t)$ for case (1).

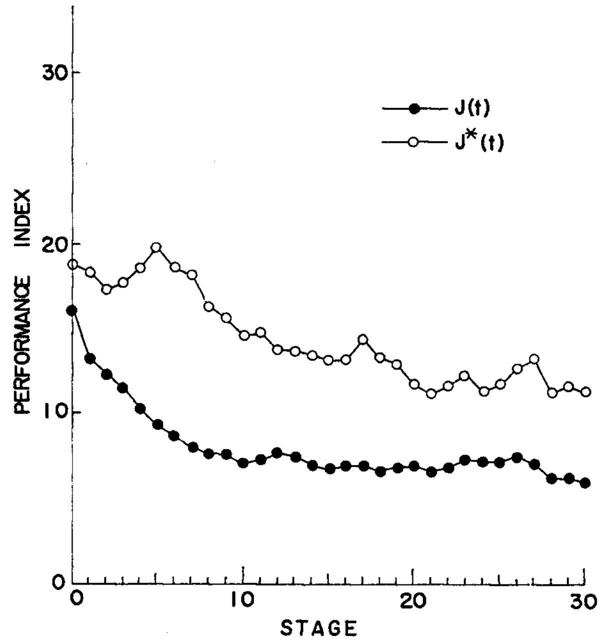

Fig. 6. Performance indices $J(t)$ and $J^*(t)$ for case (2).

$$a=0.98, \quad b=1, \quad c=1, \quad d=0,$$
$$Q(t)=Q=1.0, \quad R(t)=R=25.0, \quad S_1=0.2,$$
$$\bar{x}(0)=15, \quad \bar{P}(0)=50.$$

For case (2), the following set of numerical values is chosen:

$$a=0.98, \quad b=0, \quad c=1, \quad d=1,$$
$$Q(t)=Q=4.0, \quad R(t)=R=25.0, \quad S_2=25.0,$$
$$\bar{x}(0)=10, \quad \bar{P}(0)=50.$$

The initial value of the state $x(0)$ is sampled randomly at each experiment from the population with $N\{30,5\}$ for case (1), and with $N\{30,2\}$ for case (2). Figs. 3 and 4 display the sample paths of the estimate $\hat{x}(t)$ by the proposed algorithm, together with the estimate $\hat{x}^*(t)$ due to the linear estimator[3], and the actual state value $x(t)$. The performance of the two estimators is compared on the basis of the N sample root mean squares defined by

$$J(t) \triangleq \left\{\frac{1}{N}\sum_{i=1}^{N}[x(t)-\hat{x}(t)]^2_{(i)}\right\}^{1/2} \tag{6.3}$$

$$J^*(t) \triangleq \left\{\frac{1}{N}\sum_{i=1}^{N}[x(t)-\hat{x}^*(t)]^2_{(i)}\right\}^{1/2} \tag{6.4}$$

Here, $J(t)$ and $J^*(t)$ are the performance indices of the proposed estimator and the linear estimator[3] respectively, and the subscript i indicates the number of the simulation run. A total of 25 runs ($N=25$) with different noise samples are made in each Monte Carlo run. The results are shown in Figs. 5 and 6.

Although the justification of the assumption (4.1) has not yet been made, we can observe from Figs. 3 through 6 that the proposed estimator gives an improvement over the linear estimator given by McLane.

7. Conclusion

The estimation problem for the linear discrete-time systems with a state-dependent noise was considered. We have shown that the best estimate in the mean square sense is obtained as the weighted mean of estimates, each of which is accompanied by a specified state-dependent noise sequence. As the optimal estimator is not realistic from a practical point of view, we then have proposed an approximate method. The sub-optimal estimation algorithm has been derived by the same approach taken for the derivation of the optimal estimator. Therefore, the sub-optimal estimate is still nonlinear in the observation data, and the covariances should be obtained recursively using the observation data. Although the justification of the assumption (4.1) has not yet been made, the simulation studies demonstrate the practicability of the proposed sub-optimal estimator.

Acknowledgment

The authors would like to express their thanks to Mr. K. Yamanishi, an under-graduate student, who helped in the digital simulation.

References

1) R. E. Kalman: Trans. ASME, J. Basic Engrg., ser. D., **82**, 35 (1960).
2) R. E. Kalman and R. S. Bucy: Trans. ASME, J. Basic Engrg., ser. D., **83**, 95 (1961).
3) P. J. McLane: Int. J. Control, **10**, 41 (1969).
4) Y. C. Ho and R. C. K. Lee: IEEE Trans. Automatic Control, **9**, 333 (1964).
5) W. Feller: "An Introduction to Probability Theory and Its Applications", Vol. 2, John Wiley, New York (1966).

Appendix: Derivation of eqs. (4.14) and (4.15)

It is obvious that

$$p(x(t)|s(t+1), Y(t)) = \int p(x(t)|s(t), Y(t))p(s(t)|s(t+1), Y(t))ds(t). \tag{A.1}$$

Applying Bayes' rule again to $p(s(t)|s(t+1), Y(t))$, we have

$$p(x(t)|s(t+1), Y(t)) = \int p(x(t)|s(t), Y(t)) \frac{p(s(t)|Y(t))p(s(t+1)|s(t))}{\int p(s(t)|Y(t))p(s(t+1)|s(t))ds(t)} ds(t). \tag{A.2}$$

From this,

$$\hat{x}(t|s(t+1)) = \int \hat{x}(t|s(t)) \frac{p(s(t)|Y(t))p(s(t+1)|s(t))}{\int p(s(t)|Y(t))p(s(t+1)|s(t))ds(t)} ds(t). \tag{A.3}$$

This is eq. (4.14). By the definition of the conditional covariance,

$$\hat{P}(t|s(t+1)) = \int [x(t) - \hat{x}(t|s(t+1))][x(t) - \hat{x}(t|s(t+1))]' \times p(x(t)|s(t+1), Y(t))dx(t). \tag{A.4}$$

Substituting eq. (A.1) into eq. (A.4), we have

$$\hat{P}(t|s(t+1)) = \int \tilde{P}(t)p(s(t)|s(t+1), Y(t))ds(t) \tag{A.5}$$

where

$$\tilde{P}(t) = \int [x(t) - \hat{x}(t|s(t+1))][x(t) - \hat{x}(t|s(t+1))]'$$

$$\times p(x(t)|s(t), Y(t))dx(t), \tag{A.6}$$

which can be transformed to

$$P(t)=\int[x(t)-\hat{x}(t|s(t))][x(t)-\hat{x}(t|s(t))]'p(x(t)|s(t), Y(t))dx(t)$$
$$+[\hat{x}(t|s(t+1))-\hat{x}(t|s(t))][\hat{x}(t|s(t+1))-\hat{x}(t|s(t))]'. \tag{A.7}$$

Eqs. (A.5) and (A.7) construct eq. (4.15).