On the Identification of Variances and Adaptive Kalman Filtering

RAMAN K. MEHRA, MEMBER, IEEE

Abstract—A Kalman filter requires an exact knowledge of the process noise covariance matrix Q and the measurement noise covariance matrix R. Here we consider the case in which the true values of Q and R are unknown. The system is assumed to be constant, and the random inputs are stationary. First, a correlation test is given which checks whether a particular Kalman filter is working optimally or not. If the filter is suboptimal, a technique is given to obtain asymptotically normal, unbiased, and consistent estimates of Q and R. This technique works only for the case in which the form of Q is known and the number of unknown elements in Q is less than $n \times r$ where n is the dimension of the state vector and r is the dimension of the measurement vector. For other cases, the optimal steady-state gain K_{op} is obtained directly by an iterative procedure without identifying Q. As a corollary, it is shown that the steady-state optimal Kalman filter gain K_{op} depends only on $n \times r$ linear functionals of Q. The results are first derived for discrete systems. They are then extended to continuous systems. A numerical example is given to show the usefulness of the approach.

I. Introduction

THE OPTIMUM filtering results of Kalman and Bucy [1], [2] for linear dynamic systems require an exact knowledge of the process noise covariance matrix Q and the measurement noise covariance matrix R. In a number of practical situations, Q and R are either unknown or are known only approximately. Heffes [3] and Nishimura [4] have considered the effect of errors in Q and R on the performance of the optimal filter. Several other investigators [5]–[9] have proposed on-line schemes to identify Q and R. Most of these schemes do well in identifying R but run into difficulties in identifying Q. Moreover, their extension to continuous cases is not clear. A different approach has been taken in this paper. It is assumed that the system under consideration is time invariant, completely controllable, and observable [2]. Both the system and the filter (optimal or suboptimal) are assumed to have reached steady-state conditions. First, a correlation test is performed on the filter to check whether it is working optimally or not. The test is based on the innovation property of an optimal filter [10]. If the filter is suboptimal, the auto-correlation function of the innovation process is used to obtain asymptotically unbiased and consistent estimates of Q and R. The method has the limitation that the number of unknown elements in Q must be less than $n \times r$ where n is the dimension of the state vector and r is the dimension of the measurement vector. It is shown that in spite of this limitation, the optimal steady-state filter gain can be obtained by an iterative procedure. As a corollary, it is shown that the Kalman filter gain depends only on $n \times r$ linear relationships between the elements of Q.

A numerical example is included to illustrate the application of the results derived in the paper. The extension of the results to the continuous case is straightforward and is given in the last section.

Manuscript received August 2, 1968; revised May 16, 1969. This paper was presented at the 1969 Joint Automatic Control Conference, Boulder, Colo.

The author was with the Analytic Sciences Corporation, Reading, Mass. 01867. He is now with Systems Control, Inc., Palo Alto, Calif. 94306.

II. Statement of the Problem

System

Consider a multivariable linear discrete system

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

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

where x_i is $n \times 1$ state vector, Φ is $n \times n$ nonsingular transition matrix, Γ is $n \times q$ constant input matrix, z_i is $r \times 1$ measurement vector, and H is $r \times n$ constant output matrix.

The sequences $u_i(q \times 1)$ and $v_i(r \times 1)$ are uncorrelated Gaussian white noise sequences with means and covariances as follows:

$$E\{u_i\} = 0; \quad E\{u_i u_j^T\} = Q\delta_{ij}$$

$$E\{v_i\} = 0; \quad E\{v_i v_j^T\} = R\delta_{ij}$$

$$E\{u_i v_j^T\} = 0, \quad \text{for all } i,j$$

where $E\{\cdot\}$ denotes the expectation, and δ_{ij} denotes the Kronecker delta function.

Q and R are bounded positive definite matrices ($Q > 0$, $R > 0$). Initial state x_0 is normally distributed with zero mean and covariance P_0.

The system is assumed to be completely observable and controllable, i.e.,

$$\text{rank } [H^T, (H\Phi)^T, \cdots, (H\Phi^{n-1})^T] = n$$

$$\text{rank } [\Gamma, \Phi\Gamma, \cdots, \Phi^{n-1}\Gamma] = n.$$

Filter

Let Q_0 and R_0 be the initial estimates of Q and R ($Q_0 > 0$, $R_0 > 0$). Using these estimates, let the steady-

state Kalman filter gain be K_0 ($n \times r$ matrix)[1]

$$K_0 = M_0H^T(HM_0H^T + R_0)^{-1} \tag{3}$$

$$M_0 = \Phi[M_0 - M_0H^T(HM_0H^T + R_0)^{-1}HM_0]\Phi^T + \Gamma Q_0\Gamma^T. \tag{4}$$

M_0 may be recognized as the steady-state solution to the covariance equations of Kalman [1].

The filtering equations are

$$\hat{x}_{i+1/i} = \Phi\hat{x}_{i/i} \tag{5}$$

$$\hat{x}_{i/i} = \hat{x}_{i/i-1} + K_0(z_i - H\hat{x}_{i/i-1}) \tag{6}$$

where $\hat{x}_{i+1/i}$ is the estimate of x_{i+1} based on all the measurements up to i, i.e., $\{z_0,\cdots,z_i\}$.

In an optimal Kalman filter (i.e., when $Q_0 = Q$ and $R_0 = R$), M_0 is the covariance of the error in estimating the state. But in a suboptimal case, the covariance of the error (M_1) is given by the following equation [3]:

$$M_1 = \Phi[M_1 - K_0HM_1 - M_1H^TK_0^T + K_0(HM_1H^T + R)K_0^T]\Phi^T + \Gamma Q\Gamma^T \tag{7}$$

where $M_1 = E\{(x_i - \hat{x}_{i/i-1})(x_i - \hat{x}_{i/i-1})^T\}$.

Problem

The true values of Q and R are unknown. It is required to

1) check whether the Kalman filter constructed using some estimates of Q and R is close to optimal or not (hypothesis testing),
2) obtain unbiased and consistent estimates of Q and R (statistical estimation), and
3) adapt the Kalman filter at regular intervals using all the previous information (adaptive filtering).

To solve these problems, we make use of the innovation property of an optimal filter [10].

III. The Innovation Property of an Optimal Filter[2]

Statement

For an optimal filter, the sequence $\nu_i = (z_i - H\hat{x}_{i/i-1})$, known as the innovation sequence, is a Gaussian white noise sequence.

Proof: A direct proof is obtained using the orthogonality principle of linear estimation [10].[3] Let $e_i = x_i - \hat{x}_{i/i-1}$ denote the error in estimating the state.

