where

$$H_k(a) \triangleq \frac{\partial h_k}{\partial x_k}(a) \qquad \alpha_{ki} = \frac{\alpha_{ki}'\beta_{ki}}{\sum_{j-1}^{r_k'} \alpha_{kj}'\beta_{kj}}$$

and where

$$\beta_{ki} \triangleq N(Y_k - h_k(\hat{X}_{ki}'), H_k(\hat{X}_{ki}')P_{ki}'H_k^T(\hat{X}_{ki}') + R_k)$$

$$S_{ki} = H_k(\hat{X}_{ki}')P_{ki}'H_k^T(\hat{X}_{ki}') + R_k.$$

The number of terms in the Gaussian sum r_k are less than or equal to the number in the previous density r_k' for the case considered here. The density is propagated through the dynamics by linearizing the plant nonlinearity about each term $\hat{X}_{ki}$ in the last density. Thus

$$p(x_{k+1}|Y^k, U^k) = \sum_{i=1}^{r_{k+1}} \alpha_{(k+1)i}' N(x_{k+1} - \hat{X}_{(k+1)i}', P_{(k+1)i}')$$

where

$$\hat{X}_{(k+1)i}' = f_{k+1}(\hat{X}_{ki}, U_k)$$

$$P_{(k+1)i}' = F_{(k+1)}(\hat{X}_{ki})P_{ki}F_{(k+1)}^T(\hat{X}_{ki}) + Q_k$$

$$\alpha_{(k+1)i}' = \alpha_{ki}$$

$$r_{k+1}' = r_k$$

and

$$F_{(k+1)}(\hat{X}_{ki}) \triangleq \frac{\partial f_{k+1}}{\partial x_k}(\hat{X}_{ki}).$$

More general cases where w_j and v_j are given by Gaussian sum densities are discussed in [16].

Finally, $P(Y_k|Y^{k-1})$ of (9) is obtained from

$$P(Y_k|Y^{k-1}) = \int_{R^n} P(Y_k|X_k)P(X_k|Y^{k-1})\,dx_k$$

by linearizing the exponent of $P(Y_k|X_k)$ about each term in $P(X_k|Y^{k-1})$ and dropping higher order terms.

References

[1] M. Aoki, *Optimization of Stochastic Systems*. New York: Academic, 1967.
[2] T. L. Gunckel, II, and G. F. Franklin, "A general solution for linear sample-data control," *Trans. ASME*, J. Basic Eng., ser. D, vol. 85, pp. 197–203, June 1963.
[3] P. D. Joseph and J. T. Tou, "On linear control theory," *AIEE Trans.* (Appl. Ind.), vol. 80, pp. 193–196, Sept. 1961.
[4] C. Streibel, "Sufficient statistics in the optimal control of stochastic systems," *J. Math. Anal. Appl.*, vol. 12, pp. 576–592, 1965.
[5] R. E. Curry, "A separation theorem for nonlinear measurements," *IEEE Trans. Automat. Contr.* (Short Papers), vol. AC-14, pp. 561–564, Oct. 1969.
[6] D. L. Alspach and H. W. Sorenson, "Stochastic optimal control for linear but non-Gaussian systems," *Int. J. Contr.*, vol. 13, no. 6, pp. 1169–1181, 1971.
[7] ——, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 439–448, Aug. 1972.
[8] C. Hecht, "Digital realization of nonlinear filters," in *Proc. 2nd Symp. Nonlinear Estimation*, 1971.
[9] J. L. Center, Jr., "Practical nonlinear filtering of discrete observations by generalized least squares approximation of the conditional probability distribution," in *Proc. 2nd Symp. Nonlinear Estimation*, 1971.
[10] K. D. Senne, "Computer experiments with nonlinear estimators," in *Proc. 2nd Symp. Nonlinear Estimation*, 1971.
[11] J. P. R. de Figueiredo and Y. G. Jan, "Spline filters," in *Proc. 2nd Symp. Nonlinear Estimation*, 1971.
[12] E. Tse, L. Meier, and Y. Bar-Shalom, "Dual control of stochastic nonlinear systems," in *Proc. 1971 IEEE Decision and Control Conf.*
[13] A. A. Fel'dbaum, "Dual control theory, Part I," *Automat. Remote Contr.*, vol. 21, pp. 874–880, Apr. 1961; "Part II," vol. 21, pp. 1033–1046, May 1961; "Part III," vol. 22, pp. 1–12, Aug. 1961; "Part IV," vol. 22, pp. 109–121, Sept. 1961.
[14] R. E. Curry, "A new algorithm for suboptimal stochastic control," *IEEE Trans. Automat. Contr.* (Short Papers), vol. AC-14, pp. 533–536, Oct. 1969.
[15] E. Tse, "Parallel computation of the conditional mean state estimate for nonlinear systems," in *Proc. 2nd Symp. Nonlinear Estimation*, 1971.
[16] D. L. Alspach, "A Bayesian approximation technique for estimation and control of time-discrete stochastic systems," Ph.D. dissertation, Univ. California at San Diego, La Jolla, June 1970.

Approaches to Adaptive Filtering

RAMAN K. MEHRA, MEMBER, IEEE

***Abstract*—The different methods of adaptive filtering are divided into four categories: Bayesian, maximum likelihood (ML), correlation, and covariance matching. The relationship between the methods and the difficulties associated with each method are described. New algorithms for the direct estimation of the optimal gain of a Kalman filter are given.**

I. Introduction

The Kalman–Bucy formulation of the filtering problem assumes complete *a priori* knowledge of the process and measurement noise statistics. In most practical situations, these statistics are inexactly known. The use of wrong *a priori* statistics in the design of a Kalman filter can lead to large estimation errors or even to a divergence of errors. The purpose of an adaptive filter is to reduce or bound these errors by modifying or adapting the Kalman filter to the real data.

