Suboptimal Selection of Observation Signals in a Linear Stochastic System

By Shogo TANAKA**, Tsuyoshi OKITA**
and Yasuhide KOBAYASHI**
(Received July 11, 1981)

Abstract

A convenient suboptimal method of selection of observation signals in time and space is proposed which minimizes the mean square error of the state estimation under a restricted number of observations.

Two different approaches are taken to obtain the method, dependent on the type of the transition matrix, i.e., it is diagonalizable or not.

Several examples are also given to illustrate the effectiveness of the proposed approach.

1. Introduction

In control problems, we often encounter optimization problems of an observation system. This is because of the requirement of controllers for improvement of the control performance as economically as possible.

Kushner[2] (1964) was the first researcher to consider such problems. He treated the optimal timing of observations under the restricted number of observations for a one-dimensional linear discrete control system. The performance index to be minimized was a terminal cost. The problem of optimal timing was extended by (Sano and Terao[4], 1970) to the continuous multi-dimensional case, where the performance index to be minimized was the quadratic form in state and control at each step. Recently, the problem of optimal timing in a linear discrete system was considered for the prediction of air pollution (Sawaragi and co-workers,[5] 1978).

These studies are all concerned with the optimal timing of observations. However, analytical general solutions have not been obtained except for the property of periodical observations for a one-dimensional special case (Kushner,[2] 1964).

On the other hand, the optimization of an observation system was formulated by (Meier, Peschon and Dressler,[3] 1967) in general form for a nonlinear discrete control system, where observation gain is also controllable by the so-called measurement-control. That is, they made the operation on extraction of observation signals possible not only in time, but also in space. Athans[1] (1972) extended the problem to a continuous but linear system and derived the equations satisfied by the optimal observation policy by using a matrix minimum principle. However, an analytical solution or any concrete method for easily obtaining a desirable observation policy has not been given because of the high nonlinearity of the observation process.

This paper is concerned with a linear discrete system and considers the optimization

** Department of Electronics Engineering

of an observation system such that desirable observation signals are selected not only in time, but also in space, in order that the mean square error of the state estimation is minimized under a restricted number of observations. This problem becomes essentially a combinatorial one. So, solving the problem completely in the numerically computational manner, enormous time will be consumed.

Thus, the paper proposes a convenient method to obtain the desirable observation policy with a low computational cost. Two different approaches are taken to obtain the convenient method dependent on the type of the system, i.e., the system transition matrix is diagonalizable or not. Finally, numerical examples are presented to illustrate the effectiveness of the proposed approach.

2. Problem Statement

Consider the linear discrete dynamical system defined by

$$X_{t+1}=AX_t+W_t \quad ; \quad 0 \leqq t \leqq T-1 \tag{1}$$

$$Y_t=HX_t+V_t \quad ; \quad 0 \leqq t \leqq T \tag{2}$$

where X_t is an n-dimensional state vector, Y_t is an m-dimensional observation vector (each component is called here OS), W_t and V_t are mutually independent white gaussian noises with zero-mean and covariance

$$E[W_t W_t'] = W , \quad E[V_t V_t'] = V \tag{3}$$

where E and "′" denote respectively the expectation and the transpose. It is further assumed that the initial state X_0 is a random variable (independent of both W_t and V_t) subject to a gaussian distribution with zero-mean and covariance $S_{0/-1}$.

Our problem is then to select p optimal OSS in time and space from the m(T+1) available OSS such that the following MSE of the state estimation becomes minimim.

$$J= \sum_{t=0}^{T} E[\| X_t-\hat{X}_{t/t} \|^2] \tag{4}$$

where $\| . \|$ and $\hat{X}_{t/t}$ denote respectively the Euclidean norm and the least square unbiased estimate of X_t based on the OSS used up to time t. Of course, p is the allowable number of observations and m(T+1) is the product of the number of different types of OSS, i.e., the dimension of Y_t, and the total number of sample times.