Then

$$\nu_i = He_i + v_i$$

$$E\{\nu_i\nu_j^T\} = E\{(He_i + v_i)(He_j + v_j)^T\}. \tag{8}$$

For $i > j$, v_i is independent of e_j and v_j:

$$E\{\nu_i\nu_j^T\} = E\{He_i(He_j + v_j)^T\} = E\{He_i(z_j - H\hat{x}_{j/j-1})^T\}.$$

The orthogonality principle states that e_i is orthogonal to $\{z_k, k < i\}$. Since $\hat{x}_{j/j-1}$ depends only on $\{z_k, k < j\}$, we conclude that

$$E\{\nu_i\nu_j^T\} = 0, \quad \text{for } i > j.$$

Similarly, $E\{\nu_i\nu_j^T\} = 0$, for $i < j$.

For $i = j$, $E\{\nu_i\nu_i^T\} = HMH^T + R$. Further, since ν_i is a linear sum of Gaussian random variables, it is also Gaussian. Hence ν_i is a Gaussian white noise sequence.

Heuristically, the innovation ν_i represents the new information brought by z_i. Kailath [10] shows that ν_i and z_i contain the same statistical information and are equivalent as far as linear operations are concerned. Schweppe [12] shows that ν_i can be obtained from z_i by a Gram–Schmidt orthogonalization (or a whitening) procedure.

In this paper, we use the innovation sequence to check the optimality of a Kalman filter and to estimate Q and R. With this in mind, we investigate the effect of suboptimality on the innovation sequence.

IV. Innovation Sequence for a Suboptimal Filter

Let K denote the steady-state filter gain. We will show that under steady state, the innovation sequence ν_i is a stationary Gaussian sequence:

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

$$E\{\nu_i\nu_{i-k}^T\} = HE\{e_ie_{i-k}\}H^T + HE\{e_iv_{i-k}^T\}, \quad \text{for } k > 0.$$

A recursive relationship can be obtained for e_i by using (1), (2), (5), and (6):

$$e_i = \Phi(I - KH)e_{i-1} - \Phi Kv_{i-1} + \Gamma u_{i-1}. \tag{9}$$

Carrying (9) k steps back,

$$e_i = [\Phi(I - KH)]^k e_{i-k} - \sum_{j=1}^{k}[\Phi(I - KH)]^{j-1}\Phi Kv_{i-j} + \sum_{j=1}^{k}[\Phi(I - KH)]^{j-1}\Gamma u_{i-j}. \tag{10}$$

Postmultiplying (10) by e_{i-k}^T and taking expectations,

$$E\{e_ie_{i-k}^T\} = [\Phi(I - KH)]^k M$$

where M is the steady-state error covariance matrix. An expression for M is obtained directly from (9) or from (7):

$$M = \Phi(I - KH)M(I - KH)^T\Phi^T + \Phi KRK^T\Phi^T + \Gamma Q\Gamma^T. \tag{11}$$

Postmultiplying (10) by v_{i-k}^T and taking expectations,

$$E\{e_iv_{i-k}^T\} = -[\Phi(I - KH)]^{k-1}\Phi KR.$$

[1] The conditions of complete controllability and observability together with the positive definiteness of Q_0 and R_0 ensure the asymptotic global stability of the Kalman filter. See Deyst and Price [11].

[2] For a detailed discussion, see Kailath [10].

[3] An alternate proof will be given in Section IV.

Therefore,

$$E\{\nu_i\nu_{i-k}^T\} = H[\Phi(I - KH)]^{k-1} \cdot \Phi[MH^T - K(HMH^T + R)], \quad k > 0.$$

When $k = 0$, $E\{\nu_i\nu_i^T\} = HMH^T + R$.

It is seen that the autocorrelation function of ν_i does not depend on i. Therefore, ν_i is a stationary Gaussian random sequence (Gaussian because of linearity) and we can define

$$C_k \equiv E\{\nu_i\nu_{i-k}^T\}.$$

Then

$$C_k = HMH^T + R, \qquad k = 0 \quad (12)$$

$$= H[\Phi(I - KH)]^{k-1}\Phi[MH^T - KC_0], \quad k > 0. \quad (13)$$

Furthermore,

$$C_{-k} = C_k^T.$$

Notice that the optimal choice of K, viz. $K = MH^T(HMH^T + R)^{-1}$ makes C_k vanish for all $k \neq 0$ (the innovation property).

V. A Test of Optimality for a Kalman Filter

From the discussion of the preceding two sections, it is clear that a necessary and sufficient condition for the optimality of a Kalman filter is that the innovation sequence ν_i be white. This condition can be tested statistically by a number of different methods [13], [16]–[19]. Here we consider a particular method given in Jenkins and Watts [13].

In this method, we obtain an estimate of C_k, denoted as $\hat{C}_k$, by using the ergodic property of a stationary random sequence

$$\hat{C}_k = (1/N)\sum_{i=k}^{N} \nu_i\nu_{i-k}^T \quad (14)$$

where N is the number of sample points.

The estimates $\hat{C}_k$ are biased for finite sample sizes:

$$E\{\hat{C}_k\} = (1 - k/N)C_k. \quad (15)$$

In case an unbiased estimate is desired, we divide by $(N - k)$ instead of N in (14). However, it is shown in [13] that the estimate of (14) is preferable since it gives less mean-square error than the corresponding unbiased estimate.

An expression for the covariance of $\hat{C}_k$ can be derived by straightforward manipulation, but the general results are rather involved. We quote here approximate results for large N given in Bartlett [14]:

$$\text{cov}\,([\hat{C}_k]_{ij},[\hat{C}_l]_{pq}) \approx (1/N)\sum_{t=-\infty}^{\infty} ([C_t]_{ip}[C_{t+l-k}]_{jq} + [C_{t+l}]_{iq}[C_{t-k}]_{jp}) \quad (16)$$

where $[\hat{C}_k]_{ij}$ denotes the element in the ith row and the jth column of the matrix $\hat{C}_k$ and cov (a,b) denotes the covariance of a and b; viz.

$$\text{cov}\,(a,b) \equiv E\{[a - E(a)][b - E(b)]\}.$$

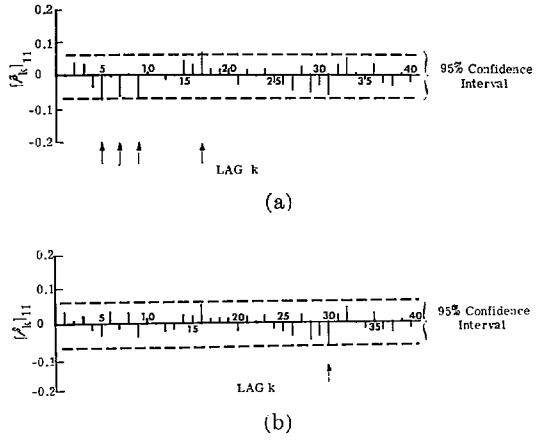

Fig. 1. Normalized autocorrelation function of innovation process. (a) Suboptimal filter. (b) Optimal filter. (Arrows indicate points for which 95 percent confidence limits do not enclose zero.)

