ADAPTIVE STATE ESTIMATION FOR SYSTEMS WITH WHITE AND COLOURED NOISE

by

ALVAN F.W. TOM, B.ENG.

A Thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Master of Engineering

McMaster University

July 1975

ADAPTIVE STATE ESTIMATION FOR SYSTEMS WITH WHITE AND COLOURED NOISE

MASTER OF ENGINEERING (1975)
(Electrical Engineering)

McMASTER UNIVERSITY
Hamilton, Ontario.

TITLE: Adaptive State Estimation for Systems with White and Coloured Noise

AUTHOR: Alvan F.W. Tom, B.ENG. (McGill University)

SUPERVISOR: N.K. Sinha, Professor of Electrical Engineering
B.Sc. (Eng.) (Banaras)
Ph.D. (University of Manchester)

NUMBER OF PAGES: viii, 70.

SCOPE AND CONTENTS:

The problem of adaptive state estimation which involves the identification of the Kalman gain matrix without *a priori* information on the noise statistics is presented. A scheme incorporating an identification algorithm and a tracking algorithm is proposed. This scheme provides a powerful approach for adaptive state estimation.

An ARMA model for system description is derived for preliminary analysis of the noise transition matrix when the observation noise is sequentially correlated.

The innovations process for systems with coloured observation noise is shown to be white for optimum filtering.

Simulations are performed on an inertial navigation system for both white and coloured observation noise. Numerical results indicate the superiority of the filter with tracking over one without. Performance of the filter

for coloured observation noise confirms the theoretical derivation of the ARMA model.

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to Dr. N.K. Sinha for guidance, encouragement, and assistance throughout the course of this work and in the preparation of this thesis.

The author wishes to thank Dr. J.F. MacGregor for his advice on time series analysis.

Sincere thanks go to A. Sen for many helpful discussions and suggestions.

It is the author's pleasure to acknowledge the inspiring discussions with his colleagues. Without being exhaustive, the author would like to name S. Abdel-Azim, J. Hickin and S. Law.

Financial support provided by the Communications Research Centre through Grant OSU4-0078 and the Department of Electrical Engineering is gratefully acknowledged.

The competence and reliability of Carolyn Zoeftig, who typed this thesis, has been much appreciated.

TABLE OF CONTENTS

		Page
CHAPTER 1 -	INTRODUCTION	1
CHAPTER 2 -	OPTIMUM LINEAR FILTERING	5
	2.1 - Introduction	5
	2.2 - Kolmogorov-Wiener Filter	6
	2.3 - Kalman Filter	7
CHAPTER 3 -	ADAPTIVE KALMAN FILTERING	14
	3.1 - Introduction	14
	3.2 - The Carew-Bélanger Algorithm	15
	3.3 - Stochastic Approximation	19
	3.3.1 - Stochastic Approximation Algorithms	19
	3.3.2 - Adaptive Kalman Filtering by Stochastic Approximation	21
	3.4 - Combined Carew-Bélanger and Stochastic Approximation Algorithm	22
	3.5 - Simulation Results	25
	3.6 - Discussions	29
CHAPTER 4 -	ADAPTIVE STATE ESTIMATION IN COLOURED OBSERVATION NOISE	30
	4.1 - Introduction	30
	4.2 - Some Results of Bryson and Henrikson	31
	4.3 - Representation of Systems by ARMA Models	36
	4.4 - Proposed Scheme	39

TABLE OF CONTENTS (Continued)

		Page
	4.5 - Simulation Results	39
	4.6 - Discussions	43
CHAPTER 5 -	CONCLUSIONS	45
APPENDIX A -	STOCHASTIC APPROXIMATION - PROOF OF CONVERGENCE	49
APPENDIX B -	WILSON'S ALGORITHM	52
APPENDIX C -	PROGRAM LISTING AND USAGE	55
BIBLIOGRAPHY		67

LIST OF FIGURES

		Page
Figure 2.1	- Block diagram of system represented in state space form.	9
Figure 2.2	- Block diagram of Kalman filter.	12
Figure 3.1	- Variation of the trace of error covariance matrix for different methods (white noise).	28
Figure 4.1	- Variation of the trace of error covariance matrix for different methods (coloured noise).	42

LIST OF TABLES

		Page
Table 3.1	- Advantages and disadvantages of the Carew-Bélanger and stochastic approximation algorithms.	23
Table 3.2	- Comparison of filter performance (white noise).	27
Table 4.1	- Comparison of filter performance (coloured noise).	43

CHAPTER 1

INTRODUCTION

Physical systems are designed to perform certain defined functions.. To determine whether a system is performing properly, its *state* must be known. In navigation, the state consists of position and velocity of the craft in question; in an AC electric power system, the state may be taken as voltages and phase angles at network nodes. In order to determine the state, observations of the system must be taken. The observations are generally contaminated with noise caused by various independent sources in the observation process.

The problem of determining the state of a system from noisy observations is called *filtering* or *state estimation*. It is of central importance in engineering, since state estimates are required in the monitoring, and for the control of systems.

Studies were first made by Kolmogorov (19) and Wiener (34) on the problem of *optimum linear filtering*. Later Kalman (15), and Kalman and Bucy (16) reformulated the problem in the state space, thus deriving the *Kalman filter*, which has as the output the optimum estimates of the state of the system. The Kalman filter, amenable in computational aspects, is still difficult to implement in practice. It requires the *a priori* knowledge of the system noise statistics to compute the *Kalman gain* matrix, which in turn determines the behaviour of the filter.

The problem of identifying the Kalman gain matrix without *a priori* knowledge of the system noise statistics, otherwise known as *adaptive state estimation*, is the main concern of the treatise.

The thesis is divided into two main parts. Chapter 2 is mainly tutorial in nature, whereas the main results of the thesis appear in Chapters 3 and 4.

In Chapter 2, the work of Kolmogorov (19), Wiener (34), Kalman (15), and Kalman and Bucy (16) are summarized.

Chapter 3 gives a brief description of the Carew-Bélanger algorithm for the identification of the Kalman gain matrix in *white observation noise*. Also discussed are the Robbins and Monro type of stochastic approximation algorithms and the assumptions which are required for convergence in the *mean-square*. Since the algorithm of Carew and Bélanger employs the *correlation technique* to compute the Kalman gain, a large amount of data must be processed to obtain results which are accurate. A method is presented which reduces the data requirement. The proposed scheme uses a small quantity of observation data to give an initial estimate of the Kalman gain, after which a Robbins and Monro type of stochastic approximation algorithm is used to track and improve the estimate of the Kalman gain matrix. Simulation results for an inertial navigation system are given. The results indicate that the proposed *closed-loop* scheme does improve the *open-loop* method of Carew and Bélanger. Theoretical analysis given in Appendix A confirms that the stochastic approximation algorithm used will converge to the optimum Kalman gain matrix in the mean-square sense.

The problem of adaptive state estimation in *coloured observation noise* is dealt with in Chapter 4. The classical results on state estimation in coloured observation noise were developed by Bryson and Henrikson (4) which are briefly reviewed. The problem of adaptive estimation in coloured noise is still an untouched problem; recently Loh and Hauser (20) have attempted to solve it using the Carew-Bélanger algorithm. However, they assumed that the *noise transition matrix* of the coloured observation process is known. In Chapter 4, a time series model called the mixed *autoregressive moving average* model is derived for the system with coloured observation noise. This time series model enables the use of a recent result of Wilson (35) to estimate the noise transition matrix. Following Bryson and Henrikson, the noise transition matrix is used to transform the observation process contaminated with coloured noise into one with white noise. Also, the optimality of the innovations process is shown to be invariant under both white and coloured observation noise. A simulation for the system of Chapter 3 with coloured observation noise is performed. It involves first the estimation of the noise transition matrix and transforming the observation process, and then using the proposed scheme of Chapter 3 to identify and track the Kalman gain matrix. The accuracy of the estimation indicates the appropriateness of this novel approach to adaptive state estimation with coloured observation noise.

Conclusions and suggestions for future investigation in the problem of adaptive state estimation are discussed in Chapter 5. The algorithm of Wilson for multivariate time series estimation is outlined in Appendix B. A complete program listing of the subroutines used in implementing the Carew-Bélanger algorithm can be found in Appendix C.

The numerical results were obtained using a CDC 6400 computer. Parts of this work have been published and appear in references (32,33).

CHAPTER 2

OPTIMUM LINEAR FILTERING

2.1 Introduction

A problem which arises in a wide variety of engineering disciplines is the so-called filtering problem. The filtering problem was first formulated in the now famous studies of Kolmogorov (19) in U.S.S.R. and Wiener (34) in U.S.A. Working independently, the two almost simultaneously solved the linear filtering problem in which the criterion of optimality requires that the estimate of the signal be a *linear transformation* of the observation that minimizes the mean-square estimation error. Within the framework of the Kolmogorov-Wiener theory, all random processes (random functions of time t) are characterized by correlation functions. The optimum linear filter whose output is the desired estimate, when the input is the observation, is specified in terms of the known correlation functions by an integral equation called the *Wiener-Hopf* equation.

In 1961 Kalman and Bucy (16) presented a new approach to the linear filtering problem. The novelty of their formulation was the representation of all random processes by differential (difference) or state equations rather than correlation functions. By restricting their attention to

Gauss-Markov processes in particular, they derived a set of differential (difference) equations for the estimate. These equations can be used to construct a linear processor that is identical to the one specified by the Wiener-Hopf equation. There is a definite practical advantage, however, to a set of differential (difference) equations for the estimate, instead of an integral equation for the processor. To be more explicit, it is much easier to solve a set of differential (difference) equations by analog (digital) techniques than to solve an integral equation and then perform a convolution.

2.2 Kolmogorov-Wiener Filter

The optimum linear filtering and prediction problem, first solved by Kolmogorov (19) and Wiener (34), may be stated as follows: Given the scalar random process

$$y(t) = x(t) + n(t) \tag{2.1}$$

where $x(t)$ is the useful signal imbedded in the noise $n(t)$, and both are assumed to be random processes, determine a filter such that its output $\hat{x}(t)$ will be the best approximation to $x(t)$ in the mean-square sense. That is, minimize $E\,[\varepsilon^2(t)]$, with

$$\varepsilon(t) \overset{\Delta}{=} x(t) - \hat{x}(t), \tag{2.2}$$

and $E[\cdot]$ is the expectation operator.

Using variational arguments, Wiener showed that the impulse response $h(t)$ of the optimum linear filter satisfies the following Wiener-Hopf integral equation,

$$R_{xy}(\tau) - \int_{-\infty}^{\infty} R_y(\tau - \sigma)\, h(\sigma)\, d\sigma = 0 , \qquad (2.3)$$

where R_{xy} is the crosscorrelation function between $x(t)$ and $y(t)$, and R_y is the autocorrelation of $y(t)$, defined as

$$R_{xy}(\tau) \triangleq E[x(t)\, y(t - \tau)] , \qquad (2.4a)$$

$$R_y(\tau) \triangleq E[y(t)\, y(t - \tau)] . \qquad (2.4b)$$

The determination of the optimum filter transfer function requires the knowledge of the correlation functions (or the corresponding spectral densities) as well as performing spectral factorization (2). Clearly the formulation is cumbersome to implement with the present day computer techniques.

2.3 Kalman Filter

Kalman (15) and Kalman and Bucy (16) reformulated the optimum filtering problem to remove some of the difficulties of the Kolmogorov-Wiener filter. The novelty of the Kalman filter consisted of combining two well-known ideas;

(i) the state transition method of describing dynamical systems (36), and

(ii) linear filtering regarded as orthogonal projection in Hilbert space (24).

Consider the *discrete-time-invariant system* shown in Figure 2.1. The *message-generating* process is modeled by the vector Gauss-Markov equation

$$x_{i+1} = Fx_i + Gu_i \quad , \tag{2.5}$$

and the observation is modeled by the linear algebraic relationship

$$y_i = Hx_i + v_i \quad , \tag{2.6}$$

with

x = the n dimensional state vector,

y = the m dimensional observation vector,

u = the p dimensional message - generating noise vector,

v = the m dimensional observation noise vector.

