Maximum Likelihood Estimates of Linear Dynamic Systems

H. E. RAUCH,* F. TUNG,* AND C. T. STRIEBEL*
Lockheed Missiles and Space Company, Palo Alto, Calif.

This paper considers the problem of estimating the states of linear dynamic systems in the presence of additive Gaussian noise. Difference equations relating the estimates for the problems of filtering and smoothing are derived as well as a similar set of equations relating the covariance of the errors. The derivation is based on the method of maximum likelihood and depends primarily on the simple manipulation of the probability density functions. The solutions are in a form easily mechanized on a digital computer. A numerical example is included to show the advantage of smoothing in reducing the errors in estimation. In the Appendix the results for discrete systems are formally extended to continuous systems.

1. Introduction

THE pioneer work of Wiener[1] on the problem of linear smoothing, filtering and prediction has received considerable attention over the past few years in fields such as space science, statistical communication theory, and many others that often require the estimates of certain variables that are not directly measurable. Many papers have appeared since then giving different solutions to this problem. A summary of these solutions can be found in a paper by Parzen[2] who gives a general treatment of the problem from the point of view of reproducing kernel Hilbert Space. The most widely used solution in practice in linear filtering and prediction is probably the one derived by Kalman[3] using the method of projections. The primary advantage of Kalman's solution is that the equations that specify the optimum filter are in the form of difference equations, so that they can be mechanized easily on the present-day digital computer. However, Kalman does not consider the important problem of smoothing. (The filtering and prediction solution allows one to estimate current and future values of the variables of interest, whereas the smoothing solution permits one to estimate past values.) The purpose of this paper is to provide a solution of the linear smoothing problem based on the principle of the maximum likelihood, and a derivation of the filtering problem based on the same principle. It is shown that the equations describing the smoothing solution also can be easily implemented on a digital computer and a numerical example is presented to show the advantage of smoothing in reducing the errors in estimation.

Solutions of the smoothing problem in different forms have been obtained recently by Rauch[4] for discrete systems and by Bryson and Frazier[5] for continuous systems. The elegant proof and the tools used by Bryson and Frazier are based on the calculus of variations and the method of maximum likelihood. Our derivation differs from their work in that the method used here depends primarily on the simple manipulation of the probability density functions and hence leads immediately to recursion equations. Our results are also different. The derivation leads directly to a smoothing solution that uses processed data instead of the original measurements.

An early version of this paper was published as a company report.[6] During the period in which the paper was being revised for publication, Cox[7] had also presented some similar results using a slightly different approach.

Received December 18, 1964; also presented at the Joint AIAA-IMS-SIAM-ONR Symposium on Control and System Optimization, Monterey, Calif., January 27–29, 1964 (no preprint number); revision received May 13, 1965.

* Research Scientist.

2. Statement of the Problem

2.1 Dynamic System

a) Given†:

$$x_{k+1} = \Phi(k+1, k)x_k + w_k \tag{2.1}$$

$$y_k = M_k x_k + v_k \tag{2.2}$$

where

x_k	=	state vector $(n \times 1)$
y_k	=	output vector $(r \times 1)$, $r \leqq n$
w_k	=	Gaussian random disturbance $(n \times 1)$
v_k	=	Gaussian random disturbance $(r \times 1)$
$\Phi(k+1, k)$	=	transition matrix $(n \times n)$
M_k	=	output matrix $(r \times n)$

and w_k and v_k are independent Gaussian vectors with zero mean and covariances

$$\mathrm{cov}(w_j, w_k) = Q_k\,\delta_{jk} \tag{2.3}$$

$$\mathrm{cov}(v_j, v_k) = R_k\,\delta_{jk} \tag{2.4}$$

$$\mathrm{cov}(w_j, v_k) = 0 \tag{2.5}$$

where δ_{jk} is the Kronecker delta, and we assume that R_k is positive definite.

b) Initial condition x_0 is a Gaussian vector with the a priori information

$$E(x_0) = \bar{x}_0 \qquad \mathrm{cov}(x_0) = \bar{P}_0 \tag{2.6}$$

c) Observations: $y_0, y_1, \ldots, y_N$ $(N = 0, 1, \ldots)$.

The problem is to find an estimate of x_k from the observations $y_0, \ldots, y_N$. Such an estimate will be denoted by $\hat{x}_{k/N} = \hat{x}_{k/N}(y_0, \ldots, y_N)$. It is commonly called the problem of 1) filtering if $k = N$, 2) prediction with filtering if $k \geqq N$, and 3) smoothing if $k \leqq N$.

2.2. Estimation Criteria

Three possible estimation criteria will be presented in this section. For the linear Gaussian case defined in Sec. 2.1 these three criteria result in the same estimate. The distinction is made here in order to see how this problem can be extended to the nonlinear case and how it compares with other work in this field.

