THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

A NON-LINEAR ADAPTIVE FILTER AND ITS APPLICATION TO A CONTROLLED STOCHASTIC CHEMICAL PROCESS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

BY

KENNETH EUGENE SANDERS

Norman, Oklahoma

1973

A NON-LINEAR ADAPTIVE FILTER AND ITS APPLICATION TO A CONTROLLED STOCHASTIC CHEMICAL PROCESS

APPROVED BY

DISSERTATION COMMITTEE

ABSTRACT

The equations for a non-linear adaptive filter are developed and applied to a continuously stirred tank reactor that is modelled by a fourth order system of non-linear equations. An algorithm for estimation of the plant and observation noise covariance matrices and of the state vector is obtained by maximizing a likelihood function. The state estimator equations turn out to be identically the extended Kalman filter and are coupled to the covariance matrix estimator equations.

How the algorithm optimally combines information regarding observational data with information about the plant model is demonstrated for several systems including the controlled chemical reactor. The effects of model error, steady state estimation, and response of the covariance matrix for the state estimate error are investigated. The response of the algorithm is compared to that of the Bayes maximum a posteriori estimator which was developed by Sage and Husa. The results indicate when use of the algorithm developed herein is preferable.

ACKNOWLEDGMENTS

I would like to express my appreciation and gratitude to the individuals and organizations which made the completion of this work possible:

Dr. M. L. McGuire, Adjunct Associate Professor, for his guidance and assistance throughout graduate school.

Professor Yoshikazu Sasaki for his assitance and interest which has been a stimulus to this research.

Dr. Cedomir Sliepcevich, George Lynn Cross Research Professor, Associate Professor James A. Payne, and Dr. Harry West for serving on my advisory committee.

Associate Professor James Christensen for the availability of the PDP 8 computer at the CEMS Department.

Dr. Harry West for his time and guidance throughout graduate school.

Dr. Harold Wesson and Dr. Harry West for the availability of the CDC 6600 computer through University Engineers, Inc.

Dr. Bill Woodard and Dr. Dan Clifton for their enlightening and helpful discussions.

Dr. Dan Rimmer for his interesting discussions and preparation of computational results on the calcomp plotter.

Mr. Jerry Bordello for special assistance and availability of the IBM 360/50 computer at Merrick Computer Center.

Ms. Bobbie Everidge for preparation of this manuscript.

Chemical Engineering and Materials Science Department through which financial support for this work was administered.

I would especially like to express my appreciation and gratitude to my family for their encouragement throughout school, to my parents for their encouragement and understanding, to my wife for her patience and understanding, and to Melissa.

TABLE OF CONTENTS

Table of Contents--Continued

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS, Continued

LIST OF ILLUSTRATIONS, Continued

CHAPTER I

INTRODUCTION

I.1 Optimal Estimation

As time has passed, numerous techniques for combining knowledge of a model with available observational data have been proposed anew, but the basic problem of how to combine these two sources of information remains fundamentally the same.

In 1809, the German mathematician Karl F. Gauss discussed in his classic treatise "Theoria Motus" (7) the problem of determining the orbital elements of a celestial body from available measurement data. He proposed that the problem could be properly undertaken only when an approximate knowledge (the model) of the orbit had been already obtained, which afterwards was to be corrected so as to satisfy all the observations in the most accurate manner possible. The technique that Gauss suggested for obtaining the approximations (or estimates) of the unknown quantities has come to be known as the method of least squares.

There have been many contributions to the estimation problem since Gauss, but the subsequent major one was that of Norbert Wiener.

The pioneering work of Wiener (35) on the problem of linear extrapolation, interpolation and smoothing has received much attention since the appearance of the original "yellow peril," as his theory of time series analysis which was written on yellow tablet paper was called by his colleagues as they struggled through it. A contemporary major impetus to the development of estimation theory was contributed by R. E. Kalman (13, 14). Much work has been conducted since then in the time domain for the development of solutions to the linear prediction, filtering and smoothing problems, the combined problem being referred to as the estimation problem. So much emphasis has been placed on the analysis of these problems in fact that an entire issue, over 300 pages in length, of the IEEE Transactions (10) has been dedicated to R. E. Kalman on the topic "Linear Quadratic Gaussian Estimation and Control."

The driving force, even from initial work, has been the cognizance of a need for optimal estimation techniques by a wide diversity of scientific fields. Industries most actively applying optimal estimation techniques to satisfy this need historically have been those of communication, aerospace, defense, and information. Currently, applications are being proposed and applied in the power (16) and chemical industries (23), both of which require complex control schemes with consequent estimation techniques.

The need for estimation in the chemical process industry has several sources. Not all state variables may be

measured to be used as feedback for process control. Consequently, filtering provides an estimate of inaccessible variables plus the degree of reliability of the estimate.

Also required for control is a model of the process. Should operating conditions change unexpectedly, the model parameters may undergo significant changes as well. Any control scheme depending upon this consequently inaccurate model may thus become ineffective. Using an on-line filtering scheme overcomes this problem by a periodic updating of time varying model parameters. The periodic updating uses a minimum of time and hardware due to the sequential nature of filtering as opposed to an off-line simultaneous model identification scheme.

A final example of filtering as applied to the chemical industry is its use for processing noisy plant data. The filter optimally weights knowledge of plant dynamics and plant data to obtain a combined estimate of the actual state. The existence of nonstationary noise as well as nonstationary plant inputs is no limitation. Both types of random inputs may be estimated concomitantly to improve filtering with a consequent improvement in control.

Although applications to dynamic processes have been emphasized, and will be throughout this work, application of filtering to steady state plant operations and observational data is equally amenable. Whether on-line or off-line, the sequential nature of filtering is advantageous because of its

efficiency and economy as well as flexibility (4). Consequently, the sequential nature can be exploited for use in analyzing large amounts of data using algebraic models such as those encountered for describing physical property relations.

I.2 Kalman Filtering

The linear filtering problem was solves in the time domain by Kalman and Bucy (13, 14) using the concept of "state variables." The statistical dynamical model of the plant and the statistical model of the measurement model comprise the system. These models are defined by the discrete, linear, vector difference equation

$$x(k+1) = \phi(k+1,k)\ x(k) + \Gamma(k+1,k)\ w(k)$$

where x is an n-vector, the state; w is a p-vector, the disturbance; k = 0, 1, ... is the discrete time index; ϕ is an n x n matrix, the state transition matrix; Γ is an n x p matrix, the disturbance transition matrix; and

$$z(k+1) = H(k)\ x(k) + v(k)$$

where z is an m-vector, the measurement; H is an m x n matrix, the measurement matrix; and v is an m-vector, the measurement error.

The random noise sequences w(k) and v(k) along with x(0) are assumed to be Gaussian with the following properties. They are zero mean, E[x(0)] = 0, E[w(k)] = 0, E[v(k)] = 0

where "E" denotes the expected value (i.e., ensemble average) operator. They are uncorrelated,

$$E[w(j)\ v^T(k)] = 0$$
$$E[w(k)\ x^T(0)] = 0$$
$$E[v(k)\ x^T(0)] = 0.$$

Finally, the stationary random sequences are "white" noise with covariance matrices

$$E[w(j)\ w^T(k)] = Q(k)\ \delta_{jk}$$
$$E[v(j)\ v^T(k)] = R(k)\ \delta_{jk}$$

where $Q(k)$ is an n x n matrix and $R(k)$ is an m x m matrix. δ_{jk} is the Kronecker delta. Both $Q(k)$ and $R(k)$ are assumed to be positive definite.

An estimate of the state $x(k)$, based upon knowledge of the measurements $Z(j)$ which denotes the set $\{z(1), z(2), \ldots, z(j)\}$, is denoted by $x(k|j)$. For the case $k > j$, $x(k|j)$ is the predicted estimate; for $k = j$, $x(k|j)$ is the filtered estimate; and for $k < j$, $x(k|j)$ is the smoothed estimate.

For estimation with minimum mean square error, Kalman showed that the optimal estimate is given by

$$\hat{x}(k|j) = E[x(k)|z(j)]$$

which denotes the conditional expected value of $x(k)$, given the knowledge of $z(j)$. If complete knowledge of the given

system, including knowledge of ϕ, Γ, H, Q and R, is available, then the Kalman filter algorithm is given by

$$\hat{x}(k|k) = \hat{x}(k|k-1) + W(k)\ \tilde{z}(k|k-1)$$

$$\hat{x}(k|k-1) = \phi(k,k-1)\ \hat{x}(k-1|k-1)$$

$$\tilde{z}(k|k-1) = z(k) - H(k)\ \hat{x}(k|k-1)$$

$$W(k) = P(k|k-1)\ H^T(k)[H(k)P(k|k-1)\ H^T(k) + R(k)]^{-1}$$

$$P(k|k-1) = \phi(k,k-1)\ P(k-1|k-1)\ \phi^T(k,k-1) + \Gamma(k,k-1)\ Q(k-1)\ \Gamma^T(k,k-1)$$

$$P(k|k) = [I - W(k)\ H(k)]\ P(k|k-1)$$

Initial values are

$$\hat{x}(0|0) \triangleq E[x(0)] = 0$$

$$P(0|0) \triangleq E[x(0)\ x^T(0)]$$

The optimal weighting matrix for assigning relative weights to new observations is W(k). The "innovation sequence" (12) is $\tilde{z}(k|k-1)$. The predicted estimate of x(k) is denoted by $\hat{x}(k|k-1)$ with its associated covariance matrix of prediction error, $P(k|k-1)$. $P(k|k-1)$ is referred to as the a priori covariance matrix and is a measure of the reliability of the prediction based on one less measurement than the filtered estimate. The covariance matrix of the error in the filtered estimate is $P(k|k)$ and is referred to as the a posteriori covariance matrix. A block diagram for information flow during processing is shown in Figure 1.1.

The resulting Kalman filter is a recursive algorithm. It processes the measurements as they become available in real

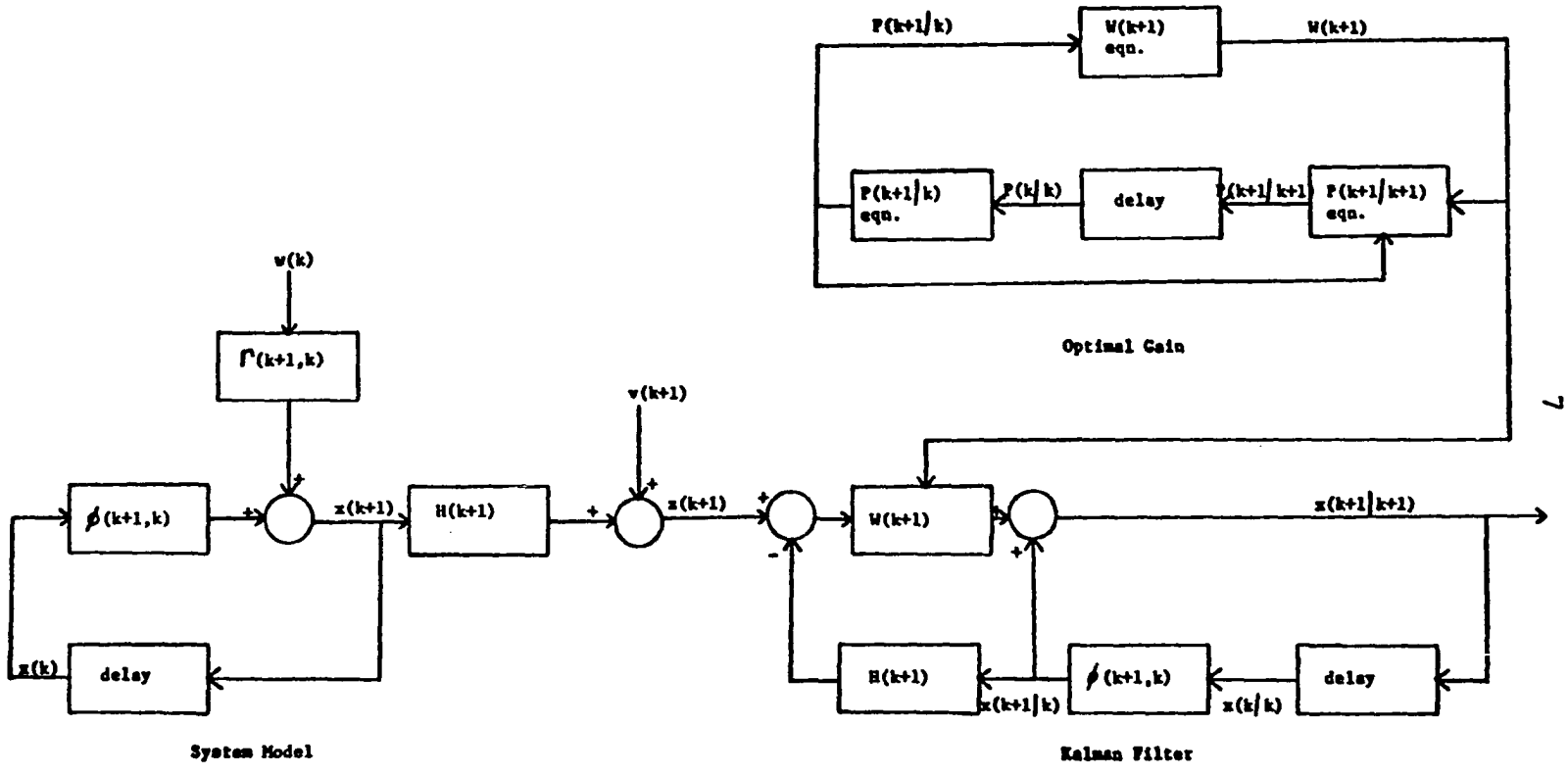


Figure 1.1. Information Flow Diagram.

time. Consequently, the growing memory problem is avoided and the algorithm is easily implemented on the digital computer for on-line estimation.

I.3 Adaptive Filtering

As useful as the Kalman filter is, a disadvantage is the requirement that the Gaussian statistical models of the plant dynamics and the measurements be completely known in order for them to be used in the filter algorithm. This is unfortunate for many chemical process applications because only portions of the model transition matrices $\phi(k+1,k)$ and $\Gamma(k+1,k)$, the measurement matrix $H(k+1)$, or input covariance matrices $Q(k)$ and $R(k)$ may be known. Consequently, parameters in these matrices as well as the state variables need to be estimated.

For the purposes of this work, adaptive estimation refers to the estimation of the statistical moment matrices $Q(k)$ and $R(k)$. The estimation by state augmentation of system parameters in matrices $\phi(k+1,k)$, $\Gamma(k+1,k)$ and $H(k+1)$ for chemical processes has been presented elsewhere (34). Estimation of the means of the inputs may be handled similarly.

There are several possible approaches to adaptively estimating the unknown matrices Q and R. The most obvious approach is to de ive an estimator that is insensitive to the unknown portion of the model. This approach includes the least squares estimator (17) and the minimax estimator (5).