The message-generating and observation noise are assumed to be zero-mean, independent white gaussian noise (WGN) with covariance matrices Q and R respectively; in other words

$$E[u_i v_j'] = 0 \ , \ E[u_i] = 0 \ , \ E[v_i] = 0 \ ,$$
$$E[u_i u_j'] = Q\delta_{ij} \ , \ E[v_i v_j'] = R\delta_{ij} \ , \tag{2.7}$$

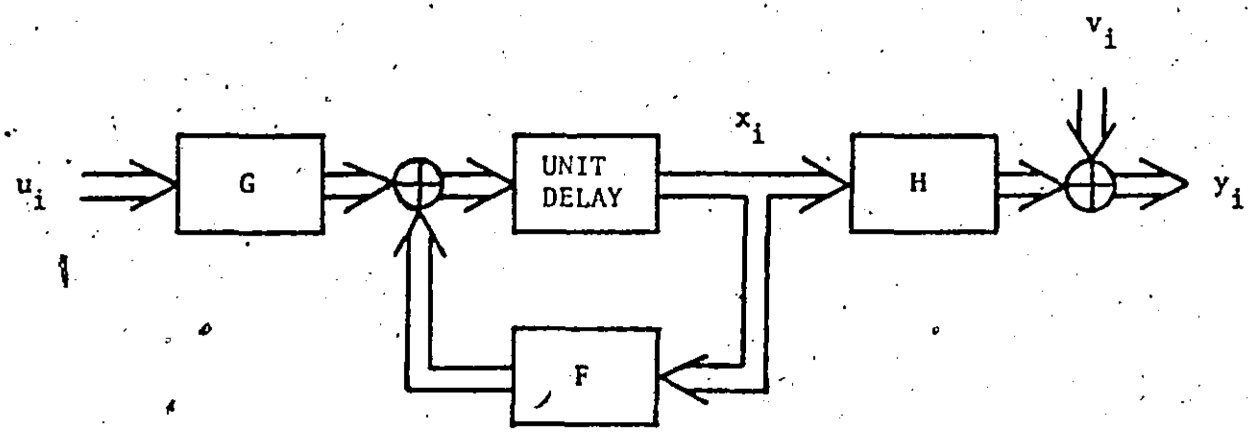

Figure 2.1   Block diagram of system represented in state space form.

where δ_{ij} is the Kronecker delta function, and the prime denotes matrix transpose.

It is required to obtain the best estimate of the state, x_i, given the fixed triple $\{F,G,H\}$ and the observation set $Y_i \triangleq \{y_k, 0 \le k \le i\}$.

At a first glance, it might seem that the state estimation problem is quite different from the filtering problem mentioned earlier. However, as pointed out by Kalman, any signal with a rational spectral density can be obtained by applying WGN to a linear system. Thus the optimum estimate can also be regarded as the signal to be obtained by filtering y_k, and the specification of the system dynamics is equivalent to specifying the autocorrelation function of the signal.

As shown by Kalman and Bucy (16), the optimal estimate $\hat{x}_{i+1|i}$, of the state vector x_{i+1}, given the subspace Y_i spanned by the random variables $\{y_k, 0 \le k \le i\}$ may be interpreted geometrically as the projection of x_{i+1} onto Y_i. Algebraically, the estimates are given by

$$\hat{x}_{i+1|i} \triangleq \text{Proj}\ \{x_{i+1}|Y_i\}$$

$$= F\hat{x}_{i|i-1} + K_i\varepsilon_i \tag{2.8}$$

where ε_i, the innovations process (14), is defined as

$$\varepsilon_i \triangleq y_i - H\hat{x}_{i|i-1} \quad , \tag{2.9}$$

and K_i, the Kalman gain matrix, is found recursively by the following set of matrix equations

$$K_i \triangleq FP_iH'W_i^{-1} \quad , \qquad (2.10)$$

$$W_i \triangleq E[\varepsilon_i\varepsilon_i']$$

$$= HP_iH'+R \quad , \qquad (2.11)$$

$$P_{i+1} \triangleq E[(x_{i+1} - \hat{x}_{i+1|i})(x_{i+1} - \hat{x}_{i+1|i})']$$

$$= FP_iF' - FP_iH'W_i^{-1} HP_iF' + GQG' \; . \qquad (2.12)$$

Kalman and Bucy (16) have also shown that if the system (2.5) and (2.6) satisfies the following conditions;

C - 2.1 uniformly *completely observable* ;

C - 2.2 uniformly *completely controllable* ;

C - 2.3 $\alpha_1 \leq ||Q|| \leq \alpha_2, \alpha_3 \leq ||R|| \leq \alpha_4$;

C - 2.4 $||F|| \leq \alpha_5$.

Then the Kalman filter is identical to the Kolmogorov-Wiener filter and that

(i) the Kalman filter is uniformly asymptotically stable,

(ii) every solution of the variance equation (2.12) starting at a symmetric, non-negative matrix P_0 converges to P as $i \to \infty$.

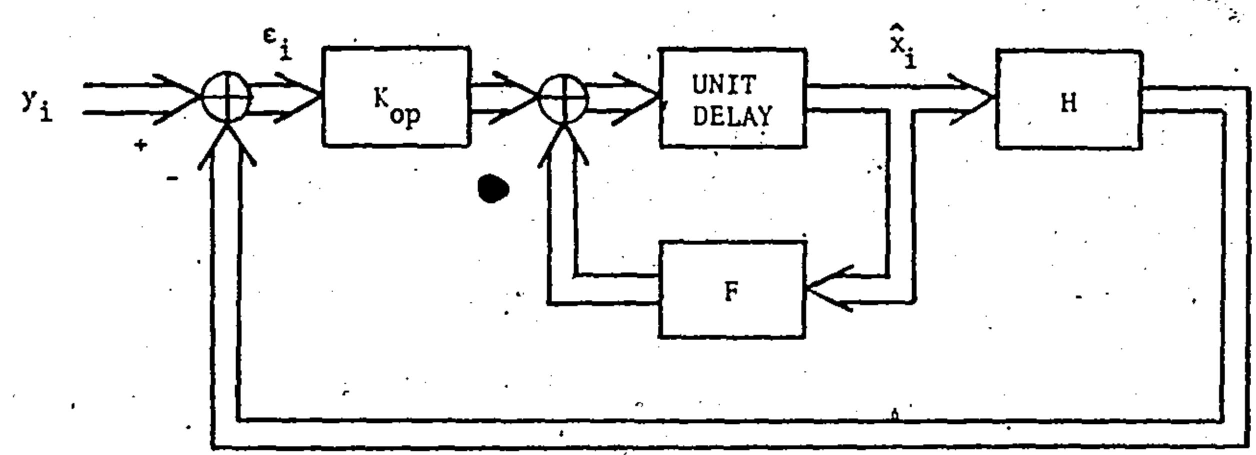

Figure 2.2 Block diagram of Kalman filter.

Therefore, the steady-state value of the Kalman filter parameters are given by

$$K_{op} \triangleq \lim_{i\to\infty} K_i = FPH' \, W^{-1}, \tag{2.13}$$

$$W \triangleq \lim_{i\to\infty} W_i = HPH' + R \,, \tag{2.14}$$

$$P \triangleq \lim_{i\to\infty} P_i = FPF' - FPH' \, W^{-1} \, HPF' + GQG' . \tag{2.15}$$

The filter described by the matrix equations (2.8), (2.9), and (2.13) - (2.15) is termed a stationary Kalman filter (from here on, the term Kalman filter will imply the stationary Kalman filter). To implement the Kalman filter would require prior knowledge of the message-generating and observation noise covariance matrices Q and R respectively. In practice, such extensive *a priori* information is seldom available, with the result that the optimum Kalman gain matrix K_{op} cannot be calculated. On the other hand, if incorrect assumptions are made about these matrices, the resulting estimates of the states are suboptimum. A more straightforward method to solve the optimum filtering problem is to determine the Kalman gain matrix directly by using only the observations in some adaptive schemes instead of equations (2.13) - (2.15). In the next chapter, the problem of Kalman filtering without prior knowledge of noise statistics is dealt with.

CHAPTER 3

ADAPTIVE KALMAN FILTERING

3.1 Introduction

The problem of optimum filtering, formulated by Kalman and Bucy (16), assumes complete *a priori* knowledge of the message-generating and observation noise covariance matrices. These statistics in most practical situations are either unknown or known only approximately; in such cases, the system performance is at best suboptimum (11). The purpose of an adaptive filter is to reduce or bound the estimation errors by modifying or adapting the Kalman filter to the real data.

A number of approaches (12, 30) have been presented with varying degrees of success for the estimation of the unknown covariance matrices, a good summary can be found in the paper by Mehra (22). The use of the innovations sequence $\{\varepsilon_i\}$ in the estimation of the unknown covariances as a criterion of optimality was introduced by Mehra (21). Carew and Bélanger (6), using the same argument as Mehra, proposed an algorithm for estimating the Kalman gain matrix directly. Other authors (10, 29, 31) have suggested the use of stochastic approximation techniques based on properties of the optimal filter to adopt the Kalman gain matrix.

In this chapter, the Carew-Bélanger algorithm and stochastic approximation algorithm are briefly examined. Also a scheme is proposed to identify and track the Kalman gain matrix combining the advantageous properties of the above algorithms.

3.2 The Carew-Bélanger Algorithm (6)

The algorithm of Carew and Bélanger is based on the correlation technique (22). Starting with an arbitrary gain matrix K_s, which may be calculated from equations (2.13) - (2.15) by using some assumed values of Q and R, the sub-optimum filter is given by

$$x^*_{i+1|i} \triangleq F x_i^*|_{i-1} + K_s(y_i - H x_i^*|_{i-1}) \quad , \tag{3.1}$$

where $x_i^*|_{i-1}$ denotes the suboptimum estimates of the state x_i given $\mathcal{Y}_{i-1}$. The steady-state error covariance of the suboptimum filter, related to the optimum, is defined by

$$P^* \triangleq E[(\hat{x}_{i|i-1} - x_i^*|_{i-1})(\hat{x}_{i|i-1} - x_i^*|_{i-1})'] . \tag{3.2}$$

Using equations (2.6), (2.8) and (3.1), it can be easily shown that

$$P^* = (F-K_sH) P^*(F-K_sH)' + (K_s-K_{op}) W(K_s-K_{op})' \quad , \tag{3.3}$$

where K_{op} is the optimum Kalman gain.

Following Mehra (21), the autocorrelation functions for the suboptimum innovations sequence in the steady-state are given by

$$C_j \triangleq E[\varepsilon_i^* \varepsilon_{i-j}^{*\prime}]$$

$$= \begin{cases} H(F-K_sH)^{j-1}[(F-K_sH)P^*H' - (K_s-K_{op})W], & j \neq 0 \\ HP^*H' + W, & j = 0 \end{cases} \tag{3.4}$$

where ε_i^*, the suboptimum innovations sequence, defined by

$$\varepsilon_i^* \triangleq y_i - Hx_i^*|_{i-1} , \tag{3.5}$$

is the innovations of the suboptimum filter. For the optimum filter $K_s = K_{op}$ and $P^* = 0$; hence from (3.4), $C_j = 0$ for $j \neq 0$.

Define,

$$B \triangleq [H' | F'H' | \ldots | (F')^{n-1} H'] \tag{3.6}$$

which may be recognized as the nrxn system *observability matrix*. From (3.4) and (3.6), define A as

$$A \triangleq B(FP^*H' + K_{op}W)$$

$$= \begin{bmatrix} C_1 + HK_sC_0 \\ C_2 + HK_sC_1 + HFK_sC_0 \\ \vdots \\ C_n + HK_sC_{n-1} + \ldots + HF^{n-1}K_sC_0 \end{bmatrix} \tag{3.7}$$

Since the system is completely observable, matrix B is of full rank, therefore its pseudoinverse $B^{\dagger}$, defined as

$$B^{\dagger} \triangleq (B'B)^{-1}B', \tag{3.8}$$

exists and the matrix equation (3.7) may be solved to determine $(FP^*H' + K_{op}W)$ from the experimentally estimated autocorrelation functions C_j ($j=0, 1, \ldots n$). The solution is given by

$$(FP^*H' + K_{op}W) = B^{\dagger}A \quad . \tag{3.9}$$

From equations (3.3), (3.5) and (3.9), these simultaneous matrix equations for W, K_{op} and P^* are established as below

$$W = C_0 - HP^*H' \quad , \tag{3.10a}$$

$$K_{op} = (B^{\dagger}A - FP^*H')\, W^{-1} \quad , \tag{3.10b}$$

$$P^* = (F-K_sH)\, P^*(F-K_sH)' + (K_s-K_{op})\, W(K_s-K_{op})' \; . \tag{3.10c}$$

Rewritting (3.10a) - (3.10c) in recursive form

$$W\,(X_k) = C_o - HX_kH' \quad , \tag{3.11a}$$

$$K_{op}(X_k) = (B^{\dagger}A - FX_kH')\, W^{-1}(X_k) \quad , \tag{3.11b}$$

$$X_{k+1} = (F-K_sH)X_k(F-K_sH)' + (K_s-K_{op}(X_k))\,W(X_k)\,(K_s-K_{op}(X_k))', \tag{3.11c}$$

where $X_k \varepsilon R^{nxn}$ is positive semidefinite. Carew and Bélanger have proved that the algorithm represented by (3.11a) - (3.11c) will converge uniquely to K_{op} by showing that these equations represent a contraction mapping. A summary of the scheme of Carew and Bélanger is given by the following steps:

(i) For an arbitrary gain matrix K_s, based on assumed values of Q and R, generate the suboptimum innovations sequence $\{\varepsilon_i^*\}$, and hence estimate the autocorrelation functions C_j from

$$\hat{C}_j = \frac{1}{N} \sum_{i=1}^{n-j} \varepsilon_{i+j}^* \varepsilon_i^{*'} \quad , \qquad j = 0,1,\ldots,n \quad (3.12)$$

where N is the size of the innovations sequence. The estimates given by (3.12) are biased for finite N, but asymptotically they are unbiased and consistent.

