J. Operations Research Soc. of Japan
Vol. 18, No.1 & 2, July 1975

MAXIMUM LIKELIHOOD IDENTIFICATION OF NOISE STATISTICS AND ADAPTIVE PREDICTION *

MASATO KODA

The University of Tokyo

(Received July 9; Revised November 21, 1973)

Abstract

In this paper, the noise adaptive prediction problem is solved for a general class of linear dynamical systems with noisy measurements when the statistics of the measurement noise, i. e. the mean and the covariance of the noise, are either unknown or known only imperfectly. The mathematical model for the measurement mechanism is described by stochastic linear difference equation. The criterion of the maximum likelihood is used to obtain the sequential algorithm for the noise adaptive prediction. Formulation is given in an optimization problem which can be decomposed in the identification of noise covariance and the simultaneous estimation of future state and noise mean. Application of the discrete maximum principle results in a two-point boundary value problem. Based upon the method of discrete invariant imbedding, the recursive solution for the noise adaptive prediction is derived. For the purpose of exploring quantitative aspects, numerical example by digital simulation is presented. It has been demonstrated that the present algorithm is preferable to existing techniques of Kalman filter.

* The parts of this paper were presented at the 10th International Symposium on Space Technology and Science, Tokyo, Sept. 3-8, 1973.

1. Introduction

In general, the estimation based on the maximization of the conditional probability density function is called the maximum likelihood (Bayesian), or most probable estimate. That is, the estimate is the peak or "mode" of the conditional probability density. In this sense, if noise is Gaussian, the optimal (minimum variance) filter which has been introduced by Kalman (6) (7) into the field of systems theory is essentially equivalent to maximum likelihood filter.

The original formulation of the Kalman filter has assumed an exact knowledge of the statistics of the measurement and plant noise. However, under a number of actual operational situations, the noise statistics that are used in the filter are in fact only a priori estimates of the noise statistics that will actually be encountered in the future. In some cases these prior noise statistics might be quite accurate, but in other cases they might be sufficiently in error to adversely affect the filter. One serious effect of this can be the large discrepancy between the computed covariance matrix of the state estimation error within the filter and the "actual" covariance matrix. Naturally this results in a growth of the estimation errors; the estimated states will diverge. This fact renders the operations of the Kalman filter unsatisfactory when the statistics of the noise are either unknown or known only imperfectly.

Possible remedy for the difficulty of divergence may be the noise adaptive estimation techniques. Kashyap (8) and Mehra (11) have considered the noise adaptive filtering by identifying the noise covariance matrix. On the other hand, Lin and Sage (10) have proposed the adaptive bias filter when the noise mean vectors are unknown. In spite of its particular importance in the actual application, however, we have found very few paper studying the estimation problem where the mean and the covariance of the noise are both unknown. Taking into account the fact that the mean and the covariance are the statistical parameters which completely characterize the distribution of Gaussian noise sequence, this problem is greatly important in the actual design of the estimator.

The object of the present paper is to develop an advanced estimation algorithm for the state of a linear dynamical system when both the mean and the covariance of the measurement noise are not sufficiently known for an adequate stable solution. We shall concentrate our attention to the noise adaptive prediction problem, that is the simultaneous estimation of the future state and the noise statistics. This problem would be particulary important in the systems theory.

The criterion of the maximum likelihood is used to obtain the recursive solution for the noise adaptive prediction. The use of the conditional probability or expectation in this work is somewhat unconventional because the parameters of the noise statistics, i. e. the mean and the covariance, are to be estimated on the basis of a realization of the measurements which is itself a function of these parameters. However, the application of the principle of maximum likelihood reveal that the noise adaptive estimation can be separable into the identification of the parameters of the noise statistics and the estimation of the future state. This principle also leads to the evaluation of cost functional by the so-called generalized variance, which is of determinant form, and its relation to "entropy" is shown.

We first reduce the original prediction problem to more general problem of adaptive filtering, and formulate it as a discrete dynamic optimization problem. Application of discrete maximum principle results in a two-point boundary value problem, then it is resolved by the method of discrete invariant imbedding as to obtain adaptive predictor of recursive structure. Numerical example is also simulated to illustrate the quantitative aspects of the present noise adaptive predictor.

2. Statement of the Problem

We consider a general class of linear discrete dynamical systems, defined by

(2.1) $$x(k+1)=\Phi(k+1,k)x(k)$$

and measurement mechanism described by

(2.2) $$z(k)=Hx(k)+v(k)$$

where $x(k)$ is $n \times 1$ state vector of the system at time k, Φ is $n \times n$ nonsingular state transition matrix, z is $m \times 1$ measurement vector, and H is $m \times n$ constant measurement matrix.

The sequence $\{v(k),\ k=1,2,\dots\}$ is $m \times 1$ white Gaussian sequence with unknown mean vector μ and unknown $m \times m$ covariance matrix R.

(2.3) $$v(k) \sim N[\mu(k),R(k)]$$ [1)]

The distribution of the initial condition is normal,

(2.4) $$x(0) \sim N[\hat{x}(0),P(0)]$$

and assumed given, and $\hat{x}(0)$ is independent of the sequence $\{v(k),\ k=1,2,\dots\}$.

The system is assumed to be completely observable [2)] and R is bounded positive definite. Here we note that the system (2.1) and (2.2) is described by linear time-varying equations; this might result from linearizing the system dynamics around some reference trajectory. And we do not consider the additive plant noise sequence in (2.1). In the actual cases such noise might be imposed, however, this can be treated by a simple modifications of the techniques which are given in the subsequent development.