It is seen from (13) that $C_k \to 0$ for large k. It can be shown[4] that the infinite series in (16) has a finite sum so that the covariance of $\hat{C}_k$ is proportional to $1/N$. Thus, the estimates $\hat{C}_k$ are asymptotically unbiased and consistent. Moreover, since all the eigenvalues of $\Phi(I - KH)$ lie inside the unit circle, ν_i belongs to the class of linear processes [14] for which Parzen [15] has shown that $\hat{C}_k$ are asymptotically normal.

For the white noise case, (16) is greatly simplified by putting $C_k = 0$, for all $k \neq 0$:

$$\text{cov}\,([\hat{C}_k]_{ij},[\hat{C}_l]_{pq}) = 0, \quad k \neq l$$

$$= (1/N)[C_0]_{ip}[C_0]_{jq}, \quad k = l > 0$$

$$= (1/N)[C_0]_{ip}[C_0]_{jq} + [C_0]_{iq}[C_0]_{jp}, \quad k = l = 0. \quad (17)$$

Estimates of the normalized autocorrelation coefficients ρ_k are obtained by dividing the elements of $\hat{C}_k$ by the appropriate elements of $\hat{C}_0$, e.g.,

$$[\hat{\rho}_k]_{ij} = \frac{[\hat{C}_k]_{ij}}{\{[\hat{C}_0]_{ii}[\hat{C}_0]_{jj}\}^{1/2}}. \quad (18)$$

Of particular interest here are the diagonal elements of $\hat{\rho}_k$ for the case of white noise. Using (17), we can show that

$$\text{var}\,[\hat{\rho}_k]_{ii} = 1/N + 0(1/N^2). \quad (19)$$

Further, $[\hat{\rho}_k]_{ii}$ like $[\hat{C}_k]_{ii}$ are asymptotically normal [15]. Therefore, the 95 percent confidence limits for $[\hat{\rho}_k]_{ii}$, $k > 0$ are $\pm(1.96/N^{1/2})$, or equivalently the 95 percent confidence limits for $[\hat{C}_k]_{ii}$ are $\pm(1.96/N^{1/2})[\hat{C}_0]_{ii}$.

Test

Look at a set of values for $[\hat{\rho}_k]_{ii}$, $k > 0$ and check the number of times they lie outside the band $\pm(1.96/N^{1/2})$. If this number is less than 5 percent of the total, the sequence ν_i is white. (Examples of a nonwhite and a white

[4] The proof is essentially similar to the one for proving the stability of a Kalman filter [11].

sequence are shown in Fig. 1. See the example in Section IX.)

This test is based on the assumption of large N. If N is small, other tests proposed by Anderson [17] and Hannan [16], etc., may be used. Jenkins and Watts [13] also give a frequency domain test which is useful if there are slow periodic components in the time series.

VI. Estimation of Q and R

If the test of Section V reveals that the filter is suboptimal, the next step will be to obtain better estimates of Q and R. This can be done using C_k computed earlier. The method proceeds in three steps.

1) Obtain an estimate of MH^T using (13). Rewriting (13) explicitly,

$$\begin{aligned} C_1 &= H\Phi MH^T - H\Phi KC_0 \\ C_2 &= H\Phi^2 MH^T - H\Phi KC_1 - H\Phi^2 KC_0 \\ &\vdots \\ C_n &= H\Phi^n MH^T - H\Phi KC_{n-1} - \cdots - H\Phi^n KC_0. \end{aligned}$$

Therefore

$$MH^T = B^* \begin{bmatrix} C_1 + H\Phi KC_0 \\ C_2 + H\Phi KC_1 + H\Phi^2 KC_0 \\ \vdots \\ C_n + H\Phi KC_{n-1} + \cdots + H\Phi^n KC_0 \end{bmatrix} \tag{20}$$

where B^* is the pseudo-inverse of matrix B [1] defined as

$$B \equiv \begin{bmatrix} H \\ H\Phi \\ \vdots \\ H\Phi^{n-1} \end{bmatrix} \cdot \Phi.$$

Notice that B is the product of the observability matrix and the nonsingular transition matrix Φ. Therefore

$$\text{rank } (B) = n$$

and

$$B^* = (B^TB)^{-1}B^T.$$

Denoting[5] by $\hat{M}\hat{H}^T$ the estimate of MH^T and using (20), we can write

$$\hat{M}\hat{H}^T = B^* \begin{bmatrix} \hat{C}_1 + H\Phi K\hat{C}_0 \\ \hat{C}_2 + H\Phi K\hat{C}_1 + H\Phi^2 K\hat{C}_0 \\ \vdots \\ \hat{C}_n + H\Phi K\hat{C}_{n-1} + \cdots + H\Phi^n K\hat{C}_0 \end{bmatrix}. \tag{21}$$

An alternate form for MH^T can be obtained directly from (13):

$$\hat{M}\hat{H}^T = K\hat{C}_0 + A^* \begin{bmatrix} \hat{C}_1 \\ \vdots \\ \hat{C}_n \end{bmatrix} \tag{22}$$

[5] The symbol $\hat{A}\hat{B}$ always implies $\widehat{AB}$ as a single symbol.

where

$$A = \begin{bmatrix} H\Phi \\ H\Phi(I - KH)\Phi \\ H[\Phi(I - KH)]^{n-1}\Phi \end{bmatrix}.$$

In numerical computation, it has been found preferable to use (22) since matrix A is better conditioned than matrix B. (This is an experimental observation.)

2) Obtain an estimate of R using (12):

$$\hat{R} = \hat{C}_0 - H(\hat{M}\hat{H}^T). \tag{23}$$

3) Obtain an estimate of Q using (11).

This step gets complicated due to the fact that only the estimate of MH^T instead of M is available. Consequently, only $n \times r$ linear relationships between the unknown elements of Q are available. If the number of unknowns in Q is $n \times r$ or less, a solution can be obtainedl But if the number of unknowns in Q is greater than $n \times r$. a unique solution cannot be obtained. However, it wil, be shown in the next section that a unique solution for the optimal gain K_{op} can still be obtained.