3. Approach to the Solution

Using a Kalman filter, equation (4) can be rewritten as

$$J = \sum_{t=0}^{T} tr \, [S_{t/t}] \tag{5}$$

where tr [•] means the trace of a matrix and $S_{t/t}$ expresses the estimation error covariance which can be computed based on the given observation policy by the following recursive equations.

$$S_{k/k}=S_{k/k-1}-S_{k/k-1}H_k'(H_kS_{k/k-1}H_k'+V_k)^{-1}H_kS_{k/k-1} \tag{6}$$

$$S_{k+1/k}=AS_{k/k}A'+W \quad ; \ k=0,1,\cdots,T \tag{7}$$

where H_k and V_k represent respectively the observation gain and the corresponding observation noise covariance at time k. It is obvious from (6) and (7) that the observation process has a historical property of high nonlinearity. So, it is in general very difficult to solve the above problem analytically.

Thus, the paper presents a convenient selection method, which is desirable from the viewpoints of both computational cost and optimality, by referring the results obtained in [6] (Tanaka, Okita and Matsuoka, 1981) for a continuous system.

We first consider the case where the transition matrix A is diagonalizable.

3.1. Case 1 (Diagonalizable Case)

Letting the eigenvalues of A be $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that $|\lambda_1|\geqq|\lambda_2|\geqq\ldots\geqq|\lambda_n|$ and the corresponding normalized eigenvectors of A be $\xi_1, \xi_2, \ldots, \xi_n$, then we have

$$A=P\Lambda P^{-1} \tag{8}$$

where

$$\Lambda=\begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix} \tag{9}$$

$$P=[\xi_1,\xi_2,\cdots,\xi_n] \tag{10}$$

Defining further the reciprocal vectors η_i (i=1, 2,..., n) such that

$$P^{-1}=[n_1,n_2,\cdots,n_n]' \tag{11}$$

,then the state X_t can be rewritten as

$$\begin{aligned} X_t&=A^tX_o+\sum_{k=0}^{t-1}A^kW_{t-1-k} \\ &=P[\Lambda^tP^{-1}X_o+\sum_{k=0}^{t-1}\Lambda^kP^{-1}W_{t-1-k}] \\ &=\sum_{i=1}^{n}[(n_i,X_o)\lambda_i^t+\sum_{k=0}^{t-1}(n_i,W_{t-1-k})\lambda_i^k]\ \xi_i \end{aligned} \tag{12}$$

where (.,.) denotes the inner product in Euclidean space. It is seen from (12) that no observation on [0, T] results in an increase of error covariance of the state vector X as time passes, caused by the initial error covariance of state and the accumulation of noises entered the system.

Considering the rate of increase of each component in ξ_i (i=1, 2,..., n), it is evident from the assumption $|\lambda_1| \geqq |\lambda_2| \geqq \cdots \geqq |\lambda_n|$ that the rate of increase is faster in the order of $\xi_1, \xi_2, \ldots, \xi_n$. Therefore, in the synthesis of the optimal observation policy which minimizes the performance (4), the following guide lines are obtained by referring to the discussions of paper [6]. First, each OS of the appropriately selected OSS $y^*_1, y^*_2, \ldots, y^*_n$, which decrease the error variance of ξ_i-component (i=1, 2,..., n) of X respectively, should be observed periodically in time. Second, it is desirable to allot more number of observations to OS in the order of $y^*_1, y^*_2, \ldots, y^*_n$. Third, the initial observation time OT should be selected appropriately for each OS y^*_i.

The above conclusions are obtained on the basis of the results of paper [6], where the observation noise was assumed to be zero for simplicity. However, noisy observation is usual in practical systems. When considering the effect of the presence of the observation noise, slight modifications are seen to become necessary in the optimal timing of observations especially when the system is stable. When the system is unstable, the policy of the optimal timing need not be changed. This is because of the relative weakness of the observation noise compared to the increasing error covariance of the system state. However, the situation is slightly different for the stable system with observation noise. As an example, we can imagine the case where observation noise is large and the initial error covariance of the state is larger than the stationary value in the case of no observation on [0, T]. In the case, a priori error covariance can not be repressed down below the stationary value at one observation at the start of [0, T], i.e., t=0. The desirable observation policy in that case is expected, considering the correspondence between the results in [6], to become one such that observation is done sequentially from the start of [0, T] until the MSE of the state estimation becomes smaller than the stationary value and that a periodical observation follows after the sequential one. However, in the case where observation noise variance is small or the initial error variance is smaller than the stationary value, the policy of the optimal timing need not be changed even when the system is stable. Although the analytical proofs are omitted here for its difficulties, these have been all verified by many numerical experiments.