A second approach is to apply the Kalman filter using a conservative guess for the unknown portion of the model. The consequence of such an empirical approach is a suboptimal estimator since the true model is not used. Nishimura (24) developed covariance and sensitivity analysis equations that determine the suboptimality of this approach. Once the effects of erroneous values of Q and R are determined then the initial guess may be adjusted to achieve less suboptimal estimation. This procedure is then repeated.

A truly adaptive method however is the direct estimation of the unknown matrices. Sequential equations using measurements as they become available, much like the state estimation equations, are obtained by concomitantly satisfying the optimization criterion with respect to the identifiable matrices as well as the state vector. The resulting system of equations then comprises the adaptive filter.

One of the advantages of using a truly adaptive filter is that a less mean square error of estimation should be obtained than the previous suboptimal approaches discussed. This additional estimating capability is achieved however at the expense of requiring implementation of a more complex algorithm, which may be prohibitive. The extent of additional complexity and/or suboptimality depends upon the procedure selected. Previously proposed adaptive estimation algorithms are reviewed in the next section, after which a new procedure is proposed.

CHAPTER II

LITERATURE REVIEW OF ADAPTIVE FILTERING

II.1 Adaptive Filtering

In any control system, the reliability of much of the knowledge of the process dynamics is often questionable. Of particular interest is the re-evaluation of a priori information about random plant inputs and observation noise. Each of these random quantities is specified by the first two statistical moments--the mean vectors and the covariance matrices Q and R. Estimation of the means may be handled by state augmentation. However, for estimation of the covariances, it is desirable to jointly obtain an algorithm which is incorporated into the overall estimation scheme and which adaptively updates the a priori knowledge of Q and R. Several of the significant approaches recently undertaken to adaptively estimate Q and R will be briefly reviewed.

Magill (18, 19) proposed an adaptive estimator, which is optimal in the mean square error sense, for Gaussian linear systems. The state estimate is a weighted sum of elemental Kalman filter estimates. The weighting coefficients, which comprise the adaptive feature since they change with the measurement sequence, are generated by operating on the

incoming measurements. The algorithm for generating the coefficients is obtained from Bayes rule for probability density functions.

The number of elemental Kalman filters required for the above approach was reduced by Sengbush and Lainiotis (29) by using an iterative technique. However, the consequent technique no longer is practical for on-line applications.

Mehra (21, 22) assumed the Kalman filter and linear system to have reached steady state. If the filter is operating suboptimally by using assumed values of Q and R, the autocorrelation function of the measurement innovation process (12) is generated, assuming stationary noise. This function is then utilized to obtain asymptotically unbiased and consistent estimates of Q and R. The updated Q and R are then used in the Kalman filter and the process repeated. Unfortunately, a large amount of data is required to ensure steady state and to generate the autocorrelation function.

Another source of information for estimating Q was proposed by Jazwinsky (11). He examined the predicted measurement residuals

$$\tilde{z}(k+j|k) \triangleq z(k+j) - H(k+j)\ \phi(k+j,k)\ \hat{x}(k|k),$$

and required the residuals to be consistent with their statistical properties,

$$E[\tilde{z}(k+j|k)\ \tilde{z}^T(k+j|k)] = \tilde{z}(k+j|k)\ \tilde{z}^T(k+j|k).$$

The resulting algorithm for Q was used with the Kalman filter for linear systems to provide improved estimates of the state.

Smith (32) developed an approximate Bayesian approach to estimate a diagonal R matrix for measurement noise assumed to be Gaussian-Inverted-Gamma distributed. He used Bayes' rule for the density function

$$p[x(k),R|Z(k)] = p[Z(k)|x(k),R[\ p[x(k),R]/p[Z(k)].$$

Distribution functions were substituted on the right side of the above equation. After regrouping terms, the approximate conditional means estimator of the state and parameters of the Gaussian-Inverted-Gamma distribution function for R were selected by inspection. The final algorithm consists of the Kalman filter for linear systems and a set of equations to sequentially estimate the distribution scaling parameter for R.

Kashyap (15) obtained a maximum likelihood estimator applicable for estimating Q and R. He separately treated the problem in two parts. First, the system model parameters θ and R may be estimated by maximizing the likelihood function $\ell n\ [p(Z(k)|R,\theta)]$ with a conjugate gradient search scheme. Second, estimate Q and R directly from maximum likelihood estimates of θ. These estimates are valid only for large amounts of data, stationary linear systems, and steady state.

Shellenbarger (31) developed two approximate maximum likelihood estimators. By maximizing $p[z(k)|Q,R]$ with

respect to Q or R, estimators were obtained independent of the filter algorithm for state estimation. He showed the asymptotic stability and unbiased properties of these estimators. On the other hand, the estimators obtained by maximizing $p[z(k)|Z(k-1),Q,R]$ at each stage k with respect to Q or R require algorithms consisting of interdependent equations for state estimation and estimation of Q or R. With the specification of certain restrictions, Q and R could be estimated concomitantly with the state. In general, the estimators for Q and R were on-line recursive and applicable to linear systems.

These methods were improved upon (30) by selecting the R and Q estimates according to a least squares fit of the set of products $\{\tilde{z}(k|k-1)\ \tilde{z}^T(k|k-1),\ k = 1, 2,...\}$ to their expected values. The Kalman filter algorithm and the equations for Q and R are again mutually interdependent however.

Finally, Sage and Husa (24) began with a Bayesian approach by attempting to maximize the a posteriori probability density function $P[X(k),Q,R|Z_{(k)}]$ with respect to X(k), Q and R at each stage k. Estimators are developed for a linear system and uniform a priori distributions for constant Q and R. For optimization, the necessary conditions are in the form of a nonlinear two-point boundary value problem which is solved via discrete invariant imbedding. However, the optimal solution which is in terms of the smoothed estimates $\{\hat{x}(j|k),\ j = 0, 1, ..., k\}$ is dismissed in preference for a suboptimal solution using only those terms generated by a Kalman filter.

Suboptimality is due to the approximation

$$\hat{x}(j|j-1) \cong \hat{x}(j|k-1), \; j = 1, 2, \ldots, k-1$$

which replaces the fixed interval smoothed estimate with the fixed single lag smoothed estimate.

CHAPTER III

MAXIMUM LIKELIHOOD ADAPTIVE FILTERING

III.1 General Problem Formulation

Consider the class of systems which can be represented by the nonlinear vector difference equation

$$x(k+1) = f[x(k),u(k)] + w(k) \qquad (1)$$

where $f[x(k),u(k)]$ is an n-vector function of the state variable $x(k)$ and a deterministic variable $u(k)$.

The observation equations can be represented by

$$z(k+1) = h[x(k+1)] + v(k+1) \qquad (2)$$

where $h[x(k+1)]$ is an m-vector function of the state $x(k+1)$. These equations hold for $k = 0, 1, 2,...$ where k is the discrete time index; x is the n-dimensional state vector; w is the n-dimensional stochastic input vector; z is the m-dimensional measurement vector; and v is the m-dimensional stochastic measurement error. Both f and h may also be functions of time varying parameters.

The stochastic sequences $\{w(k), k = 0,1,2...\}$ and $\{v(k), k = 1,2,...\}$ are uncorrelated with one another and are Gaussian with means and covariance matrices

$$E\{w(k)\} = q(k) \qquad E\{v(k)\} = r(k)$$

$$\mathrm{Cov}\{w(k),w(j)\} \equiv E\{[w(k)-q(k)][w(j)-q(j)]^T\} = Q(k)\delta_{jk}$$

$$\mathrm{Cov}\{v(k),v(j)\} \equiv E\{[v(k)-r(k)][v(j)-r(j)]^T\} = R(k)\delta_{jk}$$

$$\mathrm{Cov}\{w(k),v(j)\} \equiv E\{[w(k)-q(k)][v(j)-r(j)]^T\} = 0$$

for all discrete times $j,k \geq 0$. E denotes the expected value operator and δ_{jk} denotes the Kronecker delta. Q(k) and R(k) are the real positive definite covariance matrices of dimension n x n and m x m respectively.

The initial state x(0) is a Gaussian vector which is independent of {w(k), k=0,1,2...} and {v(k), k=1,2...} and has a positive definite n x n covariance matrix and mean,

$$\mathrm{Cov}\{x(0), x(0)\} \equiv \mathrm{Var}\{x(0)\} \equiv E\{[x(0)-\bar{x}(0)][x(0)-\bar{x}(0)]^T\} = P(0)$$

$$E\{x(0)\} = \bar{x}(0).$$

Our goal is to obtain sequentially an estimate of the state x(k), k > 0, given the sequence of measurements {z(k), k=1,2,...}. Concomitantly, we desire a sequential estimate of the stochastic inputs, w(k) and v(k), whether these are random inputs or unknown constants. To complete the adaptive estimation, we require sequential estimates of Q(k) and R(k). The resulting estimate of x(k) is denoted by $\hat{x}(k|k)$.

III.2 Solution Formulation

In particular, x(k k) denotes the estimate of x(k), given all the observations up to and including the point in time k, z(1), z(2),...,z(k). Z(k) denotes the data set {z(i), i=1,2,...,k}.

Using marginal maximum likelihood estimation, the function $L[x(k),Z(k)] = p[x(k)|Z(k)]$ has been maximized by Rauch, Tung and Striebel (26) for linear systems and by Meditch (20) for nonlinear systems. For adaptive estimation we would like to maximize the likelihood function

$$L[x(k),Q,R,Z(k)] = p[x(k),Q,R|Z(k)]$$

which is equivalent to the joint conditional probability density function. Maximization is carried out with respect to $x(k)$, Q and R such as to yield the extended Kalman (3) filtered estimate $\hat{x}(k|k)$ adapted to the estimates $\hat{Q}$ and $\hat{R}$. The means q and r likewise may be estimated. However, with no loss of generality, these are not considered here since estimates of q and r are available by using the extended Kalman filter with state augmentation.

Details of the following operations are available in the Appendices. Upon utilization of the probability density function identities, the present form of the likelihood function may be altered to a more amenable form. Specifically, by applying the definition of the conditional density thrice consecutively, one obtains

$$L[x(k),Q,R,Z(k)] = p[z(k)|x(k),Q,R] \cdot p[x(k),Q,R|Z(k-1)] \cdot p[Z(k-1)] \div p[Z(k)]$$

It is assumed that the state vector and the second moments of the random input and measurement error are statistically independent. Furthermore, it is assumed that the density functions of the input and measurement error statistics are uniform. The consequent likelihood function to be maximized reduces to

$$L = p[z(k)|x(k),Q,R]p[x(k)|Z(k-1)]p(Q)p(R) \div p[z(k)|Z(k-1)]$$

as opposed to the function maximized by Sage and Husa (27),

$$L = p[Z(k)|X(k),Q,R]\ p[X(k)]p(Q)p(R)$$

which yielded a maximum a posteriori estimate.

Each density is evaluated using Appendices D and E. The final non-Gaussian density function is

$$L = (2\pi)^{-(2m+n)/2}|R|^{-1/2}|P(k|k-1)|^{-1/2}|Y(k|k-1)|^{-1/2} \cdot \mathrm{Exp}[-1/2\ \|z(k)-\bar{z}(k)\|^2_{R^{-1}} - 1/2\ \|x(k)-\hat{x}(k|k-1)\|^2_{P^{-1}(k|k-1)} - 1/2\ \|z(k)-\hat{z}(k|k-1)\|^2_{Y^{-1}(k|k-1)}]$$

The recursive equations for the single stage prediction covariance matrix and estimate are derived in Appendix F as

$$\hat{x}(k|k-1) = f[\hat{x}(k-1|k-1)]$$

$$P(k|k-1) = f_x[\hat{x}(k-1|k-1)]\ P(k-1|k-1)\ f_x^T[\hat{x}(k-1|k-1)] + Q$$

The conditional mean of the observation z(k) given x(k), is

$$\overline{z}(k) = h[x(k)]$$

Finally, the conditional mean and covariance of the observation z(k), given Z(k-1), are

$$\hat{z}(k|k-1) = h[\hat{x}(k|k-1)]$$

and $$Y(k|k-1) = h_x[\hat{x}(k|k-1)]\ P(k|k-1)\ h_x^T[\hat{x}(k|k-1)] + R$$

Although the combined density function is non-Gaussian, we regard it as approximately Gaussian by having retained only the first term of the asymptotic expansion of a Gram-Charlier or Edgeworth series (2).

It follows that the joint maximum likelihood filtered estimate of x(k), and as many of the statistical moments Q and R as desired, may be obtained by maximizing L[x(k),Q,R, Z(k)] with respect to x(k), Q and R.

Setting the gradient of L with respect to the set of variables S

$$S \triangleq \{x(k),\ Q,\ R\}$$

equal to zero, yields the necessary conditions for maximum likelihood estimation,

$$\frac{\partial L}{\partial S} = 0$$

Evaluated at $\hat{S} = \{\hat{x}(k|k), \hat{Q}, \hat{R}\}$, one obtains a set of non-linear simultaneous equations, the details of which are presented in Appendix C. The state estimator is

$$\hat{x}(k|k) = \hat{x}(k|k-1) + W(k)\{z(k) - h[\hat{x}(k|k-1)]\}$$

which is identical to the extended Kalman filter previously obtained (3) and applied (34). The optimal filter gain matrix is defined by

$$W(k) = P(k|k-1)\ h_x[\hat{x}(k|k-1)] \cdot \{\hat{R} + h_x[\hat{x}(k|k-1)]\ P(k|k-1)\ h_x^T[\hat{x}(k|k-1)]\}^{-1}$$

The moment estimator equations are

$$\begin{aligned} 0 = {} & P^{-1}(k|k-1) + h_x^T[\hat{x}(k|k-1)]\ Y^{-1}(k|k-1)\ h_x[\hat{x}(k|k-1)] \\ & - P^{-1}(k|k-1)[\hat{x}(k|k) - \hat{x}(k|k-1)][\hat{x}(k|k) - \hat{x}(k|k-1)]^T \\ & \cdot P^{-1}(k|k-1) \\ & - h_x^T[\hat{x}(k|k-1)]\ Y^{-1}(k|k-1)[z(k) - \hat{z}(k|k-1)] \\ & \cdot [z(k) - \hat{z}(k|k-1)]^T \\ & \cdot Y^{-1}(k|k-1)\ h_x[\hat{x}(k|k-1) \end{aligned}$$

and,

$$\begin{aligned} 0 = {} & \hat{R}^{-1} + Y^{-1}(k|k-1) - \hat{R}^{-1}\{z(k) - h[\hat{x}(k|k)]\} \\ & \cdot \{z(k) - h[\hat{x}(k|k)]\}^T R^{-1} \\ & - Y^{-1}(k|k-1)[z(k) - \hat{z}(k|k-1)] \\ & \cdot [z(k) - \hat{z}(k|k-1)]^T\ Y^{-1}(k|k-1) \end{aligned}$$

The expressions for $\hat{x}(k|k-1)$, $\hat{z}(k|k-1)$, $P(k|k-1)$ and $Y(k|k-1)$ previously were given. The symmetry of $P(k|k-1)$ and $Y(k|k-1)$

has been used in deriving the above equations. Note that $P(k|k-1)$ and $Y(k|k-1)$ are implicit in $\hat{Q}$ and $\hat{R}$. Therefore a recursive solution of these equations is required. It has been suggested (20) that Newton's method be used to recursively solve a similar set of nonlinear equations for a non-adaptive filter.