Restricting ourselves to the case in which the number of unknowns in Q is $n \times r$ or less, we can solve for the unknown elements of Q by rewriting (11) as follows:

$$M = \Phi M\Phi^T + \Omega + \Gamma Q\Gamma^T \tag{24}$$

where

$$\Omega = \Phi[-KHM - MH^TK^T + KC_0K^T]\Phi^T.$$

Substituting back for M on the right-hand side of (24),

$$M = \Phi^2 M(\Phi^2)^T + \Phi\Omega\Phi^T + \Omega + \Phi\Gamma Q\Gamma^T\Phi^T + \Gamma Q\Gamma^T. \tag{25}$$

Repeating the same procedure n times and separating the terms involving Q on the left-hand side of the equation, we obtain

$$\sum_{j=0}^{k-1} \Phi^j\Gamma Q\Gamma^T(\Phi^j)^T = M - \Phi^k M(\Phi^k)^T - \sum_{j=0}^{k-1} \Phi^j\Omega(\Phi^j)^T, \quad \text{for } k = 1,\cdots,n. \tag{26}$$

Premultiplying both sides of (26) by H and postmultiplying by $(\Phi^{-k})^TH^T$, we obtain

$$\sum_{j=0}^{k-1} H\Phi^j\Gamma Q\Gamma^T(\Phi^{j-k})^TH^T = HM(\Phi^{-k})^TH^T - H\Phi^k MH^T - \sum_{j=0}^{k-1} H\Phi^j\Omega(\Phi^{j-k})^TH^T, \quad k = 1,\cdots,n. \tag{27}$$

The right-hand side of (27) is completely determined from MH^T and C_0. Substituting their estimated values,

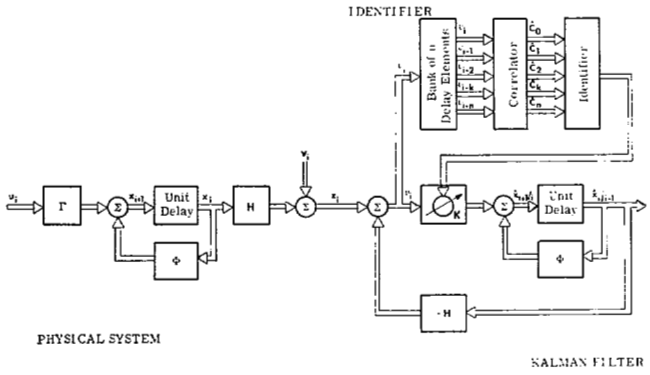

Fig. 2. Identification scheme.

we obtain

$$\sum_{j=0}^{k-1} H\Phi^j\Gamma\hat{Q}\Gamma^T(\Phi^{j-k})^TH^T = \hat{H}\hat{M}(\Phi^{-k})^TH^T - H\Phi^k\hat{M}\hat{H}^T - \sum_{j=0}^{k-1} H\Phi^j\hat{\Omega}(\Phi^{j-k})^TH^T, \quad k = 1,\cdots,n \tag{28}$$

where

$$\hat{\Omega} = \Phi[-K\hat{H}\hat{M} - \hat{M}\hat{H}^TK^T + K\hat{C}_0K^T]\Phi^T. \tag{29}$$

The set of equations (28) is not linearly independent. In any particular case, one has to choose a linearly independent subset of these equations. The procedure will be illustrated by an example in Section IX.

The preceding identification scheme is shown schematically in Fig. 2.

VII. Direct Estimation of the Optimal Gain

If the number of unknowns in Q is more than $n \times r$, or the structure of Q is unknown, the method of the previous section for estimating Q does not work. However, it is still possible to estimate the optimal gain K_{op} by an iterative procedure.

Following the notation of Section II, let K_0 denote the initial gain of the Kalman filter. Let M_1 be the error covariance matrix corresponding to K_0. Then M_1 satisfies the following equation [cf., (7)]:

$$M_1 = \Phi[M_1 - K_0HM_1 - M_1H^TK_0^T + K_0(HM_1H^T + R)K_0^T]\Phi^T + \Gamma Q\Gamma^T. \tag{30}$$

Define

$$K_1 \equiv M_1H^T(HM_1H^T + R)^{-1}.$$

Let the error covariance matrix corresponding to gain K_1 be called M_2. Then

$$M_2 = \Phi[M_2 - K_1HM_2 - M_2H^TK_1^T + K_1(HM_2H^T + R)K_1^T]\Phi^T + \Gamma Q\Gamma^T. \tag{31}$$

Subtracting (30) from (31) and simplifying

$$(M_2 - M_1) = \Phi(I - K_1H)(M_2 - M_1)(I - K_1H)^T\Phi^T - \Phi(K_1 - K_0)(HM_1H^T + R)(K_1 - K_0)^T\Phi^T. \tag{32}$$

The solution to (32) can be written as an infinite sum. Then, using observability and controllability conditions, it can be shown that[6]

$$M_2 - M_1 < 0 \quad \text{or} \quad M_2 < M_1.$$

Similarly, define $K_2 \equiv M_2H^T(HM_2H^T + R)^{-1}$ and M_3 as the corresponding error covariance matrix. Then, by a similar argument,

$$M_3 < M_2 < M_1.$$

The above sequence of monotonically decreasing matrices must converge since it is bounded from below ($M > 0$). Hence, the sequence $K_0,K_1,K_2,\cdots$ must converge to K_{op}.

Based on the preceding property of K, we now construct the following scheme for estimating K_{op}.

1) Obtain an estimate of K_1, denoted as $\hat{K}_1$ from (22):

$$\hat{K}_1 = K_0 + A^* \begin{bmatrix} \hat{C}_1 \\ \hat{C}_2 \\ \hat{C}_n \end{bmatrix} \hat{C}_0^{-1}. \tag{33}$$

Also, obtain estimates of M_1H^T and R from (22) and (23).

2) Define $\delta M_1 = M_2 - M_1$. Obtain $\hat{\delta M}_1$, an estimate of δM_1, using (32):

$$\hat{\delta M}_1 = \Phi(I - \hat{K}_1H)\hat{\delta M}_1(I - \hat{K}_1H)^T\Phi^T - \Phi(\hat{K}_1 - K_0)\hat{C}_0(\hat{K}_1 - K_0)^T\Phi^T. \tag{34}$$

$\hat{\delta M}_1$ can be calculated recursively in the same manner as M_0 is calculated for a Kalman filter. For convergence, it is sufficient that $\Phi(I - \hat{K}_1H)$ be stable, i.e., all eigenvalues be inside the unit circle.