---

1) Here and throughout the paper, the symbol $N(\mu, R)$ denotes the normal distribution with the mean vector μ and the covariance matrix R.

2) In the modern control theory, the concepts of "complete controllability" and "complete observability" play an important role. For further information, see R. E. Kalman, J. SIAM Control, pp. 152-192, vol. 1, No. 2, 1963.

Let Z_k be the sequence of observations as

$$Z_k=\{z(1),z(2),\ldots,z(k)\}. \tag{2.5}$$

Given a realization of Z_k, the noise adaptive prediction problem consists of the simultaneous computation of the optimal estimates of $\mu(k)$, $R(k)$, and $x(\bar{N})$ based on Z_k. $\bar{N}$ is the fixed future time of interest when the prediction estimate of the state $x(\bar{N})$ is required ($\bar{N} > k$).

If the mean and the covariance of measurement noise are assumed to be known, then using (2.3), it is very easy to obtain the prediction estimate of $x(\bar{N})$ given Z_k by appropriately manipulating the standard Kalman filtering equations. Let $\hat{x}(\bar{N}|k)$ be an optimal prediction estimate (obtained from a Kalman filter) of $x(\bar{N})$ based on the observations up to and including the current time k. Then the covariance matrix of the prediction error can be computed as

$$C(k)=E\{[x(\bar{N})-\hat{x}(\bar{N}|k)][x(\bar{N})-\hat{x}(\bar{N}|k)]^T\} \tag{2.6}$$

where $E\{\cdot\}$ denotes the expectation operator. Then it can be shown that $\hat{x}(N|k)$ is obtained from the following recursive algorithms.

$$\hat{x}(\bar{N}|k+1)=\hat{x}(\bar{N}|k)+K(k+1)[z(k+1)-H\Phi^{-1}(\bar{N},k+1)\hat{x}(\bar{N}|k)-\mu(k+1)] \tag{2.7}$$

$$C(k+1)=[I-K(k+1)H\Phi^{-1}(\bar{N},k+1)]C(k)[I-K(k+1)H\Phi^{-1}(\bar{N},k+1)]^T +K(k+1)R(k+1)K^T(k+1) \tag{2.8}$$

$$K(k+1)=C(k)\Phi^{-T}(\bar{N},k+1)H^T \times[H\Phi^{-1}(\bar{N},k+1)C(k)\Phi^{-T}(\bar{N},k+1)H^T+R(k+1)]^{-1} \tag{2.9}$$

We note that K in (2.9) is usually called the Kalman gain and it can be considered as a weighting factor for the measurement.

For the convenience of the formulation, it is further assumed that the mean and the covariance of the measurement noise are constant. From (2.3) this implies that

(2.10) $$\mu(k+1)=\mu(k)$$

(2.11) $$R(k+1)=R(k).$$

They are, in fact, unkown parameters which completely characterize the statistics of the measurement noise and to be estimated by maximum likelihood techniques.

3. Maximum Likelihood Identification of Noise Statistics

Maximum likelihood techniques are concerned with finding the maximum of a likelihood function defined as a natural logarithm of the conditional probability density. If a priori information about the statistics of the noise can not be used, the proper likelihood function should be

(3.1) $$L_k(Z_k,\mu,R) = \ell n\ p(Z_k|\mu,R),$$

for constant unknown mean μ and covariance R, where Z_k is defined by (2.5) using (2.1) and (2.2). Then the maximum likelihood identification of the noise statistics consists of finding μ and R such that

(3.2) $$L_k(Z_k,\hat{\mu},\hat{R}) = \max_{\mu,R} L_k(Z_k,\mu,R).$$

Using these quantities in (3.2), an optimal estimate of $x(\bar{N})$ based on Z_k can be shown to give

(3.3) $$\hat{x}(\bar{N}|k) \rightarrow \hat{x}[\bar{N}|k;\hat{\mu}(k),\hat{R}(k)].$$

This depicts that the estimate of $x(\bar{N})$ is just the maximum likelihood estimator of the state that uses the estimates of μ and R to compute the proper estimation gain. Thus, the design of a noise adaptive predictor can be separated into the identification of μ and R, and the estimation of $x(\bar{N})$.

By repeated application of Bayes' rule, the conditional probability density in (3. 1) can be rewritten in the following form.

$$p(Z_k|\mu,R)=\prod_{i=1}^{k} p(z(i)|Z_{i-1},\mu,R) \tag{3.4}$$

Therefore, in order to compute the likelihood function (3. 1), we need the conditional probability density of $z(k+1)$ given all previous observations Z_k. By Gaussian assumption, the conditional probability density has the form

$$p(z(k+1)|Z_k,\mu,R) \sim N[\hat{z}(k+1|k),V(k+1|k)], \tag{3.5}$$

where

$$\begin{aligned}\hat{z}(k+1|k)&=E\{z(k+1)|Z_k,\mu,R\}\\&=H\Phi^{-1}(\bar{N},k+1)x(\bar{N}|k)+\mu(k)\end{aligned} \tag{3.6}$$

and

$$\begin{aligned}V(k+1|k)&=E\{[z(k+1)-\hat{z}(k+1|k)][z(k+1)-\hat{z}(k+1|k)]^T\}\\&=H\Phi^{-1}(\bar{N},k+1)C(k)\Phi^{-T}(\bar{N},k+1)H^T+R(k)\end{aligned} \tag{3.7}$$

where we used system models defined by (2. 1) and (2. 2) together with (2. 6) and (2. 7).