The vectors $f[x(k)]$ and $h[x(k)]$ are approximated by the first two terms of a Taylor series expansion about $\hat{x}(k|k)$ and $\hat{x}(k|k-1)$ respectively.

$$f[x(k)] \cong f[\hat{x}(k|k)] + f_x[\hat{x}(k|k)][x(k) - \hat{x}(k|k)]$$
$$h[x(k)] \cong h[\hat{x}(k|k-1)] + h_x[\hat{x}(k|k-1)][x(k) - \hat{x}(k|k-1)]$$

$f_x[\hat{x}(k|k)]$ is the n x n Jacobian matrix of $f[x(k)]$ evaluated at $\hat{x}(k|k)$. $h_x[\hat{x}(k|k-1)]$ is the m x n Jacobian matrix of $h[x(k)]$ evaluated at $\hat{x}(k|k-1)$.

A measure of reliability of the filtered estimate $\hat{x}(k|k)$ is the filter error covariance matrix $P(k|k)$. This recursive equation has been developed from the filter error expression $[x(k) - \hat{x}(k|k)]$ in Appendix F.

$$P(k|k) = \{I - W(k)\ h_x[\hat{x}(k|k-1)]\}\ P(k|k-1)$$

where $W(k)$ and $P(k|k-1)$ are previously defined. The I is the identity matrix.

The sufficient condition for the joint maximum likelihood estimate to provide a maximum of the objective function L is

$$\left.\frac{\partial^2 L}{\partial x^2(k)}\right]_{x(k)\to\hat{x}(k|k)} < 0$$

It is shown in Appendix C that $\hat{x}(k|k)$ is a maximum if $P^{-1}(k|k-1) - g_x[x(k),R,z(k)]$ is positive definite where $g_x[x(k),R,z(k)]$ is the n x n Jacobian matrix of $h_x^T R^{-1}[z(k)-h]$, evaluated at $\hat{x}(k|k)$.

CHAPTER IV

ADAPTIVE FILTERING OF CHEMICAL PROCESSES

IV.1 A First Order System

1. The Model

In order to illustrate the mechanics of the Kalman filter and the utility of adaptively estimating the covariance matrices for the plant noise and measurement noise, consider the following simple case. The first order model is,

$$x(k+1) = \phi(k+1,k)\, x(k) + \Gamma(k+1,k)\, w(k)$$

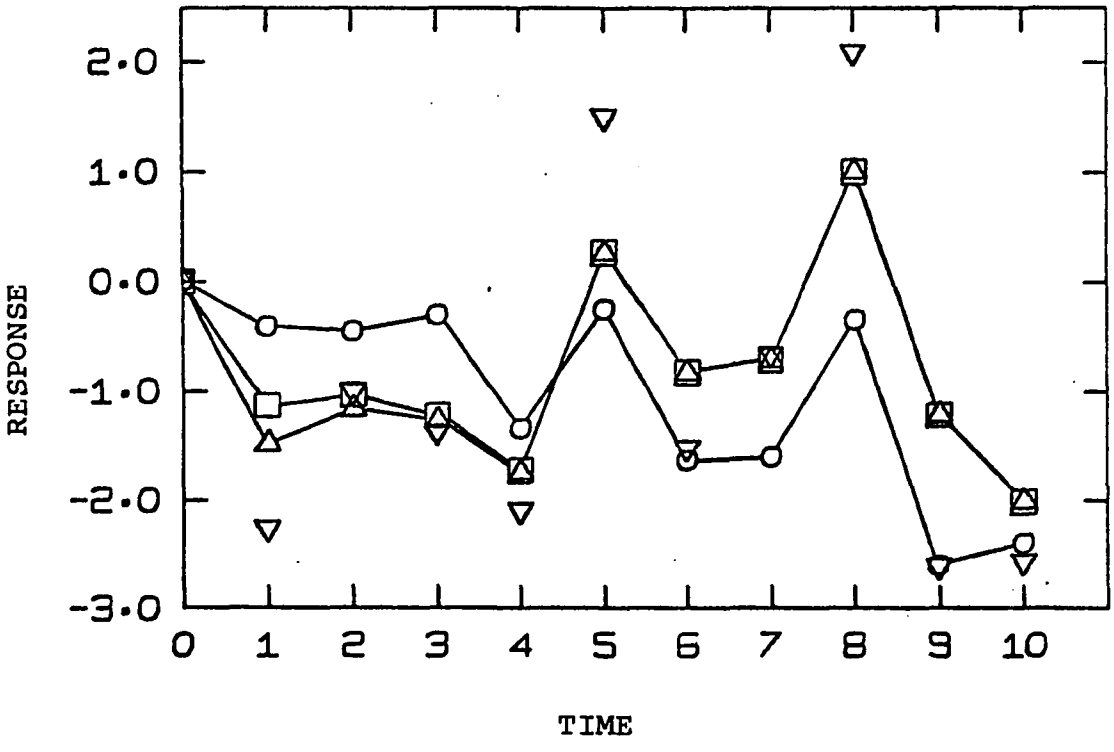
$$x(0) = 0,\ \phi(k+1,k) = 0.95,\ \Gamma(k+1,k) = 1$$

$$z(k) = H(k)\, x(k) + v(k),\ H(k) = 1$$

with $\hat{x}(0|0) = \hat{x}(0)$ and $P(0|0) = P(0)$

where $w(k)$ and $v(k)$ are zero mean independent white Gaussian random variables.

For known noise statistics, R and Q, and for initial conditions, $\hat{x}(0)$ and $P(0)$, Figure 4.1 illustrates the response of the filter to observations, $z(k)$, and for the model, $x(k+1)$. The estimated value $\hat{x}(k|k)$ generally lies between the observed value and the modelled value and represents a smoothed value by weighting $x(k)$ and $z(k)$ with the optimal gain $W(k)$.



$x(k)$ = ○

$z(k)$ = ▽

$P(0) = 0$, $\hat{x}(k|k)$ = △

$P(0) = 1$, $\hat{x}(k|k)$ = □

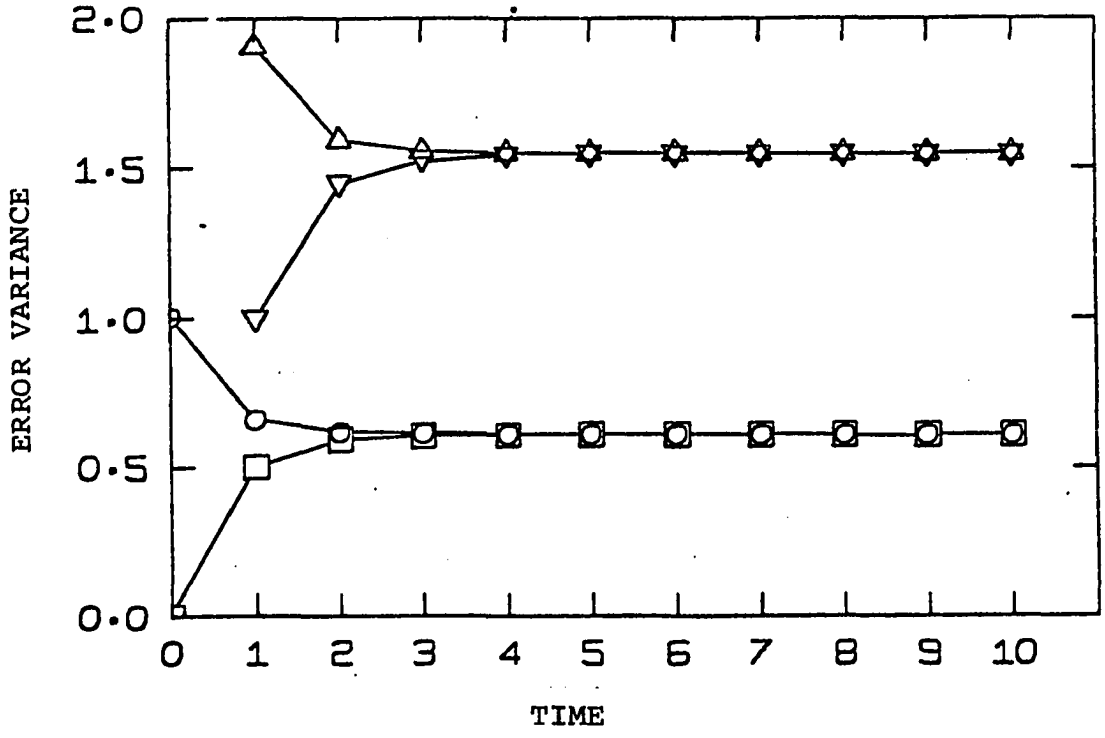
Figure 4.1. Response for First Order Model, $R = 1$, $Q = 1$.

A priori values of $\hat{x}(0)$ and $P(0)$ do have an effect on the filtered estimate normally. Figure 4.1 also illustrates how fast the filter recovers from assuming the initial estimation error variance to have a value of zero or one. Within four steps, the estimate recovers completely and remains identical thereafter whether $P(0)$ is zero or one. The effects of the value of $x(0)$ are more obvious, but equally damped out by the filtering algorithm.

Figure 4.2 illustrates the effect of the value of $P(0)$ on the filtered and predicted error variances for the filtered estimate. For an optimistic initial guess of $P(0) = 0$, $P(k|k)$ reaches the steady state value in four estimation intervals. The same is true for a conservative guess, $P(0) = 1$. The effects of a priori $P(0)$ on the predicted variance, $P(k|k-1)$, are also shown, $P(k|k-1)$ is the estimated error variance for the predicted estimate, $\hat{x}(k|k-1)$. Intuitively, the predicted error variance is expected to be larger than the filtered error variance for the estimate of the state $x(k)$, since $P(k|k)$ has the advantage of being based on one more recent observation than does $P(k|k-1)$.

2. Limiting Cases

For this model, the state transition matrix, $\phi(k+1,k)$, has a scalar value of 0.95. The disturbance transition matrix, $\Gamma(k+1,k)$, has a scalar value of 1. Throughout this work, $\Gamma(k+1,k)$ is assumed to be the identity matrix. This



P(0) = 1, P(k|k-1) = △

P(0) = 0, P(k|k-1) = ▽

P(0) = 1, P(k|k) = ○

P(0) = 0, P(k|k) = □

Figure 4.2. Filtered and Predicted Error Variance for First Order Model.

simplification imposes no restrictions to the technique, as can be seen for this model. By assuming the identity matrix, the plant noise covariance matrix, Q, must be scaled by a factor of Γ^2, however. The measurement matrix, H(k), also has a scalar value of 1 for this model.

The optimal filter equations are

$$\hat{x}(k|k) = \phi\hat{x}(k-1|k-1) + W(k)[z(k) - \phi\hat{x}(k-1|k-1)]$$

$$P(k|k-1) = \phi^2 P(k-1|k-1) + Q$$

$$W(k) = P(k|k-1)[P(k|k-1) + R]^{-1}$$

Combining,

$$W(k) = \frac{\phi^2 P(k-1|k-1) + Q}{\phi^2 P(k-1|k-1) + Q + R}$$

$$P(k|k) = [1 - W(k)] P(k|k-1)$$

Combining,

$$P(k|k) = \frac{R[\phi^2 P(k-1|k-1) + Q]}{\phi^2 P(k-1|k-1) + Q + R}$$

The equation for $P(k|k-1)$ indicates that $P(k|k-1) \geq Q$ since $P(k-1|k-1) \geq 0$. Therefore, the performance limit of the prediction accuracy is determined by the value of Q. Secondly, the optimal filter gain as calculated from the above equation for W(k) is limited as $0 \leq W(k) \leq 1$, except for the special case $P(k|k) = Q = R = 0$. Thirdly, from the equations for $P(k|k)$ and W(k), it can be seen that $P(k|k) = R\ W(k)$; therefore, the filtering error covariance is limited to the values, $0 \leq P(k|k) \leq R$.

The steady state behavior of the error covariance can be determined by setting $P(k|k) = P(k-1|k-1) = \bar{P}$ and then solving for $\bar{P}$,

$$\bar{P} = \frac{-(-R\phi^2 + R + Q) \pm \sqrt{(-R\phi^2 + R + Q)^2 + 4RQ\phi^2}}{2\phi^2}$$

Since P is a variance, only the positive root is valid. For the first order model with $\phi = 0.95$, $R = 1$, $Q = 1$, then $\bar{P} = 0.58$, the predicted error covariance is 1.52, and the optimal gain is 0.60 (see Figure 4.2). Finally, at steady state the filtered estimate is obtained by optimally weighting a new observation with the previous prediction as follows

$$\hat{x}(k|k) = 0.6\ z(k) + 0.4\ \hat{x}(k|k-1)$$

In the limiting case for $R = 0$ and nonzero Q, $\bar{P}$ is zero, W(k) becomes 1, and $P(k|k-1)$ becomes equal to Q. Therefore, the filter algorithm closely follows the observations to the extent that the optimal estimate becomes

$$\hat{x}(k|k) = z(k)$$

For the other limiting case $Q = 0$ and nonzero R,

$$\bar{P} = \frac{R\phi^2\bar{P}}{\phi^2\bar{P} + R}$$

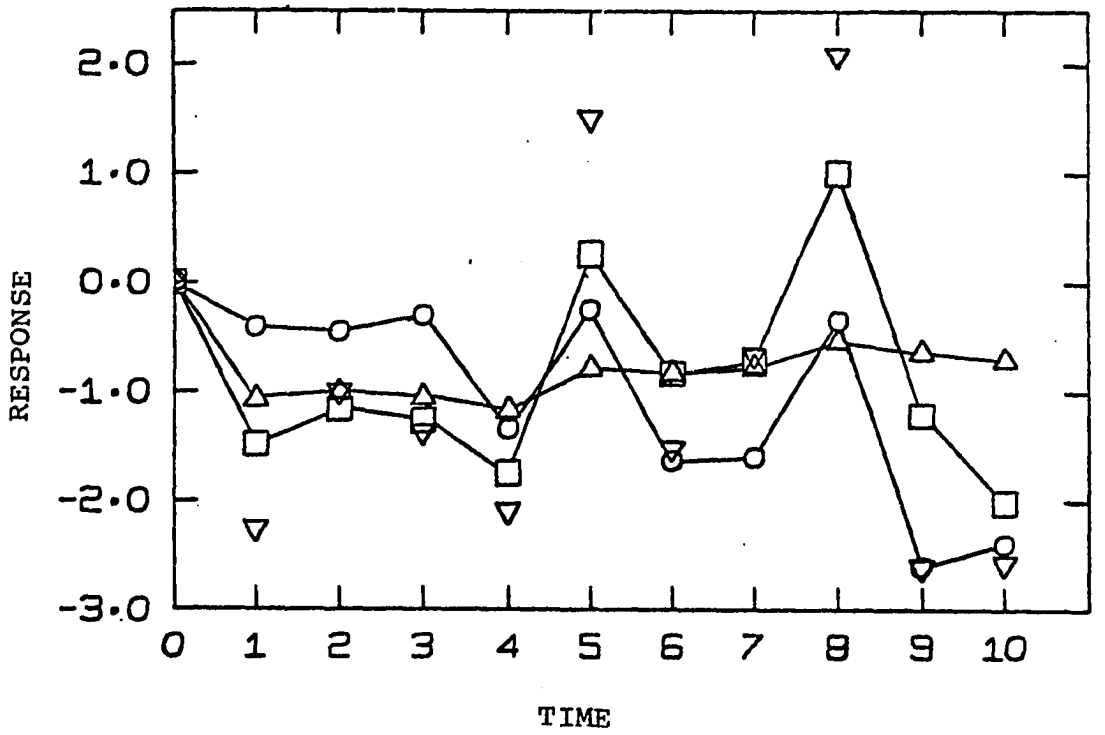
Since $\bar{P} \geq 0$ and $\phi^2 < 1$ for this example, the only value allowed is $\bar{P} = 0$; also $P(k|k-1)$ and W(k) become zero.