A number of approaches can be taken to adaptive filtering. Since the basic source of uncertainty is due to unknown *a priori* statistics of noise, one can estimate them on-line from the observed data. Another approach is to estimate the optimal Kalman gain directly without estimating the covariances of the process and measurement noise. Some of these approaches are discussed in the present short paper. An extensive Bibliography is included.

II. Statement of the Problem

Consider a discrete-time dynamic system described by vector difference equations

$$x_{i+1} = \Phi_i x_i + u_i \tag{1}$$

$$z_i = H_i x_i + v_i \tag{2}$$

where x_i is $n \times 1$ state vector, Φ_i is $n \times n$ state transition matrix, u_i is $q \times 1$ process noise vector, z_i is $r \times 1$ measurement vector, v_i is $r \times 1$ measurement noise vector, and H_i is $r \times n$ measurement matrix.

Both u_i and v_i are assumed to be uncorrelated zero-mean Gaussian white noise sequences with covariances

$$E\{u_i\} = 0, \qquad E\{v_i\} = 0$$

$$E\{u_i u_j^T\} = Q\delta_{ij}, \qquad E\{v_i v_j^T\} = R\delta_j$$

in which Q and R are nonnegative definite matrices whose true values are assumed to be unknown.

The adaptive filtering problem consists of obtaining on-line estimates of x_i based on the observation set $Z^i = \{z_1, \cdots, z_i\}$.

Let

$$\hat{x}_{i/j} = \text{an estimate of } x_i \text{ based on the observations set } Z^j = \{z_1, \cdots, z_j\}$$

$$P_{i/j} = E\{(x_i - \hat{x}_{i/j})(x_i - \hat{x}_{i/j})^T\}.$$

The solution to this problem when Q and R are exactly known is given by the Kalman filter of the form

$$\hat{x}_{i+1/i} = \Phi_i \hat{x}_{i/i} \tag{3}$$

$$\hat{x}_{i/i} = \hat{x}_{i/i-1} + K_i(z_i - H_i\hat{x}_{i/i-1}) \tag{4}$$

$$K_i = P_{i/i-1}H_i^T(H_iP_{i/i-1}H_i^T + R)^{-1} \tag{5}$$

Manuscript received March 19, 1971; revised March 30, 1972. Paper recommended by J. M. Mendel, Past Chairman of the IEEE S-CS Adaptive and Learning Systems, Pattern Recognition Committee. This paper was presented at the IEEE Symposium on Adaptive Processes, Austin, Tex., Dec. 1970.

The author is with Systems Control, Inc., Palo Alto, Calif. 94306.

$$P_{i/i} = (I - K_iH_i)P_{i/i-1} \tag{6}$$

$$P_{i+1/i} = \Phi_iP_{i/i}\Phi_i^T + Q \tag{7}$$

where K_i denotes the Kalman gain and $\nu_i = (z_i - H_i\hat{x}_{i/i-1})$ denotes the innovation or the one-step-ahead prediction error.

III. Bayesian Estimation

Let α denote the vector of unknown parameters in the system. For the time being, α may include unknowns in Φ,H,Q, and R. Later, we will specialize it to the unknowns in Q and R only.

In Bayesian estimation, one tries to obtain recursive equations for the *a posteriori* probability density of x_i and α, viz., $p(x_i,\alpha|Z^i)$, which can be written as

$$p(x_i,\alpha|Z^i) = p(x_i|\alpha,Z^i)p(\alpha|Z^i) \tag{8}$$

where $p(x_i|\alpha,Z^i)$ is Gaussian with mean $\hat{x}_{i/i}(\alpha)$ and covariance $P_{i/i}(\alpha)$.

Notice that $\hat{x}_{i/i}(\alpha)$ and $P_{i/i}(\alpha)$ are obtained from the Kalman filter, in (3)–(7), for a particular α. The calculation of $p(\alpha|Z^i)$ is performed as follows.

$$\begin{aligned} p(\alpha|Z^i) &= p(\alpha|z_i,Z^{i-1}) \\ &= \frac{p(z_i|\alpha,Z^{i-1})p(\alpha|Z^{i-1})}{\int_A p(z_i|\alpha,Z^{i-1})p(\alpha|Z^{i-1})\,d\alpha} \end{aligned} \tag{9}$$

where A is the set of all α's. Notice that $p(z_i|\alpha,Z^{i-1})$ is Gaussian with mean $H_i\hat{x}_{i/i-1}(\alpha)$ and covariance $(H_iP_{i/i-1}(\alpha)H_i^T + R)$. Equation (9) is solved recursively starting from $p(\alpha)$, the *a priori* distribution of α.

An estimate of x_i which minimizes the mean-square error $\|x_i - \hat{x}_{i/i}\|^2$ is given by the conditional mean

$$\hat{x}_{i/i} = E(x_i|Z^i). \tag{10}$$

This can be obtained from $p(x_i,\alpha|Z^i)$ by integrating out α and taking expectation over x_i:

$$\begin{aligned} \hat{x}_{i/i} &= E_{x_i}p(x_i|Z^i) \\ &= E_{x_i}\int_A p(x_i|\alpha,Z^i)p(\alpha|Z^i)\,d\alpha \\ &= \int_A \hat{x}_{i/i}(\alpha)p(\alpha|Z^i)\,d\alpha \end{aligned} \tag{11}$$

where the linear operations of expectation and integration have been interchanged.

The main difficulty in using this algorithm is due to (9) and (11), which involve integration over a large dimensional space A. By making certain special assumptions, these calculations can be simplified. For example, if Q is known and the *a priori* density of the elements of R or α is assumed to be inverted gamma, it can be shown from (9) that the *a posteriori* density of α is also inverted gamma. *A priori* distributions having this property are called conjugate prior distributions. The advantage of using these distributions is that only a few parameters need to be updated with each new measurement.

Similarly, the calculation in (11) can be simplified if the integral can be evaluated explicitly. Alternately, if α is known to come from a finite set, the integrations can be replaced by summations over this finite set. Still, this requires calculation of $\hat{x}_{i/i}(\alpha)$ for different α, which can be very time-consuming.