In general, when k is sufficiently large and "near" $\bar{N}$ ($\bar{N} > k$), one can show that

$$\lim_{k\to\infty} \Phi^{-1}(\bar{N},k)=I$$

$$\lim_{k\to\infty} V(k+1|k)=HC_\infty H^T+R_\infty=V_\infty>0$$

where V_∞ , R_∞ , are constant materices. Then, in view of (3. 1), (3. 4), and (3.5), we can easily obtain the expression for the likelihood for sufficiently large N ($N < \bar{N}$; N can be considered as the final time of observation),

$$L_N(Z_N, \mu, R) = -\frac{mN}{2}\ln 2\pi - \frac{N}{2}\ln(\det\{V_\infty\}) - \frac{1}{2}\sum_{j=1}^{N} ||\nu(j|j-1)||^2_{V_\infty^{-1}} \tag{3.8}$$

where we have chosen $z(1|0)$ arbitrary, and have defined the measurement residual $\nu(k+1|k)$ as

$$\nu(k+1|k) = z(k+1) - \hat{z}(k+1|k). \tag{3.9}$$

This measurement residual may be the most important random variables upon which the recursive maximum likelihood state and noise estimation can be based.

Now, we can identify the estimate of R by maximizing the likelihood function (3.8) with respect to R_∞. Using the concept of "gradient matrix" (1), the likelihood equation is obtained by equating the derivative of (3.8) with respect to R_∞ to zero.

$$\frac{\partial L_N}{\partial R_\infty} = \frac{1}{2}V_\infty^{-1}\{\sum_{j=1}^{N}\nu(j|j-1)\nu^T(j|j-1)\}V_\infty^{-T} - \frac{N}{2}V_\infty^{-T} = 0 \tag{3.10}$$

Regarding V_∞ as the estimate of $V(N|N-1)$, we can obtain from (3.10)

$$\hat{V}(N|N-1) = \frac{1}{N}\sum_{j=1}^{N}\nu(j|j-1)\nu^T(j|j-1). \tag{3.11}$$

In view of the ergodic property of a stationary random sequence [3], (3.11)

---

3) For the rigorous discussion of the ergodic property of the random sequence, consult Loeve, Probability Theory, 3rd ed. Van Nostrand, New York, 1963.

is quite natural as an optimal estimate of $V(N|N-1)$. Hence using (3.7), the optimal estimate of noise covariance matrix becomes

$$(3.12) \qquad \hat{R}(N)=\hat{V}(N|N-1)-H\Phi^{-1}(\bar{N},N)C(N-1)\Phi^{-T}(\bar{N},N)H^{T}.$$

Replacing V_∞ and R_∞ in (3.8) by (3.11) and (3.12), we obtain the new expression for the likelihood,

$$(3.13) \qquad L_N(Z_N,\mu,\hat{R}) = -\frac{mN}{2}\ell n 2\pi e - \frac{N}{2}\ell n(\det\{\hat{V}(N|N-1)\}).$$

Then maximizing (3.13) with respect to μ and $\hat{x}(\bar{N}|N)$ is equivalent to minimizing $\det\{\hat{V}(N|N-1)\}$. Hence, from (3.11) and (3.13), we are led to evaluate the cost functional of determinant form which is essentially different from the usual quadratic cost.

$$(3.14) \qquad J= \frac{1}{N}\sum_{j=1}^{N}\det\{\nu(j|j-1)\nu^{T}(j|j-1)\}$$

Here we note that the particular expression for (3.13) can also be derived from the definition of the entropy which is the most general measure of the amount of information (see Appendix). With this fact and also that (3.14) is related to the average volume of the error ellipsoid of the measurement residual make (3.14) the most reasonable cost for the maximum likelihood estimation. And (3.14) guarantees a solution in the limit $N\to\infty$.

4. Applications of Maximum Priciple and Invariant Imbedding

For the convenience of the formulation, we rewrite (2.7), (2.8), and (2.9) in different forms:

$$(4.1) \qquad \hat{x}(\bar{N}|k+1)=\Theta(k+1)\hat{x}(\bar{N}|k)+G(k+1)z(k+1)-G(k+1)\mu(k)$$

(4.2) $$C(k+1)=\Theta(k+1)C(k)\Theta^T(k+1)+G(k+1)\hat{R}(k+1)G^T(k+1)$$

(4.3) $$G(k+1)=C(k)\Phi^{-T}(\bar{N},k+1)H^T\hat{V}^{-1}(k+1|k)$$

(4.4) $$\Theta(k+1)=I-G(k+1)H\Phi^{-1}(\bar{N},k+1)$$

where we use the estimates (3.11) and (3.12). Then a precise statement of the noise adaptive prediction problem is to minimize the cost functional (3.14) with respect to μ and $\hat{x}(N|k)$, subject to the constraints of (2.4), (2.10), and (4.1). Thus, the problem of maximum likelihood estimation of the state and the noise mean has been reduced to an optimal control problem.