The standard procedure is to specify a loss function

$$l(x_0, \hat{x}_{0/N};\ x_1, \hat{x}_{1/N};\ \ldots;\ x_K, \hat{x}_{K/N}) \tag{2.7}$$

† If the original problem is described by nonlinear equations, the linear system can be obtained from equations governing small deviations from a reference path.

and then to find the functions $\hat{x}_{k/N}$ for $k = 0, \ldots, K$ which minimize the expected loss. In order to do this, the distribution of interest is the joint distribution of $x_0, \ldots, x_K$ conditioned on $y_0, \ldots, y_N$:

$$p(x_0, \ldots, x_K/y_0, \ldots, y_N)$$

If the loss function (2.7) is zero near $x_k = \hat{x}_{k/N}$ for $k = 0, \ldots, K$ and very large otherwise, the optimum estimating procedure is the maximum likelihood, and the estimate will be called the joint maximum likelihood estimate. It is obtained by solving the simultaneous equations

$$(\partial/\partial x_k)p(x_0, \ldots, x_K/y_0, \ldots, y_N) = 0 \qquad k = 0, \ldots, K \tag{2.8}$$

If the loss function (2.7) has the special form

$$\sum_{k=0}^{K} l_k(x_k, \hat{x}_{k/N})$$

or equivalently, if $K + 1$ distinct estimation problems with losses $l_k(x_k, \hat{x}_{k/N})$ are considered, the distribution of interest is the marginal distribution of x_k conditioned on $y_0, \ldots, y_N$

$$p(x_k/y_0, \ldots, y_N)$$

The distribution can be obtained from (2.7) by integrating out the variables x_j for $j \neq k$. If $l_k(x_k, \hat{x}_{k/N})$ is zero near $x_k = \hat{x}_{k/N}$ and very large otherwise, the optimum estimate is the marginal maximum likelihood estimate obtained as a solution to the single equation

$$(\partial/\partial x_k)p(x_k/y_0, \ldots, y_N) = 0 \tag{2.9}$$

The marginal maximum likelihood estimate (MLE) is the estimate that will be derived in this paper. The estimate used by Bryson and Frazier[4] is the joint maximum likelihood estimate; so that, although the estimates they obtain in the linear case are the same as the MLE to be derived here, it should be expected that in the nonlinear cases they would not necessarily agree.

Another estimation criterion that is often appropriate is the conditional mean given by $\hat{x}_{k/N} = \int x_k p(x_k/y_0, \ldots, y_N) dx_k$. The conditional mean has the advantage that it is the same for the joint and marginal distributions, and that it minimizes a large class of loss functions.

3. Solutions

3.1. Filtering and Prediction

We shall first consider the case of estimating x_k given all the data up to t_k, i.e., $y_0, \ldots, y_k$. The estimate will be denoted $\hat{x}_{k/k}$, whereas the data $y_0, \ldots, y_k$ will be denoted by Y_k. From the discussion in the previous section, we know that $\hat{x}_{k/k}$ is the solution of x_k which maximizes the conditional probability density function $p(x_k/Y_k)$. This is the same as maximizing the log of the density given by

$$L(x_k, Y_k) = \log p(x_k/Y_k) = \log p(x_k, Y_k) - \log p(Y_k) \tag{3.1}$$

Using the concept of conditional probabilities and the fact that the v_k are independent random vectors, we see

$$\begin{aligned} p(x_k, Y_k) &= p(y_k/x_k, Y_{k-1})p(x_k, Y_{k-1}) \\ &= p(y_k/x_k)\, p(x_k/Y_{k-1})\, p(Y_{k-1}) \end{aligned} \tag{3.2}$$

Let $\hat{x}_{k-1/k-1}$ and $\hat{x}_{k/k-1}$ be the estimates of x_{k-1} and x_k given Y_{k-1}, respectively, and let $\tilde{x}_{k-1/k-1}$ and $\tilde{x}_{k/k-1}$ be the errors in these estimates. Define

$$\text{cov}(\tilde{x}_{k-1/k-1}) = P_{k-1/k-1} \tag{3.3}$$

and

$$\text{cov}(\tilde{x}_{k/k-1}) = P_{k/k-1} \tag{3.4}$$

Since all the random disturbances v_k are statistically independent, it follows that

$$\hat{x}_{k/k-1} = \Phi(k, k-1)\hat{x}_{k-1/k-1} \tag{3.5}$$

and

$$P_{k/k-1} = \Phi(k, k-1)P_{k-1/k-1}\Phi'(k, k-1) + Q_{k-1} \tag{3.6}$$

This is, in fact, the solution of the prediction problem. Using (2.1–2.4) and the assumption that the random disturbances are normally distributed, we see that the conditional random vector x_k given Y_{k-1} has a mean