IV. Maximum Likelihood Estimation

Three different kinds of maximum likelihood (ML) estimates can be defined, depending upon the density function used.

1) Joint ML Estimate: If x_i and α are to be estimated simultaneously, then $p(x_{ik},\alpha|Z^i)$ is maximized jointly with respect to x_i and α.

2) Marginal ML Estimate of α: The marginal density $p(\alpha|Z^i)$ is maximized with respect to α.

3) Conditional Mode Estimate: $p(x_i|Z^i)$ is maximized with respect to x_i. This estimate can be obtained from the Bayesian approach of Section III.

We first consider the marginal ML estimate of α since it is easiest to obtain. Using Baye's rule,

$$p(\alpha|Z^i) = \frac{p(Z^i/\alpha)p(\alpha)}{p(Z^i)} \tag{12}$$

where

$$\begin{aligned} p(Z^i|\alpha) &= p(Z^{i-1},z_i|\alpha) \\ &= p(z_i|Z^{i-1},\alpha)p(Z^{i-1}|\alpha) \\ &\cdots \\ &= p(z_i|Z^{i-1},\alpha)p(z_{i-1}|Z^{i-2},\alpha)\cdots p(z_1|\alpha). \end{aligned} \tag{13}$$

The ML estimate of α can now be obtained by maximizing

$$\begin{aligned} L(\alpha) &= \log p(\alpha|Z^i) \\ &= -\frac{1}{2}\sum_{k=1}^{i}\{\|z_k - H_k\hat{x}_{k/k-1}(\alpha)\|^2_{(H_kP_{k/k-1}(\alpha)H_k^T+R)^{-1}} \\ &\quad + \log|H_kP_{k/k-1}(\alpha)H_k^T + R|\} \\ &\quad + \log p(\alpha) + \text{constants}. \end{aligned} \tag{14}$$

If no *a priori* distribution of α is given, the log $p(\alpha)$ term is dropped. In other words, $p(Z^i|\alpha)$ or the joint probability of the observed sequence, given α, is maximized.

If x_i and α are to be estimated simultaneously, we maximize

$$\begin{aligned} L'(x_i,\alpha) &= \log p(x_i,\alpha|Z^i) \\ &= -\frac{1}{2}\{\log|P_{i/i}(\alpha)| + \|x_i - \hat{x}_{i/i}(\alpha)\|^2_{P^{-1}_{i/i}(\alpha)} \\ &\quad - \frac{1}{2}\sum_{k=1}^{i}\|z_k - H_k\hat{x}_{k/k}(\alpha)\|^2_{(H_kP_{k/k-1}(\alpha)H_k^T+R)^{-1}} \\ &\quad + \log|H_kP_{k/k-1}(\alpha)H_k^T + R|\} + \log P(\alpha) \\ &\quad + \text{constant}. \end{aligned} \tag{15}$$

Setting $\partial L'/\partial x_i = 0$ and $\partial L'/\partial\alpha = 0$, then

$$x_i = \hat{x}_{i/i}(\hat{\alpha}) \tag{16}$$

where $\hat{\alpha}$ is the root of the equation

$$\begin{aligned} \operatorname{Tr}\left(P_{i/i}^{-1}\frac{\partial P_{i/i}}{\partial\alpha^j}\right) + \sum_{k=1}^{i}\Big[\operatorname{Tr}(B_k^{-1} - B_k^{-1}\nu_k\nu_k^TB_k^{-1})\frac{\partial B_k}{\partial\alpha^j} \\ - 2B_k^{-1}\nu_k\frac{\partial\hat{x}_{k/k-1}^T}{\partial\alpha^j}H_k^T\Big] + \frac{\partial\log p(\alpha)}{\partial\alpha^j} = 0. \end{aligned} \tag{17}$$

Here α^j denotes the jth component of α, and

$$B_k = H_kP_{k/k-1}H_k^T + R \tag{18}$$

$$\nu_k = z_k - H_k\hat{x}_{k/k-1}. \tag{19}$$

Since (17) is nonlinear in α, an iterative method is required to find its root. A Newton–Raphson iteration can be used:

$$\alpha(l+1) = \alpha(l) - \left(\frac{\partial^2L'}{\partial\alpha^2}\right)^{-1}\frac{\partial L'}{\partial\alpha}\bigg|_{\alpha=\alpha(l)} \tag{20}$$

where l denotes the iteration number.

The evaluation of $\partial L'/\partial\alpha^j$ and $\partial^2L'/\partial(\alpha^j)^2$ is quite complicated since it involves evaluating terms like

$$\frac{\partial\hat{x}_{k/k-1}}{\partial\alpha^j},\quad \frac{\partial^2\hat{x}_{k/k-1}}{\partial(\alpha^j)^2},\quad \frac{\partial P_{k/k-1}}{\partial\alpha^j},\quad \frac{\partial P_{k/k-1}}{\partial(\alpha^j)^2},\ \text{etc.}$$

The number of difference equations to be solved increases rapidly with the dimension of α. There is thus a need for simplifying the

preceding calculations. This can be done by making the following assumptions.

1) The system is time invariant, i.e., matrices Φ_i and H_i are constant. They will be denoted by Φ and H.

2) The system is completely controllable and observable.

3) The filter has reached a steady state, i.e., the gain K_i and the covariances B_i, $P_{i/i}$, and $P_{i/i-1}$ are constants, viz., $K_i \to K$, $B_i \to B$, $P_{i/i} \to P$, and $P_{i/i-1} \to M$.

4) No *a priori* information on α is available, or the last term in (17) is dropped.

5) The first term in (17) is neglected since for large i it is much smaller than the second term in (17). This assumption is equivalent to maximizing $L(\alpha)$ instead of $L'(x_i,\alpha)$.