The problem under consideration falls naturally into the framework of the discrete maximum principle (3). We assume that the problem is nonsingular and that the convexity conditions associated with the discrete maximum principle are met. Thus we are free to apply the discrete maximum principle. Let us define the Hamiltonian

$$\begin{aligned}\mathcal{H}(k)=&\det\{\nu(k+1|k)\nu^T(k+1|k)\}\\&+\lambda^T(k+1)[\hat{x}(\bar{N}|k+1)-\hat{x}(\bar{N}|k)]\\&+\omega^T(k+1)[\mu(k+1)-\mu(k)]\end{aligned}$$

We can then assert that there exist adjoint vectors λ and ω such that

$$\lambda(k+1)-\lambda(k)=-\frac{\partial\mathcal{H}(k)}{\partial\hat{x}(\bar{N}|k)}$$

$$\omega(k+1)-\omega(k)=-\frac{\partial\mathcal{H}(k)}{\partial\mu(k)}$$

Thus we obtain the adjoint equations

(4.5) $$\begin{aligned}\lambda(k+1)=&\Theta^{-T}(k+1)\lambda(k)-\Gamma(k+1)H\Phi^{-T}(\bar{N},k+1)\hat{x}(\bar{N}|k)\\&-\Gamma(k+1)\mu(k)+\Gamma(k+1)z(k+1)\end{aligned}$$

$$\omega(k+1)=\omega(k)+G^T(k+1)\Theta^{-T}(k+1)\lambda(k) \tag{4.6}$$
$$-\Lambda(k+1)H\Phi^{-1}(\overline{N},k+1)\hat{x}(\overline{N}|k)$$
$$-\Lambda(k+1)\mu(k)+\Lambda(k+1)z(k+1)$$

where

$$\Gamma(k+1)=2\Delta\Theta^{-T}(k+1)\Phi^{-T}(\overline{N},k+1)H^TW^{-1}(k+1) \tag{4.7}$$

$$\Lambda(k+1)=G^T(k+1)\Gamma(k+1)+2\Delta W^{-1}(k+1) \tag{4.8}$$

$$W(k+1)=\nu(k+1|k)\nu^T(k+1|k) \tag{4.9}$$

$$\Delta=\det\{W(k+1)\}. \tag{4.10}$$

Then the problem turns out to a two-point boundary value problem with the following associated boundary condition

$$\begin{aligned} &\hat{x}(\overline{N}|0)=\Phi(\overline{N},0)\hat{x}(0) \\ &C(0)=\Phi(\overline{N},0)P(0)\Phi^T(\overline{N},0) \\ &\mu(0)=\hat{\mu}(0) \\ &\lambda(0)=0, \qquad \lambda(N)=0 \\ &\omega(0)=0, \qquad \omega(N)=0. \end{aligned} \tag{4.11}$$

If this two-point boundary value problem is solved for $k \in [0,N]$, then the fixed interval smoothing solutions for $\hat{x}(N|k)$ and $\mu(k)$ are obtained.

In order to treat this two-point boundary value problem, we put it into the vector-matrix form such that (2.10) and (4.1) become

$$y(k+1)=A(k)y(k)+a(k) \tag{4.12}$$

where

$$y(k)=[\hat{x}(\bar{N}|k) \qquad \mu(k)]^T$$

$$a(k)=[G(k+1)z(k+1) \qquad 0\ ]^T$$

$$A(k)=\begin{bmatrix} \Theta(k+1) & -G(k+1) \\ 0 & I \end{bmatrix}. \tag{4.13}$$

Also, (4.5) and (4.6) become in vector-matrix notation

$$e(k+1)=B(k)e(k)+D(k)y(k)+b(k) \tag{4.14}$$

where

$$e(k)=[\lambda(k) \qquad \omega(k)]^T$$

$$b(k)=[\Gamma(k+1)z(k+1) \qquad \Lambda(k+1)z(k+1)]^T$$

$$B(k)=\begin{bmatrix} \Theta^{-T}(k+1) & 0 \\ G^T(k+1)\Theta^{-T}(k+1) & I \end{bmatrix} \tag{4.15}$$

$$D(k)=\begin{bmatrix} -\Gamma(k+1)H\Phi^{-1}(\bar{N},k+1) & -\Gamma(k+1) \\ -\Lambda(k+1)H\Phi^{-1}(\bar{N},k+1) & -\Lambda(k+1) \end{bmatrix}.$$

It is now desired to formulate the solution for the maximum likelihood estimates of $\hat{x}(\bar{N}|k)$ and $\mu(k)$ in a recursive manner. In order to obtain the recursive solution, we adopt the invariant imbedding (2). Suppose we solve the problem (4.11), (4.12), (4.14), and obtain the missing terminal condition on $y(N)$, denote it $y(N) = [\ \hat{x}(\bar{N}|N) \quad \mu(N)\ ]^T$. Then the two-point boundary value problem has to be resolved to produce

$$e(N)=0.$$

However, in general, $e(N)=c \neq 0$ and clearly $y(N)$ is the fuction of c and N:

(4.16) $$y(N)=r(c,N)$$

and naturally we have

$$\hat{y}(N)=r(0,N).$$

Thus we have imbedded the original problem in a family of problems parameterized by c.

When the method of discrete invariant imbedding is employed to (4.12) and (4.14), the evolution of r is governed by the invariant imbedding partial difference equation

(4.17) $$\frac{\delta r(c,N)}{\delta N}+\left\{\frac{\delta r(c,N)}{\delta c}+\frac{\delta^2 r(c,N)}{\delta c\,\delta N}\right\}\times[E(N)c+D(N)r(c,N)+b(N)-c]$$
$$=A(N)r(c,N)+a(N)-r(c,N)$$

where N is now considered as a running time k. We assume a solution for (4.17) of the form

(4.18) $$r(k)=\hat{y}(k)-S(k)c.$$

This is motivated from (4.16), since $\hat{y}(N)$ is obtained by setting $c=0$. Substitution of (4.18) into (4.17) and separating terms involving c from those not involving c results in the recursive formulations

(4.19) $$S(k+1)=A(k)S(k)[E(k)-D(k)S(k)]^{-1}$$

(4.20) $$\hat{y}(k+1)=A(k)\hat{y}(k)+a(k)+S(k+1)[D(k)\hat{y}(k)+b(k)].$$

Equations (4.19) and (4.20) give the recursive solutions to our two-point boundary value problem. When the appropriate interpretation of the measurement residual is used, we obtain the noise adaptive linear prediction algorithms.