Therefore the filter algorithm ignores the observations and closely follows the plant model,

$$\hat{x}(k|k) = \hat{x}(k|k-1)$$

3. Modelling Error

Modelling error occurs in several ways. One common type is inherent error in the coefficients of a deterministic model. As constants, the coefficients can be handled by augmenting the state vector and then updating the value each time the state vector is estimated from a new observation. Biases and trends similarly can be handled. On the other hand, model error may occur when one assumes a deterministic model but a stationary random component exists which is not accounted for. Figure 4.3 illustrates the case where R and Q actually have values of 1, but it was assumed that there was no plant noise (i.e., Q = 0). The result is filter divergence. Instead of following information from the observations, the filtered estimate, $\hat{x}(k|k)$, diverges to a nearly constant value of -0.5 after 20 time increments. The reason for this response becomes apparent from the previous equations. As a result of assuming Q = 0 when actually Q = 1, the filtered error covariance matrix is erroneously calculated to be less than when it would be if it were known that Q = 1. In particular, the erroneous steady state value is 0 instead of 0.58 (see Figure 4.4). Even as the estimate degenerates, the covariance calculation appears satisfactory because the algorithm relies upon



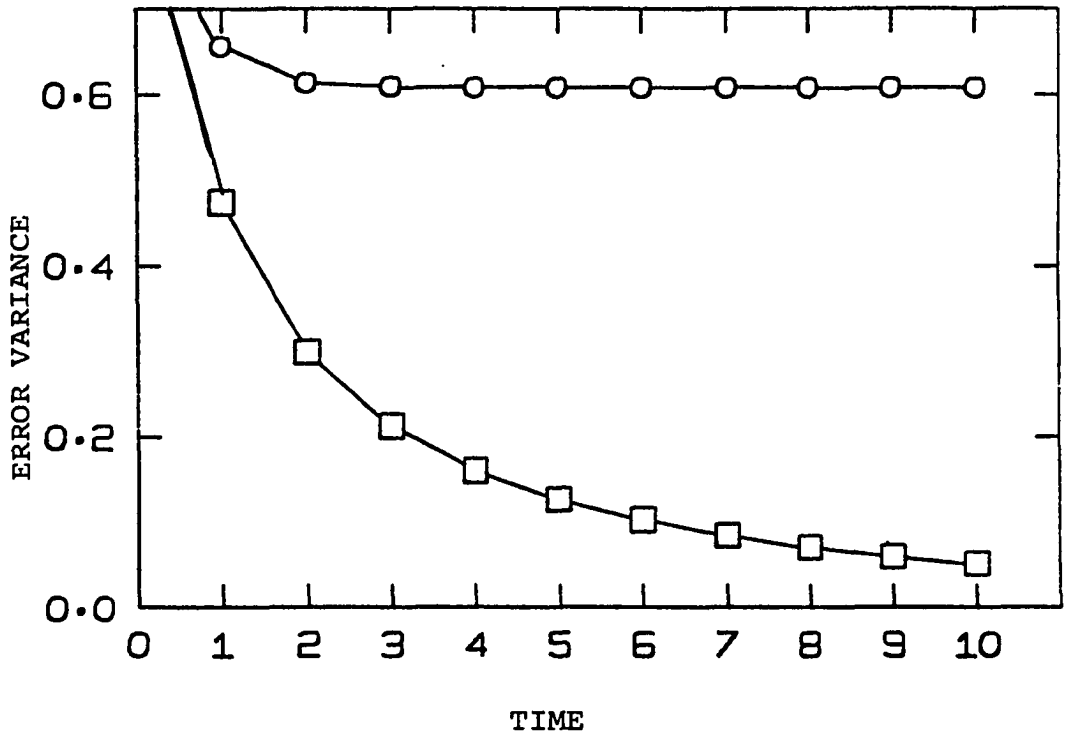
$x(k)$ = ○

$z(k)$ = ▽

$P(0) = 1$, $x(k|k)$ = □, $R = 1$, $Q = 1$

$P(0) = 1$, $x(k|k)$ = △, $R = 1$, $Q = 0$

Figure 4.3. Response for First Order Model with Model Error.



P(k|k) = ○ , R = 1, Q = 1

P(k|k) = □ , R = 1, Q = 0

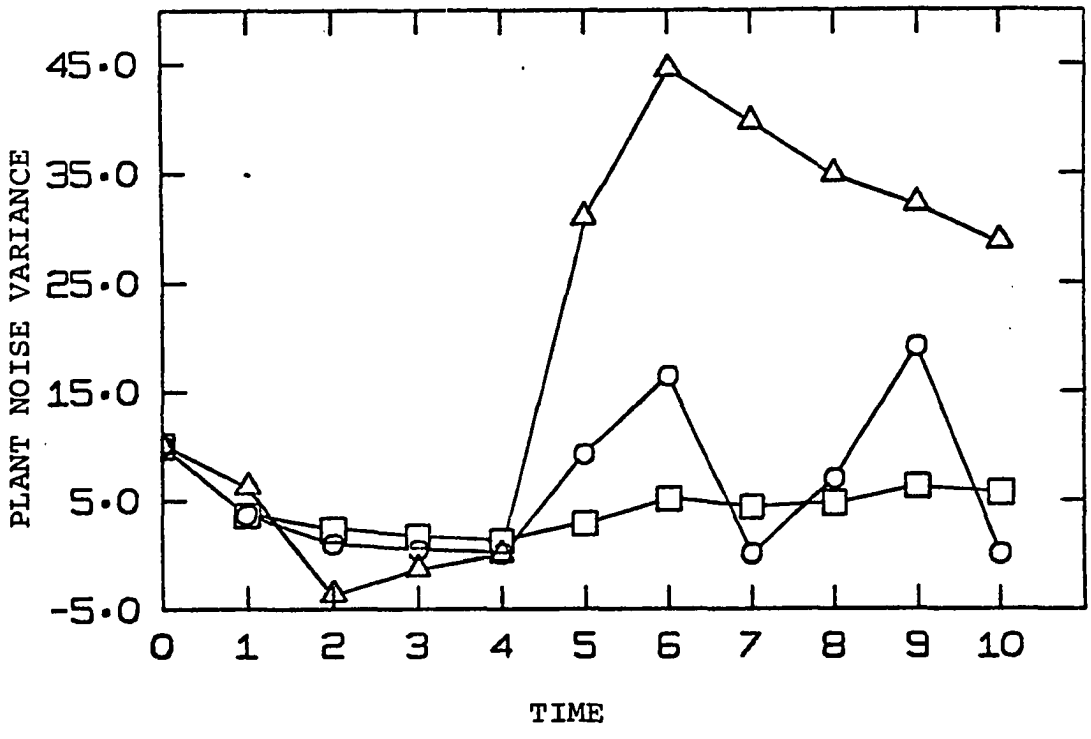
Figure 4.4. Filtered Error Variance for First Order Model with Model Error, P(0) = 1.

the assumption that the model is known exactly a priori. As a result, W(k) becomes zero and eventually uncouples the observations from the filtered estimate completely as the error covariance approaches zero. The observations are ignored and $\hat{x}(k|k) = \hat{x}(k|k-1)$ with the result that $\hat{x}(k|k)$ diverges from the "true" estimate. This example points out the importance of assuming that knowledge of the plant model is available (including random components as well as deterministic components).

4. Adaptive Estimation of Plant and Measurement Noise Covariances

Filter divergence due to the above reason can be avoided by not necessitating the assumption of knowing the covariances Q and R, but by adaptively estimating the covariances along with other state variables beginning with an initial guess.

The adaptive filter algorithm derived in Chapter 3 was applied to the scalar system in this section. Figure 4.5 compares the results from using that algorithm with the results from using an algorithm derived by Sage and Husa (27). Three curves are shown in Figure 4.5. The curve with the largest oscillation is calculated by using the Sage, Husa method for suboptimal Bayes estimation. The suboptimality arises in this method due to the replacement of the fixed interval smoothed estimate, $\hat{Q}(j|k-1)$, with the predicted estimate, $\hat{Q}(j|j-1)$. The second, less oscillatory, curve is



○ = $\hat{Q}(k|k)$, estimated (this work)

□ = $\hat{Q}(k|k)$, smoothed (this work)

△ = $\hat{Q}(k|k)$, estimated (Sage, Husa)

Figure 4.5. Adaptive Estimation of Plant Noise Variance, Q. $P(0) = 0$, $R = 1$, $Q = 1$, $\hat{R}(0|-1) = 10$, $\hat{Q}(0|0) = 10$.

calculated using the method of maximum likelihood estimation derived in this work. The least oscillatory curve in the figure is a smoothed function of the second curve.

It should be pointed out that throughout this work, the algorithms for estimating the second moments of the plant and measurement noise were derived and used. But the algorithms for the corresponding first moments were not derived nor used. The reason for this is that the filtered estimate of the mean for each noise can be obtained by proper structuring of the mathematical model of the system, application of state augmentation, and then using the state estimation filter algorithm. In this way the non-zero mean noise is handled as a zero mean noise term plus a constant bias, which is estimated along with the state vector. It also should be noted that the noise covariance matrices enter into the pre-exponential factor of the likelihood function, whereas the noise means do not.

Figure 4.5 illustrates that in general, the estimate of Q from this work is less oscillagory than the estimate using the Sage, Husa algorithm, and the algorithm derived in this work more rapidly tracks the value of Q. The smoothed estimate, $\bar{Q}$, has been obtained by using an exponential filter which combines the instantaneous estimate with the previous smoothed value. The following relation was used,

$$\bar{Q}(k|k) = [(k-1)\ \bar{Q}(k-1|k-1) + \hat{Q}(k|k)]/k$$

where the filter constant, α, is assigned a value between zero and one.

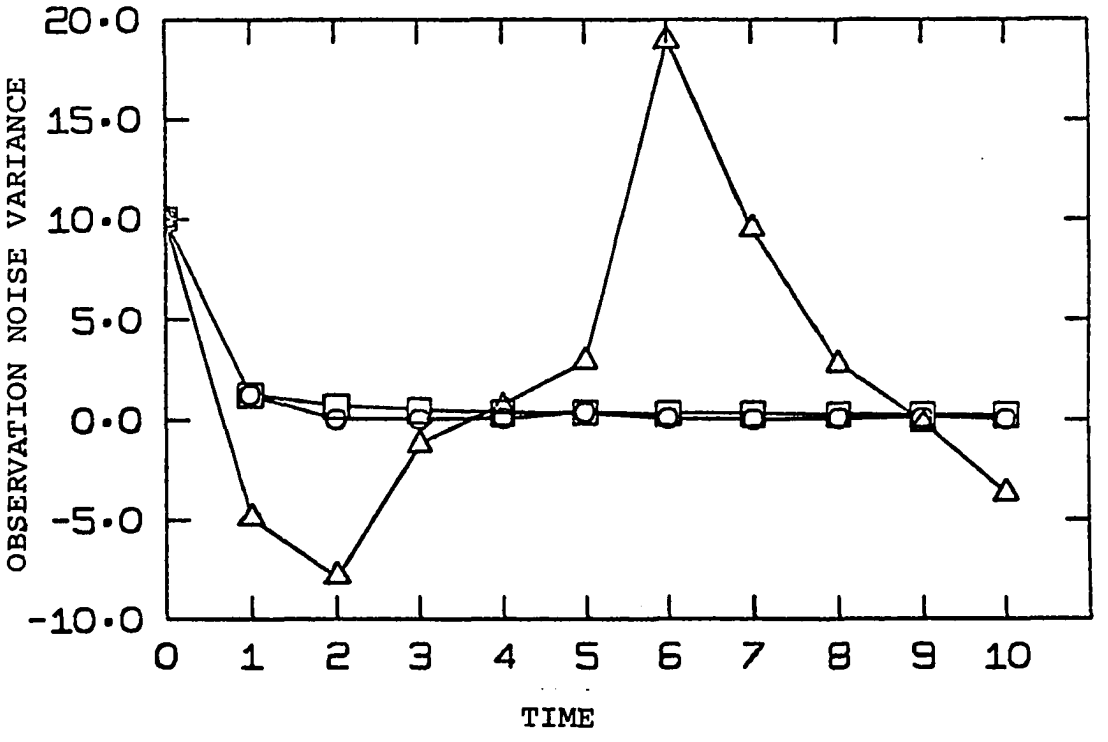
Although often requiring a large number of time increments for convergence (27) (100-5000 depending on the desired final accuracy of estimation), the Sage, Husa filter converges ofter a lengthy period of oscillation. The method proposed herein converges more rapidly, but after about 100 increments the two estimates become essentially equivalent for this example.

Figure 4.6 illustrates similar results for calculating the measurement noise covariance. The Sage, Husa estimate is more oscillatory than the method of this work, and the smoothed estimate is nearly equivalent to the instantaneous estimate of this work. The smoothing algorithm used was

$$\bar{R}(k|k) = [(k-1)\ \bar{R}(k-1|k-1) + \hat{R}(k|k)]/k$$

The main point of comparison of the two discussed methods is the rate of convergence. After 100 increments, the Sage, Husa estimate was still converging to the true value of R, but very slowly.

Although the estimation of R and W were demonstrated separately for the simple model, they can of course be estimated simultaneously using either the derived algorithm or the Sage, Husa algorithm. However, with both inexactly known simultaneously, more time is necessary for convergence.