$$E(x_k/Y_{k-1}) = \hat{x}_{k/k-1} \tag{3.7}$$

and a covariance

$$\text{cov}(x_k/Y_{k-1}) = P_{k/k-1} \tag{3.8}$$

whereas the conditional vector y_k given x_k has a mean

$$E(y_k/x_k) = M_k x_k \tag{3.9}$$

and a covariance

$$\text{cov}(y_k/x_k) = R_k \tag{3.10}$$

Substituting (3.7) to (3.10) into (3.2) and using the fact that all the vectors are normally distributed, we find‡

$$\begin{aligned} p(x_k, Y_k) = {} & (2\pi)^{-r/2}(\det R_k)^{-1/2} \times \\ & \exp(-1/2\|y_k - M_k x_k\|^2_{R_k^{-1}})\cdot(2\pi)^{-n/2}(\det P_{k/k-1})^{-1/2} \times \\ & \exp(-1/2\|x_k - \hat{x}_{k/k-1}\|^2_{(P_{k/k-1})^{-1}})\cdot p(Y_{k-1}) \end{aligned} \tag{3.11}$$

Substitution of (3.11) into (3.1) shows that the terms in L which depend on x_k can be written as

$$J = \|y_k - M_k x_k\|^2_{R_k^{-1}} + \|x_k - \hat{x}_{k/k-1}\|^2_{(P_{k/k-1})^{-1}} \tag{3.12}$$

Setting the gradient of J to zero, we find

$$\hat{x}_{k/k} = (M_k'R_k^{-1}M_k + P_{k/k-1}^{-1})(M_k'R_k^{-1}y_k + P_{k/k-1}^{-1}\hat{x}_{k/k-1}) \tag{3.13}$$

which is essentially the solution of the filtering problem. Equation (3.13) may be put into a more convenient form by using a well-known matrix inversion lemma.§ This lemma, for instance, has beeen used by Ho[8] to show the relations between the stochastic approximation method and the optimal filter theory.

Lemma: If $S_{k+1}^{-1} = S_k^{-1} + M_k'R_k^{-1}M_k$ where S_k and R_k are symmetric and positive definite, then S_{k+1} exists and is given by $S_{k+1} = S_k - S_kM_k'(M_kS_kM_k' + R_k)^{-1}M_kS_k$. The proof is by direct substitution. By making use of this lemma, it is seen that (3.13) also can be written as

$$\begin{aligned} \hat{x}_{k/k} &= \hat{x}_{k/k-1} + B_k(y_k - M_k\hat{x}_{k/k-1}) \\ &= \Phi(k, k-1)\hat{x}_{k-1/k-1} + B_k[y_k - M_k\Phi(k, k-1)\hat{x}_{k-1/k-1}] \end{aligned} \tag{3.14}$$

where

$$B_k = P_{k/k-1}M_k'(M_kP_{k/k-1}M_k' + R_k)^{-1} \tag{3.15}$$

Remark: The computation of $\hat{x}_{k/k}$ by way of (3.13) requires the inversion of a $n \times n$ matrix $(M_k'R_k^{-1}M_k + P_{k/k-1}^{-1})$ whereas Eqs. (3.14) and (3.15) require only the inversion of the matrix $(M_kP_{k/k-1}M_k' + R_k)$ which is $r \times r$ $(r \leqq n)$. Hence, the representation given by (3.14) and (3.15) appears to be more desirable for the purpose of computation.

Substituting (2.1) and (2.2) into (3.14) shows that the estimation error satisfies the recursive equation

$$\tilde{x}_{k/k} = (I - B_kM_k)[\Phi(k, k-1)\tilde{x}_{k-1/k-1} + w_{k-1}] - B_kv_k \tag{3.16}$$

‡ $\|a\|^2_R = a'Ra$.

§ The authors wish to acknowledge Y. C. Ho of Harvard for pointing out this identity.

where I is the identity matrix. Since $\tilde{x}_{k-1/k-1}$, v_k, and w_{k-1} are statistically independent, it follows that

$$P_{k/k} = \operatorname{cov}(\tilde{x}_{k/k}) = (I - B_k M_k) P_{k/k-1} \tag{3.17}$$

where use is made of (3.15). Equations (3.14–3.17) are the same as those derived originally by Kalman.[2] To start the recursive equation, we need $\hat{x}_{0/-1}$ and $P_{0/-1}$. From the a priori information about x_0, we see

$$\hat{x}_{0/-1} = \bar{x}_0 \tag{3.18}$$

and

$$P_{0/-1} = \bar{P}_0 \tag{3.19}$$

This completes the solution of the filtering problem. The solution of the prediction problem has already been obtained. For any $N \geqq k$,

$$\hat{x}_{N/k} = \Phi(N, k)\hat{x}_{k/k} \tag{3.20}$$

3.2 Smoothing

From the principle of the MLE, we know that the estimate of x_k given Y_N, denoted by $\hat{x}_{k/N}$, is that value of x_k which maximizes the function

$$L(x_k, Y_N) = \log p(x_k/Y_N) \tag{3.21}$$

Similarly, $\hat{x}_{k/N}$ and $\hat{x}_{k+1/N}$ are the values of x_k and x_{k+1} which maximize

$$L(x_k, x_{k+1}, Y_N) = \log p(x_k, x_{k+1}/Y_N) \tag{3.22}$$

Let us now inspect the joint probability density function $p(x_k, x_{k+1}, Y_N)$. Using the concept of conditional probabilities, we see

$$p(x_k, x_{k+1}, Y_N) = p(x_k, x_{k+1}, y_{k+1}, \ldots, y_N/Y_k)p(Y_k) \tag{3.23}$$