5. Formulation of the Noise Adaptive Predictor

Using the solution of (4.20), we can redefine the measurement residual (3.9) as

$$\nu(k+1|k)=z(k+1)-H\Phi^{-1}(\bar{N},k+1)\hat{x}(\bar{N}|k)-\hat{\mu}(k). \tag{5.1}$$

And defining

$$\begin{aligned}\Sigma(k) &= \sum_{j=1}^{k}\{\nu(j|j-1)\nu^{T}(j|j-1)\} \\ &= \sum_{j=1}^{k} W(j)\end{aligned} \tag{5.2}$$

we can obtain the recursive relation,

$$\Sigma(k+1)=\Sigma(k)+W(k+1) \tag{5.3}$$

and also the expression for (3.11),

$$\hat{V}(k+1|k)=\frac{1}{k+1}\Sigma(k+1). \tag{5.4}$$

Then, using the newly defined measurement residual (5.1) together with (5.3) and (5.4), we can choose $y(k)$, $S(k)$, $C(k)$, and $\Sigma(k)$ as the augmented "state" of the noise adaptive linear predictor. Initial value for S is arbitrary, however, taking into account the fact that it is closely related to the estimation error covariance matrix of $\hat{y}$, it should be chosen adequately. Initial value of Σ is also arbitrary, but in view of (5.2) it can be automatically generated within the predictor without the initial value. This completely concludes the formulation. We summarize it in the following theorem.

THEOREM (Noise Adaptive Linear Predictor). Given $y(k)$, $S(k)$, $C(k)$, and $\Sigma(k)$, with the initial conditions

$$\hat{y}(0)=[\hat{x}(\bar{N}|0) \quad \hat{\mu}(0)]^T=[\Phi(\bar{N},0)\hat{x}(0) \quad \hat{\mu}(0)]^T$$

(5.5) $$C(0)=\Phi(\bar{N},0)P(0)\Phi^T(\bar{N},0)$$

$$S(0), \quad \Sigma(0).$$

Then the noise adaptive linear predictor for the system (2.1) and (2.2) consists of the following consecutive equations. Noise covariance identification by means of

(6.6) $$\nu(k+1|k)=z(k+1)-H\Phi^{-1}(\bar{N},k+1)\hat{x}(\bar{N}\ k)-\mu(k)$$

(5.7) $$W(k+1)=\nu(k+1|k)\nu^T(k+1|k)$$

(5.8) $$\Sigma(k+1)=\Sigma(k)+W(k)$$

(5.9) $$\hat{V}(k+1|k)=\frac{1}{k+1}\Sigma(k+1)$$

(5.10) $$\hat{R}(k+1)=V(k+1|k)-H\Phi^{-1}(\bar{N},k+1)C(k)\Phi^{-T}(\bar{N},k+1)H^T$$

(5.11) $$G(k+1)=C(k)\Phi^{-T}(\bar{N},k+1)H^T\hat{V}^{-1}(k+1|k)$$

(5,12) $$\Theta(k+1)=I-G(k+1)H\Phi^{-1}(\bar{N},k+1)$$

and

(5.13) $$C(k+1)=\Theta(k+1)C(k)\Theta^T(k+1)+G(k+1)R(k+1)G^T(k+1).$$

And defining $A(k)$, $B(k)$, $D(k)$, $a(k)$, and $b(k)$ as (4.13) and (4.15), the recursive equations for $\hat{y}(k)$ and $S(k)$ are given by

(5.14) $$\hat{y}(k+1)=[A(k)+S(k+1)D(k)]\hat{y}(k)+a(k)+S(k+1)b(k)$$

(5.15) $$S(k+1)=A(k)S(k)[B(k)-D(k)S(k)]^{-1}$$

where we have defined

(5.16) $$\hat{y}(k)=[\hat{x}(\bar{N}|k) \quad \hat{\Omega}(k)]^T.$$

The block diagram of the noise adaptive linear predictor is given in Fig. 1. The basic feature of this adaptive predictor is that the predictor can change (adapt) its structural parameters of the predictor itself in accordance with the changes in measurement residual. It should be noted that $\hat{x}(N|k)$, the solution of (5.14), is not the solution of (4.1) or (2.7) (which is obtained on the basis of the Kalman filter equation). Additive "adaptive" gain which is computed from S in (5.15) is imposed on the equation (5.14).

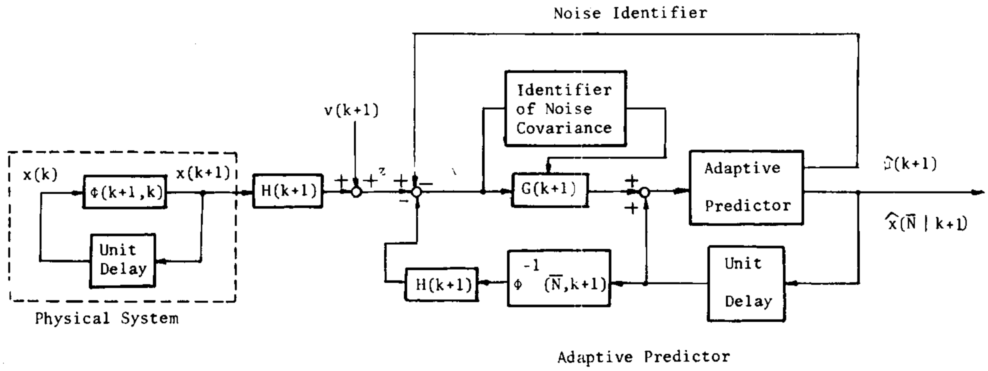

Fig. 1 Block Diagram of the Linear Adaptive Predictor

6. Numerical Example

An example has been presented to demonstrate the usefulness and effectiveness of the present noise adaptive predictor. For simplicity, the example treats the special case of scalar system.