○ = R(k+1|k), estimated (this work)

□ = R(k+1|k), smoothed (this work)

△ = R(k+1|k), estimated (Sage, Husa)

Figure 4.6. Adaptive Estimation of Observation Noise Variance R. $P(0) = 0$, $R = 1$, $Q = 1$, $\hat{R}(0|-1) = 10$, $\hat{Q}(0|0) = 10$.

IV.2 A Second Order System

1. The Model

For all but the simplest cases, the adaptive filter algorithm cannot be solved directly. For vector models, an iterative method of solution is necessary since the equations of the algorithm are so interrelated and cannot be uncoupled as the equations of the Kalman filter were in the previous section to analyze the steady state behavior of the algorithm. At each estimation time when a new observation is available to update the previous estimate, use of an iterative technique is necessary to solve the adaptive filter equations.

A technique which was found to be applicable and was used to obtain adaptive estimates was the successive approximation method of Wegstein (33). This method does not require use of any derivatives, which could be very complex for a set of matrix equations. The Wegstein method is based on a projection technique and is similar to the mathematical technique called "false position." Using the most recent two trial values, a projection is made to obtain the next trial value. Details of the computational aspects are available elsewhere (6, 9, 25).

A second order plant model with redundant observations ("vertical" measurement matrix) is considered before applying the adaptive technique to higher order systems in the next section. The system is modelled by the following equations.

$$x(k+1) = \phi(k+1,k)\ x(k) + w(k),\ x(0) = 0$$

$$x^T(k) = [x_1(k)\ \ x_2(k)]$$

$$\phi(k+1,k) = \begin{bmatrix} 1.0 & 0.05 \\ 0 & 1.0 \end{bmatrix}$$

and $$z(k) = H(k)\ x(k) + v(k)$$

$$z^T(k) = [z_1(k)\ \ z_2(k)\ \ z_3(k)]$$

$$H^T(k) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

with $\hat{x}(0|0) = \hat{x}(0)$ and $P(0|0) = P(0)$.

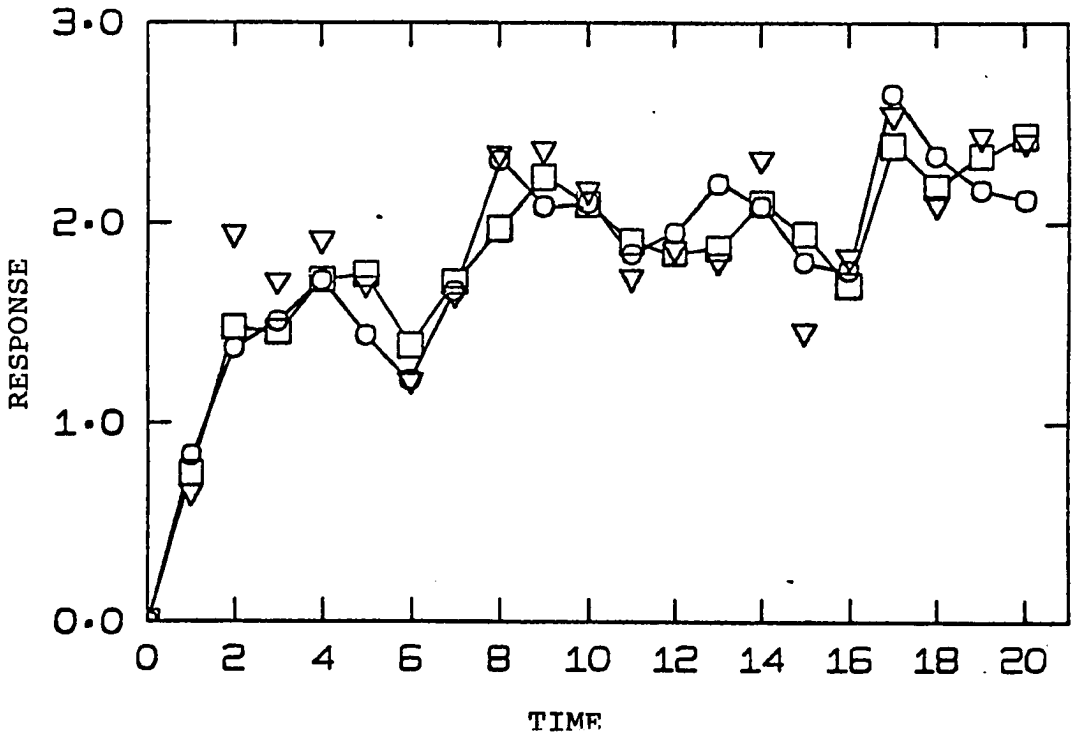
The known noise statistics, R and Q, and zero mean independent white Gaussian random variables,

$$w^T(k) = [w_1(k)\ \ w_2(k)]$$

$$v^T(k) = [v_1(k)\ \ v_2(k)\ \ v_3(k)]$$

with initial conditions for the filter, $\hat{x}(0)$ and $P(0)$, complete the specifications for filtering. The filter response to the observations and model dynamics is shown in Figures 4.7 and 4.8. The figures show that the estimates neither predominantly follow the model nor the observations when the noise in each is equivalent, i.e., $Q_{ii} = R_{ii} = 0.1$

The filter is very insensitive to the initial condition $P(0|0)$. Within four estimation intervals, the covariance matrices have reached steady state. Figure 4.9 shows



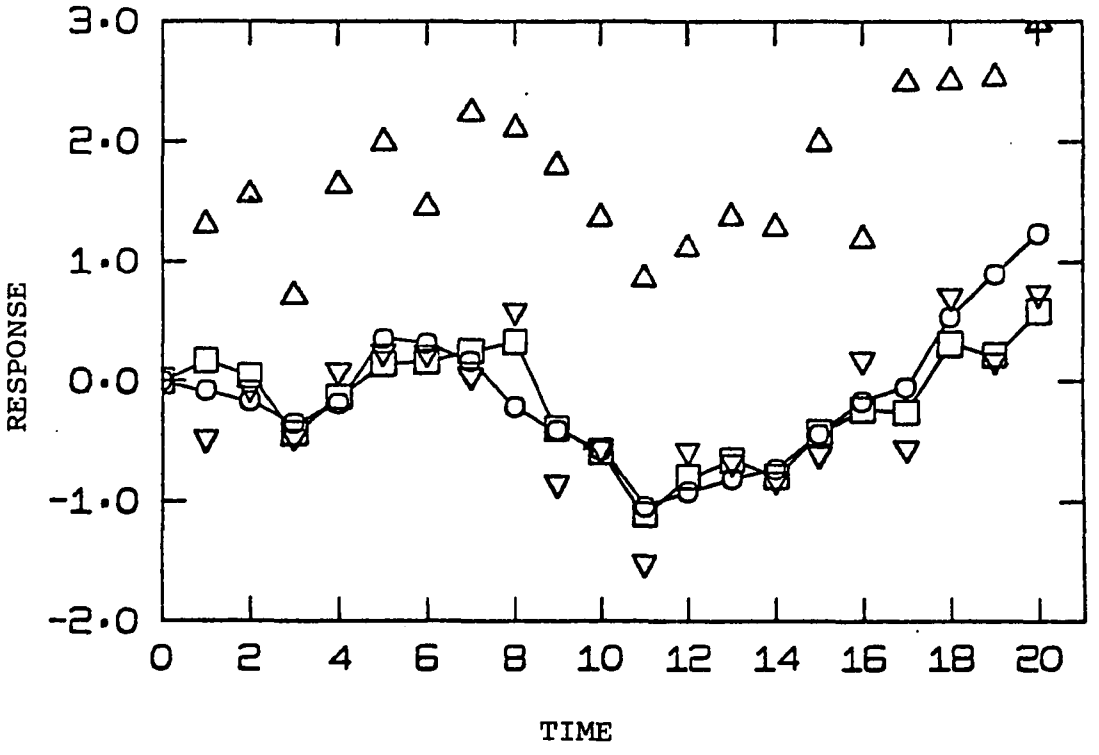
$x_1(k)$ = ○

$z_1(k)$ = ▽

$\hat{x}_1(k|k)$ = □

$$Q(k) = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \qquad R(k) = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 1.0 \end{bmatrix}$$

Figure 4.7. Response for Second Order Model.



$x_2(k)$ = ○

$z_2(k)$ = ▽

$z_3(k)$ = △

$\hat{x}_2(k|k)$ = □

Figure 4.8. Response for Second Order Model.

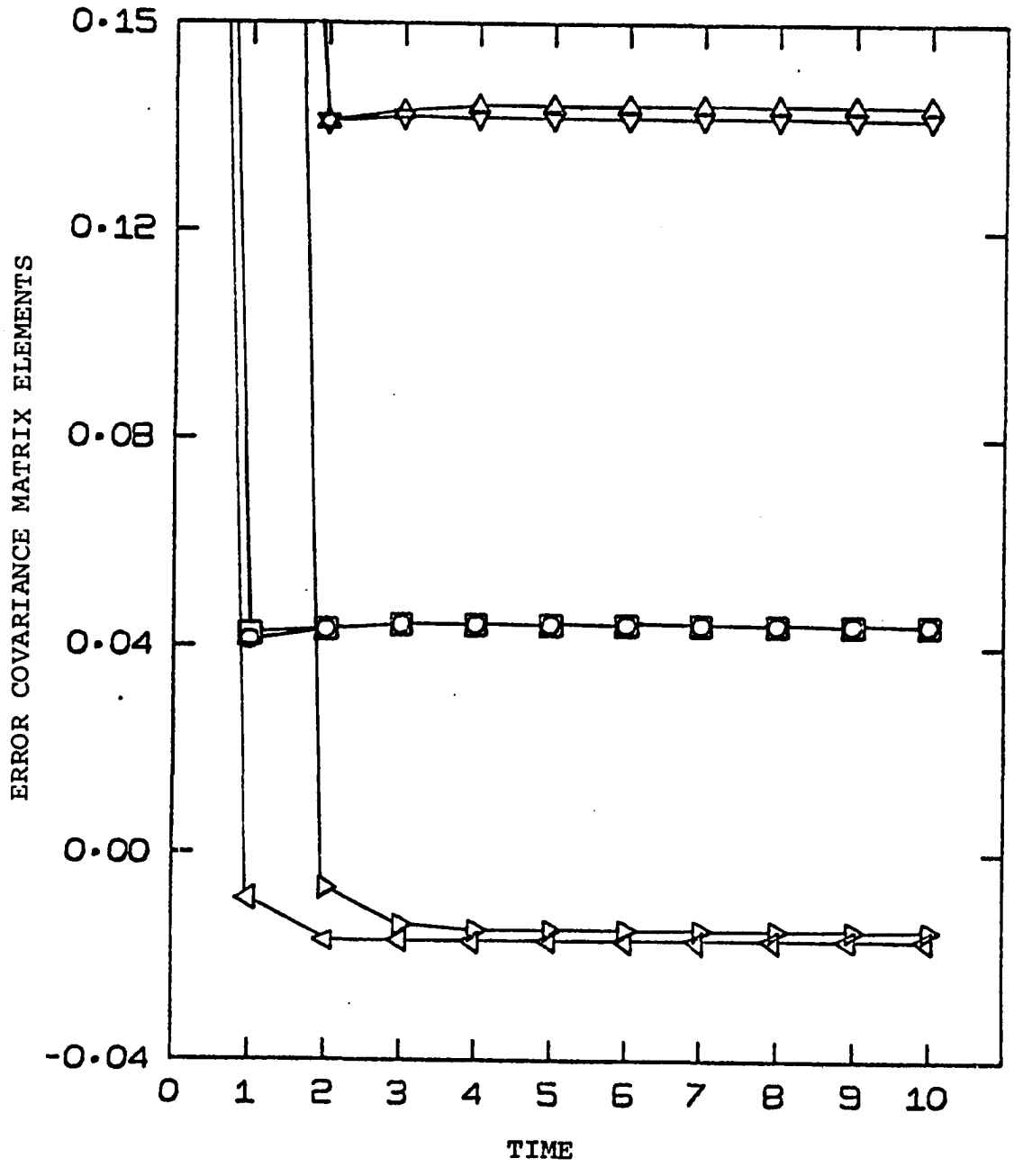


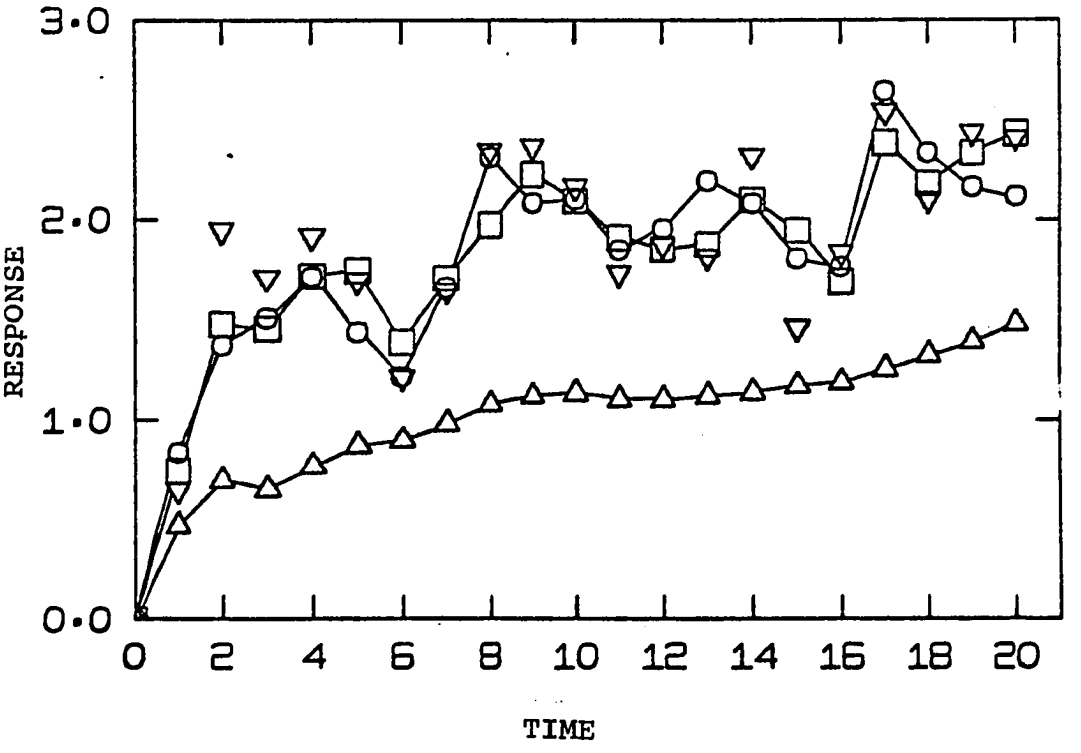
Figure 4.9. Filtered and Predicted Error Covariance Matrix for Second Order Model.

that the covariance matrix has decreased significantly from the initial $P(0|0)$ estimate. The prediction variances (P_{11} and P_{22}) are much larger (about three times) than the filter variances, as was previously seen for the scalar model. However, the prediction and filter covariances (P_{12} and P_{21}) are equivalent and negative.