Notice that assumptions 3)–5) are valid asymptotically. Under these assumptions, (17) can be written as

$$\sum_{k=1}^{i} \operatorname{Tr}\left[(B^{-1} - B^{-1}\nu_k\nu_k^T B^{-1}) \frac{\partial B}{\partial \alpha^j} - 2B^{-1}\nu_k \frac{\partial \hat{x}_{k/k-1}^T}{\partial \alpha^j} H^T \right] = 0. \tag{21}$$

Up to this time, we have not specified the vector α of unknowns. The solution to (21) is greatly simplified if α is chosen from the elements of the matrix $\begin{bmatrix} B \\ K \end{bmatrix}$ or from a unique one-to-one transformation of $\begin{bmatrix} B \\ K \end{bmatrix}$. Since in a number of adaptive filtering situations, the exact form of Q is unknown and the important variable to estimate is the Kalman gain K, this choice of α is particularly appropriate. Moreover, it will be shown later that B and K can be estimated uniquely even though Q may not be. The set in (21) can now be simplified since $\hat{x}_{k/k-1}$ does not depend on B. We get the following two sets of equations:

$$\operatorname{Tr}\left[iB^{-1} - B^{-1}\left(\sum_{k=1}^{i} \nu_k\nu_k^T \right) B^{-1} \right] = 0 \tag{22}$$

and

$$\sum_{k=1}^{i} \operatorname{Tr}\left[\nu_k \frac{\partial \hat{x}_{k/k-1}^T}{\partial K^{jm}} H^T \right] = 0, \qquad \begin{matrix} j = 1, \cdots, n \\ m = 1, \cdots, r \end{matrix} \tag{23}$$

where K^{jm} denotes the (j,m) element of K. Equation (22) gives

$$\hat{B} = \frac{1}{i} \sum_{k=1}^{i} \nu_k\nu_k^T. \tag{24}$$

To solve (23), we must evaluate $\partial \hat{x}_{k/k-1}/\partial K^{jm}$. Consider (3) and (4) under steady state. Taking partials with respect to K^{jm},

$$\begin{aligned} \frac{\partial \hat{x}_{k/k-1}}{\partial K^{jm}} &= \Phi \frac{\hat{x}_{k-1/k-1}}{\partial K^{jm}} \\ &= \Phi\left[\frac{\partial \hat{x}_{k-1/k-2}}{\partial K^{jm}} + I_{jm}\nu_{k-1} - KH \frac{\partial \hat{x}_{k-1/k-2}}{\partial K^{jm}} \right] \\ &= \Phi(I - KH) \frac{\partial \hat{x}_{k-1/k-2}}{\partial K^{jm}} + I_{jm}\nu_{k-1} \end{aligned} \tag{25}$$

where I_{jm} is a matrix of all zeros except one in the (j,m) element. Equation (25) is a linear difference equation that can be solved recursively starting from $\partial \hat{x}_{0/-1}/\partial K^{jm} = 0$ for $j = 1, \cdots, n$ and $m = 1, \cdots, r$. However, since $\partial \hat{x}_{k/k-1}/\partial K^{jm}$ is nonlinear in K, an iterative scheme must be used to solve (23). We propose here a modified Newton–Raphson iteration in which the K^{jm} is updated as

$$\hat{K}^{jm}(l+1) = \hat{K}^{jm}(l) + \left[\sum_{k=1}^{i} \operatorname{Tr}\left(H \frac{\partial \hat{x}_{k/k-1}}{\partial K^m} \frac{\partial \hat{x}_{k/k-1}^T}{\partial K^{jm}} H^T \right) \right]^{-1} \cdot \left[\sum_{k=1}^{i} \operatorname{Tr}\left(\nu_k \frac{\partial \hat{x}_{k/k-1}^T}{\partial K^{jm}} H^T \right) \right] \Bigg|_{K^{jm} = \hat{K}^{jm}(l)} \tag{26}$$

where l denotes the iteration number.

For adaptive filtering, (24)–(26) must be solved recursively. This is easily done for (24), which can be written as

$$\hat{B} = \frac{1}{i-1} \sum_{k=1}^{i-1} \nu_k\nu_k^T + \frac{1}{i}\left[\nu_i\nu_i^T - \frac{1}{i-1} \sum_{k=1}^{i-1} \nu_k\nu_k^T \right]. \tag{27}$$

Equation (25) is already in a recursive form. Equation (26) can be made recursive only if a single iteration is used. To show this, let us denote

$$\Lambda_i = \sum_{k=1}^{i} \operatorname{Tr}\left(H \frac{\partial \hat{x}_{k/k-1}}{\partial K^{jm}} \frac{\partial \hat{x}_{k/k-1}^T}{\partial K^{jm}} H^T \right) \tag{28}$$

$$g_i = \sum_{k=1}^{i} \operatorname{Tr}\left(\nu_k \frac{\partial \hat{x}_{k/k-1}^T}{\partial K^{jm}} H^T \right). \tag{29}$$

It is clear that

$$\Lambda_{i+1} = \Lambda_i + \operatorname{Tr}\left(H \frac{\partial \hat{x}_{i+1/i}}{\partial K^{jm}} \frac{\partial \hat{x}_{i+1/i}}{\partial K^{jm}} H^T \right) \tag{30}$$

$$g_{i+1} = g_i + \operatorname{Tr}\left(\nu_{i+1} \frac{\partial \hat{x}_{+1/i}^T}{\partial K^{jm}} H^T \right) \tag{31}$$

and

$$\hat{K}_{i+1}{}^{jm} = \hat{K}_i{}^{jm} + \Lambda_{i+1}{}^{-1} g_{i+1} \tag{32}$$

where $\hat{K}_i{}^{jm}$ denotes the estimate of K^{jm} based on measurements up to i.

We now have the following *suboptimal ML adaptive filter.*

$$\hat{x}_{i+1/i} = \Phi[\hat{x}_{i/i-1} + \hat{K}_i\nu_i], \ \hat{x}_{0/-1} \text{ given} \tag{33}$$

$$\nu_i = z_i - H\hat{x}_{i/i-1} \tag{34}$$

$$\frac{\partial \hat{x}_{i+1/i}}{\partial K^{jm}} = \Phi(I - \hat{K}_i H) \frac{\partial \hat{x}_{i/i-1}}{\partial K^{jm}} + I_{jm}\nu_i \tag{35}$$

$$\frac{\partial \hat{x}_{0/-1}}{\partial K^{jm}} = 0, \qquad j = 1, \cdots, n \text{ and } m = 1, \cdots, r.$$

$\hat{K}_i$ is updated according to (30)–(32), and $\hat{B}_i$ is updated according to (27).