(ii) Using K_s and $\hat{C}_j$ $(j = 0,\ldots n)$, estimates of P*,W, and K_{op} are obtained iteratively from equations (3.11a) - (3.11c) with $X_0 \varepsilon \zeta = \{X : X$ positive semidefinite, $X \le P^*+P\}$.

The iterative scheme (3.11a) — (3.11c) converges uniquely to the optimum Kalman gain K_{op} if the autocorrelation functions, C_j $(j = 0,1,\ldots,n)$ are known accurately. However, due to the finite size of the innovations sequence and other experimental errors, the accuracy with which K_{op} can be determined is limited.

3.3 Stochastic Approximation

3.3.1. Stochastic Approximation Algorithms

Stochastic approximation methods may be considered as recursive estimation methods, updated by an appropriately weighted, arbitrarily chosen error corrective term, with the only requirement that, in the limit, it converges to the true parameter sought. Applications of stochastic approximation algorithms have been proposed in adaptive and learning systems (23), systems identification (26), adaptive communication (27).

Historically, stochastic approximation was first treated by Robbins and Monro (25) and Kiefer and Wolfowitz (17), who were concerned with solution to two specific problems; finding the root of a regression function, and finding the value that minimizes a regression function given only pertinent random observations. It was Dvoretzky (8) who generalized stochastic approximation to any sort of iterative solution algorithm, which is convergent, when direct observations of a regression function can be adopted successfully. Excellent surveys of stochastic approximation can be found in papers by Sakrison (27) and Saridis (28).

In general, stochastic approximation algorithms of the Robbins and Monro type are used in adaptive filtering. They are of the form

$$K_{i+1} = K_i + \gamma_i \left[f(y_i, K_i) - m_o \right], \qquad (3.12)$$

where $\{\gamma_i\}$ is a sequence of suitably chosen smoothing values, and $\{f(y_i, K_i) - m_o\}$ is an error correction sequence generated at every time instant i by measuring the deviation from an appropriate goal. The iterative scheme (3.12) approaches the optimal parameter value, K_{op}, where $E[f(y_i, K_{op})] = m_o$, in the mean-square sense provided the following assumptions are satisfied (see Saridis (28)):

A - 3.1 $\exists\ \alpha, \beta$, $-\infty < \alpha \leq \beta < 0$, such that
$\alpha \, ||K-K_{op}||^2 \leq \langle K-K_{op}, E[f(y_i, K_i)] - m_o \rangle \leq \beta \, ||K-K_{op}||^2$,
where $\langle \cdot , \cdot \rangle$ denotes the matrix inner product operator:

A - 3.2 The γ_i's are positive monotone decreasing; and

$$\sum_{i=1}^{\infty} \gamma_i = \infty \quad ; \quad \sum_{i=1}^{\infty} \gamma_i^2 < \infty \quad .$$

Heuristically, A - 3.1 requires that the regression function $f(y_i, K) - m_o$ be bounded on all sides of a true solution by a rectangular set in the solution space such that it is not possible to overshoot the solution, K_{op}, which cannot be corrected by a γ_i satisfying A-3.2. Assumption A - 3.2 provides, smoothing effect on the regression function, unlimited correction effort, and mutual cancellation of individual errors for a large number of iterations.

3,3.2 Adaptive Kalman Filtering by Stochastic Approximation

The successful application of stochastic approximation to search for the optimum gain matrix, K_{op}, requires some suitable method for testing if the value presently being used is optimum. Hampton and Schultz (10) have proposed a Robbins and Monro type algorithm which uses the orthogonal condition

$$E\left[\{x_i - \hat{x}_{i|i-1}\} y_j'\right] = 0 \quad , \ \forall\, j<i \tag{3.13}$$

for this test. Since the actual x_i is not known, equation (3.13) can only be approximated in a rather involved manner. More recently, Sinha and Muhkerjee (31) have also proposed a Robbins and Monro type algorithm which makes use of the property that the innovations process is white (14). They used, for the test of optimality

$$E\left[\varepsilon_i \varepsilon_j'\right] = 0 \quad , \quad i \neq j \quad . \tag{3.14}$$

Although their idea is conceptually more direct and works quite well for the scalar case, it is unsuitable for the multivariate case since the error correction term is restricted to a subspace of the solution set in most instances. An algorithm which is more suitable would be that proposed by Scharf and Alspach (29), who utilized the orthogonality between the innovations process and the estimated state, that is,

$$E[\hat{x}_{i|i-1}\,\varepsilon_i'] = 0 \qquad (3.15)$$

Note that the product $\hat{x}_{i|i-1}\varepsilon_i'$ is contained in the same space as that of the Kalman gain, making it applicable for multivariate problems. The stochastic approximation algorithm thus can be written as

$$K_{i+1} = K_i + \gamma_i \cdot \frac{\hat{x}_{i|i-1}\varepsilon_i'}{||\hat{x}_{i|i-1}\varepsilon_i'||^2} , \qquad (3.16)$$

where γ_i is choosen so as to satisfy assumption A-3.2. Scharf and Alspach (29) have shown, in the scalar case of (3.16), that the regression function, $\hat{x}_{i|i-1}\varepsilon_i'$, satisfies assumption A-3.1, thereby showing that (3.16) converges in the mean-square to K_{op}. The proof of mean-square convergence of (3.16) in the multivariate case is shown in Appendix A.

3.4 Combined Carew-Bélanger and Stochastic Approximation Algorithm

The algorithms discussed in the previous two sections have their advantages and disadvantages listed in Table 3.1. It would appear logical to combine the two methods in such a manner as to retain their relative advantages, while disposing with their basic drawbacks.

ADVANTAGES		DISADVANTAGES	
CAREW-BELANGER	STOCHASTIC APPROXIMATION	CAREW-BELANGER	STOCHASTIC APPROXIMATION
(i) fast convergence	(i) minimum computation	(i) opened-loop scheme	(i) very slow convergence
(ii) minimum *a priori* information	(ii) closed-loop adaptivity	(ii) complex computation	(ii) required "good" starting values
		(iii) large computer storage required for accurate results.	

TABLE 3.1 Advantages and disadvantages of the Carew-Bélanger and stochastic approximation algorithms

Basically the Carew-Bélanger algorithm is an open-loop estimator. It gathers a sample of the observations and processes it. Should the noise statistic be "slowly-varying", however, then the estimated gain will in all likelihood be suboptimal and the estimates of the state might even diverge after some time.

More appropriate would be a controller which monitors the observations and the estimates, and be able to determine whether the filter is optimum or not. If the filter is suboptimum, the said controller should be able to adjust the filter again to bring the filter back towards the optimum state. Such a controller can be implemented by using the stochastic approximation algorithm (3.16) discussed in Section 3.3.

By the above arguments, it is proposed to implement the Carew-Bélanger algorithm to arrive at a one-shot estimate of the Kalman gain, followed by using stochastic approximation to track any change in the gain matrix which may bring further improvement in the estimates. As the system noise is slowly varying, these changes will usually be small, therefore a slight improvement would result as compared with the open-loop estimation of the gain matrix using only the method of Carew and Bélanger.

The steps of the proposed scheme can be stated as follows:

(i) Using the algorithm of Carew and Bélanger described in section 3.2 and a finite observation sample, calculate $\hat{K}$, the estimate of K_{op};

(ii) Using $\hat{K} = K_0$, track the Kalman gain with equation (3.16).

3.5 Simulation Results

To test the proposed scheme, it was applied to the same system from inertial navigation as was used by Mehra (21) as well as Carew and Bélanger (6). For this case, the system matrices are

$$F = \begin{bmatrix} 0.75 & -1.74 & -0.3 & 0.0 & -0.15 \\ 0.09 & -0.31 & -0.0015 & 0.0 & -0.008 \\ 0.0 & 0.0 & 0.95 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.55 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.905 \end{bmatrix},$$

$$G = \begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \\ 24.64 & 0.0 & 0.0 \\ 0.0 & 0.835 & 0.0 \\ 0.0 & 0.0 & 1.83 \end{bmatrix},$$

$$H = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 1.0 \\ 0.0 & 1.0 & 0.0 & 1.0 & 0.0 \end{bmatrix} .$$

The noise sequence $\{u_i\}$ and $\{v_i\}$ were generated on the computer, and their actual covariances, obtained from 2000 samples are

$$Q = \text{diag}\ (0.941,\ 1.050,\ 0.980),$$
$$R = \text{diag}\ (1.040,\ 1.024)\ .$$

The optimal Kalman gain matrix for this system, found by solving (2.13) - (2.15), is

$$K_{op} = \begin{bmatrix} 1.563 & 0.557 \\ 0.092 & 0.387 \\ -2.718 & -1.416 \\ 0.0 & 0.137 \\ 0.029 & -0.702 \end{bmatrix} . \qquad (3.17)$$

The first 1000 values of $\{y_i\}$ were used with the Carew Bélanger algorithm to estimate the Kalman gain, and determined to be

$$\hat{K} = \begin{bmatrix} 1.439 & 1.070 \\ 0.130 & 0.306 \\ -2.767 & -2.043 \\ -0.034 & 0.197 \\ 0.139 & -0.793 \end{bmatrix} . \qquad (3.18)$$

The next 1000 values of $\{y_i\}$ were filtered by three methods;

(1) using the optimum gain matrix (3.17),

(2) using the estimated gain matrix (3.28) found by the Carew-Bélanger algorithm,

(3) starting with the Carew-Bélanger gain matrix, and using the stochastic algorithm (3.16), with $\gamma_i = \frac{1}{i+1}$ and $K_0 = \hat{K}$.

The performance of the three filters are summarized in Table 3.2. The trace of the error covariance matrix P was plotted for each of the three methods, against iterations in Figure 3.1. It was perceived that initially, these values were very close, but moved apart considerably as more time elapsed.

TABLE 3.2

COMPARISON OF FILTER PERFORMANCE (WHITE NOISE)

Method	P_{11}	P_{22}	P_{33}	P_{44}	P_{55}	trace P.
1)	75.893	1.012	1251.941	0.897	11.725	1341.468
2)	81.850	1.260	1253.537	1.031	17.185	1354.863
3)	80.232	1.174	1252.451	1.022	15.600	1350.479

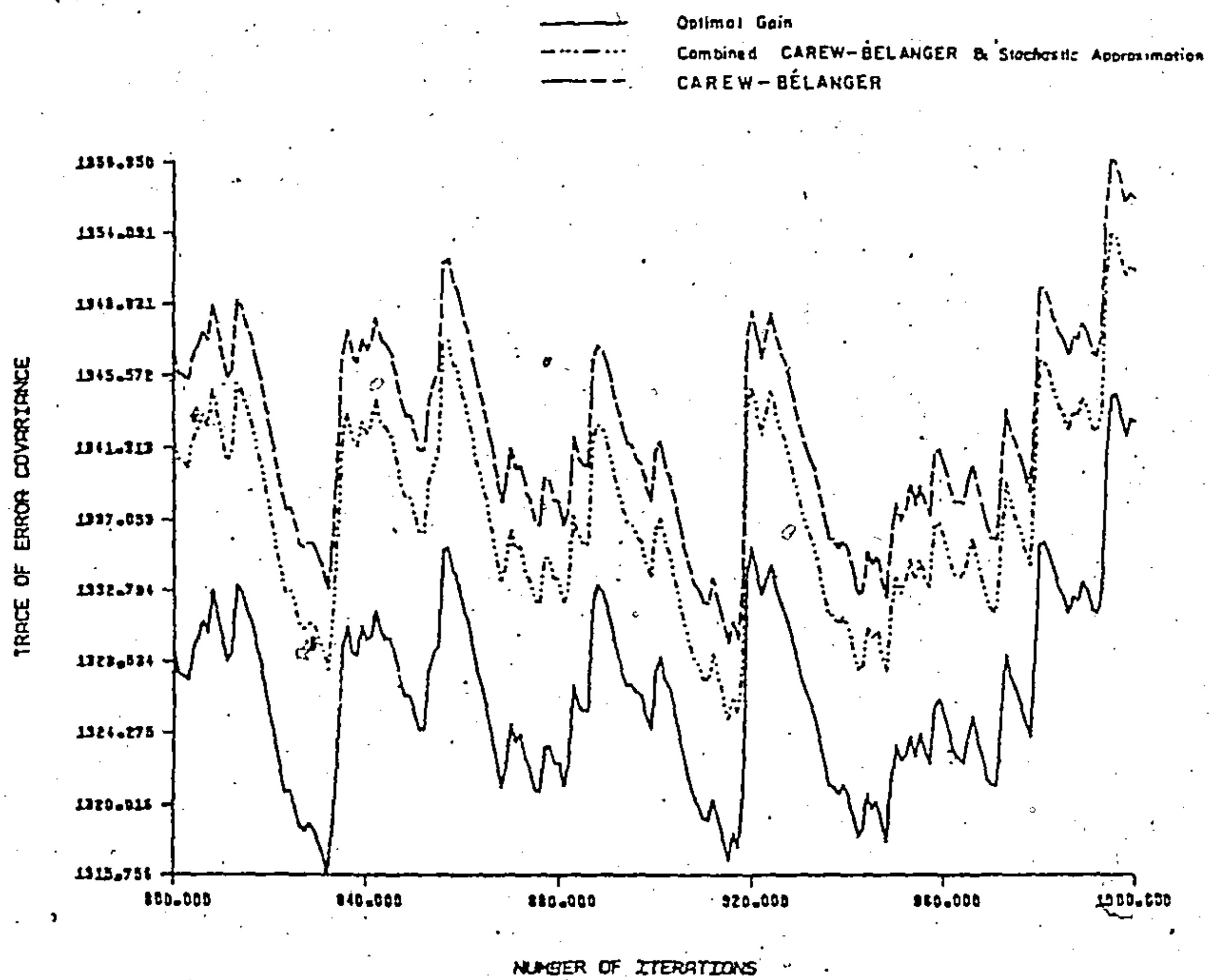

Figure 3.1 Variation of the trace of error covariance matrix for different methods (white noise).