2. Modelling Error

Figures 4.10 and 4.11 illustrate the consequences of assuming incorrect knowledge of the plant model dynamics. If it is assumed that no random component exists in the plant (i.e., $Q = 0$), then the filter diverges when there actually is a noise component (i.e., $Q = 1$). Both the estimate for x_1 and the estimate for x_2 diverge to erroneous values; whereas with the correct model, the filter correctly follows the model dynamics and observations. The divergence is easily apparent here because of the large discrepancy between the values for the true and assumed plant noise covariance matrices. Generally, one expects the value of the filtered estimate to be between the value of the observation and the model since the noise covariance matrices were identical and the filter is essentially an optimal smoothing device. A small amount of fluctuation outside these limits occurs due to the random nature of the noise vectors w and v.

Figure 4.12 more clearly shows the reason for the filter divergence. For zero plant noise and complete knowledge of the plant dynamics, intuitively the best estimate is the



$x_1(k)$ = ○

$z_1(k)$ = ▽

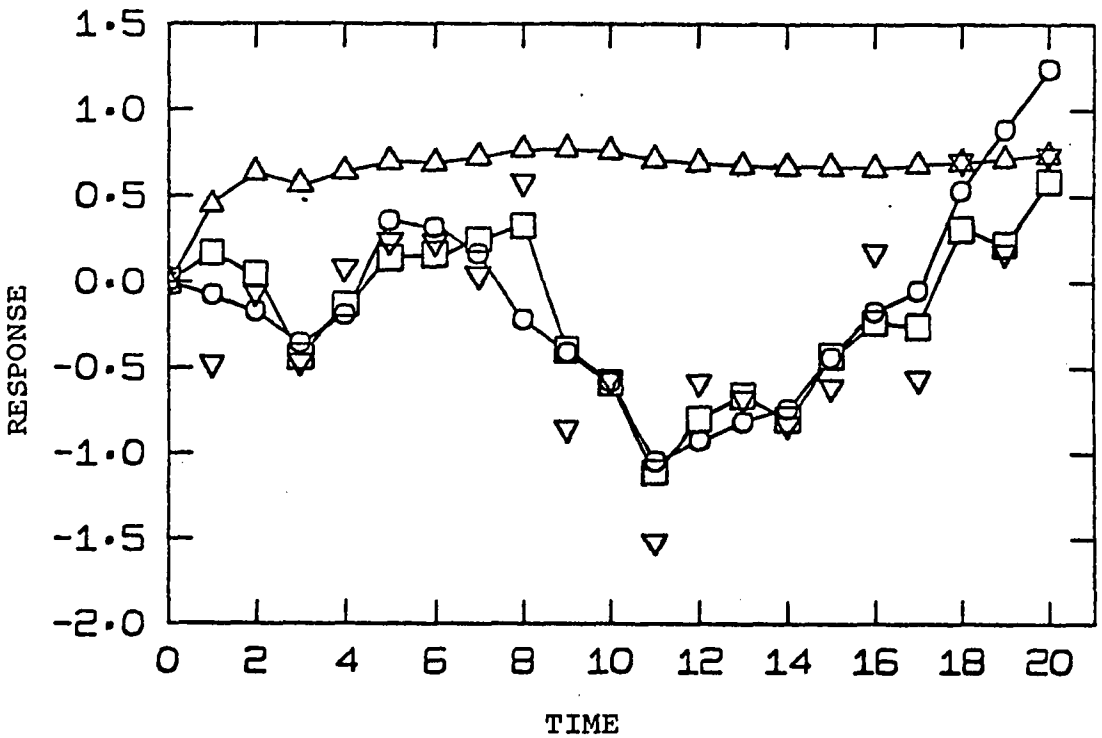
$\hat{x}_1(k|k)$ = □ , correct model

$$Q(k) = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \qquad R(k) = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}$$

$\hat{x}_1(k|k)$ = △ , incorrect model

$$Q(k) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad R(k) = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}$$

Figure 4.10. Response, x_1, for Second Order Model with Model Error.



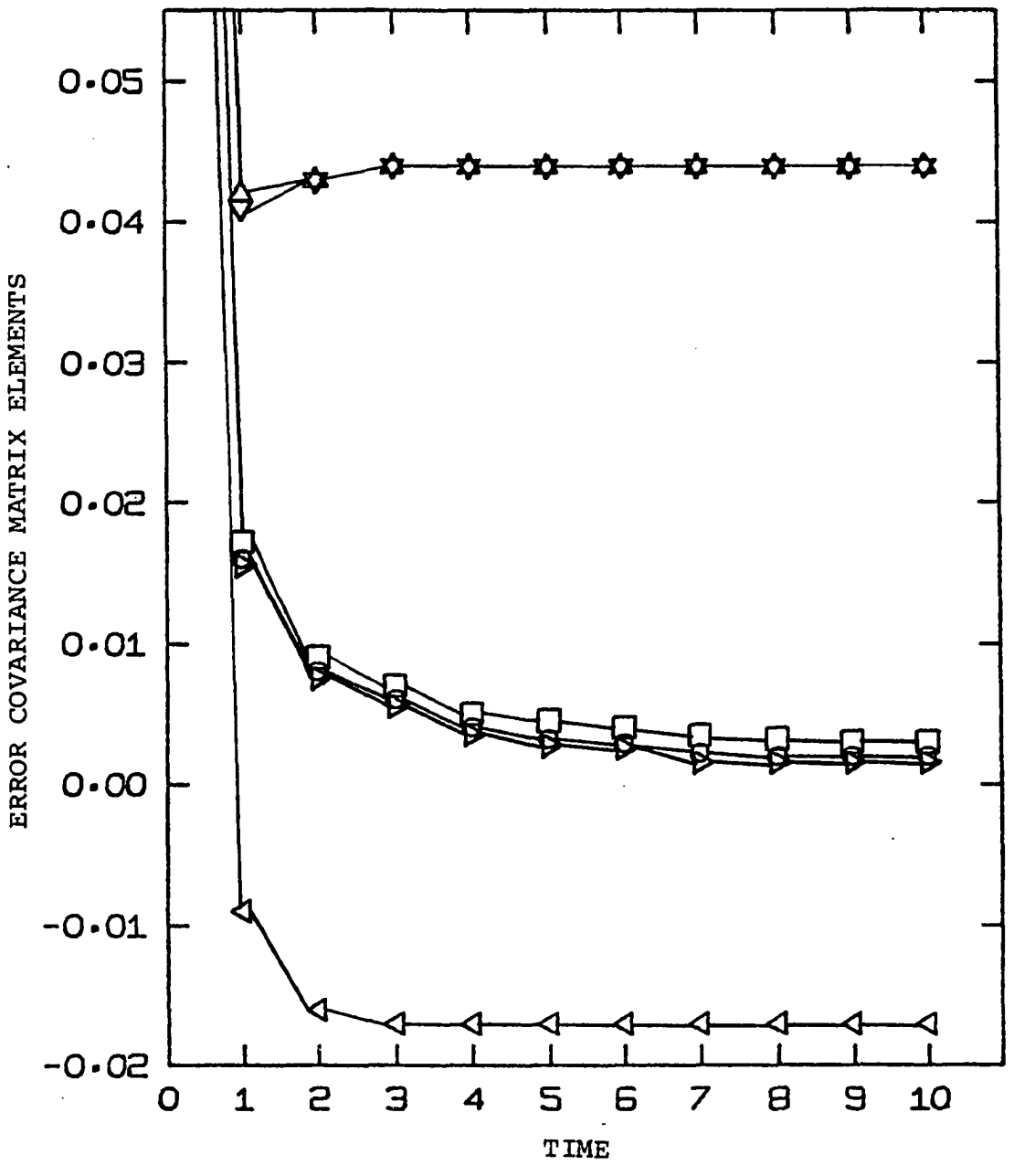
$x_2(k)$ = ○

$z_2(k)$ = ▽

$\hat{x}_2(k|k)$ = □ , correct model

$\hat{x}_2(k|k)$ = △ , incorrect model

Figure 4.11. Response, x_2, for Second Order Model with Model Error.



$$P(0) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$P_{11}(k|k) = \Delta$, correct model
$P_{22}(k|k) = \nabla$, correct model
$P_{11}(k|k) = \square$, incorrect model
$P_{12}(k|k) = O$, incorrect model
$P_{22}(k|k) = \triangleright$, incorrect model
$P_{12}(k|k) = \triangleleft$, correct model

Figure 4.12. Filter Error Covariance Matrix for Second Order Model.

value generated by the model. This value furthermore would be exact (i.e., zero variance). Computationally, this occurs as well. The covariance matrix is computed based on knowledge of the model, including knowledge of noise inputs. If knowledge of the inputs is erroneous, a false reliability is generated by the covariance equations. In fact, the erroneous covariance matrix becomes smaller in magnitude than the true covariance matrix. As a result, an erroneous weighting factor is calculated (but is optimal for the assumption, $Q = 0$). The consequences are propagated in the filter by providing an estimate which diverges from both the true model and the observations to the model assumed to have no noise component.

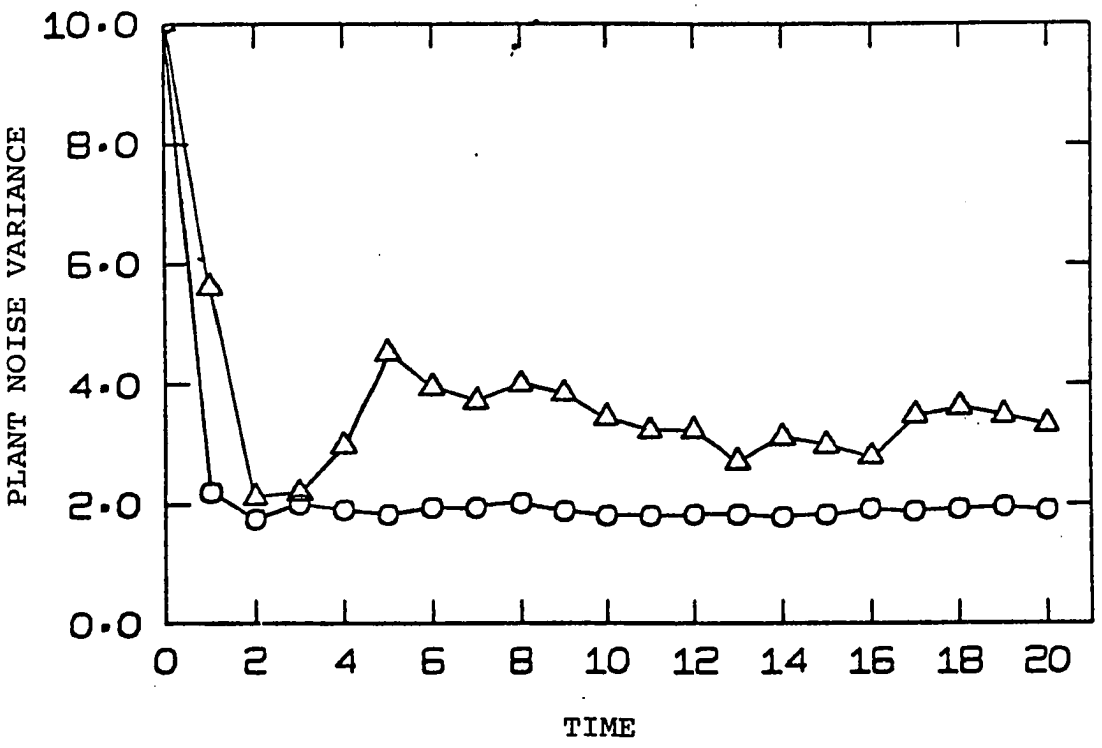
The resultant anomaly is that observations are totally ignored by the filter because the covariance matrix becomes zero (see Figure 4.12) when the observations, which constitute a source of new information, are most needed to correct the model. A consequent need arises for updating the a priori assumptions for the random noise components described by Q and R. Of course, the variance of observational data can be statistically calculated with no a priori assumptions regarding the dynamics of the phenomena which was observed when all the data is simultaneously available. The following illustrates the sequential estimation of the noise covariances during evolution of a phenomena described by second order dynamics.

3. Adaptive Estimation

Two cases are considered. The results for the two are similar and the conclusions are identical, but the differences are in the relative order of magnitudes of the covariance matrices.

Figures 4.13 and 4.14 illustrate estimation of the plant noise variances. The estimate of Q_{11} converges very slowly for both methods illustrated. The estimate of Q_{22} converges more rapidly, but the Sage, Husa method permits much larger oscillation initially and converges relatively more slowly. Results for the observation noise variance estimator (Figures 4.15, 4.16 and 4.17) are similar. Large initial oscillations are characteristic of the Sage, Husa estimator with slow convergence, from the negative side however; whereas, the adaptive estimator of this work does not have oscillatory behavior and converges relatively more rapidly, especially in the early stages of estimation.