The convergence of this filter has not been proved as yet.

Estimation of Q and R

Under steady-state filtering conditions, K and B are related to Q and R by the following equations [see (5)–(7)]:

$$K = MH^TB^{-1} \tag{36}$$

$$B = HMH^T + R \tag{37}$$

$$P = (I - KH)M \tag{38}$$

$$M = \Phi P \Phi^T + Q. \tag{39}$$

Solving (36)–(39) for M, Q, and R, we get

$$M = KB(H^T)\dagger \tag{40}$$

$$Q = M - \Phi(I - KH)M\Phi^T \tag{41}$$

$$R = (I - KH)B$$

where $(H^T)^\dagger$ denotes the pseudo-inverse of H^T. Notice that a unique solution is obtained for R, but not, in general, for M and Q. However, if $r \geq n$, i.e., if the dimension of the observation vector is greater than or equal to the dimension of the state vector and H is of full rank, a unique solution for M and Q is obtained. In another case, one must put additional restrictions (besides positive semidefiniteness and symmetry) on M or Q to obtain a unique solution. One such restriction can be the minimization of a suitable norm of M which might lead to a Penrose type of pseudo-inverse.

V. Correlation Methods

Correlation methods have been used for estimation in time series analysis for a long time. The basic idea is to correlate the output of the system either directly or after a known linear operation on it. A set of equations is derived relating the system parameters to the observed autocorrelation function and these equations are solved simultaneously for the unknown parameters. These methods are mainly applicable to constant coefficient systems. Two different methods can be developed in the present case by considering either the autocorrelation function of the output z_i or the autocorrelation function of the innovations ν_i.

It turns out that the estimates obtained from the second method are more efficient than those obtained from the first method since the innovations ν_i are less correlated than the output z_i. This is not surprising since the ML estimates of Section IV also use correlations of the innovations [cf. (24)–(26)]. Moreover, the first method is applicable only if the output z_i is stationary or Φ is a stable matrix, whereas the second method is applicable to cases where Φ is not stable. Both methods require the system to be completely controllable and observable. It should be mentioned that stochastic approximation methods are special cases of the correlation methods.

Output Correlation Method

Let C_k be the kth lag autocorrelation of the output z_i:

$$C_k = E\{z_i z_{i-k}^T\}. \tag{42}$$

It is assumed here that the output z_i is stationary so that the autocorrelation is only a function of lag. An expression for C_k is easily derived from (1) and (2):

$$C_k = \begin{cases} H\Sigma H^T + R, & k = 0 \quad (43) \\ H\Phi\Sigma H^T, & k > 0 \quad (44) \end{cases}$$

where $\Sigma = E\{x_i x_i^T\}$ and satisfies

$$\Sigma = \Phi\Sigma\Phi^T + Q. \tag{45}$$

We now show that (43), (44), and (45) can be solved for Q and R provided the system is observable. Writing (44) for $k = 1, \cdots, n$,

$$\begin{bmatrix} C_1 \\ \cdot \\ \cdot \\ \cdot \\ C_n \end{bmatrix} = A\Sigma H^T \tag{46}$$

where $A^T = [\Phi^T H^T, \cdots, (\Phi^T)^n H^T]$.

Since Φ is nonsingular and the system is observable, rank $(A) = n$. Solving (46) for ΣH^T,

$$\Sigma H^T = (A^T A)^{-1} A^T \begin{bmatrix} C_1 \\ \cdot \\ \cdot \\ \cdot \\ C_n \end{bmatrix}. \tag{47}$$

Using (44) for $k = 0$,

$$R = C_0 - H(\Sigma H^T). \tag{48}$$

To obtain Q, we must solve (45) and (47), which are of the same type as (40) and (41). As before, a unique solution for Q cannot be obtained, in general. However, the optimal steady-state filter gain K can still be determined uniquely as follows.

Define

$$\pi = E\{\hat{x}_{i/i-1}\hat{x}_{i/i-1}^T\}.$$

From (3) and (4),

$$\hat{x}_{i+1/i} = \Phi[\hat{x}_{i/i-1} + K\nu_i]. \tag{49}$$

Therefore,

$$\pi = \Phi[\pi + KBK^T]\Phi^T \tag{50}$$

where use has been made of the fact that ν_i is white and uncorrelated to $\hat{x}_{i/i-1}$. Another relationship for π can be derived from the fact that

$$x_i = \hat{x}_{i/i-1} + e_i \tag{51}$$

where the error e_i has covariance M and is uncorrelated to $\hat{x}_{i/i-1}$ by virtue of the orthogonality principle:

$$E\{\hat{x}_{i/i-1}e_i^T\} = 0. \tag{52}$$

Therefore,

$$\Sigma = \pi + M. \tag{53}$$

Now the Kalman gain K is given by

$$\begin{aligned} K &= MH^T B^{-1} \\ &= (\Sigma - H)H^T B^{-1} \\ &= (\Sigma - \pi)H^T(C_0 - H\pi H^T)^{-1}. \end{aligned} \tag{54}$$

Substituting for K in (50),

$$\pi = \Phi[\pi + (\Sigma - \pi)H^T(C_0 - H\pi H^T)^{-1}H(\Sigma - \pi)]\Phi^T \tag{55}$$

which can be solved for π since ΣH^T is known from (47). Then K is obtained from (54).

We have now derived all the equations needed to calculate Q, R, and K from the autocorrelations $C_0, \cdots, C_n$. Using the ergodic property of the random sequence z_i, the autocorrelations are estimated as

$$\hat{C}_k^N = \frac{1}{N}\sum_{i=k}^{N} z_i z_{i-k}^T \tag{56}$$

where N is the sample size.