3.6 Discussion

In the example considered, the performance of the Carew-Bélanger algorithm for estimating the Kalman gain matrix is adequate. Due to computer storage limitations and the possibility of non-stationarity of the noise in the system, however, a closed-loop scheme employing stochastic approximation is used. This scheme, which requires minimal storage of past data and computation once it is initiated, is very suitable for on-line tracking of the Kalman gain.

As can be seen from Figure 3.1, the use of stochastic approximation method in conjunction with that of Carew and Bélanger gives a slight improvement in the estimates. From Table 3.2, the improvement is approximately 48% with respect to the optimum gain after 1000 iterations of stochastic approximation. Furthermore, this improvement becomes more marked in time even if the noise statistics are stationary. If the noise statistics should deviate slowly, the closed-loop nature of the proposed scheme would adjust to these changes. Thus, the proposed scheme is very suitable for on-line adaptive Kalman filtering.

CHAPTER 4

ADAPTIVE STATE ESTIMATION IN COLOURED OBSERVATION NOISE

4.1 Introduction

In the previous two chapters, discussion of state estimation via Kalman filtering was restricted to systems with white noise. More general is the problem of state estimation in coloured noise. By coloured noise, it is implied that the autocorrelation function of the noise is nonzero for nonzero time lags.

Coloured input or message-generating noise actually presents no problem since it can be simply taken into account by regarding it as the output of a linear system with white noise as input and augmenting the state vector accordingly. Hence, the basic problem of interest is coloured observation noise.

The problem of state estimation in discrete-time with coloured observation noise is essentially the same as for the white noise case - to find a filter which will minimize the state estimation error in the mean-square sense. The problem has been studied by Cox (7) in 1963 and later by several others. However, at the present time, perhaps the best known results on such problems are those of Bryson and Henrikson (4). Motivated by the solution of the corresponding continuous time problem (5), Bryson and Henrikson transformed the observation process into one with white noise disturbances and then solving for a filter of the Kalman type for estimating the state.

Though their solution is elegant, Bryson and Henrikson's results require the prior knowledge of the statistics and behaviour of the system noise. In practical situations, such *a priori* knowledge if often not at hand. Hence, some type of learning filter must be used to adaptively estimate the states without requiring such *a priori* information.

In this chapter, a brief look at Bryson and Henrikson's results is taken. Also, the innovations process of the optimum filter for coloured observation noise is shown to be white, thus optimality is shown to be invariant for both the white and coloured observation noise. An approach is presented in which the state space model is rewritten as an autoregressive moving average (ARMA) model, allowing for preliminary analysis of a class of coloured noise to be made. The combined results above are used in conjunction with the methods of Chapter 3 to identify and track the optimum Kalman gain matrix.

4.2 Some Results of Bryson and Henrikson (4)

Consider the system as described by the state equations (2.5) and (2.6). It is assumed that the message-generating disturbance $\{u_i\}$ is a white gaussian noise sequence with statistics given by (2.7). The observation noise $\{v_i\}$ is a coloured noise which will be assumed to be of the Gauss-Markov type, modeled by the following equation,

$$v_{i+1} = Av_i + w_i . \tag{4.1}$$

The noise transition matrix A, is assumed to be diagonal; and $\{w_i\}$ is a white gaussian noise sequence of dimension m independent of $\{u_i\}$. The statistics of $\{w_i\}$ are

$$E[w_i] = 0, \quad E[w_i w_j'] = R\,\delta_{ij} . \tag{4.2}$$

The problem is to determine the best estimate (in the mean-square sense) of the state vector, x_i, from the record of the noisy input data sequence $\mathcal{Y}_i = \{y_k, 0 \leq k \leq i\}$.

Following the track of the corresponding continuous time problem (5), Bryson and Henrikson began by employing a *measurement-differencing* procedure on the observation process. Thus they define

$$y_i^c \triangleq y_{i+1} - Ay_i \tag{4.3a}$$

$$= H_c x_i + v_i^c , \tag{4.3b}$$

where

$$H_c \triangleq HF - AH , \tag{4.4}$$

$$v_i^c \triangleq HGu_i + Bw_i . \tag{4.5}$$

Clearly $\{v_i^c\}$ is a white gaussian process, thus the transformed observation process $\{y_i^c\}$ now has a white disturbance noise instead of a sequentially correlated one. Looking closely at (4.5) will show that this white disturbance is now correlated with the message-generating noise $\{u_i\}$, and is given by

$$E[u_i v_j^{c\prime}] = QG'H'\delta_{ij} \quad . \tag{4.6}$$

Hence, modifying the original Kalman equations (2.8), (2.9) and (2.13) - (2.15), a Kalman type filter can be applied to the transformed system. The equations governing such a filter are

$$\hat{x}_{i+1} \stackrel{\Delta}{=} F\hat{x}_i + K_{op}^c \, \varepsilon_i^c \quad , \tag{4.7}$$

$$\varepsilon_i^c \stackrel{\Delta}{=} y_i^c - H_c \hat{x}_i \quad , \tag{4.8}$$

where the Kalman gain, K_{op}^c, is given by

$$K_{op}^c \stackrel{\Delta}{=} (FP_c H_c' + GQG'H') W_c^{-1} \quad . \tag{4.9}$$

As can be seen, to solve for the Kalman gain requires the knowledge of P_c and W_c which are the covariance matrices of the state estimation error and innovations process respectively. They can be found by solving simultaneously the following matrix equations:

$$P_c \overset{\Delta}{=} E\,[\tilde{x}_{i+1}\; \tilde{x}_{i+1}']$$
$$= FP_cF' - (FP_cH_c' + GQG'H')\, W_c^{-1}(FP_cH_c' + GQG'H')' + GQG' , \tag{4.10}$$

$$W_c \overset{\Delta}{=} E\,[\varepsilon_i^c\; \varepsilon_i^{c\prime}]$$
$$= H_cP_cH_c' + HGQG'H' + R, \tag{4.11}$$

where

$$\tilde{x}_i \overset{\Delta}{=} x_i - \hat{x}_i \tag{4.12}$$

is the state estimation error.

Also of interest is the behaviour of the autocorrelation functions of the innovations process $\{\varepsilon_i^c\}$ with respect to the filter gain K.

Using (4.3b), (4.8), and (4.12), the innovations process can be reexpressed in the following form:

$$\varepsilon_i^c \overset{\Delta}{=} H_c\tilde{x}_i + \tilde{v}_i^c . \tag{4.13}$$

Thus, the autocorrelation functions of $\{\varepsilon_i^c\}$ are given by

$$C_j \overset{\Delta}{=} E[\varepsilon_i^c\varepsilon_{i-j}^{c\prime}]$$
$$= H_c\{E[\tilde{x}_i\; \tilde{x}_{i-j}]\, H_c' + E\,[\tilde{x}_i v_{i-j}^c]\} . \tag{4.14}$$

The expectation values of (4.14) can be easily shown to be

$$E\left[\tilde{x}_i \, \tilde{x}_{i-j}'\right] = (F - KH_c)^j P_c , \tag{4.15a}$$

and

$$E\left[\tilde{x}_i \, v^{c\,\prime}_{i-j}\right] = (F - KH_c)^{j-1}\left[GQG'H' - K(HGQG'H' + R)\right] . \tag{4.15b}$$

Therefore combining (4.11), (4.14), (4.15a) and (4.15b) the autocorrelation functions of the innovations expressed in terms of the filter gain are

$$C_j = H_c \, (F-KH_c)^{j-1} \left[FP_cH_c' + GQG'H' - KW_c \right] , \quad j \neq 0. \tag{4.16}$$

Substituting the value of K_{op}^{c} from (4.9) for K in (4.16), it is obvious that

$$C_j = 0 \quad , \quad \forall \, j \neq 0 . \tag{4.17}$$

Therefore, as in the case considered in Chapter 3, the optimum Kalman gain matrix for the coloured noise case will result in a Kalman filter which produces a white innovations process. Thus, the optimality of the innovations for the coloured observation noise system, is the same as that of the white noise system.

The solution of the Kalman gain in coloured noise requires the values Q, R and A. In the next section, a model is developed for estimating A, the noise transition matrix.

4.3 Representation of Systems by ARMA Models

The general system as represented by equations (2.5), (2.6) and (4.1) can be rewritten as a mixed autoregressive moving average (ARMA) model (3).

Equations (2.5) and (4.1) are both Gauss-Markov type equations, and can be expressed in terms of the backward shift operator Z^{-1} as below

$$x_{i+1} = (I_n - FZ^{-1})^{-1}Gu_i \,, \tag{4.18}$$

$$v_{i+1} = (I_m - AZ^{-1})^{-1}w_i \,, \tag{4.19}$$

where I_n is an nxn identity matrix.

Substituting the above equations into (2.6), the observation process becomes

$$\begin{aligned} y_{i+1} &= Hx_{i+1} + v_{i+1} \\ &= H(I_n - FZ^{-1})^{-1}Gu_i + (I_m - AZ^{-1})^{-1}w_i \,. \end{aligned} \tag{4.20}$$

Premultiplying both sides of the equation by $(I_m - AZ^{-1})$ results in

$$(I_m - AZ^{-1})y_{i+1} = (I_m - AZ^{-1})\, H\, (I_n - FZ^{-1})^{-1}Gu_i + W_i \,. \tag{4.21}$$

The annoying fact about (4.21) is the matrix inverse $(I_n - FZ^{-1})^{-1}$. This can be removed by employing Fadeeva's scheme (9),

$$(I_n - FZ^{-1})^{-1} = \frac{\Lambda_1 Z^{-1} + \ldots + \Lambda_n Z^{-n}}{1 + \lambda_1 Z^{-1} + \ldots + \lambda_n Z^{-n}} \,, \quad (4.22)$$

where
$$\Lambda_1 = I,$$
$$\lambda_1 = - \text{trace } F,$$
$$\Lambda_k = F\Lambda_{k-1} + \lambda_{k-1} I \;,\; k = 2,\ldots,n$$
$$\lambda_k = - \frac{1}{k} \text{ trace } F\Lambda_k \;,\; k = 2,\ldots,n.$$

Thus from (4.21)and (4.22), the system model is

$$(I_m - AZ^{-1})\ \xi_{i+1} = (I_m - AZ^{-1})\ H\ (\Lambda_1 Z^{-1} + \ldots + \Lambda_n Z^{-n})\ Gu_i$$
$$+ (1 + \lambda_1 Z^{-1} + \ldots + \lambda_n Z^{-n})\ w_i \;, \quad (4.23)$$

where $$\xi_{i+1} = (1 + \lambda_1 Z^{-1} + \ldots + \lambda_n Z^{-n})\ y_{i+1}, \quad (4.24)$$

and $$\{\Lambda_k,\ k=1,\ldots,n\}\ \varepsilon\ R^{nxn} .$$

The right hand side of (4.23) may be recognized as the sum of two moving average terms. Box and Jenkins (3) have shown that the sum can be expressed as a single moving average term in the following manner

$$(I_m - AZ^{-1})\, H\, (\Lambda_1 Z^{-1} + \ldots + \Lambda_n Z^{-n})\, Gu_i + (1 + \lambda_1 Z^{-1} + \ldots + \lambda_n Z^{-n}) w_i$$

$$= (I_m - \theta_1 Z^{-1} - \ldots - \theta_q Z^{-q}) a_{i+1} \quad , \qquad (4.25)$$

where $\{\theta_k,\ k = 1,\ldots,q\} \in R^{m\times m}$.