3) Obtain $\hat{M}_2\hat{H}^T$ and $\hat{K}_2$ as follows:

$$\hat{M}_2\hat{H}^T = \hat{M}_1\hat{H}^T + \hat{\delta M}_1H^T \tag{35}$$

$$\hat{K}_2 = \hat{M}_2\hat{H}^T(H\hat{M}_2\hat{H}^T + \hat{R})^{-1}. \tag{36}$$

4) Repeat steps 2) and 3) until $\|\hat{\delta M}_i\|$ or $\|\hat{K}_i - \hat{K}_{i-1}\|$ become small compared to $\|\hat{M}_i\|$ or $\|K_i\|$ where $\|\cdot\|$ denotes a suitable matrix norm. An alternative way to get K_2 would be to filter data z again using K_1 and then use (33).

This procedure for obtaining K_{op} reveals an interesting relationship between K_{op} and Q. It is seen that the equa-

[6] The proof is similar to the one by Kalman [1] for showing the positive definiteness of M in (11).

tion for $(M_2 - M_1)$ does not involve Q. We need Q only to calculate M_1H^T. This leads us to the following corollary.

Corollary: It is sufficient to know $n \times r$ linear functions of Q in order to obtain the optimal gain of a Kalman filter.

Proof: Consider (30) which can be written as

$$M_1 = \Phi(I - K_0H)M_1(I - K_0H)^T\Phi^T + \Phi K_0RK_0^T\Phi^T + \Gamma Q\Gamma^T. \quad (37)$$

Writing the solution as an infinite series,

$$M_1H^T = \sum_{j=0}^{\infty} [\Phi(I - K_0H)]^j(\Phi K_0RK_0^T\Phi^T + \Gamma Q\Gamma^T) \cdot[(I - K_0H)^T\Phi^T]^jH^T. \quad (38)$$

M_1H^T depends on $n \times r$ linear functions of Q; viz.,

$$\sum_{j=0}^{\infty} [\Phi(I - K_0H)]^j\Gamma Q\Gamma^T[(I - K_0H)^T\Phi^T]^jH^T.$$

If these linear functions are given, we do not need to know Q itself to obtain M_1H^T. Furthermore, since the equation for $(M_2 - M_1)$ does not involve Q explicitly, the optimal gain K_{op} can be obtained by knowing the preceding $n \times r$ linear functions of Q only.

Notice that a complete knowledge of Q is required to obtain the covariance matrix M of a Kalman filter. If one is interested only in K_{op}, the preceding corollary shows that a complete knowledge of Q is not essential. Since our iterative scheme tries to identify K_{op} by whitening the residuals ν_i, it fails to identify the complete Q matrix if the unknowns in Q are more than $n \times r$.

VIII. Statistical Properties of the Estimates

It was shown in Section V that the estimates $\hat{C}_k$ are asymptotically normal, unbiased, and consistent. Since $\hat{M}\hat{H}^T$, $\hat{R}$, and $\hat{Q}$ are linearly related to $\hat{C}_k$, it is easy to show that they are also asymptotically normal, unbiased, and consistent.

The general expressions for the mean and covariance of the estimates are rather involved. We, therefore, specialize to the case of a scalar measurement.

Using (22), (23), and (28),

$$E[\hat{M}\hat{H}^T] = KC_0 + A^* \begin{bmatrix} C_1 \\ \vdots \\ C_n \end{bmatrix} - \frac{A^*}{N} \begin{bmatrix} C_1 \\ 2C_2 \\ \vdots \\ nC_n \end{bmatrix}$$

or

$$E[\hat{M}\hat{H}^T] = MH^T - (1/N)A^* \begin{bmatrix} C_1 \\ 2C_2 \\ \vdots \\ nC_n \end{bmatrix}. \quad (39)$$

For $N \gg n$, the bias in $\hat{M}\hat{H}^T$ is negligible. The covariance of $\hat{M}\hat{H}^T$ for large N is

$$\text{cov}\,(\hat{M}\hat{H}^T) \approx K \text{ var}\,(\hat{C}_0)K^T + A^* \text{ cov}\left(\begin{bmatrix} \hat{C}_1 \\ \vdots \\ \hat{C}_n \end{bmatrix}\right) A^{*T} + K \text{ cov}\left(\hat{C}_0, \begin{bmatrix} \hat{C}_1 \\ \vdots \\ \hat{C}_n \end{bmatrix}\right) A^{*T} + A^* \text{ cov}\left(\begin{bmatrix} \hat{C}_1 \\ \vdots \\ \hat{C}_n \end{bmatrix}, C_0\right) K^T. \quad (40)$$

Expressions for

$$\text{cov}\left(\begin{bmatrix} \hat{C}_1 \\ \vdots \\ \hat{C}_n \end{bmatrix}\right), \quad \text{etc.}$$

can be obtained from (16). It can be seen that cov $(\hat{M}\hat{H}^T)$ decreases as $1/N$ for large sample sizes. Similarly,

$$E(\hat{R}) = R + (1/N)HA^* \begin{bmatrix} C_1 \\ 2C_2 \\ \vdots \\ nC_n \end{bmatrix} \quad (41)$$

$$\text{var}\,(\hat{R}) = \text{var}\,(\hat{C}_0) + H \text{ cov}\,(\hat{M}\hat{H}^T)H^T - \text{cov}\,(\hat{C}_0,\hat{M}\hat{H}^T)H^T - H \text{ cov}\,(\hat{M}\hat{H}^T,\hat{C}_0). \quad (42)$$

The expressions for $E[\hat{Q}]$ and var $([\hat{Q}]_{ij})$ can be obtained similarly.

The usefulness of the preceding expressions is limited by the fact that they depend on the actual values of Q and R which are unknown. If the values of Q and R are known to lie within a certain range, one might use these expressions to plot curves of var $(\hat{R})$ and var $([\hat{Q}]_{ij})$ versus N for different values of Q and R. The dependence on Q and R may be removed by considering the covariance of $\hat{K}$:

$$\text{cov}\,(\hat{K}) = \text{cov}\,(\hat{M}\hat{H}^T\hat{C}_0^{-1}) \approx A^* \text{ cov}\left(\begin{bmatrix} \hat{\rho}_1 \\ \vdots \\ \hat{\rho}_n \end{bmatrix}\right) A^{*T}. \quad (43)$$

It can be shown [14], [15] that $\hat{\rho}_k$ are asymptotically normal with mean ρ_k and covariance

$$\text{cov}\,(\hat{\rho}_k,\hat{\rho}_l) \approx (1/N) \sum_{j=-\infty}^{\infty} \rho_j\rho_{j+l-k}.$$

A satisfactory estimate of cov $(\hat{\rho}_k,\hat{\rho}_l)$ is provided by

$$(1/2N) \sum_{j=-(N-1)}^{N-1} \hat{\rho}_j\hat{\rho}_{j+l-k}$$

which can be used in (43) to calculate cov $(\hat{K})$. For the special case of an optimal filter, (43) reduces to

$$\text{cov}\,(\hat{K}) \approx (1/N)A^*A^{*T} = (1/N)(A^TA)^{-1}. \quad (44)$$

Equation (44) gives us a simple expression for the minimum variance in estimating K. It can be used in deciding upon the minimum sample size N.