Estimates $\hat{C}_k^N$ can be computed recursively for $k = 0, \cdots n$:

$$\hat{C}_k^N = \hat{C}_k^{N-1} + \frac{1}{N}(z_N z_{N-k}^T - \hat{C}_k^{N-1}). \tag{57}$$

It can be shown that the estimates $\hat{C}_k$ are asymptotically normal, unbiased, and consistent. Since Q and R are linearly related to $\hat{C}_k$, their estimates are also asymptotically normal, unbiased, and consistent. Similarly, the estimate of K is asymptotically unbiased and consistent.

Innovation Correlation Method

It is known from the theory of Kalman filtering that the innovation sequence

$$\nu_i = z_i - H\hat{x}_{i/i-1} = He_i + \nu_i \tag{58}$$

is a zero-mean Gaussian white noise sequence for an optimal filter. However, for a suboptimal filter, the innovation sequence is correlated as shown below.

Let

$$\Gamma_k = E\{\nu_i \nu_{i-k}^T\}.$$

Substituting from (58) and considering $k > 0$,

$$\begin{aligned} \Gamma_k &= E\{(He_i + v_i)(He_{i-k} + v_{i-k})^T\} \\ &= HE\{e_i e_{i-k}^T\}H^T + HE\{e_i v_{i-k}^T\}. \end{aligned} \tag{59}$$

For $k = 0$,

$$\Gamma_0 = HMH^T + R.$$

The expectation terms in (59) can be evaluated by writing the difference equation for e_i using (1)–(4). We assume that the filter uses a suboptimal *a priori* gain K_0:

$$e_i = \Phi(I - K_0 H)e_{i-1} - \Phi K_0 v_{i-1} + u_{i-1}. \tag{60}$$

Carrying (60) k steps back,

$$e_i = [\Phi(I - K_0H)]^k e_{i-k} - \sum_{j=1}^{k} [\Phi(I - K_0H)]^{j-1}\Phi K_0 v_{i-j} + \sum_{j=1}^{k} [\Phi(I - K_0H)]^{j-1} u_{i-j}. \quad (61)$$

Therefore,

$$E\{e_i e_{i-k}^T\} = [\Phi(I - K_0H)]^k M_1 \quad (62)$$

where $M_1 = E\{e_i e_i^T\}$. Similarly,

$$E\{e_i v_{i-k}^T\} = -[\Phi(I - K_0H)]^{k-1}\Phi K_0 R. \quad (63)$$

Substituting in (59),

$$\Gamma_k = H[\Phi(I - K_0H)]^{k-1}\Phi[M_1H^T - K_0\Gamma_0]. \quad (64)$$

For an optimal filter,

$$K = MH^T(HMH^T + R)^{-1}$$

so that Γ_k for $k > 0$ vanishes identically. When there are uncertainties in Q and R, the covariance M_c calculated from (5)–(7) will be different from the true covariance M_1 and the filter gain K_0 will be suboptimal. Therefore, Γ_k will be nonzero. The optimal gain K, Q, and R can be obtained as follows.

1) Obtain M_1H^T by solving (64) for $k = 1, \cdots, n$.

$$M_1H^T = (A^TA)^{-1}A^T \begin{bmatrix} \Gamma_1 + H\Phi K_0\Gamma_0 \\ \Gamma_2 + H\Phi K_0\Gamma_1 + H\Phi^2 K_0\Gamma_0 \\ \vdots \\ \Gamma_n + H\Phi K_0\Gamma_{n-1} + \cdots H\Phi_k K_0\Gamma_0 \end{bmatrix} \quad (65)$$

where A is as defined in (46).

2) Calculate R by using Γ_0 and M_1H^T.

$$R = \Gamma_0 - H(M_1H^T). \quad (66)$$

3) Denote by M the error covariance associated with optimal gain K. Then from (38) and (39),

$$M = \Phi(M - KHM)\Phi^T + Q. \quad (67)$$

An equation for M_1 is derived easily from (60):

$$M_1 = \Phi[M_1 - K_0HM_1 - M_1H^TK_0^T + K_0(HM_1H^T + R)K_0^T]\Phi^T + Q. \quad (68)$$

Subtracting (68) from (67),

$$(M - M_1) = \Phi[M - M_1 - KHM + K_0HM_1 + M_1H^TK_0^T - K_0(HM_1H^T + R)K_0^T]\Phi^T. \quad (69)$$

Let $\delta M = M - M_1$.

The optimal gain K can be written as

$$K = (M_1H^T + \delta MH^T)(C_0 + H\delta MH^T)^{-1}. \quad (70)$$

Substituting in (69),

$$\delta M = \Phi[\delta M - (M_1H^T + \delta MH^T)(C_0 + H\delta MH^T)^{-1}(HM_1 + H\delta M) + K_0HM_1 + M_1H^TK_0^T - K_0\Gamma_0K_0^T]\Phi^T. \quad (71)$$

Notice the similarity between (70)–(71) and (54)–(55). Solving (71) for δM by using M_1H^T from (65), optimal gain K can be obtained in one step from (70). In actual practice, if a batch of observations z_i can be stored, these calculations might be repeated to improve the estimates. The innovation sequence ν_i will become more and more white with each iteration, resulting in better and better estimates for the autocorrelations $\Gamma_0, \Gamma_1, \cdots$, and, therefore, for R and K. The estimation of Q can be made from (40) and (41).

The preceding algorithm can be used on-line since the autocorrelations

$$\hat{\Gamma}_k = \frac{1}{N}\sum_{i=k}^{N} \nu_i\nu_{i-k}^T$$

can be computed on-line. It can be shown that the estimates $\hat{\Gamma}_k$, $\hat{R}$, and $\hat{K}$ are asymptotically unbiased and consistent.

The efficiency of the correlation methods can be improved by using higher order correlations, i.e., $k > n$ to estimate ΣH^T or M_1H^T. Moreover, the two correlation methods can be combined by using the output correlation method as a start-up procedure since it does not require *a priori* estimates of R and Q (or K).