Note that, as is expected, in the selection of desirable OSS $y^*_1, y^*_2, \ldots, y^*_n$, it is of course required not only to select OSS which decrease the error covariance of ξ_i-component (i=1,2,...,n) of X respectively, but also to select OSS whose redundancies are mutually weak. We present next a convenient method by which such desirable OSS $y^*_1, y^*_2, \ldots, y^*_n$ are easily obtained. First, we select y^*_1 such that the MSE of ξ_1-component of X

$$\begin{aligned} J\ (y^*_1\,;0 \leqq t \leqq T) &= \underset{1 \leqq j \leqq m}{Min}\quad J\ (y_j\ ;0 \leqq t \leqq T) \\ &= \underset{1 \leqq j \leqq m}{Min}\quad E\,[\,\sum_{t=0}^{T} \{(X_t - \hat{X}_{t/t})'\xi_1\}^2 / y_j;\ 0 \leqq t \leqq T\,] \qquad (13) \\ &= \underset{1 \leqq j \leqq m}{Min}\quad \sum_{t=0}^{T} tr\ [[\xi_1 \xi_1']\ S_{t/t}] \end{aligned}$$

is minimized under the assumption that OS is used throughout [0,T]. Next, y^*_2 is selected from the remaining OSS except for y^*_1 such that the MSE of $\{\xi_1, \xi_2\}$-component of X

$$J\,(y^*_1, y^*_2,\ 0 \leqq t \leqq T) = \underset{1 \leqq j \leqq m\,(Y_j \neq Y^*_1)}{Min} J\,(y^*_1, y_j;\ 0 \leqq t \leqq T)$$
$$= \underset{1 \leqq j \leqq m\,(Y_j \neq Y^*_1)}{Min} E\left[\sum_{t=0}^{T} \| (X_t - \hat{X}_{t/t})' \,[\xi_1 \xi^*_2]\, \|^2 / y^*_1,\ y_j;\ 0 \leqq t \leqq T\right] \tag{14}$$
$$= \underset{1 \leqq j \leqq m\,(Y_j \neq Y^*_1)}{Min} \sum_{t=0}^{T} tr\,[[\xi_1 \xi^*_2][\xi_1 \xi^*_2]'\ S_{t/t}]$$

is minimized. This time, the two OSS of a candidate of y^*_2 and OS y^*_1 are used together in [0,T] in order that the redundancy between them can be checked. By this procedure, OS which has much redundancy with y^*_1 is necessarily eliminated from the possible candidates of y^*_2. Note also that $\{\xi_1, \xi^*_2\}$ in equation (14) are orthonormal vectors constructed from $\{\xi_1, \xi_2\}$. Applying the Gram-Schmidt's orthogonalization technique, we obtain

$$\xi^*_2 = (\xi_2 - (\xi_2, \xi_1)\xi_1) \,/\, \| \xi_2 - (\xi_2, \xi_1)\xi_1 \| \tag{15}$$

It is obvious that $[\xi_1\ \xi_1']$ in (13) and $[\xi_1\, \xi^*_2]\,[\xi_1\, \xi^*_2]'$ in (13) are both the projectors on the subspace spanned by $\{\xi_1\}$ and the subspace spanned by $\{\xi_1, \xi_2\}$, respectively. Continuing the above procedure, we finally obtain n desirable OSS $y^*_1, y^*_2, \ldots, y^*_n$. This procedure allow us to pick up n desirable OSS with large signal-to-noise ratios and weak mutual redundancies. The procedure is also effective in the case of $m<n$, although $m \geqq n$ has been assumed so far for clarification of the foregoing discussions.