In terms of the notation used in the previous development, we have

$$x(k+1) = 0.99x(k)$$
$$z(k) = x(k)+v(k)$$
$$v(k) \sim N(\mu, R) .$$

The true statistics for the measurement noise are

$$\mu = 0.5 , \quad R=1 .$$

The following set of values are specified throughout all the computations.

$$\bar{N}=N= 100$$
$$x(0)=10 , \quad x(100)=3.660323$$
$$\hat{\mu}(0)= 0$$

And three different sets of values for the initial conditions are selected.

Case 1.
$$\hat{x}(0)=10$$
$$\Sigma(0)= C(0)= 0$$
$$S(0)= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Case 2.
$$\hat{x}(0)=9.7268$$
$$\Sigma(0)=1$$
$$C(0)=10^{-4}$$
$$S(0)= \begin{bmatrix} C(0) & 0 \\ 0 & \Sigma(0) \end{bmatrix}$$

Case 3.
$$\hat{x}(0)=9$$
$$\Sigma(0)= C(0)=0$$
$$S(0)= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

The computations were carried out by HITAC 5020 system at the Computer Center of the University of Toky. Double precision arithmetic was used, and the Gaussian noise was generated by a standard subroutine. Results are shown in Fig. s 2 ∿ 5. The estimation of the noise covariance

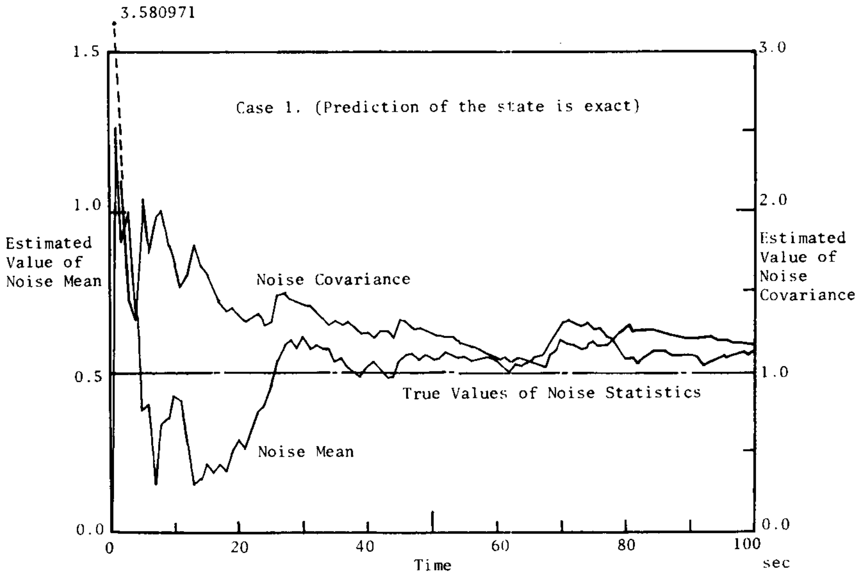

Fig. 2 Estimated Mean and Covariance of Measurement Noise (Case 1.)