We now consider the asymptotic convergence (N large) of the iterative scheme of Section VII. Equation (34) shows that $E[\hat{\delta}\hat{M}_1]$ depends on the second- and higher order moments of $\hat{K}_1$ which for a normal process are finite and tend to zero asymptotically. Therefore, $\lim_{n\to\infty} E[\hat{\delta}\hat{M}_1] = \delta M_1$.

Similarly, the covariance of $\hat{\delta}\hat{M}_1$ asymptotically tends to zero. Thus, $\hat{\delta}\hat{M}_1$ tends to δM_1 with probability one. Extending the same argument, $\hat{K}_2 \to K_2$, $\hat{K}_3 \to K_3$, $\cdots$, $\hat{K}_{op} \to K_{op}$ with probability one.

IX. A Numerical Example from Inertial Navigation

The results of Sections V and VI are applied to a damped Schuler loop forced by an exponentially correlated stationary random input. Two measurements are made on the system, both of which are corrupted by exponentially correlated as well as white noise type errors. The state of the system is augmented to include all the correlated random inputs so that the augmented state vector x is 5×1, the random input vector u is 3×1, and the measurement noise vector v is 2×1. The system is discretized using a time step of 0.1 and the resultant system matrices are

$$\Phi = \begin{bmatrix} 0.75 & -1.74 & -0.3 & 0 & -0.15 \\ 0.09 & 0.91 & -0.0015 & 0 & -0.008 \\ 0 & 0 & 0.95 & 0 & 0 \\ 0 & 0 & 0 & 0.55 & 0 \\ 0 & 0 & 0 & 0 & 0.905 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 24.64 & 0 & 0 \\ 0 & 0.835 & 0 \\ 0 & 0 & 1.83 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

$$Q = \begin{bmatrix} q_1 & 0 & 0 \\ 0 & q_2 & 0 \\ 0 & 0 & q_3 \end{bmatrix}, \qquad R = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix}.$$

The actual values of q_1, q_2, q_3, r_1, and r_2 are unity, but they are assumed unknown. It is required to identify these values using measurements $\{z_i, i = 1,N\}$.

The starting values of Q and R are taken as

$$Q_0 = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.75 \end{bmatrix}, \quad R_0 = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.6 \end{bmatrix}.$$

Using these values, the innovation sequence $\nu_i = (z_i - H\hat{x}_{i/i-1})$ is generated from (3) to (6). The estimates $\hat{C}_0, \hat{C}_1, \cdots, \hat{C}_k$ of the autocorrelation are calculated using (14). For a typical sample of 950 points, Fig. 1(a) shows a plot of the first diagonal element of $\hat{\rho}_k$ for $k = 0,40$. The 95 percent confidence limits are ± 0.0636 and four points lie outside this band (i.e., 10 percent of the total). Therefore, we reject the hypothesis that ν_i is white. The same conclusion is reached by looking at the second diagonal element of $\hat{\rho}_k$.

We now proceed to the identification of Q and R. Since the number of unknowns in Q is less than $n \times r = 10$, we can identify Q completely. The set of equations (28) gives us a large number of linear equations for $\hat{q}_1$, $\hat{q}_2$, and $\hat{q}_3$. However, the most important of these occur along the diagonal for $k = 1$ and $k = 5$.

For $k = 1$ the left-hand side of (28) is

$$\begin{bmatrix} 4.37\hat{q}_3, & -0.0326\hat{q}_3 \\ 0, & 1.27\hat{q}_2 \end{bmatrix}.$$

For $k = 5$ the left-hand side of (28) is

$$\begin{bmatrix} -8.38\hat{q}_1 + 22.3\hat{q}_3, & 1.22\hat{q}_1 - 1.47\hat{q}_2 \\ -1.25\hat{q}_1 - 0.87\hat{q}_3, & 0.141\hat{q}_1 + 20\hat{q}_2 + 0.023\hat{q}_3 \end{bmatrix}.$$

The diagonal elements of the first equation are used to calculate $\hat{q}_3$ and $\hat{q}_2$. The first diagonal element of the second equation is then used to calculate $\hat{q}_1$.

It is possible to use a few other equations and to make a least-squares fit for $\hat{q}_1$, $\hat{q}_2$, and $\hat{q}_3$. This, however, does not alter the results significantly in the present example.

The results obtained by using the identification scheme repeatedly on the same batch of data are shown in Table I. It is seen that most of the identification is done during the first iteration. Further iterations do not increase the likelihood function[7] much, even though the changes in Q and R are significant. A check case using true values of Q and R is also shown in Table I. It is seen that the value of the likelihood function in the check case is very close to that in the first iteration. This indicates that the estimates obtained are quite close to the maximum likelihood estimates. It was further noticed that even if different starting values are used for Q and R, the identification scheme converges to the same values.

[7] The likelihood function $L(Q,R)$ has been given by Schweppe [12]:

$$L(Q,R) = -(1/N) \sum_{i=1}^{N} \nu_i^T (HMH^T + R)^{-1} \nu_i - \ln |HMH^T + R|.$$

TABLE I
ESTIMATES OF Q AND R BASED ON A SET OF 950 POINTS

Number of Iterations	$\hat{q}_1$	$\hat{q}_2$	$\hat{q}_3$	$\hat{r}_1$	$\hat{r}_2$	Likelihood Function $L(Q,\hat{R})$	Percentage of Points Lying Outside the 95 Percent Confidence Limits: First Measurement (percent)	Percentage of Points Lying Outside the 95 Percent Confidence Limits: Second Measurement (percent)	Estimate of Actual Mean-Square Error*	Calculated Mean-Square Error†
0	0.25	0.5	0.75	0.4	0.6	−5.17	10	10	2915	902
1	0.731	1.31	0.867	1.444	0.776	−4.676	2.5	5	2755	2390
2	0.87	1.39	0.797	1.537	0.767	−4.673	2.5	5	2720	2725
3	0.91	1.40	0.776	1.565	0.765	−4.672	2.5	5	2714	2814
4	0.92	1.41	0.77	1.573	0.7646	−4.671	2.5	5	2712	2840
Check case	1.0	1.0	1.0	1.0	1.0	−4.669	2.5	5	2720	2900

* Estimate of mean-square error is

$$(1/N)\sum_{i=1}^{N}(x_i - \hat{x}_{i\,i-1})^T(x_i - \hat{x}_{i/i-1})$$

where x_i is obtained by actual simulation.
† Calculated mean-square error is tr (M_c) where M_c is obtained from the variance equation using $\hat{Q}$ and $\hat{R}$ [cf. (4)].