From (4.23) and (4.25), the system, rewritten in terms of the backward shift operator becomes

$$\Phi(Z^{-1})\xi_{i+1} = \theta(Z^{-1}) a_{i+1} \quad , \qquad (4.26)$$

with

$$\Phi(Z^{-1}) \triangleq (I_m - AZ^{-1}) \quad , \qquad (4.27)$$

$$\theta(Z^{-1}) \triangleq (I_m - \theta_1 Z^{-1} - \ldots - \theta_q Z^{-q}) \, , \qquad (4.28)$$

which may be recognized as an ARMA model of order (1, q) where $q \leq n$ is to be determined.

The estimation of the ARMA parameter matrices $B \triangleq \{A,\ \theta_k,\ k=1,\ldots,q\}$ is accomplished by a non-linear least-squares estimation technique proposed recently by Wilson (35) (see appendix B for outline of algorithm). The value of q is the minimum number of moving average terms to obtain a residual sequence $\{a_i\}$ which is white, i.e. $E\,[a_i a_j'] = C\delta_{ij}$.

4.4 Proposed Scheme

In a recent paper (20), Loh and Hauser have considered the use of the Carew-Bélanger algorithm for message-generating and observation noise which are correlated. They have shown that the convergence property of the Carew and Bélanger algorithm is not affected by such correlation. A system with coloured noise was simulated and the estimation of the Kalman gain was made. Although the results were very good, an *a priori* knowledge of the noise transition matrix was assumed.

The proposed scheme relaxes the above restriction. The approach therefore, is to use the derived ARMA model of the previous section and a finite observation sample $\{y_i\}$ to estimate the noise transition matrix A by the algorithm of Wilson (35). From the estimated value of the matrix A, the transformation (4.3a) can be performed on the observation process. Since the optimality of the innovations process is invariant, the scheme of Carew and Bélanger can be applied to the transformed system (2.5) and (4.3b) to obtain an initial estimate of the Kalman gain K_{op}^{c}. A stochastic approximation algorithm is then used, as proposed in Chapter 3, for tracking the filter gain matrix in a closed-loop fashion.

4.5 Simulation Results

To test the proposed scheme, it was applied to the same system used in Section 3.5. The white noise sequence

$\{u_i\}$ and $\{w_i\}$ were generated on the computer. Their covariances, calculated from 2000 samples are

$$\text{diag } Q^{\rho} = (0.999 \; , \; 0.986 \; , \; 1.012 \;) \; ,$$
$$\text{diag } R = (0.245 \; , \; 0.040 \;) \; .$$

The coloured observation noise $\{v_i\}$ was calculated from $\{w_i\}$ using (4.1) and a noise transition matrix given by

$$\text{diag } A = (0.5 \; , \; 0.2) \; .$$

The optimum Kalman gain derived from equations (4.97) - (4.11) is

$$K_{op}^{c} = \begin{bmatrix} 0.963 & 1.294 \\ 0.002 & 0.476 \\ -2.943 & -2.102 \\ -0.002 & 0.490 \\ 0.035 & -1.279 \end{bmatrix} . \qquad (4.29)$$

From the first 1000 samples of $\{y_i\}$, 250 were used to obtain an estimate of A‡. A value of $q = 3$ was found to be the minimum moving average order to result in a white residual sequence. Using the estimated A, the 1000

---

‡Acknowledgement to M.A. Lauzon is extended here for the use of his program to implement Wilson's algorithm.

samples of $\{y_i\}$ were transformed by (4.3a) to give 999 values for $\{y_i^c\}$. Applying the scheme of Carew and Bélanger to $\{y_i^c\}$, an initial estimate of the gain matrix is found to be

$$\hat{K} = \begin{bmatrix} 0.897 & 1.360 \\ 0.053 & 0.553 \\ -2.909 & -3.620 \\ -0.095 & 0.157 \\ 0.146 & -1.058 \end{bmatrix} \qquad (4.30)$$

The remaining 1000 samples of the transformed observations $\{y_i^c\}$ were then filtered by three methods as described in Section 3.5.

The trace of the error covariance matrix P_c is plotted for each of the three methods against the number of iterations, and is shown in Figure 4.1. It was found that the behaviour of the error covariance matrix in coloured noise was essentially the same as the white noise case in Chapter 3. That is, better results were obtained when the stochastic approximation algorithm was used to track the filter gain. The final values for the three methods (see Section 3.5) are given below in Table 4.1.

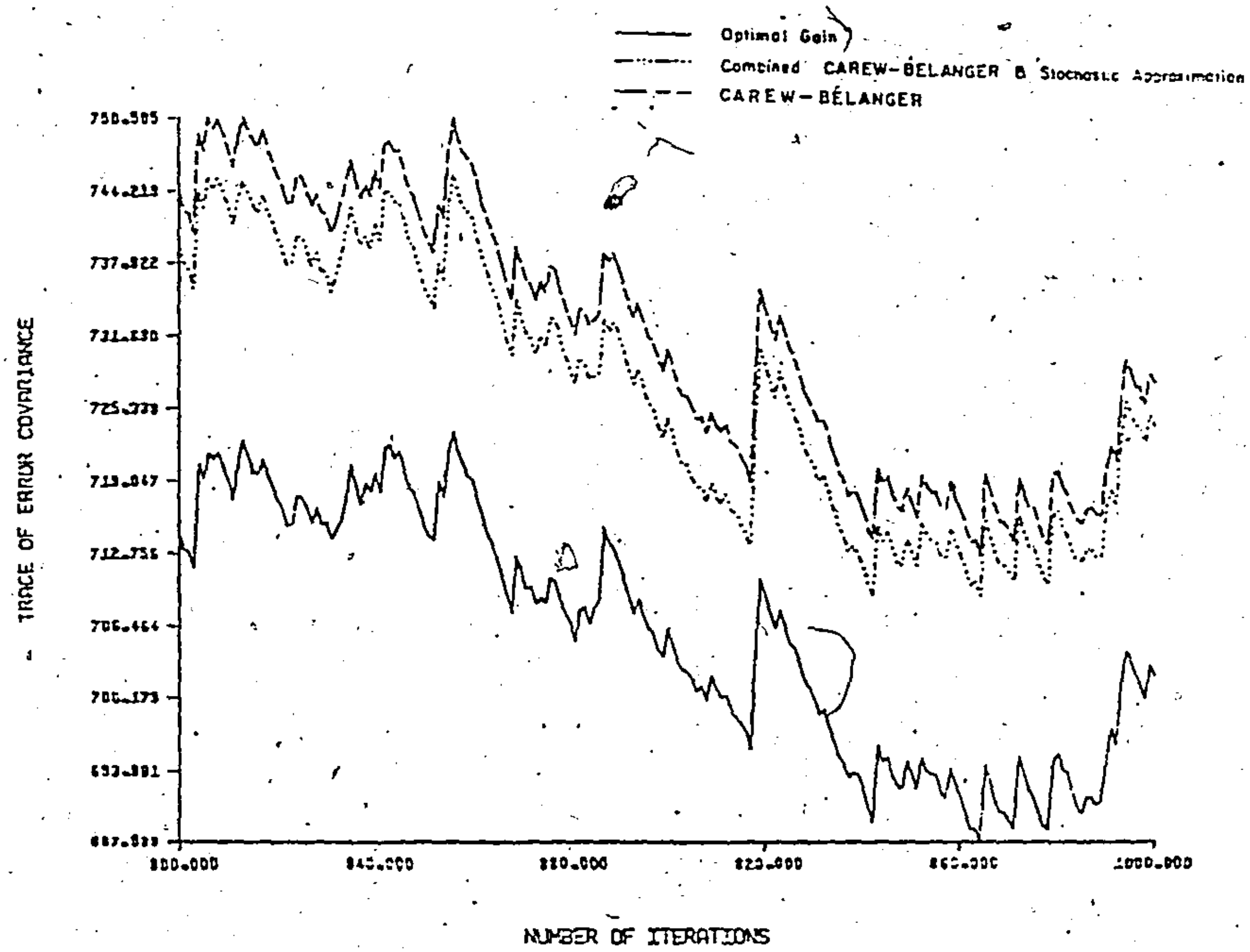

Figure 4.1 Variation of the trace of error covariance matrix for different methods (coloured noise).

TABLE 4.1

COMPARISON OF FILTER PERFORMANCE (COLOURED NOISE)

Method	P_{c11}	P_{c22}	P_{c33}	P_{c44}	P_{c55}	trace P_c
(1)	10.10	0.54	681.14	0.55	9.65	702.02
(2)	14.04	0.89	695.52	1.51	15.10	727.44
(3)	11.82	1.26	697.74	1.80	11.12	723.78

4.6 Discussions

The method proposed in Chapter 3 of combining stochastic approximation with the algorithm of Carew and Bélanger has been extended to the case of coloured observation noise. The scheme proposed in this chapter relaxes the conditions for adaptive state estimation in coloured observation noise (cf.(20)). The scheme makes use of two results derived in the Chapter:

(1) The development of the ARMA model for the system which allows the algorithm of Wilson to be used to estimate the noise transition matrix A:

(2) The invariance of the optimum innovations process which allows the use of the methods described in Chapter 3 to be used, that is the Carew-Bélanger algorithm and the stochastic approximation algorithm.

The results shown in Table 4.1 indicate that the scheme used to estimate the noise transition matrix A is valid because filter 2 (the Carew-Bélanger gain using the estimated A), is different from filter 1 (the optimum gain using the true value of A) by only 3.6%.

As can be seen, when the stochastic approximation algorithm (3.16) is used to track the filter gain (filter 3), an improvement of 17% was achieved over that of filter 2. Once again, the scheme of closed-loop tracking is shown to be more advantageous without much burden on the computation and data storage in the computer.

CHAPTER 5

CONCLUSIONS

The problem of adaptive state estimation which involves the identification of the Kalman gain matrix without *a priori* information on the noise statistics have been considered in this thesis.

For systems with white observation noise, two methods for solving the adaptive state estimation problem have been examined:

(1) The open-loop method of Carew and Bélanger, which is a contraction-mapping algorithm based on a correlation technique:

(2) The closed-loop method of the Robbins and Monro stochastic approximation algorithms, which are basically the stochastic counterpart of the steepest descend algorithms.

A scheme which combines the two algorithms has been proposed in Chapter 3. It essentially used the fast convergence (8 iterations in approximately 13 seconds of computer execution time) of the Carew-Bélanger algorithm and a finite observation sample to identify a Kalman gain within some epsilon neighbourhood of the optimum. Once inside this neighbourhood, it used the closed-loop method of the stochastic approximation algorithm (3.16) to track and improve the Kalman gain. This tracking, (especially useful for noise which are slow time-varying) has been shown in Chapter 3 to give an improvement of approximately 48% over the filter with no tracking. This improvement occuring after 1000 iterations of the stochastic approximation algorithm in which the ten parameters of the gain matrix were updated leads to the confirmation that with tracking, the optimum filter will be realized as more time elapses. This statement has been proved theoretically in Appendix A, where the stochastic approximation algorithm (3.16) has been shown to converge in the mean-square sense to K_{op}.

The adaptive state estimation in coloured observation noise problem has been investigated in Chapter 4. This investigation has led to two results:

(1) The optimum innovations process is invariant for both the white and coloured observation noise system.

(2) Derivation of a mixed autoregressive moving average model to represent the system.

Result 1 has allowed the methods of Chapter 3 to be incorporated into the solution of the coloured noise problem. Result 2 has allowed preliminary analysis of the noise transition matrix using the result of Wilson. A novel approach has been presented for the problem of adaptive state estimation in coloured noise. It involved the estimation of the noise transition matrix and the transformation of the observation process; after which the scheme of Chapter 3 was applied. This formulation as far as the author is aware of, is a first attempt of its kind. Numerical results obtained have shown that the prescribed approach gives approximately a 3.6% increase in the trace of the error covariance matrix for the open-loop estimation scheme of Carew and Belanger. Also, the closed-loop approach was shown to be, once again, a better method since it improved the open-loop filter by approximately 17% without much effort in computation or storage of data. These results confirm the appropriateness of the proposed general approach to solve the adaptive state estimation problem for coloured observation noise.

Throughout this thesis, stochastic approximation has been used for the closed-loop tracking of the filter. The main difficulty, which has yet to be overcome, is the slow rate of convergence. It takes approximately 1000 iterations per perameter to arrive at a relatively good estimate (31). Acceleration techniques for stochastic approximation algorithms have been studied by Kesten (18). It is left for future workers in this area to investigate techniques which will accelerate convergence of stochastic approximation as applied to adaptive state estimation. Thus, the overall efficiency of the closed-loop controller may be enhanced.