VI. Covariance-Matching Techniques

The basic idea behind the covariance-matching techniques is to make the residuals consistent with their theoretical covariances. For example, consider the innovation sequence ν_i, which has a theoretical covariance of $(HP_{i/i-1}H^T + R)$. If it is noticed that the actual covariance of ν_i is much larger than $(HP_{i/i-1}H^T + R)$ obtained from the Kalman filter, then the process noise Q should be increased. This has the effect of increasing $P_{i/i-1}$ and bringing the actual covariance of ν_i closer to $(HP_{i/i-1}H^T + R)$. The actual covariance of ν_i is approximated by its sample covariance, viz.,

$$\frac{1}{m}\sum_{i=1}^{m} \nu_i\nu_i^T$$

where m is chosen empirically to give some statistical smoothing. An equation for Q is obtained by setting

$$HP_{i/i-1}H^T + R = E\{\nu_i\nu_i^T\} \quad (72)$$

or

$$H(\Phi P_{i-1/i-1}\Phi^T + Q)H^T + R = E\{\nu_i\nu_i^T\}$$

or

$$HQH^T = E\{\nu_i\nu_i^T\} - H\Phi P_{i-1/i-1}\Phi^T H^T - R. \quad (73)$$

Equation (73) does not give a unique solution for Q if H is of rank less than n. However, if the number of unknowns in Q are restricted, a unique solution can be obtained. Notice, however, that (73) is only approximate since $P_{i/i-1}$ and $P_{i-1/i-1}$ do not represent the actual error covariances when the true values of Q and R are unknown. For this reason, the convergence of the covariance matching techniques is doubtful.

It is possible to write more equations like (73) by considering predicted residuals, viz., $(x_{i+j} - \hat{x}_{i+j/i})$, $j = 2,3,\cdots$, and matching their estimated covariance with the theoretical covariance. This gives extra equations in the unknowns of Q and R. Certain authors have also considered filtering residuals $(x_i - \hat{x}_{i/i})$ and smoothing residuals $(x_i - \hat{x}_{i/i-1})$, etc. However, the convergence of these methods has not been proved.

The case in which Q is known but R is unknown has been handled more successfully by these methods. In this case, R is estimated from (72) as

$$\hat{R}_i = \frac{1}{m}\sum_{j=1}^{m} \nu_{i-j}\nu_{i-j}^T - HP_{i/i-1}H^T. \quad (74)$$

Notice that (74) is of the same type as (24) or (66) except that $P_{i/i-1}$ is obtained from the Kalman filter. In certain applications where Q is known, this turns out to be a good approximation.

VII. Conclusions

Different approaches to adaptive filtering are described as belonging to one of four categories: Bayesian, maximum likelihood, correlation, or covariance matching. The relationship between these

methods is shown. Three new algorithms for the adaptive estimation of the steady-state Kalman gain K are given. It is shown that the optimal gain K can be estimated uniquely, even when a unique estimate for the process noise covariance Q cannot be obtained.

Bibliography

Kalman Filter

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME*, J. Basic Eng., vol. 82, pp. 35–44, 1960.
[2] T. Kailath, "An innovations approach to least-squares estimation—Part I: Linear filtering in additive white noise," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 646–655, Dec. 1968.

Error Analysis of Kalman Filter

[3] H. Heffes, "The effect of erroneous models on the Kalman filter response," *IEEE Trans. Automat. Contr.* (Short Papers), vol. AC-11, pp. 541–543, July 1966.
[4] T. Nishimura, "Error bounds of continuous Kalman filters and the application to orbit determination problems," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 268–275, June 1967.
[5] R. J. Fitzgerald, "Error divergence in optimal filtering problems," presented at the 2nd IFAC Symp. Automatic Control in Space, Vienna, Austria, 1967.
[6] C. F. Price, "An analysis of the divergence problem in the Kalman filter," *IEEE Trans. Automat. Contr.* (Short Papers), vol. AC-13, pp. 699–702, Dec. 1968.
[7] F. H. Schlee, C. J. Standish, and N. F. Toda, "Divergence in the Kalman filter," *AIAA J.*, vol. 5, pp. 1114–1120, June 1967.

Bayesian Estimation

[8] D. T. Magill, "Optimal adaptive estimation of sampled stochastic processes," *IEEE Trans. Automat. Contr.*, vol. AC-10, pp. 434–439, Oct. 1965.
[9] C. G. Hilborn, Jr., and D. G. Lainiotis, "Optimal estimation in the presence of unknown parameters," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-5, pp. 38–43, Jan. 1969.
[10] F. L. Sims, D. G. Lainiotis, and D. T. Magill, "Recursive algorithm for the calculation of the adaptive Kalman filter weighting coefficients," *IEEE Trans. Automat. Contr.* (Tech. Notes and Corresp.), vol. AC-14, pp. 215–218, Apr. 1969.
[11] C. G. Hilborn, Jr., and D. G. Lainiotis, "Optimal adaptive filter realizations for sampled stochastic processes with unknown parameters," Dep. Elec. Eng., Univ. Texas, Austin, Communications Systems Research Group Tech. Rep. 67-3, July 1967.
[12] G. L. Smith, "Sequential estimation of observation error variances in a trajectory estimation problem," *AIAA J.*, vol. 5, pp. 1964–1970, Nov. 1967.
[13] H. Raiffa and R. Schlaifer, "Applied statistical decision theory," Grad. School Business Admin., Harvard Univ., Cambridge, Mass., 1961.

Maximum Likelihood Estimation

[14] K. J. Åström and S. Wenmark, "Numerical identification of stationary time series," presented at the 6th Int. Instruments and Measurements Congr., Section on Automatic Control, Stockholm, Sweden, Sept. 1964.
[15] R. L. Kashyap, "Maximum likelihood identification of stochastic linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 25–34, Feb. 1970.
[16] R. K. Mehra, "Identification of stochastic linear dynamic systems using Kalman filter representation," in *Proc. 8th IEEE Symp. Adaptive Processes* (Pennsylvania State Univ., University Park, Nov. 1969); also *AIAA J.*, Oct. 1970.
[17] P. D. Abramson, Jr., "Simultaneous estimation of the state and noise statistics in linear dynamic systems," M.I.T. Rep. TE-25, May 10, 1968.