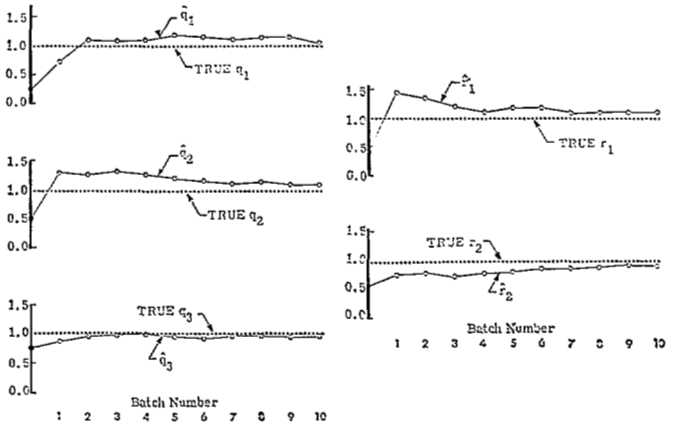

Fig. 3. On-line identification of Q and R.

We now check the optimality of the filter after identification. Fig. 1(b) shows a plot of the first diagonal element of $\hat{p}_k$, for $k = 0,40$. It is seen that only one point lies outside the band of 95 percent confidence limits (2.5 percent of the total). This supports the hypothesis that ν_i is white.

The asymptotic convergence of Q and R towards their actual values is shown in Fig. 3. The estimates of Q and R are updated after every batch of N points ($N = 950$). In the absence of any knowledge about the variances of the estimates, a simple averaging of all the previous values is performed. This is equivalent to the following stochastic approximation scheme [5]:

$$\hat{Q}_{k+1} = \hat{Q}_k + [1/(k+1)](\hat{Q}_{k+1,k} - \hat{Q}_k) \tag{45}$$

where k denotes the batch number,

$\hat{Q}_k$ the estimate of Q after k batches
$\hat{Q}_{k+1,k}$ the estimate of Q based on the $(k+1)$th batch
$\hat{Q}_{k+1}$ the estimate of Q after $(k+1)$ batches.

Similarly,

$$\hat{R}_{k+1} = \hat{R}_k + [1/(k+1)](\hat{R}_{k+1,k} - \hat{R}_k). \tag{46}$$

X. CONTINUOUS SYSTEM

The results of the previous sections can be easily extended to continuous systems. We simply state the results below.[8]

System

$$\dot{x} = Fx + Gu \tag{47}$$

$$z = Hx + v. \tag{48}$$

Filter

$$\dot{\hat{x}} = F\hat{x} + K_0(z - H\hat{x}) \tag{49}$$

where

$$K_0 = P_0H^TR_0^{-1} \tag{50}$$

and

$$FP_0 + P_0F^T + GQ_0G^T - P_0H^TR_0^{-1}HP_0 = 0. \tag{51}$$

The error covariance P_1 is given as

$$(F - K_0H)P_1 + P_1(F - K_0H)^T + GQG^T + K_0RK_0^T = 0. \tag{52}$$

Innovation Process

$$\nu = z - H\hat{x} = He + v \tag{53}$$

where $e = (x - \hat{x})$. For an optimal filter, ν is white with the same covariance as v [10]. For a suboptimal filter,

$$\dot{e} = (F - K_0H)e + Gu - K_0v \tag{54}$$

[8] These results have not been applied to a practical problem so far.

and the autocorrelation function $C(\tau)$ of ν is given as

$$\begin{aligned} C(\tau) &= E\{\nu(t)\nu^T(t-\tau)\} \\ &= HE\{e(t)e^T(t-\tau)\}H^T \\ &\quad + HE\{e(t)v^T(t-\tau)\} + R\delta(\tau), \quad \tau > 0 \\ &= He^{F'\tau}[P_1H^T - K_0R] + R\delta(\tau), \qquad F' = F - K_0H. \end{aligned} \tag{55}$$

Let $S(w)$ denote the Fourier transform of $C(\tau)$

$$S(w) = H(iw - F')^{-1}(P_1H^T - K_0R) + (HP_1 - RK_0^T) \cdot(-iw - F'^T)^{-1}H^T + R. \tag{56}$$

Test of Optimality and the Estimation of Q and R

We may use either the estimates of $C(\tau)$ or of $S(w)$ to test the optimality of the Kalman filter and to identify Q and R. These estimates are obtained by using methods given in [13].

P_1H^T and R may be obtained from the set of equations (55) or (56) by using methods very similar to the discrete case. If the number of unknowns in Q is $n \times r$ or less, Q can be obtained using (52). We obtain expressions for

$$\sum_{j=0}^{k-1} (-1)^j HF^jGQG^TF^{k-j}H^T, \quad \text{for } k = 0,1,\cdots$$

[the set of equations analogous to (28)].

If the number of unknowns in Q is more than $n \times r$, K_{op} is obtained directly without identifying Q. The procedure is as follows. Define

$$K_1 = P_1H^TR^{-1}. \tag{57}$$

Let P_2 be the error covariance corresponding to K_1. Then it can be shown that

$$(F - K_1H)(P_2 - P_1) + (P_2 - P_1)(F - K_1H)^T - (K_1 - K_0)R(K_1 - K_0)^T = 0. \tag{58}$$

Therefore

$$P_2 < P_1.$$

Similarly, define $K_2 = P_2H^TR^{-1}$ and let P_3 be the error covariance for K_2. Then

$$P_3 < P_2 < P_1.$$

In this way, P is decreased at each step and the sequence $K_0,K_1,K_2\cdots$ converges to K_{op}.

Equation (58) is now used to obtain $\hat{K}_{op}$, an estimate of K_{op}. After obtaining $\hat{K}_1 = \hat{P}_1\hat{H}^T\hat{R}^{-1}$, we substitute it in (58) to get an estimate of $\delta P_1 = P_2 - P_1$:

$$(F - \hat{K}_1H)\hat{\delta}\hat{P}_1 + \hat{\delta}\hat{P}_1(F - \hat{K}_1H)^T - (\hat{K}_1 - K_0)\hat{R}(\hat{K}_1 - K_0)^T = 0. \tag{59}$$

Then

$$\hat{P}_2\hat{H}^T = \hat{P}_1\hat{H}^T + \hat{\delta}\hat{P}_1H^T \tag{60}$$

and

$$\hat{K}_2 = \hat{P}_2\hat{H}^T\hat{R}^{-1} \tag{61}$$

and so on until the relative changes in $\hat{K}$ become small.