3.2. Case 2 (Non-diagonalizable Case)

We consider here the case where the transition matrix A is not diagonalizable. So, A is similar to the following so-called Jordan's canonical from

$$\Delta = \begin{pmatrix} \Delta_1 & & & O \\ & \Delta_2 & & \\ & & \ddots & \\ O & & & \Delta_\sigma \end{pmatrix} (n \times n);\quad \Delta_j = \begin{pmatrix} \Delta_{j1} & & & O \\ & \Delta_{j2} & & \\ & & \ddots & \\ O & & & \Delta_{j\alpha_j} \end{pmatrix} \begin{matrix} (m_j \times m_j) \\ j = 1, 2, \cdots, \sigma \end{matrix}$$
$$\Delta_{jk} = \begin{pmatrix} \lambda_j & 1 & & O \\ & \lambda_j & 1 & \\ & & \ddots & \ddots \\ & & & 1 \\ O & & & \lambda_j \end{pmatrix} (n_{jk} \times n_{jk}) \quad \begin{matrix} k = 1, 2, \cdots, \alpha_j \\ j = 1, 2, \cdots, \sigma \end{matrix} \tag{16}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_\sigma$ are the mutually distinct eigenvalues of A, α_j and $m_j = \sum_{k=1}^{\alpha_j} n_{jk}$ are respectively the geometric and algebraic multiplicities of the eigenvalue λ_j. We first consider the simplest case where A is similar to the following matrix for clarification of the succeeding discussions.

$$\Delta = \begin{pmatrix} \lambda & 1 & & 0 \\ & \lambda & 1 & \\ & & \ddots & \ddots \\ & & & 1 \\ 0 & & & \lambda \end{pmatrix} \qquad (n \times n) \tag{17}$$

By the similarity of A to (17), A is represented as follows.

$$A = Q \Delta Q^{-1} \tag{18}$$

where Q is defined by

$$Q = [q_1, q_2, \cdots, q_n] \tag{19}$$

where $q_1, q_2, \ldots, q_n$ are eigenvector or generalized eigenvectors of A corresponding to the eigenvalue λ. By similar expansions as used in obtaining (12), we have

$$X_t = Q[\Delta^t Q^{-1} X_0 + \sum_{k=0}^{t-1} \Delta^k Q^{-1} W_{t-1-k}] \tag{20}$$

From (20) and the property of the matrix Δ, it is easily seen that no observation on [0, T] yields not only the increase of error covariance of each component of X in q_1, $q_2, \ldots, q_n$ along the time axis, but also the accelerative spatial propagations of error covariance from the lower components to the upper components in the order of $q_n, q_{n-1}, \ldots, q_1$ (see Fig. 1). Here, by the upper components are meant the q_i-component whose subscript i is small.

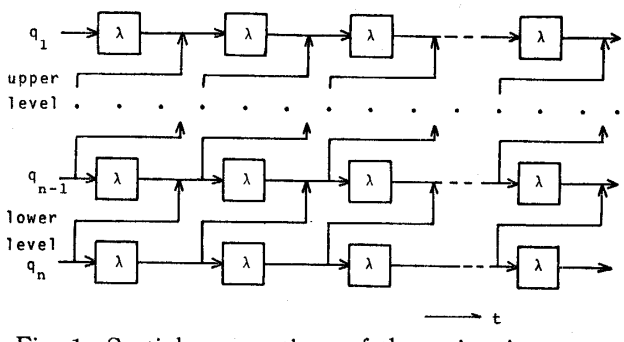

Fig. 1 Spatial propagations of the estimation error.

Therefore, it seems to be reasonable to cut off the spatial propagations periodically as well as the periodical repressions of the increasing error covariance along the time axis, in order to repress the performance (4) effectively under the restricted number of observations. Desirable OSS which cut off the spatial propagations effectively are those that minimize the MSE of q^*_i-component of $X(i=1, 2, \ldots, n^*)$, where $\{q^*_i\}$ are eigenvectors picked up from the eigenvector-set $\{q_1, q_2, \dots, q_n\}$ with each equal intervals such as $\{q^*_i\}=\{q_2\}$ for $n=3$ and $\{q^*_i\}=\{q_3\}$ or $\{q^*_i\}=\{q_2, q_4\}$ for $n=5$. Each OS thus selected then must be used periodically in time in order that the increase of the error covariance along the time axis can be repressed as well as the repressions of the dynamical spatial propagations. In the case where many OSS are available, the procedure descrived in 3.1 may serve as the method to pick up such n^* desirable OSS.