Output Correlation Method

[18] W. N. Anderson *et al.*, "Consistent estimates of the parameters of a linear system," *Ann. Math. Statist.*, Dec. 1969.
[19] R. K. Mehra, "On-line identification of linear dynamic systems with applications to Kalman filtering," in *1970 Proc. Joint Automat. Contr. Conf.* (Atlanta, Ga.); also *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 12–21, Feb. 1971.
[20] P. Faurre and J. P. Maumarat, "Une algorithme de realization stochastique," *C. R. Acad. Sci.*, vol. 268, pp. 978–981, Apr. 28, 1969.

Innovation Correlation Method

[21] R. K. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 175–184, Apr. 1970.

Covariance-Matching Methods

[22] J. C. Shellenbarger, "Estimation of covariance parameters for an adaptive Kalman filter," in *Proc. Nat. Electronics Conf.*, vol. 22, 1966, pp. 698–702.
[23] ——, "A multivariance learning technique for improved dynamic system performance," in *Proc. Nat. Electronics Conf.*, vol. 23, 1967, pp. 146–151.
[24] A. P. Sage and G. W. Husa, "Adaptive filtering with unknown prior statistics," in *1969 Proc. Joint Automat. Contr. Conf.*, pp. 760–769.
[25] A. H. Jazwinski and A. E. Bailie, "Adaptive filtering," Analytical Mechanics Associates, Inc., Rep. 67-6, Mar. 1967.
[26] A. E. Baker, "Variance of the state estimation error of an adaptive Kalman filter," Ph.D. dissertation, Iowa State Univ., Ames, 1967.
[27] A. R. Dennis, "Functional updating and adaptive noise variance determination in recursive-type trajectory estimators," presented at the Special Projects Branch Astrodynamics Conf., NASA Goddard Space Flight Center, Greenbelt, Md., May 1967.

Survey Papers

[28] A. P. Sage and G. W. Husa, "Algorithms for sequential adaptive estimation of prior statistics," in *Proc. 8th IEEE Symp. Adaptive Processes* (Pennsylvania State Univ., University Park, Nov. 1969).
[29] I. M. Weiss, "A survey of discrete Kalman–Bucy filtering with unknown noise covariances," presented at the AIAA Guidance, Control, and Flight Mechanics Conf., Santa Barbara, Calif., Aug. 1970, Paper 70-995.

A New Control Approach Using the Inverse System

S. S. GODBOLE AND C. F. SMITH, MEMBER, IEEE

***Abstract*—A novel control scheme is presented. It utilizes the inverse system concept along with the basic structure of the conventional model following control scheme. It is a direct approach that does not involve any optimization and is simple to apply. The scheme is illustrated by several examples. An extension of this scheme to adaptive control is suggested and areas for further research are indicated.**

I. Introduction

The parallel of the idea of pole-zero cancellation in classical control theory is the idea of (pre-) inverse in modern control theory. The concept of precascading a given multivariable plant with its inverse for control purposes is probably not new. However, the evolution of a practical scheme using this concept has become possible only recently as a result of the study of inverses by several authors such as Sain and Massey [1] and Silverman [2] (for a good survey of inverse systems, see Singh and Liu [3]). The use of an inverse for control is a rather direct approach in contrast to the usual one employing optimization theory.

This paper presents such a control scheme, i.e., one that uses an inverse. The scheme has the basic structure of a model following scheme, which is one of the earliest adaptive control schemes [4]. Such a structure was chosen for two main reasons, namely, 1) it is very natural for this application of inverse, and 2) it makes the new scheme easily extendable to adaptive control (such an extension is suggested later).

II. Preliminary Results

Model-Follower Scheme

The basic scheme is shown in Fig. 1. The controller may typically be the highest possible gain consistent with overall stability [4]. In a more recent approach [5], the controller consists of a linear combination of the model input, states, and the plant states. The details of this combination are determined by an optimization technique. In the scheme proposed in this paper, the controller consists of the plant inverse, which is determined from the plant state equations.

Inversion Algorithm

The inversion algorithm used in this paper and summarized below is basically the same as that of Silverman [2]. However, it should be pointed out that the various inverses reported in recent literature by different authors are closely related [3].

Let the plant be described by

$$\dot{x} = Ax + Bu \tag{1a}$$

$$y = Cx + Du \tag{1b}$$

where $x \in R^n$, $u \in R^m$, $y \in R^r$, $r \leq m$, and A,B,C,D are constant matrices of compatible order. If $r > m$, then only m outputs can be functionally reproduced, if an inverse exists. In such a case, the y vector should contain only the m most important outputs. The inversion algorithm as used in this paper consists of sequences $\{\mathcal{R}_0,\mathcal{R}_1,\cdots,\mathcal{R}_\alpha\}$ of elementary row operations and $\{\mathcal{D}_1,\mathcal{D}_2,\cdots,\mathcal{D}_\alpha\}$ of differentiations, $\alpha \leq n$, applied to (1b) and $\{\mathcal{C}\} = \{\mathcal{C}_0,\mathcal{C}_1,\cdots,\mathcal{C}_\alpha\}$ of elementary column operations applied to D, u^T, and B, all in a mixed order. At the end of these operations, (1a), (1b) become

$$\dot{x} = Ax + B_\alpha u_\alpha \tag{2a}$$

Manuscript received March 2, 1971; revised October 22, 1971, and May 9, 1972.

S. S. Godbole is with Babcock and Wilcox Company, Lynchburg, Va. 24505.
C. F. Smith is with the Department of Electrical Engineering, Virginia Polytechnic Institute, Blacksburg, Va. 24061.