The difference in the performance of tracking between the white and coloured observation noise systems is quite large. This is due to the manner in which the noise transition matrix has been estimated. Further analysis might reveal a relationship between the Kalman gain and noise transition matrices which is explicit enough to allow a stochastic approximation algorithm to be used to track the couple in a closed-loop manner.

A large portion of the computation time of Wilson's algorithm is used to calculate derivatives of the conditional residuals. The number of derivatives is dependent on the parameter and observation size of the time series. It is hoped that the present (and the only one to date) algorithm for vector ARMA model estimation can be improved with further research.

APPENDIX A

STOCHASTIC APPROXIMATION - PROOF OF CONVERGENCE

From equations (2.6), (2.8) and (2.9), the filter can be described by the following equation

$$\hat{x}_{i+1|i} = F^{i}\hat{x}_{0|0^-} + \sum_{j=0}^{i-1} F^{j}K\varepsilon_{i-j}. \tag{A.1}$$

Define,

$$m_i(k) \triangleq E[\hat{x}_{i+1|i}\varepsilon'_{i+1}]$$

$$= F^{i}E[\hat{x}_{0|0^-}\varepsilon'_{i+1}] + \sum_{j=0}^{i-1} F^{j}KE[\varepsilon_{i-j}\varepsilon'_{i+1}]. \tag{A.2}$$

From Mehra (21), the autocorrelation functions of the innovations process are

$$E[\varepsilon_{i-j}\varepsilon'_{i+1}] = H[F - KH]^{j}[FPH' - KW], \quad j = 0,1,\ldots \tag{A.3}$$

Since the filter is required to be stable for any reasonable choice of K, then the admissible values are $K \varepsilon \kappa = \{K_i : \rho(F-KH)<1\}$, where $\rho(.)$ denotes the spectral radius of a matrix. Assuming that $\hat{x}_{0|0^-} = x_0 = 0$, then the limiting value of (A.2), using

(A.3) is

$$m(K) \triangleq \lim_{i\to\infty} m_i(K) = \Phi\left[FPH' - KW\right] , \qquad (A.4)$$

where

$$\Phi \triangleq \sum_{j=0}^{\infty} F^j KH\left[F - KH\right]^j . \qquad (A.5)$$

For F asymptotically stable and $K\epsilon\kappa$, the matrix sum (A.5) forms a finite positive definite convergent sum (29). Premultiplying (A.4) by $K_\epsilon' \triangleq (K - K_{op})'$, then

$$K_\epsilon' m(K) = K_\epsilon' \Phi\left[FPH' - KW\right]$$

$$= - K_\epsilon' \Phi K_\epsilon W. \qquad (A.6)$$

Let Φ_{max} and Φ_{min} be the maximum and minimum positive eigenvalues of Φ respectively, thus

$$- \Phi_{max} ||K_\epsilon||^2 W \leq K_\epsilon' m(k) \leq - \Phi_{min}|| K_\epsilon||^2 W. \qquad (A.7)$$

Since Φ is dependent on $K = K_{op} + K_\epsilon$, then Φ_{max} and Φ_{min} are dependent on K_ϵ. This dependence can be removed by defining

$$\Phi^S_{max} \triangleq \operatorname*{Sup}_{K\epsilon\kappa} \Phi_{max} , \qquad (A.8a)$$

$$\Phi^I_{min} \triangleq \operatorname*{Inf}_{K\epsilon\kappa} \Phi_{min} . \qquad (A.8b)$$

Therefore (A.7) can be restated as

$$\alpha||K_\varepsilon||^2 \leq K_\varepsilon' \, \dot{m}(K) \leq \beta||K_\varepsilon||^2 \, , \qquad \text{(A.9)}$$

where $\alpha = -\phi_{max}^{S} W \, ; \, \beta = -\phi_{min}^{I} W,$

and $-\infty < \alpha \leq \beta < 0.$

Thus, assumption A-3.1 is satisfied for all admissible gain $K \varepsilon \kappa$ if the regression term is $\hat{x}_{i|i-1}\varepsilon_i'$. Assunption A-3.2 can be satisfied by choosing $\gamma_i = \frac{1}{i+1}$. Therefore, the stochastic approximation algorithm

$$K_{i+1} = K_i + \frac{1}{i+1} \cdot \frac{\hat{x}_{i|i-1}\varepsilon_i'}{||\hat{x}_{i|i-1}\varepsilon_i'||^2} \qquad \text{(A.10)}$$

converges to K_{op} in the mean-square sense.

APPENDIX B

WILSON'S ALGORITHM (35)

Consider the mixed autoregressive moving average (ARMA) model for a stationary zero-mean multivariate time series $\{\chi_i\}$ of the form

$$\chi_i = \Phi_1\chi_{i-1} + \ldots + \Phi_p\chi_{i-p} + a_i - \Theta_1 a_{i-1} - \ldots - \Theta a_{i-q} . \tag{B.1}$$

where the $\{a_i\}$ are independent and identically distributed vector random variable with zero-mean and a finite covariance matrix D. Both χ_i and a_i are of dimension m.

The unknown parameters are the mxm matrices Φ_k ($k = 1, 2, \ldots, p$), Θ_k ($k = 1, 2, \ldots, q$) and D. The parameters, excluding D, are collectively referred to by the parameter vector $B = \{\Phi_{rs,k}, \theta_{vs,k}\} = \{\beta_1, \ldots, \beta_k\}$.

Define,

$$\Phi(Z^{-1}) \overset{\Delta}{=} I_m - \Phi_1 Z^{-1} - \ldots - \Phi_p Z^{-p} \tag{B.2}$$

$$\Theta(Z^{-1}) \overset{\Delta}{=} I_m - \Theta_1 Z^{-1} - \ldots - \Theta_q Z^{-q}, \tag{B.3}$$

where I_m is an mxm identity matrix and Z^{-1} is a complex variable.

Then the algorithm for estimating β and D can be summarized by the following steps:

(a) Assume some starting values for the constant parameter $\lambda > 0$, and $\hat{\beta} \varepsilon \Omega$, where Ω is a parameter space determined by the condition:

$$\det \Phi (Z^{-1}) \neq 0 \quad \text{for} \quad |Z^{-1}| \geq 1 \, , \qquad \text{(B.4)}$$

$$\det \Theta (Z^{-1}) \neq 0 \quad \text{for} \quad |Z^{-1}| \geq 1 \, . \qquad \text{(B.5)}$$

(b) Set $\hat{D} = D(\hat{\beta}) = \frac{1}{N} \sum_{i=1}^{N} \hat{a}_i \, \hat{a}_i'$

(c) Set $\hat{Q} = \hat{D}^{-1}$

(d) Form a matrix $\hat{A}$ with elements $\hat{A}_{kl} = \frac{1}{N} \sum_{i=1}^{N} (\frac{\partial \hat{a}_i}{\partial \beta_k})' \hat{Q} (\frac{\partial \hat{a}_i}{\partial \beta_l})$, the vector $\hat{g}$ with elements

$\hat{g}_k = \frac{1}{N} \sum_{i=1}^{N} (\frac{\partial \hat{a}_i}{\partial \beta_k})' \hat{Q} \, \hat{a}_i$, and the scaling quantities

$\alpha_k = (\hat{A}_{kk})^{\frac{1}{2}}$.

(e) Construct the scaled matrix $\hat{B}$ with elements $\hat{B}_{kl} = \hat{A}_k / (\alpha_k \alpha_l)$ and the scaled vector $\hat{h}$ with elements $\hat{h}_k = \hat{g}_k / \alpha_k$.

(f) Set the diagonal elements $\hat{B}$ to the value of $1 + \lambda$, and solve the equations $\hat{B}\Pi = h$, and evaluate a new set of parameters
$\hat{\beta}_k = \hat{\beta}_k - (\Pi_k/\alpha_k)$.

(g) Set $\hat{D} = D(\hat{B})$, and test whether the trace $(\hat{D}\hat{Q})<m$.

(i) If this conditions is satisfied, reduce the constraint parameter λ by a predetermined factor ν, return to step c.

(ii) If the condition is not satisfied, increase the constraint parameter λ by ν, and return to step f.

APPENDIX C

PROGRAM LISTING AND USAGE

The subroutines used to simulate the system with white observation noise described in Section 3.5 and the implementation of the Carew-Belanger algorithm are listed in this Appendix.

The calling steps of the subroutine are as follows:

(1) Using SUBROUTINE SYSTEM, a set of observation samples $\{y_i\}$ are generated for a certain triple $\{F,G,H\}$ and, sequence $\{u_i\}$ and $\{v_i\}$:

(2) From SUBROUTINE DRIC , a suboptimum Kalman gain (variable SR5 transposed) gain can be calculated using some assumed values for Q and R:

(3) The suboptimum innovation sequence can be generated using the generated observation, suboptimum gain, and SUBROUTINE INNOV:

(4) The various values for the autocorrelation functions of the suboptimum innovations sequence are the outputs of SUBROUTINE COVINN:

(5) Using the calculated autocorrelation functions and the system matrices, the values of matrix A and the observability matrix B are obtained from

SUBROUTINE MATXA and SUBROUTINE OBSERV respectively.