It is to be expected that the number of different type of OSS, i.e., n^*, depends on the total allowable number of observations p. It should be, however, noted that the number n^* for the system with small λ is larger than (or equal to) that of the system with large λ. This is because that in the former it is desirable to cut off the spatial propagations as many times as possible, since the cut-off in this case can be smoothly done only with a few periodical observations. On the contrary, in the latter, since the propagations can not be cut off by a few periodical observations, it is necessary to cut off the propagations completely with the allowable number of observations p concentrated on a few desirable OSS.

In the general case where A is similar to the matrix in (16), we need first to pick up desirable OSS for each minimum block Δ_{jK} as the same way as done for (17) previously and then to use them periodically in time respectively.

4. Numerical Examples

4.1 Diagonalizable Case

We consider the system

$$X_{t+1}=\begin{pmatrix} 1.5 & 0 & 0 \\ 0.5 & 1 & -0.5 \\ 0 & 0 & 0.5 \end{pmatrix} X_t+W_t \qquad ; 0\leqq t\leqq T-1 \tag{21}$$

$$Y_t=\begin{pmatrix} 0 & 1 & 0 \\ 0.65 & 0.65 & 0 \\ 0 & 0 & 1 \\ 0.6 & 0.5 & 0 \end{pmatrix} X_t+V_t \qquad ; 0\leqq t\leqq T \tag{22}$$

where $\Sigma_{0/-1}=0.01\ I_3$, $W=0.1\ I_3$, $V=0.025\ I_4$, $T+1=20$. The transition matrix A has three mutually distinct eigenvalues $\lambda_1=1.5$, $\lambda_2=1.0$, $\lambda_3=0.5$. Hence, A is diagonalizable. The corresponding eigenvectors to the three eigenvalues are respectively

$$\xi_1=(1/\sqrt{2},\ 1/\sqrt{2},\ 0)',\ \xi_2=(0,\ 1,\ 0)'$$

and

$$\xi_3=(0,\ 1/\sqrt{2},\ 1/\sqrt{2})' \tag{23}$$

Following the procedure descrived in section 3.1, we obtain the following three desirable

OSS

$$y_1^* = y_2, \quad y_2^* = y_1, \quad y_3^* = y_3 \tag{24}$$

This procedure does not allow us to pick y_4, which has much redundancy between the OS $y^*_1(=y_2)$, as a candidate of such desirable OSS. Now, having obtained the three desirable OSS y^*_1, y^*_2, y^*_3, we next optimize the allotment of observations to these OSS under the assumption of periodical observation for each OS and the constraint concerning the observation allotment $m_1 \geq m_2 \geq m_3$. Performance indices are plotted in Fig. 2 for

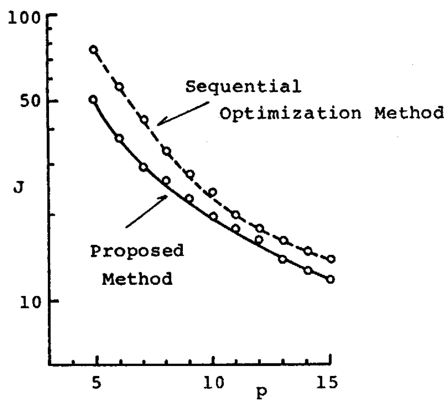

Fig. 2 Comparison of the mean square estimation errors for the two observation policies.

number of observations $p=5,6,\ldots,15$. Observation policy, for example, for $p=8$ is the observations of y_2 $(=y^*_1)$ and y_1 $(=y^*_2)$ at the times {3,6,9,12,15,18} and {2,11}, respectively.

Next, we try to apply the sequential optimization method for the selection of such desirable OSS to verify the effectiveness of the proposed approach. This sequential optimization method is a convenient one by which desirable OSS are selected sequentially one by one such that performance index becomes minimum. Observation policy, for example, for $p=8$ by this method is the observations of y_4, y_1 $(=y^*_2)$, and y_2 $(=y^*_1)$ at the times {4,7,11,16,19}, {8,15}, and {2}, respectively. We see that OS y_4 is observed with a high frequency, while the signal y_4 is not entirely used in the corresponding observation policy by the proposed method. The optimal performance indices by this sequential optimization method are also shown in Fig. 2 for $p=5,6,\ldots,15$. This figure shows us that proposed one is far superior to the sequential optimization one.