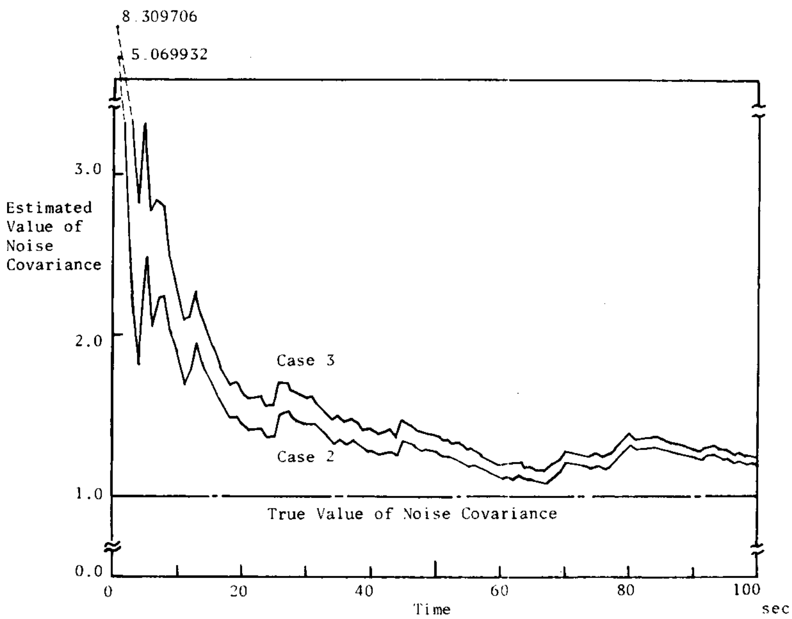

Fig. 3 Comparison of Estimated Covariances

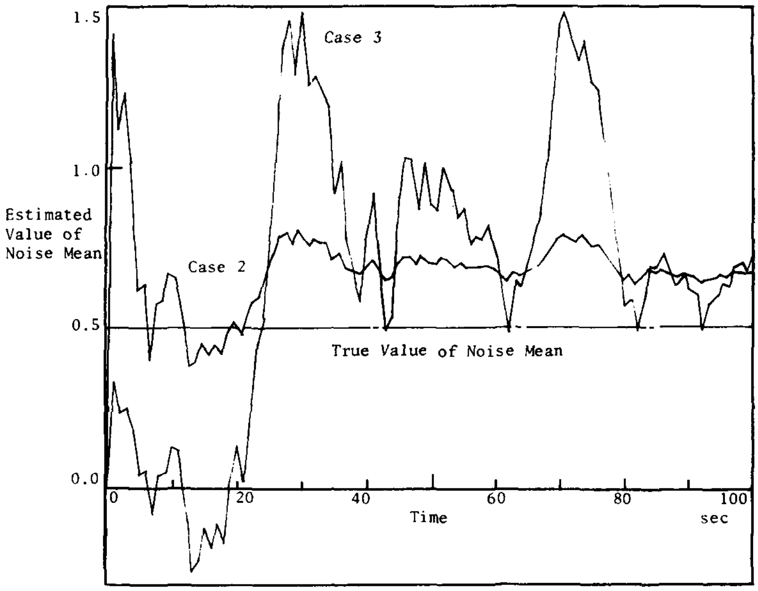

Fig. 4 Comparison of Estimated Means

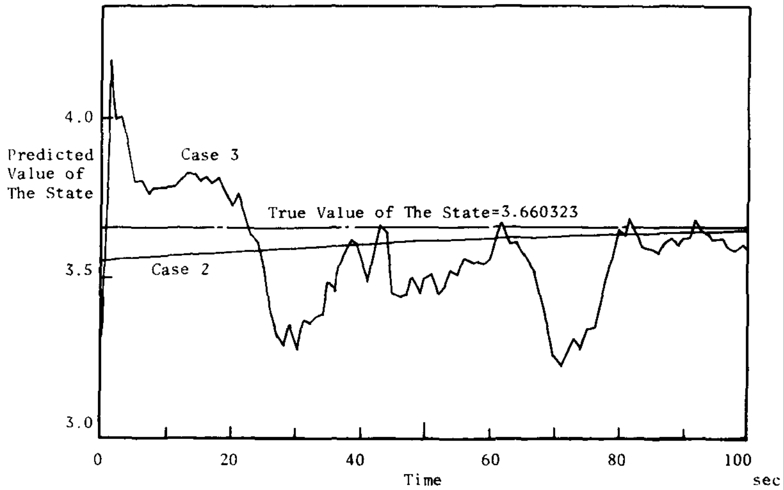

Fig. 5 Comparison of Predicted States

is very satisfactory, and displays almost the same characteristics for all the cases. The estimation of the noise mean is most satisfactory in Case 1. This is, of course, to be expected since there the prediction of the state is exact. Although it is extremely difficult to justify analytically the accuracy of the proposed algorithms, it can be concluded that the present noise adaptive algorithms are essentially as effective as the optimal (Kalman) algorithms that use the exact knowledge of the noise statistics.

By the digital simulations, it has been demonstrated that the present algorithms can be applied successfully to problems that lack the complete information of the noise statistics; these are problems on which usual formulation of the Kalman estimation procedure is often of little value. In fact, the application of the Kalman estimation algorithm has resulted in a large discrepancy between the estimates and the real values for all the cases. Thus it should be emphasized that the present algorithms provide most powerful countermeasure for divergence problems.

7. Conclusions

In this paper we have derived a new algorithm for the noise adaptive linear prediction. The result is an extension of the maximum likelihood identification and estimation techniques in general.

Jazwinski (4) has shown that the effects of errors in the dynamical system model can often be characterized as an additional noise driving the system, where the statistics of this noise are unknown. If a maximum likelihood estimator of the type which we have derived in this paper is employed in estimating the mean and the covariance of the "modeling error noise" then there is a good reason to believe that the performance of the estimator can be considerably improved. Thus we believe that the method of approach which we have adopted in this paper can improve the design of the estimator and minimize possible divergence problem within it.

We have also pointed out the relationship between the cost functional of determinant form which is derived from the principle of maximum likelihood and the entropy (see Appendix). This suggests that the analysis of the estimation problem from the information theoretical view point may give fruitful results.

Acknowledgement

Parts of this paper are based on the author's M. S. thesis (9) submitted to the University of Tokyo in February 1973. The author wishes to express his sincere appreciation to his thesis supervisor Professor Jiro Kondo of the University of Tokyo for his deep understanding and kind encouragement toward the development of the study.

References

(1) Athans, M., "The Matrix Minimum Principle," Information and Control, 11 (1968), 592-606.

(2) Bellman, R. E., H. H. Kagiwada, R. E. Kalaba, and R. Sridhar, "Invariant Imbedding and Nonlinear Filtering Theory," J. Astronaut. Sci., 13 (1966), 110-115.

(3) Holtzman, J. M. and H. Halkin, "Directional Convexity and the Maximum Principle for Discrete Systems," J. SIAM Control, 4 (1966), 263-275.

(4) Jazwinski, A. H., Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.

(5) Kailath, T., "An Innovation Approach to Least-Squares Estimation, Part I: Linear Filtering in Additive White Noise," IEEE Trans. Aut. Control, AC-13 (1968), 646-655.

(6) Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME, Ser. D: J. Basic Eng., 82 (1960), 35-45.