(6). The SUBROUTINE CARBEL will process the previous calculated values of A, B, $C_0 \triangleq E[\varepsilon_1^* \varepsilon_1^{*\prime}]$, and the suboptimum gain to give an estimate of the optimum Kalman gain.

```
      SUBROUTINE SYSTEM (X,Y,F,G,H,U,V,N,NP,NR,NS,NO,V1,V2,V3,V4,V5,V6)
C
C        THIS SUBROUTINE WILL CALCULATE THE OBSERVATIONS OF SYSTEM BELOW
C
C        X(I+1) = FX(I)+GU(I)
C        Y(I) = HX(I)+V(I)
C
C        WHERE X(I) IS THE STATES OF THE SYSTEM
C              Y(I) IS THE OBSERVATIONS
C              U(I) AND V(I) IS THE PROCESS AND OBSERVATION NOISE RESP.
C
C        X = N X NO
C        Y = NR X NS
C        F = N X N
C        H = NR X N
C        G = N X NP
C        U = NP X NS
C        V = NR X NS
C        NO = NS+1
C        NS = NUMBER OF OBSERVATION SAMPLES
C
C        WORKING VECTORS ARE-
C        V1-V3  N X 1
C        V4 AND V5  NR X 1
C        V6 IS NP X 1
C
C        NOTE THE INITIAL VALUE OF X MUST BE DEFINED.
C
C        EXTERNAL SUBROUTINES REQUIRED-
C                                 MPRD,MADD.
C
      DIMENSION X(N,NO), Y(NR,NS), F(N,N), H(NR,N), G(N,NP), U(NP,NS), V
     1(NR,NS), V1(N), V2(N), V3(N), V4(NR), V5(NR), V6(NP)
      DO 5 K=1,NS
      DO 1 I=1,N
      V1(I)=X(I,K)
    1 CONTINUE
      DO 2 I=1,NP
      V6(I)=U(I,K)
    2 CONTINUE
      DO 3 I=1,NR
      V4(I)=V(I,K)
    3 CONTINUE
      CALL MPRD (F,V1,V2,N,N,1)
      CALL MPRD (G,V6,V3,N,NP,1)
      CALL MADD (V2,V3,V2,N,1)
      DO 4 I=1,N
      X(I,K+1)=V2(I)
    4 CONTINUE
      CALL MPRD (H,V1,V5,NR,N,1)
      CALL MADD (V5,V4,V5,NR,1)
      DO 5 I=1,NR
    5 Y(I,K)=V5(I)
      RETURN
      END
```

```
      SUBROUTINE DRIC (F,G,H,R,Q,P,N,NR,NP,EPS,LIMIT,S1,S2,S3,S4,S5,S6,S
     17,SR1,SR2,SR3,SR4,SR5,SR6,SR7,SP1,SP2,V1,V2)
C
C        THIS SUBROUTINE SOLVES FOR P OF THE DISCRETE RICATTI EQUATION-
C
C           P=FPF(T)-FPH(T)(HPH(T)+R)(INVERSE)HPF(T)+GQG(T)
C
C        WHERE (T) INDICATES TRANSPOSE OF MATRIX.
C        F = N X N
C        G = N X NP
C        H = NR X N
C        R = NR X NR
C        Q = NP X NP
C        P = N X N
C
C        WORKING MATRICES REQUIRED ARE-
C
C        S1 TO S7 ARE N X N
C        SR1 TO SR3 ARE N X NR
C        SR4 IS NR X NR
C        SR5 TO SR7 ARE NR X N
C        SP1 IS NP X N
C        SP2 IS N X NP
C
C        WORKING VECTORS ARE-
C
C        V1 AND V2 ARE 1 X N
C
C
C        EXTERNAL SUBROUTINES REQUIRED ARE-
C                             MTRN,MPRD,MSUB,DLYA,MADD,MINV,NORM.
C
      DIMENSION F(N,N), G(N,NP), H(NR,N), R(NR,NR), Q(NP,NP), P(N,N)
      DIMENSION S1(N,N), S2(N,N), S3(N,N), S4(N,N), S5(N,N), S6(N,N), S7
     1(N,N)
      DIMENSION SP1(NP,N), SP2(N,NP), V1(N), V2(N)
      DIMENSION SR1(N,NR), SR2(N,NR), SR3(N,NR), SR4(NR,NR), SR5(NR,N),
     1SR6(NR,N), SR7(NR,N)
      SUM1=0.
      DO 1 I=1,NR
      DO 1 J=1,N
    1 SR5(I,J)=0.
      DO 2 K=1,LIMIT
      CALL MTRN (H,SR1,NR,N)
      CALL MPRD (SR1,SR5,S3,N,NR,N)
      CALL MTRN (F,S2,N,N)
      CALL MSUB (S2,S3,S1,N,N)
      CALL MTRN (SR5,SR3,NR,N)
      CALL MPRD (SR3,R,SR2,N,NR,NR)
      CALL MPRD (SR2,SR5,S3,N,NR,N)
      CALL MPRD (G,Q,SP2,N,NP,NR)
      CALL MTRN (G,SP1,N,NP)
      CALL MPRD (SP2,SP1,S4,N,NP,N)
      CALL MADD (S3,S4,S3,N,N)
      CALL MTRN (S1,S6,N,N)
      CALL DLYA (S6,S3,P,N,EPS,S4,S5)
      CALL MPRD (H,P,SR7,NR,N,N)
      CALL MPRD (SR7,SR1,SR4,NR,N,NR)
```

```
      CALL MADD (SR4,R,SR6,NR,NR)
      CALL MINV (SR4,NR,TEST,V1,V2)
      IF (TEST.EQ.0.) GO TO 3
      CALL MPRD (SR7,S2,SR6,NR,N,N)
      CALL MPRD (SR4,SR6,SR5,NR,NR,N)
      CALL NORM (P,SUM2,N,N)
      DELTA=ABS(SUM1-SUM2)
      IF (DELTA.LT.EPS) GO TO 4
      SUM1=SUM2
    2 CONTINUE
      GO TO 4
    3 PRINT 5
    4 RETURN
C
    5 FORMAT (////,10X,*NO SOLUTION FOR THE DISCRETE RICATTI EQUATION*,/
     1///)
      END

      SUBROUTINE DLYA (A,Q,P,N,EPS,T1,T2)
C
C        THIS SUBROUTINE SOLVES THE DISCRETE LYAPUNOV EQUATION.
C                                    P = APA(T)+Q
C
C
C        WHERE (T) INDICATES TRANSPOSE
C        A,P,AND Q ARE N X N MATRICES
C
C        T1 TO T2 ARE N X N WORKING MATRICES
C
C        EXTERNAL SUBROUTINES REQUIRED-
C                                     MPRD.
      DIMENSION A(N,N), Q(N,N), P(N,N)
      DIMENSION T1(N,N), T2(N,N)
      DO 1 J=2,N
      M=J-1
      DO 1 I=1,M
    1 Q(J,I)=Q(I,J)
      DO 2 J=1,N
      DO 2 I=1,N
    2 T1(J,I)=A(I,J)
      K=0
      DO 3 I=1,N
      DO 3 J=1,N
    3 P(I,J)=Q(I,J)
      SUM1=0.
    4 CALL MPRD (A,Q,T2,N,N,N)
      CALL MPRD (T2,T1,Q,N,N,N)
      DO 5 I=1,N
      DO 5 J=1,N
    5 P(I,J)=P(I,J)+Q(I,J)
      CALL NORM (P,SUM2,N,N)
      ERROR=ABS(SUM2-SUM1)
      SUM1=SUM2
      K=K+1
      IF (ERROR.GE.EPS) GO TO 4
      IF (K.EQ.1) GO TO 4
      RETURN
```

```
      END

      SUBROUTINE INNOV (Y,H,F,GAIN,XH,GAMMA,N,NR,NS,NO,V1,V2,V3,V4)
C
C        THIS SUBROUTINE WILL CALCULATE THE INNOVATION SEQUENCE FROM
C        THE SAMPLED OBSERVATIONS.
C
C        GAMMA(I) = Y(I)-HXH(I/I-1)
C        XH(I+1/I) = FXH(I/I-1)+GAIN GAMMA(I)
C
C        WHERE NS IS THE NUMBER OF SAMPLED POINTS
C        XH IS THE ESTIMATE OF THE STATES X
C        Y = NR X NS
C        H = NR X N
C        F = N X N
C        GAIN = N X NR
C        GAMMA = NR X NS
C        XH = N X NO
C        NO = NS+1
C
C        WORKING VECTORS ARE-
C        V1 AND V2 ARE N X 1
C        V3 AND V4 ARE NR X1
C
C        EXTERNAL SUBROUTINES REQUIRED ARE-
C                                    MPRD,MSUB,MADD.
C
      DIMENSION Y(NR,NS), H(NR,N), F(N,N), GAIN(N,NR), GAMMA(NR,NS), XH(
     1N,NO), V1(N), V2(N), V3(NR), V4(NR)
      DO 1 I=1,N
      V1(I)=XH(I,1)
 1    CONTINUE
      DO 4 K=1,NS
      DO 2 I=1,NR
      V3(I)=Y(I,K)
 2    CONTINUE
      CALL MPRD (H,V1,V4,NR,N,1)
      CALL MSUB (V3,V4,V4,NR,1)
      DO 3 I=1,NR
      GAMMA(I,K)=V4(I)
 3    CONTINUE
      CALL MPRD (F,V1,V2,N,N,1)
      CALL MPRD (GAIN,V4,V1,N,NR,1)
      CALL MADD (V2,V1,V1,N,1)
      DO 4 I=1,N
 4    XH(I,K+1)=V1(I)
      RETURN
      END

      SUBROUTINE MATXA (C,H,F,GAIN,A,N,NN,NR,NNR,S1,S2,S3,S4,S5,S6,S7)
C
C        C = NR X NR X NN TENSOR
C        H = NR X N MATRIX
C        A = NNR X NR
C        F = N X N
C        GAIN = N X NR
```

```
C     NN = N+1
C
C     WORKING MATRICES-
C     S1 AND S2 ARE N X N
C     S3 TO S5 ARE NR X NR
C     S6 IS NR X N
C     S7 IS NR X NR X N TENSOR
C
C     EXTERNAL SUBROUTINES REQUIRED-
C                               MPRD.
C
      DIMENSION C(NR,NR,NN), H(NR,N), GAIN(N,NR), A(NNR,NR), S1(N,N), S2
     1(N,N), S3(NR,NR), S4(NR,NR), S5(NR,NR), S6(NR,N), S7(NR,NR,N), F(N
     2,N)
      DO 1 I=1,N
      DO 1 J=1,N
      S1(I,J)=0.
    1 S1(I,I)=1.
      DO 3 K=1,N
      CALL MPRD (H,S1,S6,NR,N,N)
      CALL MPRD (S6,GAIN,S3,NR,N,NR)
      CALL MPRD (S1,F,S2,N,N,N)
      DO 2 I=1,N
      DO 2 J=1,N
    2 S1(I,J)=S2(I,J)
      DO 3 I=1,NR
      DO 3 J=1,NR
    3 S7(I,J,K)=S3(I,J)
      DO 4 K=2,NN
      DO 4 I=1,NR
      DO 4 J=1,NR
      IN=(K-2)*NR+I
    4 A(IN,J)=C(I,J,K)
      DO 6 K=1,N
      DO 6 KI=K,N
      KK=KI-K+1
      DO 5 II=1,NR
      DO 5 JJ=1,NR
      S3(II,JJ)=S7(II,JJ,K)
    5 S4(II,JJ)=C(II,JJ,KK)
      CALL MPRD (S3,S4,S5,NR,NR,NR)
      DO 6 I=1,NR
      DO 6 J=1,NR
      IN=(KI-1)*NR+I
    6 A(IN,J)=S5(I,J)+A(IN,J)
      RETURN
      END

      SUBROUTINE OBSERV (H,F,B,N,NR,NNR,S1,S2,S3)
C
C     THIS SUBROUTINE CALCULATES THE OBSERVATION MATRIX OF (F,H).
C     WHERE B = OBSERV(F,H).
C
C     B = NNR X N
C     H = NR X N
C     F = N X N
C
```

```
C        WORKING MATRICES-
C        S1 AND S2 ARE N X N
C        S3 IS NR X N
C
C        EXTERNAL SUBROUTINE REQUIRED-  MPRD.
C
      DIMENSION B(NNR,N), H(NR,N), F(N,N), S1(N,N), S2(N,N), S3(NR,N)
      DO 1 I=1,NR
      DO 1 J=1,N
    1 B(I,J)=H(I,J)
      DO 2 I=1,N
      DO 2 J=1,N
      S1(I,J)=0.
    2 S1(I,I)=1.
      DO 4 K=2,N
      CALL MPRD (S1,F,S2,N,N,N)
      CALL MPRD (H,S2,S3,NR,N,N)
      DO 3 I=1,NR
      DO 3 J=1,N
      IN=(K-1)*NR+I
    3 B(IN,J)=S3(I,J)
      DO 4 I=1,N
      DO 4 J=1,N
    4 S1(I,J)=S2(I,J)
      RETURN
      END

      SUBROUTINE COVINN (C,GAMMA,NR,NS,NN,V1,V2,V3,V4,S1,S2)
C
C        THIS SUBROUTINE WILL CALCULATE THE COVARIANCE OF THE
C        INNOVATION SEQUENCE.
C
C        C(J) = SUM(I=1,NS-J) GAMMA(I+J)GAMMA(I)  /NS
C
C        WHERE C = NR X NR X NN TENSOR
C              GAMMA =  NR X NS MATRIX
C        NN = N+1
C
C        WORKING MATRICES ARE-
C        S1 AND S2 ARE NR X NR
C
C        WORKING VECTORS ARE-
C        V1 TO V4 ARE NR X 1
C
C        EXTERNAL SUBROUTINES REQUIRED ARE-
C                                   MTRN,MPRD,MADD.
C
      DIMENSION C(NR,NR,NN), GAMMA(NR,NS), S1(NR,NR), S2(NR,NR), V1(NR),
     1 V2(NR), V3(NR), V4(NR)
      DO 4 I=1,NN
      NL=NS+1-I
      DO 1 J=1,NR
      DO 1 KK=1,NR
    1 S2(J,KK)=0.
      DO 3 K=1,NL
      DO 2 J=1,NR
    2 V2(J)=GAMMA(J,I+K-1)
```

```
      V3(J)=GAMMA(J,K)
    2 CONTINUE
      CALL MTRN (V3,V4,1,NR)
      CALL MPRD (V2,V4,S1,NR,1,NR)
      CALL MADD (S2,S1,S2,NR,NR)
    3 CONTINUE
      DO 4 K=1,NR
      DO 4 KK=1,NR
    4 C(K,KK,I)=S2(K,KK)/NS
      RETURN
      END

      SUBROUTINE CARBEL (F,H,CO,B,A,GAIN,GAINK,BI,BIA,LIMIT,EPS,N,NR,NNR
     1,S1,S2,S3,S4,S5,S6,S7,S8,S9,S10,S11,V1,V2,V3,V4,V5,V6)
C
C        THIS SUBROUTINE WILL CALCULATE THE ESTIMATE KALMAN GAIN
C        USING THE CAREW AND BELANGER ALGORITHM.
C
C        F IS N X N
C        H IS NR X N
C        CO IS NR X NR
C        B IS NNR X N
C        A IS NNR X NR
C        GAIN AND GAINK ARE N X NR
C        BI IS N X NNR
C        BIA IS N X NR
C
C        WORKING MATRICES
C        S1 TO S4 N X N
C        S5 AND S6 NR X NR
C        S7 AND S8 N X NR
C        S9 IS NR X N
C        S10 IS NNR X NNR
C        S11 IS N X NNR
C
C        WORKING VECTORS
C        V1 AND V2 NNR X 1
C        V3 AND V4 N X 1
C        V5 AND V6 NR X 1
C
C        EXTERNAL SUBROUTINES REQUIRED-
C                              PSEINV,MPRD,MTRN,MSUB,MINV,MADD,NORM,
C
      DIMENSION F(N,N), H(NR,N), CO(NR,NR), B(NNR,N), A(NNR,NR), GAIN(N,
     1NR), GAINK(N,NR), BI(N,NNR), BIA(N,NR)
      DIMENSION S1(N,N), S2(N,N), S3(N,N), S4(N,N), S5(NR,NR), S6(NR,NR)
     1, S7(N,NR), S8(N,NR), S9(NR,N), S10(NNR,NNR), S11(N,NNR)
      DIMENSION V1(NNR), V2(NNR), V3(N), V4(N), V5(NR), V6(NR)
      SUM=0.
      CALL PSEINV (B,NNR,N,BI,S10,S1,S11,V1,V2,V3,V4)
      CALL MPRD (BI,A,BIA,N,NNR,NR)
      DO 1 I=1,N
      DO 1 J=1,N
    1 S1(I,J)=0.
      DO 3 K=1,LIMIT
      CALL MPRD (H,S1,S9,NR,N,N)
      CALL MTRN (H,S7,NR,N)
```

```
      CALL MPRD (S9,S7,S6,NR,N,NR)
      CALL MSUB (CO,S6,S5,NR,NR)
C     *************
      DO 2 I=1,NR
      DO 2 J=1,NR
    2 S6(I,J)=S5(I,J)
      CALL MINV (S6,NR,TEST,V5,V6)
      IF (TEST.EQ.0.) GO TO 5
      CALL MPRD (F,S1,S2,N,N,N)
      CALL MPRD (S2,S7,S8,N,N,NR)
      CALL MSUB (BIA,S8,S8,N,NR)
      CALL MPRD (S8,S6,GAINK,N,NR,NR)
C     *************
      CALL MPRD (GAIN,H,S2,N,NR,N)
      CALL MSUB (F,S2,S2,N,N)
      CALL MPRD (S2,S1,S3,N,N,N)
      CALL MTRN (S2,S4,N,N)
      CALL MPRD (S3,S4,S2,N,N,N)
      CALL MSUB (GAIN,GAINK,S7,N,NR)
      CALL MTRN (S7,S9,N,NR)
      CALL MPRD (S7,S5,S8,N,NR,NR)
      CALL MPRD (S8,S9,S3,N,NR,N)
      CALL MADD (S2,S3,S1,N,N)
      CALL NORM (S1,ANORM,N,N)
      DELTA=ABS(ANORM-SUM)
      IF (DELTA.LT.EPS) GO TO 4
      SUM=ANORM
    3 CONTINUE
    4 PRINT 6, K
      RETURN
    5 PRINT 7
      RETURN
C
C
    6 FORMAT (///,T40,*NUMBER OF ITERATIONS OF THE ALGORITHM=*,I5)
    7 FORMAT (////,40X,*CAREW AND BELANGER ALGORITHM STOPS*,///)
      END

      SUBROUTINE PSEINV (A,N,M,AI,S1,S2,S3,V1,V2,V3,V4)
C
C        THIS SUBROUTINE CALCULATES THE PSEUDOINVERSE
C        OF MATRIX A(N X M).
C
C        IF A IS RANK M
C        A+=(ATA)(INVERSE)AT
C
C        IF A IS RANK N
C        A+=AT(AAT)INVERSE
C
C        WHERE A+ INDICATES PSEUDOINVERSE OF A
C
C        WORKING MATRICES
C        S1 IS N X N
C        S2 IS M X M
C        S3 IS M X N
C
C        WORKING VECTORS
```

```
C         V1 AND V2 ARE N X 1
C         V3 AND V4 ARE M X 1
C
C         EXTERNAL SUBROUTINES REQUIRED ARE-
C                                       MPRD,MINV,
C
      DIMENSION A(N,M), AI(M,N), S1(N,N), S2(M,M), S3(M,N), V1(N), V2(N)
     1, V3(M), V4(M)
      DO 1 I=1,N
      DO 1 J=1,N
    1 S3(J,I)=A(I,J)
      IF (M.GT.N) GO TO 2
      CALL MPRD (S3,A,S2,M,N,M)
      CALL MINV (S2,M,TEST,V3,V4)
      IF (TEST.EQ.0.) GO TO 3
      CALL MPRD (S2,S3,AI,M,M,N)
      RETURN
    2 CALL MPRD (A,S3,S1,N,M,N)
      CALL MINV (S1,N,TEST,V1,V2)
      IF (TEST.EQ.0.) GO TO 3
      CALL MPRD (S3,S1,AI,M,N,N)
      RETURN
    3 PRINT 4
      RETURN
C
C
    4 FORMAT (////,40X,*PSEUDOINVERSE DOES NOT EXIST*,///)
      END

      SUBROUTINE NORM (A,ANORM,NR,NC)
C
C         THIS SUBROUTINE CALCULATES THE NORM OF MATRIX A.
C
C         A IS A NR X NC MATRIX
C
      DIMENSION A(NR,NC)
      SUM=0.
      DO 1 I=1,NR
      DO 1 J=1,NC
    1 SUM=SUM+A(I,J)*A(I,J)
      ANORM=SQRT(SUM)
      RETURN
      END

      SUBROUTINE MSUB (A,B,C,NR,NC)
C
C         THIS SUBROUTINE SUBTRACTS TWO MATRICES AS BELOW-
C                                       C = A-B
C
C         WHERE A,B,AND C ARE NR X NC MATRICES
C
      DIMENSION A(NR,NC), B(NR,NC), C(NR,NC)
      DO 1 I=1,NR
      DO 1 J=1,NC
    1 C(I,J)=A(I,J)-B(I,J)
      RETURN
```

```
      END

      SUBROUTINE MTRN (A,B,NR,NC)
C
C        THIS SUBROUTINE DOES THE OPERATION BELOW-
C                                     B = A(TRANSPOSE)
C
C        WHERE A IS A NR X NC MATRIX
C
      DIMENSION A(NR,NC), B(NC,NR)
      DO 1 I=1,NR
      DO 1 J=1,NC
    1 B(J,I)=A(I,J)
      RETURN
      END

      SUBROUTINE MPRD (A,B,C,NR,NC,N)
C
C        THIS SUBROUTINE MULTIPLIES TWO MATRIX AS BELOW-
C                                     C = AB
C
C        WHERE A = NR X NC MATRIX
C              B = NC X N  MATRIX
C              C = NR X N  MATRIX
C
      DIMENSION A(NR,NC), B(NC,N), C(NR,N)
      DO 1 I=1,NR
      DO 1 J=1,N
      C(I,J)=0.
      DO 1 K=1,NC
    1 C(I,J)=A(I,K)*B(K,J)+C(I,J)
      RETURN
      END

      SUBROUTINE MADD (A,B,C,NR,NC)
C
C        THIS SUBROUTINE ADDS TWO MATRICES AS BELOW-
C                                     C = A+B
C
C        WHERE A,B,AND C ARE NR X NC MATRICES
C
      DIMENSION A(NR,NC), B(NR,NC), C(NR,NC)
      DO 1 I=1,NR
      DO 1 J=1,NC
    1 C(I,J)=A(I,J)+B(I,J)
      RETURN
      END
```