Now

$$\begin{aligned} p(x_k, x_{k+1}, y_{k+1}, \ldots, y_N/Y_k) &= p(x_{k+1}, y_{k+1}, \ldots, y_N/x_k, Y_k)\, p(x_k/Y_k) \\ &= p(x_{k+1}, y_{k+1}, \ldots, y_N/x_k)\, p(x_k/Y_k)\P \\ &= p(y_{k+1}, \ldots, y_N/x_{k+1}, x_k)\, p(x_{k+1}/x_k)\, p(x_k/Y_k) \\ &= p(y_{k+1}, \ldots, y_N/x_{k+1})\, p(x_{k+1}/x_k)\, p(x_k/Y_k) \end{aligned} \tag{3.24}$$

Substituting (3.24) into (3.23) shows that

$$p(x_k, x_{k+1}, Y_N) = p(x_{k+1}/x_k)\, p(x_k/Y_k)\, p(y_{k+1}, \ldots, y_N/x_{k+1}) \cdot p(Y_k) \tag{3.25}$$

Let us assume that $\hat{x}_{k/k}$ has already been obtained. Substituting (3.25) into (3.22) and using the same reasoning as that given in the previous section, we see

$$\max_{x_k, x_{k+1}} L(x_k, x_{k+1}, Y_N) = \max_{x_k, x_{k+1}} \{-\|x_{k+1} - \Phi(k+1, k)x_k\|^2_{Q_k^{-1}} - \|x_k - \hat{x}_{k/k}\|^2_{P_{k/k}^{-1}}\} + \text{terms which do not involve } x_k \tag{3.26}$$

It follows immediately that $\hat{x}_{k/N}$ is the solution that minimizes the expression

$$J = \|\hat{x}_{k+1/N} - \Phi(k+1, k)x_k\|^2_{Q_k^{-1}} + \|x_k - \hat{x}_{k/k}\|^2_{P_{k/k}^{-1}} \tag{3.27}$$

Setting the gradient of J to zero and using the matrix inversion lemma, we find

$$\hat{x}_{k/N} = \hat{x}_{k/k} + C_k[\hat{x}_{k+1/N} - \Phi(k+1, k)\hat{x}_{k/k}] \tag{3.28}$$

where

$$\begin{aligned} C_k &= P_{k/k}\Phi'(k+1, k) \times [\Phi(k+1, k)P_{k/k}\Phi'(k+1, k) + Q_k]^{-1} \\ &= P_{k/k}\Phi'(k+1, k)P_{k+1/k}^{-1} \end{aligned} \tag{3.29}$$

This is the solution of the smoothing problem. It is in the form of a backward recursive equation that relates the MLE of x_k given Y_N in terms of the MLE of x_{k+1} given Y_N and the MLE of x_k given Y_k. Hence, the smoothing can be obtained from the filtering solution by computing backwards using (3.29).

Subtracting x_k from both sides of (3.28) and rearranging the terms, we find

$$\tilde{x}_{k/N} + C_k\hat{x}_{k+1/N} = \tilde{x}_{k/k} + C_k\,\Phi(k+1, k)\hat{x}_{k/k} \tag{3.30}$$

Using the facts that

$$\begin{aligned} E(\tilde{x}_{k/N}\hat{x}_{k+1/N}') &= E(\tilde{x}_{k/k}\,\hat{x}_{k/k}') = 0^{**} \\ \operatorname{cov}(\hat{x}_{k+1/N}) &= \operatorname{cov}(x_{k+1}) - P_{k+1/N} \\ \operatorname{cov}(\hat{x}_{k/k}) &= \operatorname{cov}(x_k) - P_{k/k} \end{aligned}$$

and

$$\operatorname{cov}(x_{k+1}) = \Phi(k+1, k)\operatorname{cov}(x_k)\,\Phi'(k+1, k) + Q_k$$

we see from (3.30) that $P_{k/N}$ satisfies the recursive equation

$$P_{k/N} = P_{k/k} + C_k(P_{k+1/N} - P_{k+1/k})C_k' \tag{3.31}$$

The computation is initiated by specifying $P_{N/N}$. This essentially completes the solution for the smoothing problem. It should be noted that the estimates $\hat{x}_{k/k}$ $(k \leqq N)$ are assumed to have been obtained in the process of computing $\hat{x}_{N/N}$ and hence can be made available by storing them in the memory. The covariance $P_{k/k}$ also may be stored. However, it can be easily computed. We will now give a formula for computing $P_{k/k}$ from $P_{k+1/k+1}$ and hence eliminate the storage problem for $P_{k/k}(k = 0, \ldots, N)$.