(7) Kalman, R. E. and R. S. Bucy, "New Results in Linear Filtering and

Prediction Theory," Trans. ASME, Ser. D: J. Basic Eng., 83 (1961), 95-108.

(8) Kashyap, R. L., "Maximum Likelihood Identification of Stochastic Linear Systems," IEEE Trans. Aut. Control, AC-15 (1970), 25-34.

(9) Koda, M., "Estimation Theory for Iterative Nonlinear Filtering and Adaptive Prediction," Master Thesis, The Univ. of Tokyo, 1973.

(10) Lin, J. L. and A. P. Sage, "Algorithms for Discrete Sequential Maximum Likelihood Bias Estimation and Associated Error Analysis," IEEE Trans. Systems, Man, Cybern., SMC-1 (1971), 314-324.

(11) Mehra, R. K., "On the Identification of Variances and Adaptive Kalman Filtering," IEEE Trans. Aut. Control, AC-15 (1970), 175-184.

Appendix

Here we shall show that the particular expression for the likelihood (3.13) can be also derived from the definition of the entropy as the general measure of uncertainty or inaccuracy.

In order to do this, we must clarify the general stochastic property of the measurement residual defined by (3.9). In terms of the notation in this paper, we have for the measurement residual,

$$\text{(A. 1)} \qquad \nu(k+1|k)=z(k+1)-\hat{z}(k+1|k).$$

The sequence of (A. 1) is often reffered to the "innovation process" (5) of z, and is a white Gaussian sequence with zero mean and the covariance as in (3.7),

$$\text{(A. 2)} \qquad \nu(k+1|k) \sim N[0,V(k+1|k)]$$

$$\text{(A. 3)} \qquad V(k+1|k)=H\Phi^{-1}(\bar{N},k+1)C(k)\Phi^{-T}(\bar{N},k+1)H^{T}+R(k).$$

The quantity of (A. 1) may be regarded as defining the "new information" brought by the current observation $z(k+1)$, being given all the past observations Z_k, and the old information deduced therefrom. Thus, as previously stated, the measurement residual may be the most important random variables upon which the maximum likelihood estimation can be based.

For the convenience of the derivation, we rewrite (A. 1) and (A. 3) in the following forms

(A. 4) $$\nu^T(k+1|k)=(\nu_1,\nu_2,\ldots,\nu_m)$$

(A. 5) $$V(k+1|k)= [E\{\nu_i\nu_j\}]=[\sigma_{ij}]$$

(A. 6) $$V^{-1}(k+1|k)=[\sigma^{ij}].$$

Then (A. 2) implies that

(A. 7) $$p(\nu_1,\ldots,\nu_m)=\frac{1}{(2\pi)^{m/2}[\det\{V(k+1|k)\}]^{1/2}} \times\exp\left(-\frac{1}{2}\sum_{i,j}^{m}\sigma^{ij}\nu_i\nu_j\right).$$

Using (A. 7), the entropy for the random variable ν becomes

(A. 8) $$\begin{aligned} S= &-\int_{-\infty}^{\infty}\cdots\int p(\nu_1,\ldots,\nu_m)d\nu_1 d\nu_2\ldots d\nu_m \\ &= \ell n\{(2\pi)^{m/2}[\det\{V(k+1|k)\}]^{1/2}\} \\ &\quad\times\int_{-\infty}^{\infty}\cdots\int p(\nu_1,\ldots,\nu_m)d\nu_1 d\nu_2\ldots d\nu_m \\ &\quad+\frac{1}{2}\sum_{i,j}^{m}\sigma^{ij}\int_{-\infty}^{\infty}\cdots\int \nu_i\nu_j p(\nu_1,\ldots,\nu_m)d\nu_1\ldots d\nu_m \\ &= \ell n\{(2\pi)^{m/2}[\det\{V(k+1|k)\}]^{1/2} + \frac{1}{2}\sum_{i,j}^{m}\sigma^{ij}\sigma_{ij}. \end{aligned}$$

On the other hand, we have

$$\sigma_{ij}=\sigma_{ji}$$

hence

$$\sum_{i,j}^{m}\sigma^{ij}\sigma_{ij}=\sum_{i,j}^{m}\sigma^{ij}\sigma_{ji}=\sum_{i,j}^{m}1= m.$$

Therefore, (A. 8) becomes

$$\text{(A. 9)} \qquad S = \frac{1}{2}m\ \ln 2\pi + \frac{1}{2}\ln(\det\{V(k+1|k)\}) + \frac{1}{2}m$$

$$= \frac{1}{2}m\ \ln 2\pi e + \frac{1}{2}\ln(\det\{V(k+1|k)\}).$$

Thus we can obtain the relationship between (3. 13) and (A. 9),

$$\text{(A. 10)} \qquad L_N(Z_N, \mu, \hat{R}) = -N \cdot S$$

where we have replaced $V(k+1|k)$ in (A. 9) by its maximum likelihood estimate $\hat{V}(N+1|N)$. The result (A. 10) seems quite natural, since the entropy (A. 8) can be considered as the expectation of the likelihood function of (A. 7). Then maximizing the likelihood function $L_N(Z_N, \mu, \hat{R})$ with respect to μ and $\hat{x}(\bar{N}|k)$ is essentially equivalent to minimizing the entropy S. And, from the appropriate interpretation of the measurement residual and the entropy, minimizing the entropy essentially implies decreasing the uncertainty of the measurements. Thus we have verified the absolute legitimacy of the cost functional of determinant form (3. 14) for the maximum likelihood estimation of the state and noise statistics.