It was found that the algorithm proposed in this work generally brought the initial estimate of the noise covariance matrices within an order of magnitude of the true value within 20 estimation increments; whereas, the Sage, Husa algorithm required more time and was largely dependent on the initial estimate (27). Based on this observation and depending upon the system modelled and the computational facilities available to a user, an optimal approach may be to use the algorithm of this work for an initial segment of time and then



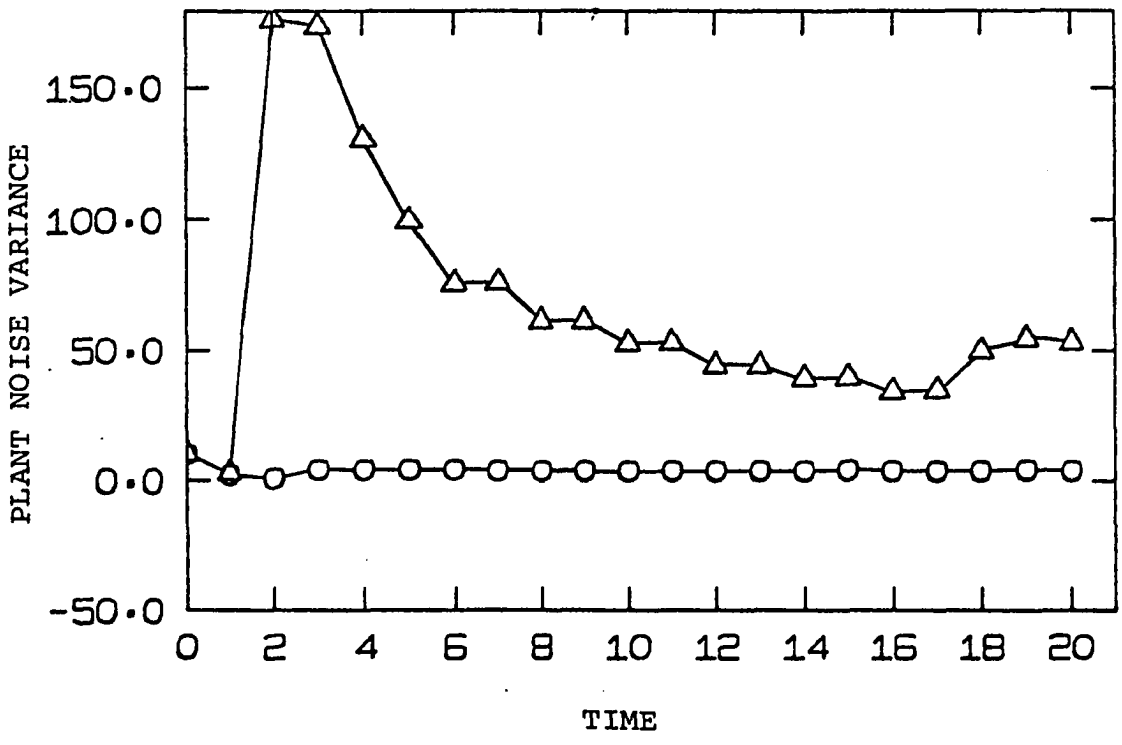
$\hat{Q}_{11}(k|k)$ = O , estimated (this work)

$\hat{Q}_{11}(k|k)$ = Δ , estimated (Sage, Husa)

$$Q(k) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad R(k) = \begin{bmatrix} 0.01 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}$$

$$Q(0|0) = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix} \qquad R(0|-1) = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 10 \end{bmatrix}$$

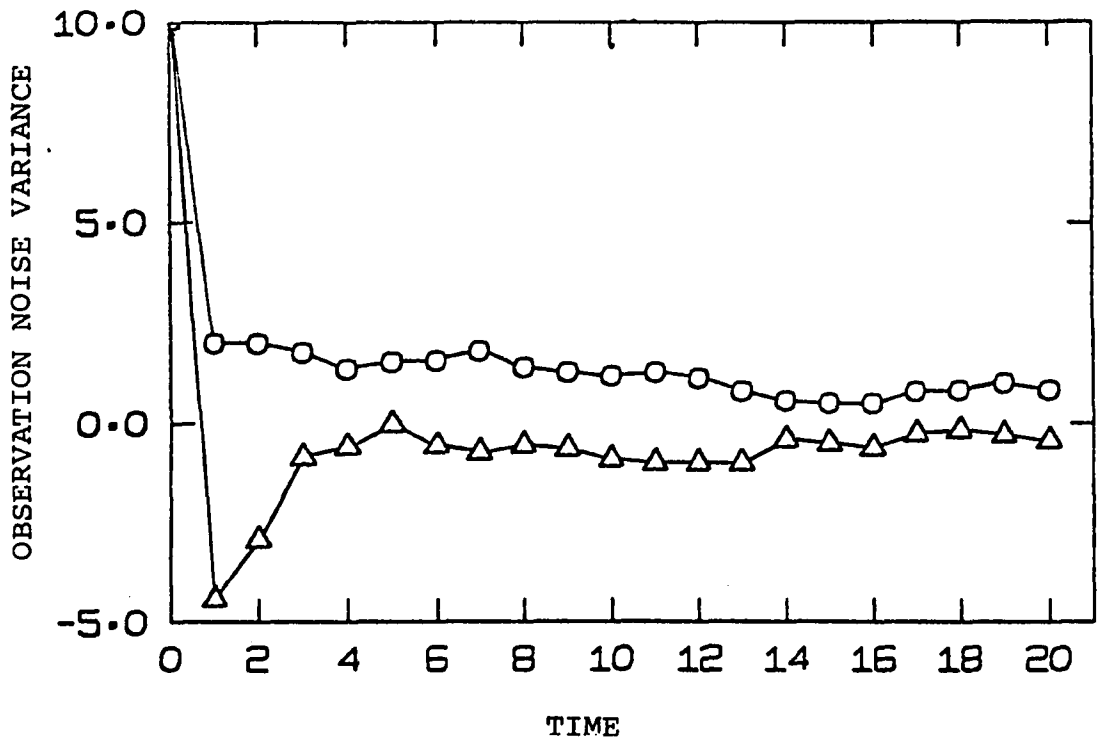
Figure 4.13. Adaptive Estimation of Plant Noise Variance, Q_{11}.



$\hat{Q}_{22}(k|k)$ = O , estimated (this work)

$\hat{Q}_{22}(k|k)$ = Δ , estimated (Sage Husa)

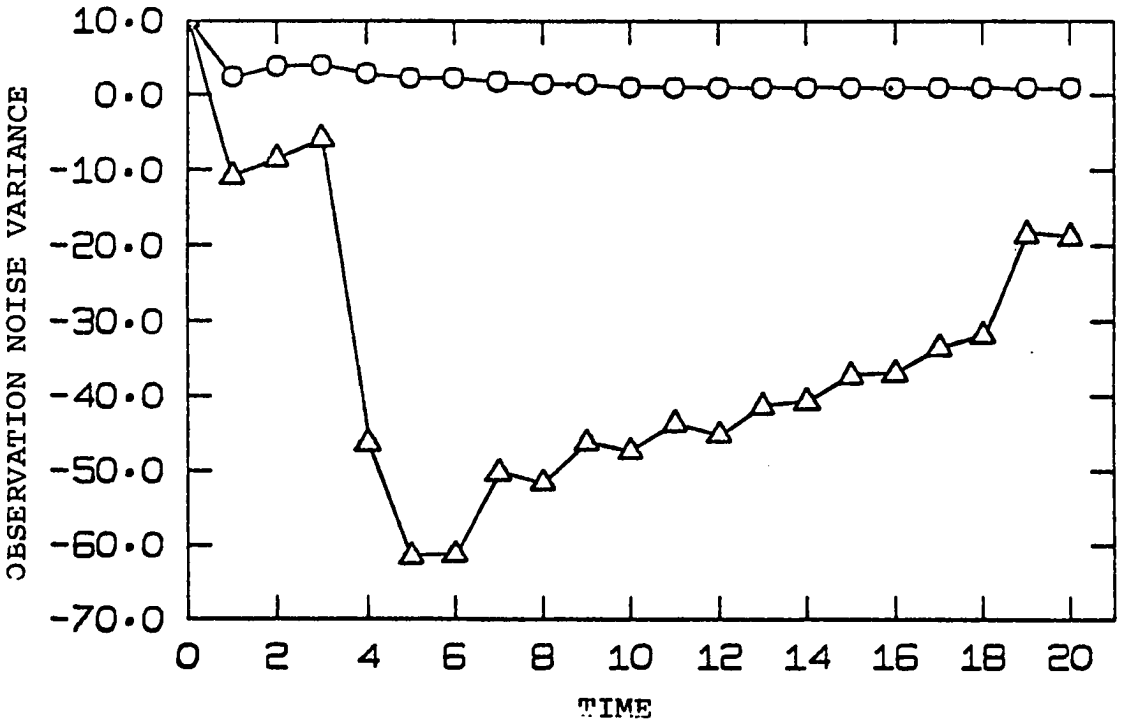
Figure 4.14. Adaptive Estimation of Plant Noise Variance, Q_{22}.



$\hat{R}_{11}(k+1|k) =$ O , estimated (this work)

$\hat{R}_{11}(k+1|k) =$ Δ , estimated (Sage, Hu a)

Figure 4.15. Adaptive Estimation of Observation Noise Variance, R_{11}.



$\hat{R}_{22}(k+1|k)$ = O , estimated (this work)

$\hat{R}_{22}(k+1|k)$ = Δ , estimated (Sage, Husa)

Figure 4.16. Adaptive Estimation of Observation Noise Variance, R_{22}.

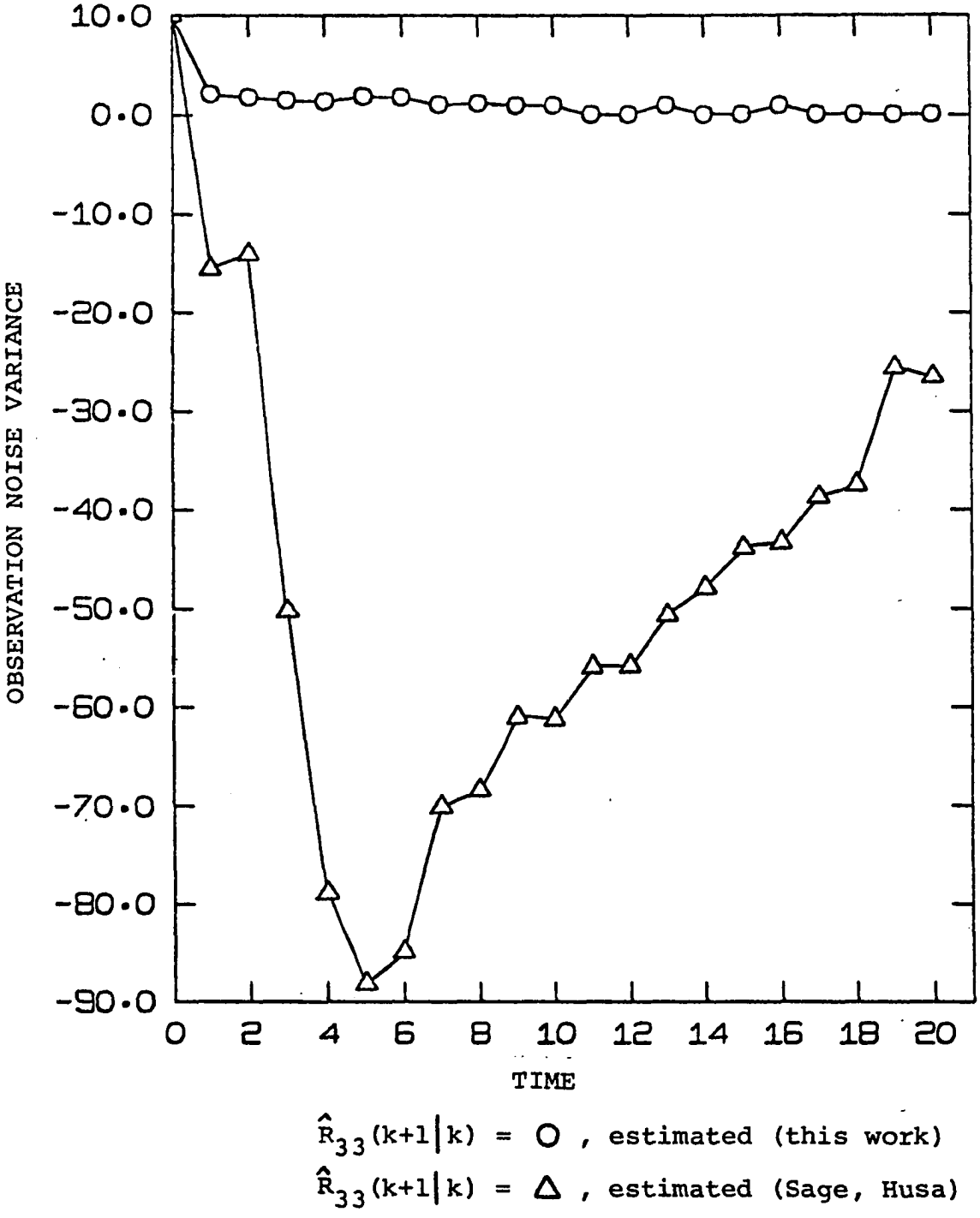


Figure 4.17. Adaptive Estimation of Observation Noise Variance, R_{33}.

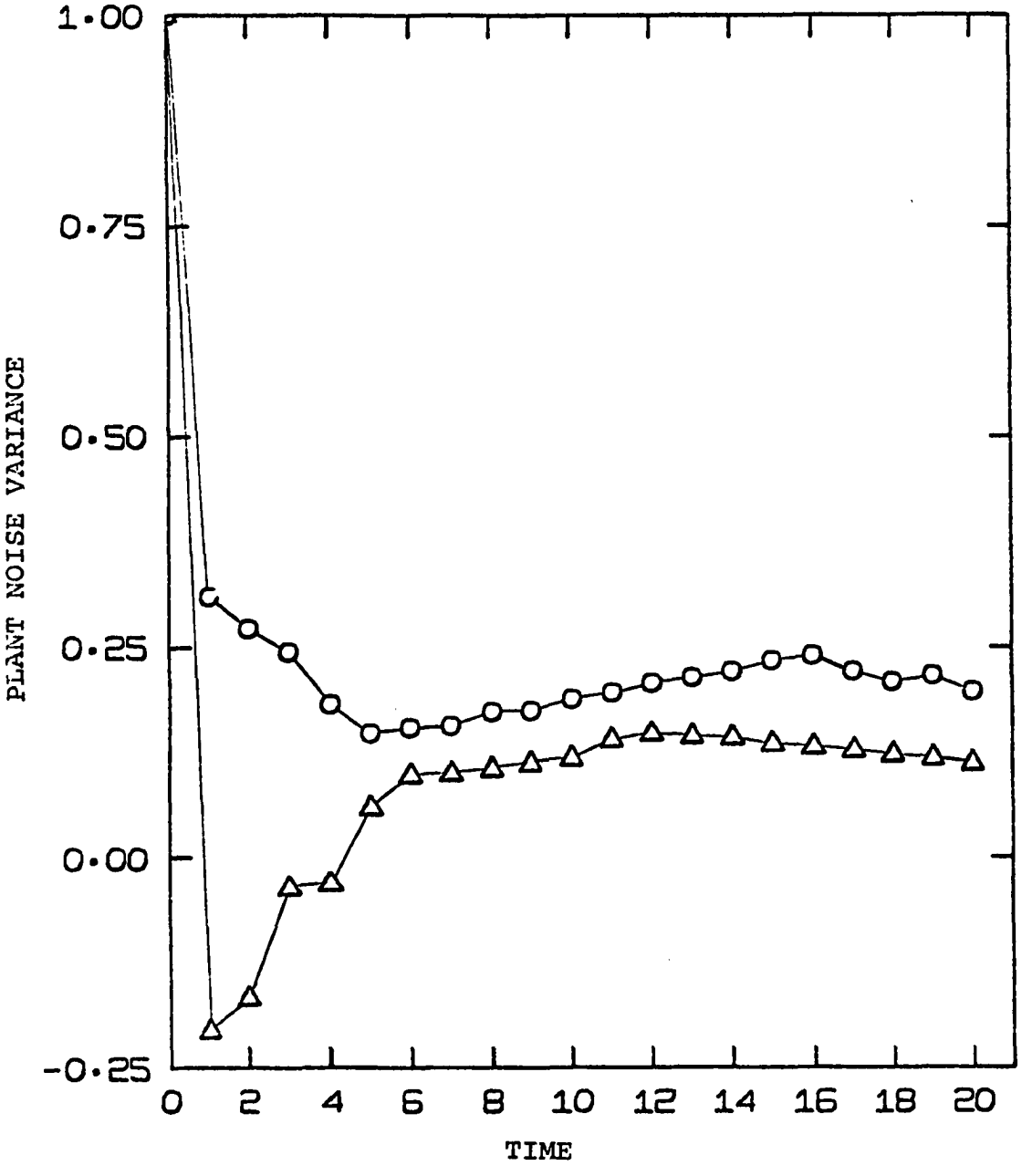
revert to the Sage, Husa algorithm to further refine the estimate after the order of magnitude is determined. In this manner, rapid convergence is achieved and less computational time is necessary. In general the proposed algorithm requires more time to process a single observation than the Sage, Husa algorithm, although the user can control the time consuming iteration necessary to solve the nonlinear filter equations. Because of the iteration, however, it is possible to obtain a more refined estimate at each estimation step, or time interval, which may explain the more rapid initial convergence of the proposed method.