Substituting (3.15) into (3.17) shows

$$P_{k/k-1} = (P_{k/k}^{-1} - M_k'R_k^{-1}M_k)^{-1} \tag{3.32}$$

which can be written as

$$P_{k/k-1} = P_{k/k} - P_{k/k}M_k'(M_kP_{k/k}M_k' - R_k)^{-1}M_kP_{k/k} \tag{3.33}$$

after applying the matrix inversion lemma. From $P_{k/k-1}$, $P_{k-1/k-1}$ can be computed by using (3.6) which can be written as

$$P_{k-1/k-1} = \Phi^{-1}(k-1, k)(P_{k/k-1} - Q_{k-1})\Phi'^{-1}(k-1, k) \tag{3.34}$$

The terminal condition for (3.33) is again $P_{N/N}$. It is of interest to note from (3.33) that the computation for $P_{k/k}$ requires only the inversion of a $r \times r$ matrix.

Remark:

1) Another formulation of the smoothing problem which relates $\hat{x}_{k/N}$ to $\hat{x}_{k+1/N}$ and all the data $y_j (j \geq k+1)$ and hence requires the storage of the data can be obtained by noting that $\hat{x}_{i/N}(i = 0, 1, \ldots, N)$ is the solution which maximizes the function

$$L(x_0, x_1, \ldots, x_N, Y_N) = \log p(x_0, x_1, \ldots, x_N/Y_N) \tag{3.35}$$

Now

$$\begin{aligned} p(x_0, x_1, \ldots, x_N, Y_N) &= p(Y_N/x_0, x_1, \ldots, x_N)\, p(x_0, x_1, \ldots, x_N) \\ &= p(Y_N/x_0, x_1, \ldots, x_N)\, p(x_N/x_{N-1})\, p(x_{N-1}/x_{N-2}) \ldots p(x_1/x_0)\, p(x_0) \end{aligned} \tag{3.36}$$

where use is made of the fact that x is a Markov process,

$$p(x_k/x_{k-1}, \ldots, x_0) = p(x_k/x_{k-1}) \tag{3.37}$$

Substituting (3.36) into (3.35) shows that maximizing L is

¶ This is because $x_{k+1}, y_{k+1}, \ldots, y_N$ given x_k is independent of y_i, $i \leqq k$, and $p(a/bc) = p(a/b)$ if a/b is independent of c.

** This can be verified after somewhat lengthy manipulation of Eqs. (3.16) and (3.30) using the properties of $\hat{x}_{k/k}$ and $\hat{x}_{k/N}$.

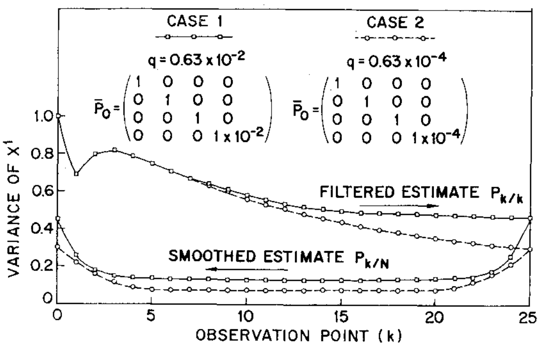

Fig. 1 Variance history for two levels of random disturbances (q).

equivalent to

$$\min_{x_0, \ldots x_N} \left\{ \sum_{i=0}^{N} ||y_i - M_i x_i||^2_{R_i^{-1}} + \sum_{i=0}^{N} ||x_i - \Phi(i, i-1)x_{i-1}||^2_{Q_{i-1}^{-1}} \right\} \quad (3.38)$$

with the initial condition

$$\Phi(0, -1)x_{-1} = \bar{x}_0 \quad (3.39)$$

and

$$Q_{-1} = \bar{P}_0 \quad (3.40)$$

This is the equivalent discrete formulation of the continuous smoothing problem recently given by Bryson and Frazier.[5] The scalar version of (3.38) may also be found in a book by Bellman.[9]

To show the equivalence of our solution with the results of Bryson and Frazier and to obtain the solution of (3.38) in terms of the observations y_k, we define a new variable

$$w_k = P_{k+1/k}^{-1}[\hat{x}_{k+1/N} - \Phi(k+1, k)\hat{x}_{k/k}] \quad (3.41)$$

It follows that

$$w_N = 0 \quad (3.42)$$

Substituting (3.41) into (3.28) and using (3.14, 3.17, and 3.32), we obtain, after many algebraic manipulation, a set of $2n$ difference equations

$$\hat{x}_{k+1/N} = \Phi(k+1, k)\hat{x}_{k/N} + Q_k w_k$$

$$w_k = \Phi'(k, k+1)M_k'R_k^{-1}M_k\,\hat{x}_{k/N} + \Phi'(k, k+1)w_{k-1} - \Phi'(k, k+1)M_kR_k^{-1}y_k \quad (3.43)$$

Notice that if $\hat{x}_{N/N}$ is given, then the set of equations given by (3.43) may be computed backwards from the index N. Otherwise, it involves the solution of a two point boundary value problem.