4.2. Non-diagonalizable case

We consider the system

$$X_{t+1} = \begin{pmatrix} 1.02 & 1 & 0 \\ 0 & 1.02 & 1 \\ 0 & 0 & 1.02 \end{pmatrix} X_t + W_t \quad ; \quad 0 \leq t \leq T-1 \quad ; \tag{25}$$

$$Y_t = X_t \quad ; \quad 0 \leq t \leq T \tag{26}$$

where $\Sigma_{0/-1}=O$, $W=0.05\ I_3$ and $T+1=100$. The a priori error covariance and the observation noise are assumed to be zero for simplicity. We see that the transition matrix A is the same matrix as in (17). Hence, A has only one distinct eigenvalue, i.e., $\lambda=1.02$, and the geometric multiplicity of it is also one. Therefore, the eigenvector of A is

$$q_1=(1,0,0)'$$

and the remaining generalized eigenvectors are

$$q_2=(0,1,0)' \text{ and } q_3=(0,0,1)'$$

Since $n=3$, it is obvious from the discussions in 3.2 that it is necessary to observe OS y_2 with a high frequency in order to repress the spatial propagations of the estimation

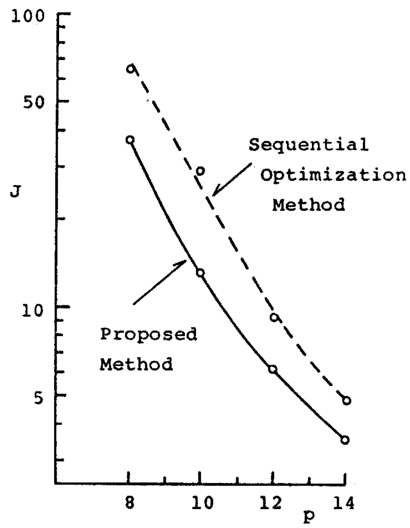

Fig. 3 Comparison of the mean square estimation errors for the two observation policies.

Table 1 Suboptimal observation policy by proposed method.

Observation Number	Observation Signals	Observation Times
	y_1	50
8	y_2	12 25 37 50 62 75 87
	y_3	
	y_1	50
10	y_2	10 20 30 40 50 60 70 80 90
	y_3	
	y_1	33 67
12	y_2	9 18 27 36 46 55 64 73 82 91
	y_3	
	y_1	50
14	y_2	7 14 21 28 35 43 50 57 64 72 78 86 93
	y_3	

error. This is because y_2 is the OS which observes the q_2-component of X. To check the validity of the argument, the allotment of the number of observations p to every OSS y_1, y_2 and y_3 is optimized for $p=8,10,12,14$ under the assumption that each OS is periodically observed. The optimized observation times for each OS and the performance indices are shown in Table 1 and Fig. 3, respectively. It is apparent from Table 1 that OS y_2 is observed in practice with a high frequency, confirming our conjecture. However, it should be noted that OS y_1, which observes the q_1-component of X, is also used, although the number of observations is very small. This phenomenon has been seen in many other numerical experiments to occure to unstable system with a small allowable number of observations. The situation does not occure to stable systems. This is interpreted by assuming that the observation of y_1 plays a role of repressing the escaped equivocations from the observation of y_2 caused by the small number of allowable observations p.

For comparison, the performance indices by the sequential optimization method are also shown in Fig. 3. From Table 1 and Fig. 3, the effectiveness of the proposed approach is verified, if allowance is made for the slight modification mentioned above to the es-

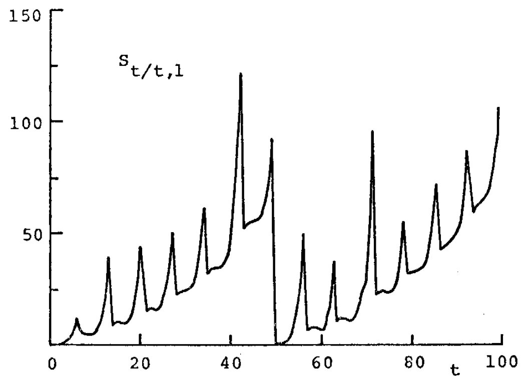

Fig. 4 Mean square error of q_1-component (proposed method).