We omit the proof of the asymptotic convergence of these estimates since they are essentially similar to the discrete case. All the estimates obtained are asymptotically unbiased and consistent.

XI. Summary and Conclusions

The problem of optimal filtering for a linear time-invariant system with unknown Q (process noise covariance matrix) and R (measurement noise covariance matrix) is considered. Based on the innovation property of an optimal filter, a statistical test is given to check whether a particular filter is working optimally or not. In case the filter is suboptimal, an identification scheme is given to obtain asymptotically unbiased and consistent estimates of Q and R. For the case in which the form of Q is unknown or the number of unknowns in Q is more than $n \times r$ (n is the dimension of the state vector and r is the dimension of the measurement vector), the preceding scheme fails and an alternate scheme is given to obtain an estimate of the optimal gain directly without identifying Q. A numerical example is given to illustrate the results and to show the usefulness of the approach. The results are first derived for a discrete system. They are then extended to continuous systems.

Nomenclature

Matrices

Φ,Γ,H,F,G,B,A	System matrices
Q,R,M,P,M_1,P_1	covariance matrices
Q_0,R_0,M_0,P_0	initial values of Q,R,M,P
$K,K_0,K_1,K_{op},\cdots$	Kalman filter gains
$\hat{Q},\hat{R},\hat{M},\hat{K},\cdots$	estimated values of Q,R,M,K (caret over any quantity denotes an estimate)
$C_k,\hat{C}_k,C(\tau)$	autocorrelation function
$\rho_k,\hat{\rho}_k$	normalized autocorrelation function
$S(w)$	power spectral density
$\delta M,\delta P,\hat{\delta M},\hat{\delta P}$	increment matrices.[5]

Vectors

$x_i,\hat{x}_{i/i-1}$	Actual and estimated states
u_i,v_i	white noise sequences
z_i	measurements
ν_i	innovations
e_i	error in state estimation.

Scalars

n,q,r	Dimension variables
N	sample size
$[C_k]_{ij}$	element in the ith row and the jth column of the matrix C_k
$\delta_{ij},\delta(\tau)$	Kronecker delta and the delta function
$L(Q,R)$	likelihood function.

Operations

$E\{\cdot\}$	Expected value operator
cov $(\cdot,\cdot)$	covariance operator
var $(\cdot)$	variance operator
$(\ )^T$	transpose of a matrix
$(\ )^\#$	pseudo-inverse of a matrix
$(\ )^{-1}$	inverse of a matrix
$\|\cdot\|$	norm of a matrix.

Acknowledgment

The author wishes to thank all his colleagues at The Analytic Sciences Corporation for their help during the course of this research, and C. L. Bradley in particular for many stimulating discussions and ideas.

References

[1] R. E. Kalman, "New methods and results in linear prediction and filtering theory," *Proc. Symp. on Engineering Applications of Random Function Theory and Probability*. New York: Wiley, 1961.
[2] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Trans. ASME, J. Basic Engrg.*, ser. D, vol. 83, pp. 95–108, March 1961.
[3] H. Heffes, "The effects of erroneous models on the Kalman filter response," *IEEE Trans. Automatic Control* (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. Automatic Control*, vol. AC-12, pp. 268–275, June 1967.
[5] R. C. K. Lee, "Optimal estimation, identification and control," Massachusetts Institute of Technology, Cambridge, Res. Mono. 28, 1964.
[6] D. T. McGill, "Optimal adaptive estimation of sampled stochastic process," Stanford Electronics Labs., Stanford, Calif., Tech. Rept. SEL-63-143(TR 6302-3), December 1963.
[7] J. S. Shellenberger, "A multivariance learning technique for improved dynamic system performance," *1967 Proc. NEC*, vol. 23, pp. 146–151.
[8] G. L. Smith, "Sequential estimation of observation error variances in a trajectory estimation problem," *AIAA J.*, vol. 5, pp. 1964–1970, November 1967.
[9] R. L. Kashyap, "Maximum likelihood identification of stochastic linear systems," Purdue University, Lafayette, Ind., Tech. Rept. TR-EE 68-28, August 1968.
[10] T. Kailath, "An innovations approach to least-squares estimation, pt. I: linear filtering in additive white noise," *IEEE Trans. Automatic Control*, vol. AC-13, pp. 646–655, December 1968.
[11] J. J. Deyst, Jr., and C. F. Price, "Conditions for asymptotic stability of the discrete minimum variance linear estimator," *IEEE Trans. Automatic Control* (Short Papers), vol. AC-13, pp. 702–705, December 1968.
[12] F. C. Schweppe, "Evaluation of likelihood functions for Gaussian signals," *IEEE Trans. Information Theory*, vol. IT-11, pp. 61–70, January 1965.
[13] G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Applications*. San Francisco: Holden Day Publ., 1968.
[14] M. S. Bartlett, *An Introduction to Stochastic Processes*. London: Cambridge University Press, 1962.
[15] E. Parzen, "An approach to time series analysis," *Ann. Math. Statist.*, vol. 32, no. 4, December 1961.
[16] E. J. Hannan, *Time Series Analysis*. New York: Wiley, 1960.
[17] R. L. Anderson, "Distribution of the serial correlation coefficient," *Ann. Math. Statist.*, vol. 13, 1942.
[18] G. S. Watson and J. Durbin, "Exact tests of serial correlation using noncircular statistics," *Ann. Math. Statist.*, vol. 22, pp. 446–451, 1951.
[19] C. W. J. Granger, "A quick test for serial correlation suitable for use with non-stationary time series," *Am. Statist. Assoc. J.*, pp. 728–736, September 1963.

Raman K. Mehra (S'67–M'68) was born in Lahore, Pakistan, on February 10, 1943. He received the B.S. degree in electrical engineering from Punjab Engineering College, Chandigarh, India, in 1964, and the M.S. and Ph.D. degrees in engineering from Harvard University, Cambridge, Mass., in 1965 and 1968, respectively.

He worked at Bell Telephone Laboratories, Inc., Andover, Mass., during the summer of 1965. From 1966 to 1967, he was a Research Assistant at Harvard University. From 1967 to 1969, he was employed by the Analytic Sciences Corp., Reading, Mass., where he applied modern estimation and control theory to problems in inertial navigation. He is presently with Systems Control, Inc., Palo Alto, Calif. His interests lie in the areas of trajectory optimization; linear, nonlinear, and adaptive filtering; smoothing and system identification; stochastic control; and pattern recognition.