2) It has been shown[10] that by simple manipulations of the results derived in this paper, namely Eqs. (3.28) and (3.31), the smoothing solution can be written in still another form that directly relates the smoothed estimate at a particular time to the new observations as they are received. This form is preferable for the class of problems where one is only interested in the smoothing solution for the state at a particular time.

3) The problem of interpolation is concerned with estimating the state between measurement points. If it is desired to estimate the state x_k at a point where no measurement was taken, the equations presented here for the smoothing solution can be used by assuming that a measurement is taken at that point with covariance R_k very large and with B_k equal to zero.

4. Numerical Example

Consider the dynamical system given by

$$x_{k+1} = \begin{pmatrix} 1 & 1 & 0.5 & 0.5 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.606 \end{pmatrix} x_k + w_k$$

$$y_k = (1, 0, 0, 0)x_k + v_k \quad (4.1)$$

where x_k is the (4×1) state vector composed of four state variables (x^1, x^2, x^3, and x^4), and y_k is the (1×1) output vector that is a noisy measurement of the state variable x^1. The disturbances w_k and v_k are independent Gaussian vectors with zero mean and covariances

$$\operatorname{cov}(w_k) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q \end{pmatrix}$$

$$\operatorname{cov}(v_k) = 1 \quad (4.2)$$

The initial condition x_0 is a Gaussian vector with a priori information such that the covariance of x_0 is given by $\bar{P}_0$.

The entire dynamic system can be considered as a linearized version of the in-track motion of a satellite traveling in a circular orbit. The satellite motion is affected by both constant and stochastic drag.[11] The state variables x^1, x^2, and x^3 can be considered as angular position, velocity, and (constant) acceleration, respectively. The state variable x^4 is a stochastic component of acceleration generated by a first-order Gauss-Markov process.

Three cases will be considered:

Case 1:

$$q = 0.63 \times 10^{-2}$$

$$\bar{P}_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \times 10^{-2} \end{pmatrix}$$

Table 1 Diagonal elements of the covariance

Observation point (k)	Filtered estimate ($P_{k/k}$) cov(x^1)	cov(x^2)	cov(x^3)	cov(x^4)
0	1.00	1.00	1.00	0.0100
1	0.69	1.31	0.92	0.0100
2	0.80	1.31	0.54	0.0100
3	0.82	0.96	0.26	0.0100
4	0.79	0.68	0.13	0.0100
5	0.75	0.49	0.07	0.0100
10	0.58	0.15	0.008	0.00995
15	0.50	0.10	0.004	0.0099
20	0.48	0.093	0.0026	0.0099
25	0.47	0.089	0.0020	0.0099
Observation point (k)	**Smoothed estimate ($P_{k/N}$) cov(x^1)**	**cov(x^2)**	**cov(x^3)**	**cov(x^4)**
25	0.47	0.089	0.0020	0.0099
24	0.26	0.058	0.0020	0.0096
23	0.18	0.036	0.0020	0.0091
22	0.15	0.023	0.0020	0.0085
21	0.15	0.017	0.0020	0.0080
20	0.15	0.015	0.0020	0.0078
15	0.135	0.014	0.0020	0.0078
10	0.135	0.014	0.0020	0.0078
5	0.14	0.015	0.0020	0.0078
1	0.26	0.053	0.0020	0.0089
0	0.45	0.082	0.0020	0.0094

Case 2:

$$q = 0.63 \times 10^{-4}$$

$$\bar{P}_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \times 10^{-4} \end{pmatrix}$$

Case 3:

$$q = 0.63 \times 10^{-2}$$

$$P_0 = \begin{pmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 1 \times 10^{-2} \end{pmatrix}$$

In each case 25 measurements are taken starting with y_1. The diagonal elements of the covariance of the estimates of the state for case 1 are presented in Table 1 for both the filtered and smoothed estimate. Notice how smoothing the estimate decreases the errors. In Fig. 1 the variance of the filtered and smoothed estimates of the state variable x^1 are plotted for both case 1 and case 2. Reducing the variance of the random disturbance reduces the variance of the estimates. In Fig. 2 the variance of the estimates of x^1 are plotted for both case 1 and case 3. Notice how the effect of initial conditions (the a priori information about the state) rapidly dies out.