BIBLIOGRAPHY

(1) H.B. Aasnaes and T. Kailath, "An innovations approach to least-squares estimation - Part VII : Some applications of vector autoregressive - moving average models", IEEE Trans. on Automatic Control, vol. AC - 18, Dec. 1973, pp. 601-607.

(2) K.J. Astrom, Introduction to Stochastic Control Theory. New York: Academic Press, 1970.

(3) G.E.P. Box and G.M. Jenkins, Time Series Analysis, Forecasting and Control. San Francisco : Holden - Day, 1970.

(4) A.E. Bryson, Jr. and L.J. Henrikson, "Estimation using simple data containing sequentially correlated noise", J. Spacecraft and Rocket, vol. 5, June 1968, pp. 662 - 665.

(5) A.E. Bryson, Jr. and D.E. Johansen, "Linear filtering for time-varying systems using measurements containing coloured noise", IEEE Trans. on Automatic Control, vol. AC - 10, Jan. 1965, pp. 4 - 10.

(6) B. Carew and P.R. Bélanger, "Identification of optimum filter steady-state gain for systems with unknown noise covariances", IEEE Trans. on Automatic Control, Dec. 1973, pp. 582-587.

(7) H. Cox, "Estimation of State-variables for noisy dynamic systems", Ph.D. dissertation, Dept. of Elect. Eng., MIT, Cambridge, Mass., May 1963.

(8) A. Dvoretzky, "On stochastic approximation", Proc. 3rd Berkeley Symp. on Math., Stat. and Prob., 1956, pp. 39-55.

(9) V.N. Fadeeva, Computational Methods in Linear Algebra. New York: Dover Publications, Inc., 1959.

(10) R.L.T. Hampton and D.G. Schultz, "Stochastic algorithm for self-adaptive filtering and prediction", Proc. Int. IEEE Conf. in Systems, Computers and Networks, 1971, pp. 463-467.

(11) H.Heffes, "The effect of erroneous models on the Kalman filter response", IEEE Trans. on Automatic Control, vol. AC - 11, July 1966, pp. 541 - 543.

(12) C.G. Hilborn and D.G. Lainiotis, "Optimum estimation in the presence of unknown parameters", IEEE Trans. on Systems, Science, and Cybernetics, vol. SSC-5, Jan. 1969, pp. 38-43.

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

(14) T. Kailath, "An innovations approach to least-squares estimation - Part I: Linear filtering in additive white noise", IEEE Trans. on Automatic Control, vol. AC-13, Dec. 1968, pp. 646-655.

(15) R.E. Kalman, "New methods and results in linear prediction and filtering theory", RIAS Tech. Rept. 61-1, Baltimore, Md., 1961.

(16) R.E. Kalman and R. Bucy, "New results in linear filtering and prediction theory", ASME J. Basic Engineering, 83, March 1961, pp. 95-108.

(17) J. Keifer and J. Wolfowitz, "Statistical estimation of the maximum of a regression function", Ann. Math. Statistics, vol. 23, 1952, pp. 462-466.

(18) H. Kesten, "Accelerated stochastic approximation method", Ann. Math. Statistics, vol. 29, 1958, pp. 41-59.

(19) A.N. Kolmogorov,, "Interpolation and extrapolation of stationary random sequences", Bull, Acad. Sci., USSR, Ser. Math., 5 (1941); transl: Memorandum RM-3090-PR, Rand Corp., Santa Monica, Calif., Apr. 1962.

(20) N.K. Loh and A. Hauser, "Identification of optimal filter gain with application to coloured noise", Sixth Annual Pittsburgh Conf. on Modeling and Simulation, Apr. 1975.

(21) R.K. Mehra, "On the identification of variances and adaptive Kalman filtering", IEEE Trans. On Automatic Control, vol. AC-15, Apr. 1970, pp. 175-184.

(22) R.K. Mehra, "Approaches to adaptive filtering", IEEE Trans. on Automatic Control, vol. AC-17, Oct. 1972, pp. 693-698.

(23) J.M. Mendel and K.S. Fu, Eds., Adaptive Learning and Pattern Recognition Systems. New York: Academic Press, 1970.

(24) A.W. Naylor and G.R. Sell, Linear Operator Theory. New York: Holt, Rinehart and Winston, Inc., 1971.

(25) H. Robbins and S. Monro, "A stochastic approximation method", Ann. Math. Statistics, vol. 22, 1951, pp. 400 - 407.

(26) A.P. Sage and J.L. Melsa, System Identification, New York: Academic Press, 1971.

(27) D.J. Sakrison, "Stochastic approximation: A recursive method for solving regression problems", Advances in Communications Systems, vol. 2, 1966, pp. 51 - 106.

(28) G.N. Saridis, "Stochastic approximation method for identification and control - a survey", IEEE Trans. on Automatic Control, Vol. AC-19, Dec. 1974, pp. 798-809.

(29) L.L. Scharf and D.L. Alspach, "On stochastic approximation and adaptive Kalman filter", Proc. IEEE Decision and Control Conf., 1972, pp. 253-257.

(30) J.C. Shellenbarger, "Estimation of covariance parameters for an adaptive Kalman filter", Proc. Nat. Electronics Conf., vol. 22, 1966, pp. 698-702.

(31) N.K. Sinha and S.K. Muhkerjee, "An adaptive scheme for Kalman filtering", Proc. IFAC Symp. on Stochastic Control, 1974, pp. 363-368.

(32) N.K. Sinha and A. Tom, "State estimation for systems with unknown noise covariances", Sixth Annual Pittsburgh Conf. on Modeling and Simulation, 1975.

(33) N.K. Sinha and A. Tom, "Identification of optimum filter gain with unknown white and coloured noise", to be presented at 18th Midwest Symposium on Circuits and System, (Montréal, Que.), Aug. 1975. Also available as SOC Int. Rept. SOC-95, McMaster Univ., Hamilton, Ont., June 1975.

(34) N. Wiener, The Extrapolation, Interpolation, and Smoothing of Stationary Time Series, New York: John Wiley and Sons., Inc., 1949.

(35) G. Tunnicliffe Wilson, "The estimation of parameters in multivariable time series models", J. Royal Statistical Society, vol. 35, 1973, pp. 76-85.

(36) L.A. Zadeh and C.A. Desoer, Linear System Theory, New York: McGraw-Hill, 1963.