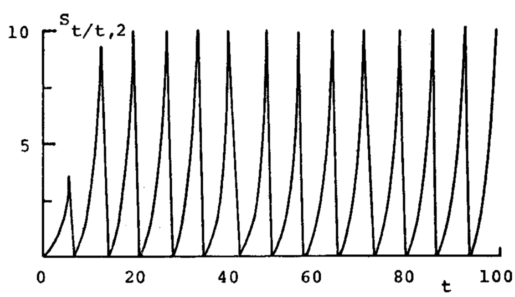

Fig. 5 Mean square error of q_2-component (proposed method).

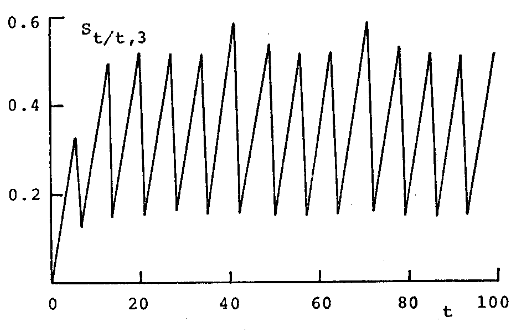

Fig. 6 Mean square error of q_3-component (proposed method).

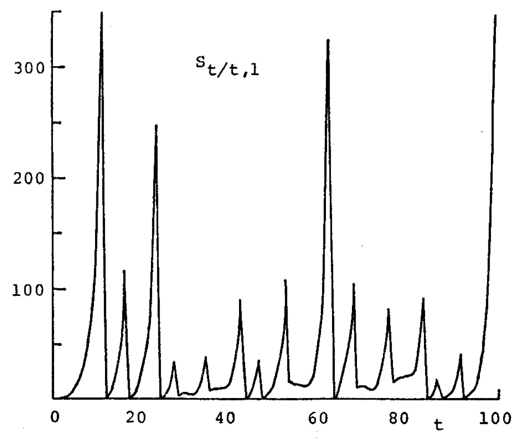

Fig. 7 Mean square error of q_1-component (sequential optimization method).

caped error covariance.

For reference, the transitions of $S_{t/t}$ for p=14 are shown in Fig. 4-7 for the two methods. These figures tell us that performance index is influenced by the behaviour of $s_{t/t,1}$ (MSE of q_1-component of X). Furthermore the performance index for the sequential optimization method is remarkably deteriorated by the irregularity of $s_{t/t,1}$, while performance index for the proposed method is considerably minimized because of the well-balanced repression of $s_{t/t,1}$, which is caused by the many regular (or periodical) observations of q_2-component of X. From these figures, it is confirmed that the periodical observations of OSS which cut off the spatial propagations are very effective in minimizing the performance index.

6. Conclusions

Two convenient methods to select the desirable p observation signals in time and space for both diagonalizable and nondiagonalizable systems have been proposed.

The proposed methods were not only desirable because the synthesis of the observation policy was non time-consuming (of the periodical observations and the convenient ordering method of observation signals to be used), but as numerical examples showed desirable from the viewpoint of optimality.

The proposed approach is expected to give a guide to the synthesis of an optimal observation policy for the systems where observation gain is controllable by observation control (Meier and co-workers[3], 1967; Athans[1], 1972).

References

1) Athans, M. "On the determination of optimal costly measurement strategies for linear stochastic systems", Automatica, 8, 397-412 (1972).
2) Kushner, H. J. "On the optimum timing of observations for linear control systems with unknown initial state", IEEE Trans. Aut. Control, 9, 144-155 (1964).
3) Meier, L., Peschon, L. and Dressler, R. M. "Optimal control of measurement subsystems", IEEE Trans. Aut. Control, 12, 528-536 (1967).
4) Sano, A. and Terao, M. "Measurement optimization in optimal process control", Automatica, 6, 705-714 (1970).
5) Sawaragi, Y., Soeda, T., Tomita, Y. and Imai, I. "Optimal timing of the observation for the state estimation and control of the stochastic discrete linear system", Int. J. Control, 27, 621-637 (1978).
6) Tanaka, S., Okita, T. and Matsuoka, S. "Optimal timing of observations in a linear discrete dynamical system", IEE of Japan, 101-C, 65-72 (1981).