5. Conclusions

The solution to the discrete version of the filtering and smoothing problem has been derived using the principal of maximum likelihood and simple manipulation of the probability density function. The filtered estimate is calculated forward point by point as a linear combination of the previous filtered estimate and the current observation. The smoothing solution starts with the filtered estimate at the last point and calculates backward point by point determining the smoothed estimate as a linear combination of the filtered estimate at that point and the smoothed estimate at the previous point. A numerical example has been presented to illustrate the advantage of smoothing in reducing the error in the estimate.

Appendix: Extension to the Continuous Case

The MLE of the states with continuous observations can be obtained formally from the MLE of the discrete system. The difference equations in the previous section become differential equations in the limit as the time between observations approaches zero. No rigorous proof of the limiting process is attempted here. A discussion of the conditions under which it is valid can be found elsewhere.[3]

Let us assume that the discrete indices k and $k + 1$ in all the variables have been replaced by t and $t + q$, and let the disturbances w_k be replaced by $qu(t)$. A Taylor series expansion in q is made of the transition matrix $\Phi(t + q, t)$ so that Eqs. (2.1) and (2.2) can be written as

$$\begin{aligned} x(t + q) &= \Phi(t + q, t)x(t) + qu(t) \\ &= [I + qF(t) + 0(q^2)]x(t) + qu(t) \end{aligned} \tag{A1}$$

and

$$y(t) = M(t)x(t) + v(t) \tag{A2}$$

where $0(q^2)$ represents the terms of the order of q^2. The covariances Q_k and R_k are replaced by $qQ(t)$ and $R(t)/q$,

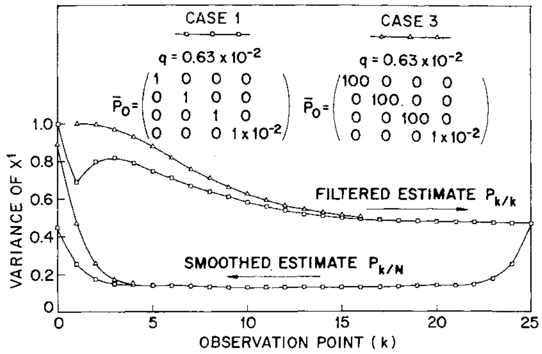

Fig. 2 Variance history for two sets of initial conditions ($\bar{P}_0$).

respectively, so that††

$$\mathrm{cov}[u(t)] = Q(t)/q \tag{A3}$$

and

$$\mathrm{cov}[v(t)] = R(t)/q$$

In the limit as q approaches zero, we find that (A1) and (A2) become

$$[dx(t)]/dt = F(t)x(t) + u(t) \tag{A4}$$

and

$$y(t) = M(t)x(t) + v(t) \tag{A5}$$

where $u(t)$ and $v(t)$ are white noises such that

$$\mathrm{cov}[u(t), u(s)] = Q(t)\,\delta(t - s) \tag{A6}$$

$$\mathrm{cov}[v(t), v(s)] = R(t)\,\delta(t - s) \tag{A7}$$

$\delta(t - s)$ being the Dirac delta function.

The same limiting process will now be applied to the solutions of the MLE derived for the discrete system in the previous section. For the purpose of clarification, the following notation will be used: $\hat{x}_t(t)$ = estimate of $x(t)$ using the data over the interval $(0, t)$, $\hat{x}_T(t)$ = estimate of $x(t)$ using the data over the interval $(0, T)$, $P_t(t) = \mathrm{cov}[x(t) - \hat{x}_t(t)]$, and $P_T(t) = \mathrm{cov}[x(t) - \hat{x}_T(t)]$.

Filtering Solution

Applying the limiting process to Eqs. (3.14) and (3.15) and the corresponding covariances given by (3.6) and (3.17), we find that the filtering solution for the continuous case can be written as

$$[d\hat{x}_t(t)]/dt = F(t)\hat{x}_t(t) + P_t(t)M'(t)R^{-1}(t)[y(t) - M(t)\hat{x}_t(t)] \tag{A8}$$

†† The replacement in (A3) keeps the statistical properties of the random disturbances nearly the same as can be shown by the following explanation. Divide the interval between k and $k + 1$ (which is of length T_k) into n equally spaced intervals with an observation made at each interval. The time between observations is $q = T_k/n$. Assume, for the moment, that there are no dynamics between k and $k + 1$. Because the errors in the observations are Gaussian, the accuracy obtained from n observations, each with covariance nR_k, would be the same as the accuracy obtained from one observation with covariance R_k. Therefore, if $v(t)$ is the noise on the observation at time t, $\mathrm{cov}[v(t)] = nR_k = R_kT_k/q = R(t)/q$. Furthermore, the sum of n identically distributed independent Gaussian random inputs with covariance Q_k/n would have the same distribution as one random input with covariance Q_k. Therefore, if $qu(t)$ is the random input at time t, $\mathrm{cov}[u(t)] = q^{-2}\,Q_k/n = q^{-1}\,Q_k/T_k = Q(t)/q$.

and

$$[dP_t(t)]/dt = F(t)P_t(t) + P_t(t)F'(t) - P_t(t)M'(t)R^{-1}(t)\,M(t)P_t(t) + Q(t) \quad \text{(A9)}$$

with the initial conditions

$$\hat{x}_0(0) = \bar{x}_0 \text{ and } P_0(0) = \bar{P}_0 \quad \text{(A10)}$$

Equations (A8) and (A9) are the same as those given by Kalman.[3]

Smoothing Solution

In a similar manner, the continuous version of the MLE for the smoothing problem given by Eqs. (3.28, 3.29, and 3.31) can be written as

$$[d\hat{x}_T(t)]/dt = F(t)\hat{x}_T(t) + Q(t)P_t^{-1}(t)[\hat{x}_T(t) - \hat{x}_t(t)] \quad \text{(A11)}$$

and

$$[dP_T(t)]/dt = [F(t) + Q(t)P_t^{-1}(t)]P_T(t) + P_T(t)[F(t) + Q(t)P_t^{-1}(t)]' - Q(t) \quad \text{(A12)}$$

with the terminal condition $\hat{x}_T(T)$ and $P_T(T)$.

To show the equivalence of our solution with the results of Bryson and Frazier, we define a new variable

$$\omega(t) = P_t^{-1}(t)[\hat{x}_T(t) - \hat{x}_t(t)] \quad \text{(A13)}$$

It follows that

$$\omega(T) = 0 \quad \text{(A14)}$$

Substituting (A13) into (A11) and using (A8), (A7), as well as (A11), we obtain a set of $2n$ differential equations

$$[d\hat{x}_T(t)]/dt = F(t)\hat{x}_T(t) + Q(t)\omega(t) \quad \text{(A15)}$$

$$[d\omega(t)]/dt = M'(t)R^{-1}(t)M(t)\hat{x}_T(t) - F'(t)\omega(t) - M'(t)R^{-1}(t)y(t) \quad \text{(A16)}$$

which are precisely those derived by Bryson and Frazier. Hence, we have given a physical interpretation of the Lagrange multipliers $\omega(t)$ used in their derivation. Moreover, it can be readily shown that

$$\text{cov}[\tilde{x}_T(t), \omega(t)] = P_T^{-1}(t)P_t(t) - I \quad \text{(A17)}$$

and

$$\text{cov}[\omega(t)] = P_t^{-1}(t)P_T(t)P_t^{-1}(t) - P_t^{-1}(t) \quad \text{(A18)}$$

where

$$\tilde{x}_T(t) = x(t) - \hat{x}_T(t)$$

References

[1] Wiener, N., *The Extrapolation, Interpolation and Smoothing of Stationary Time Series* (John Wiley & Sons, Inc., New York, 1949).

[2] Parzen, E., "An approach to time series analysis," Ann. Math. Statist. **32**, 951–989 (1961).

[3] Kalman, R. E., "New methods and results in linear prediction and filtering theory," Research Institute for Advanced Studies Rept. 61-1, Martin Co., Baltimore, Md. (1960).

[4] Rauch, H. E., "Linear estimation of sampled stochastic processes with random parameters," TR 2108-1, Stanford Electronics Lab., Stanford Univ. (April 1962).

[5] Bryson, A. E. and Frazier, M., "Smoothing for linear and nonlinear dynamic systems," Wright-Patterson Air Force Base, Ohio, Aeronautical Systems Division TDR-63-119, pp. 353–364 (September 1962).

[6] Rauch, H. E., Tung, F., and Striebel, C. T., "On the maximum likelihood estimates for linear dynamic systems," Lockheed Missiles and Space Co., Palo Alto, Calif., TR 6-90-63-62 (June 1963).

[7] Cox, H., "On the estimation of state variables and parameters for noisy dynamic systems," *IEEE Transactions on A.C.* (Institute of Electrical and Electronic Engineering, New York, January 1964), Vol. AC-9, pp. 5–12.

[8] Ho, Y. C., "On the stochastic approximation method and the optimal filtering theory," *Math. Analysis and Applications* (February 1963), Vol. 6, pp. 152–155.

[9] Bellman, R., *Introduction to Matrix Analysis* (McGraw-Hill Book Co., New York, 1950), pp. 154–155.

[10] Rauch, H. E., *Solutions to the Linear Smoothing Problem*, *IEEE Transactions on A.C.* (Institute of Electrical and Electronic Engineering, New York, October 1963), Vol. AC-8, pp. 371–372.

[11] Rauch, H. E., "Optimum estimation of satellite trajectories including random fluctuations in drag," AIAA J. **3**, 717–722 (April 1965).