Results of the second case are presented in Figures 4.18 through 4.19 for the plant noise covariance matrix and Figures 4.20 through 4.22 for the observation noise covariance matrix. Although additional runs were made, the two cases presented here represent a wide range of values of Q, R, and initial conditions for these matrices. The behavior of the algorithms for this case varies somewhat from the previously described case, but the conclusions remain the same as previously discussed.

IV.3 A Fourth Order System

1. The Model and Modelling Error

A final demonstration of the applicability of the proposed adaptive filter is presented in this section. A higher order system is considered which previously was



$Q_{11}(k|k)$ = O , estimated (this work)

$Q_{11}(k|k)$ = Δ , estimated (Sage, Husa)

$$Q(k)=\begin{bmatrix}0.1 & 0\\ 0 & 0.1\end{bmatrix};\quad R(k)=\begin{bmatrix}0.1 & 0 & 0\\ 0 & 0.1 & 0\\ 0 & 0 & 0.1\end{bmatrix};\quad Q(0|0)=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix};\quad R(0|-1)=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

Figure 4.18. Adaptive Estimation of Plant Noise Variance, Q_{11}:

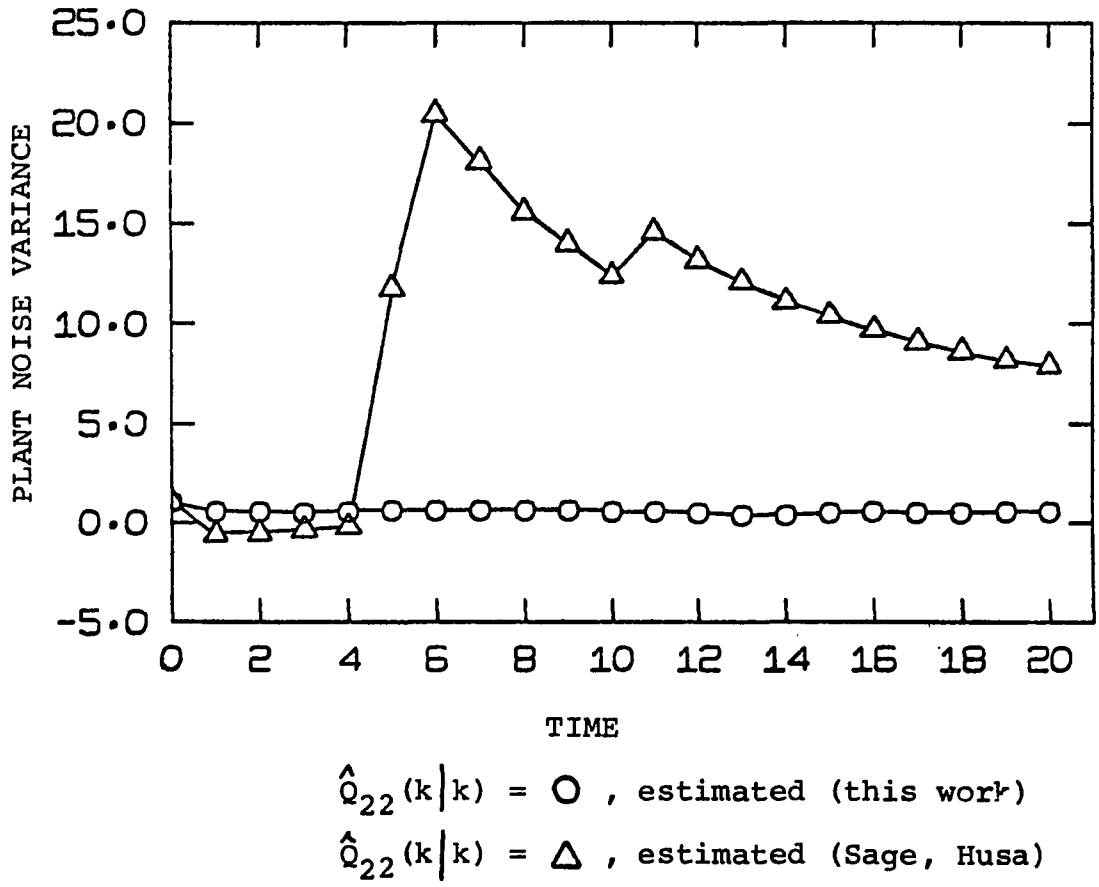
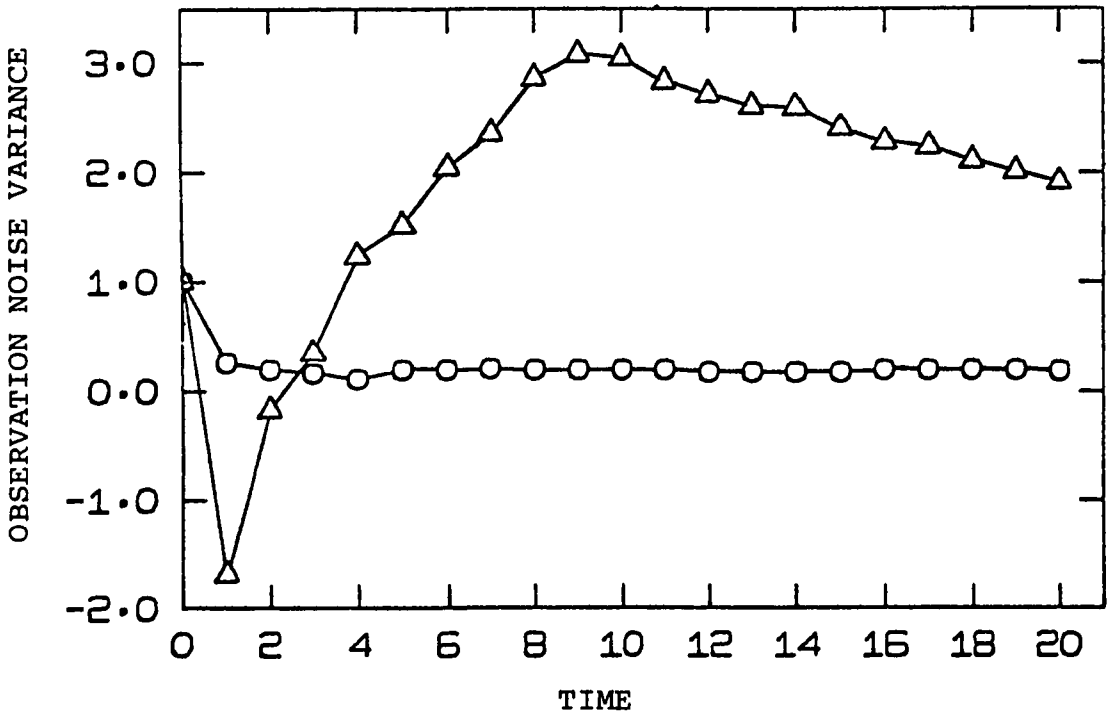


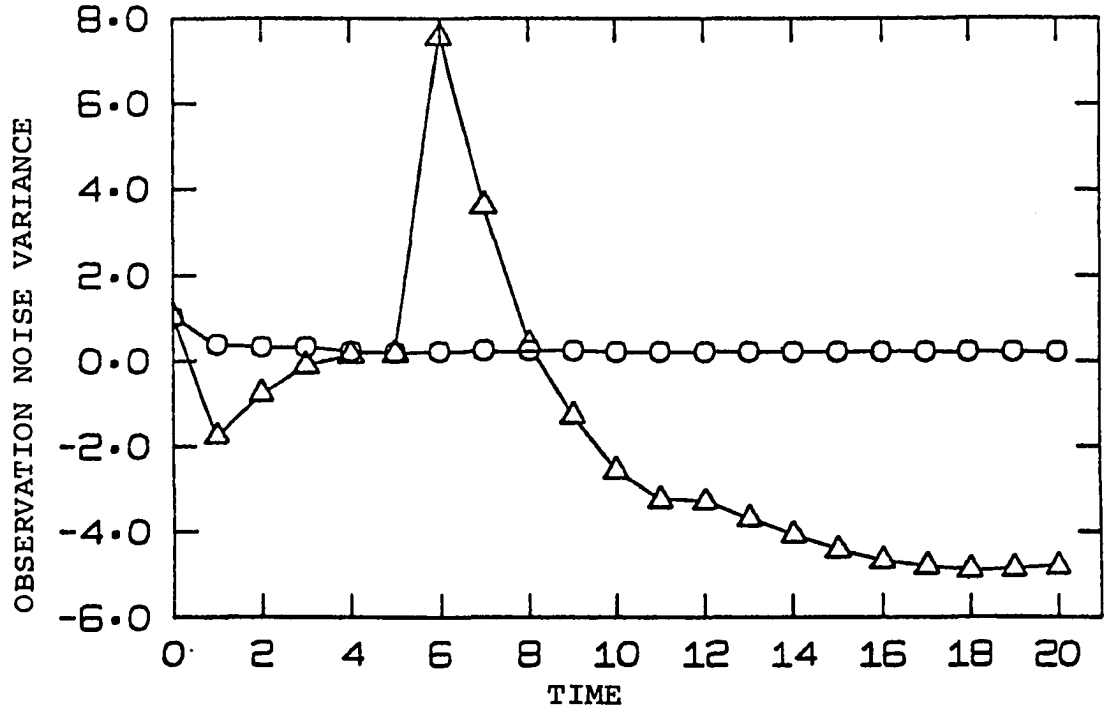
Figure 4.19. Adaptive Estimation of Plant Noise Variance, Q_{22}.



$\hat{R}_{11}(k+1|k)$ = O , estimated (this work)

$\hat{R}_{11}(k+1|k)$ = Δ , estimated (Sage, Husa)

Figure 4.20. Adaptive Estimation of Observation Noise Variance, R_{11}.



$\hat{R}_{22}(k+1|k)$ = O , estimated (this week)

$\hat{R}_{22}(k+1|k)$ = Δ , estimated (Sage, Husa)

Figure 4.21. Adaptive Estimation of Observation Noise Variance, R_{22}.

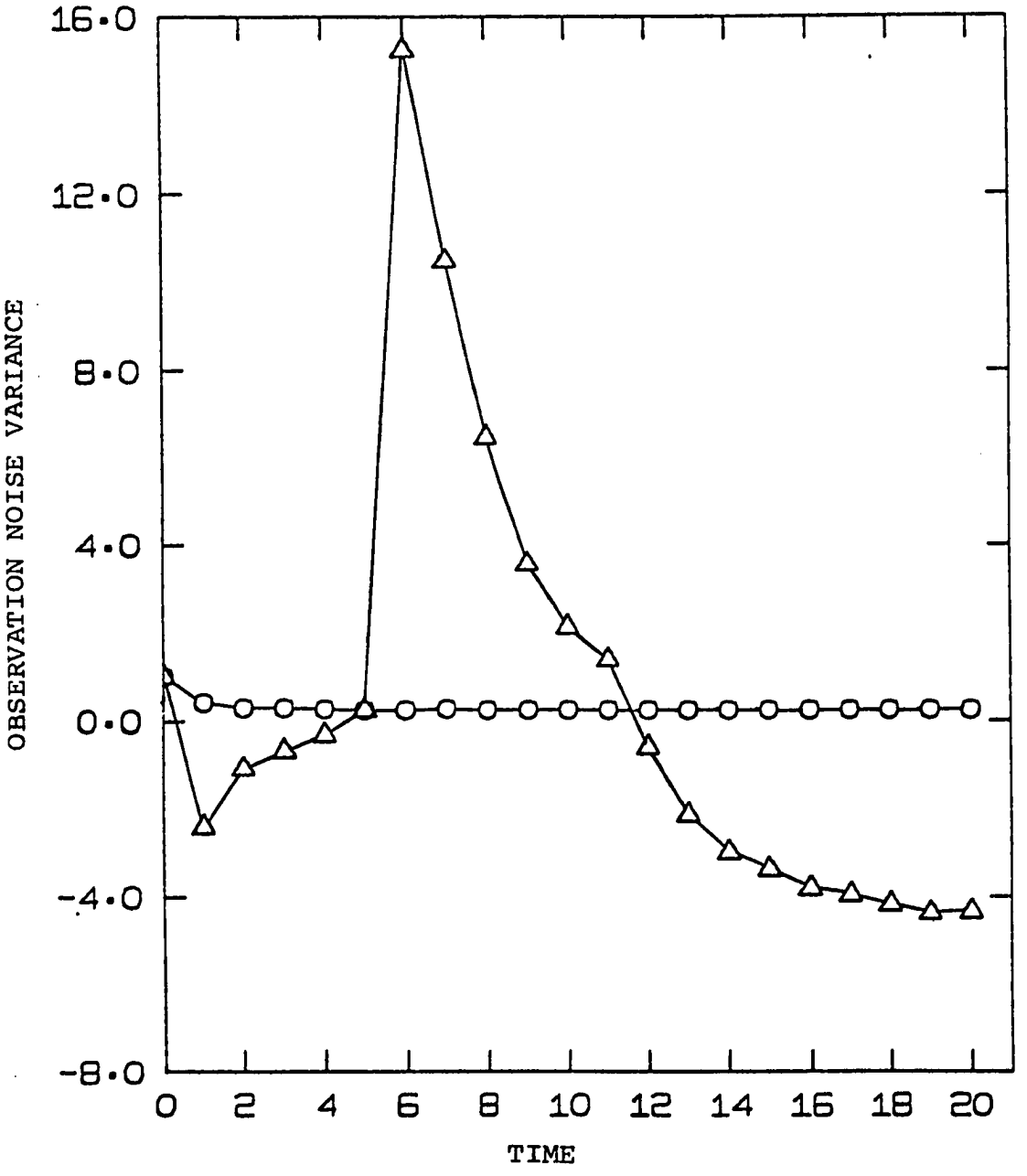


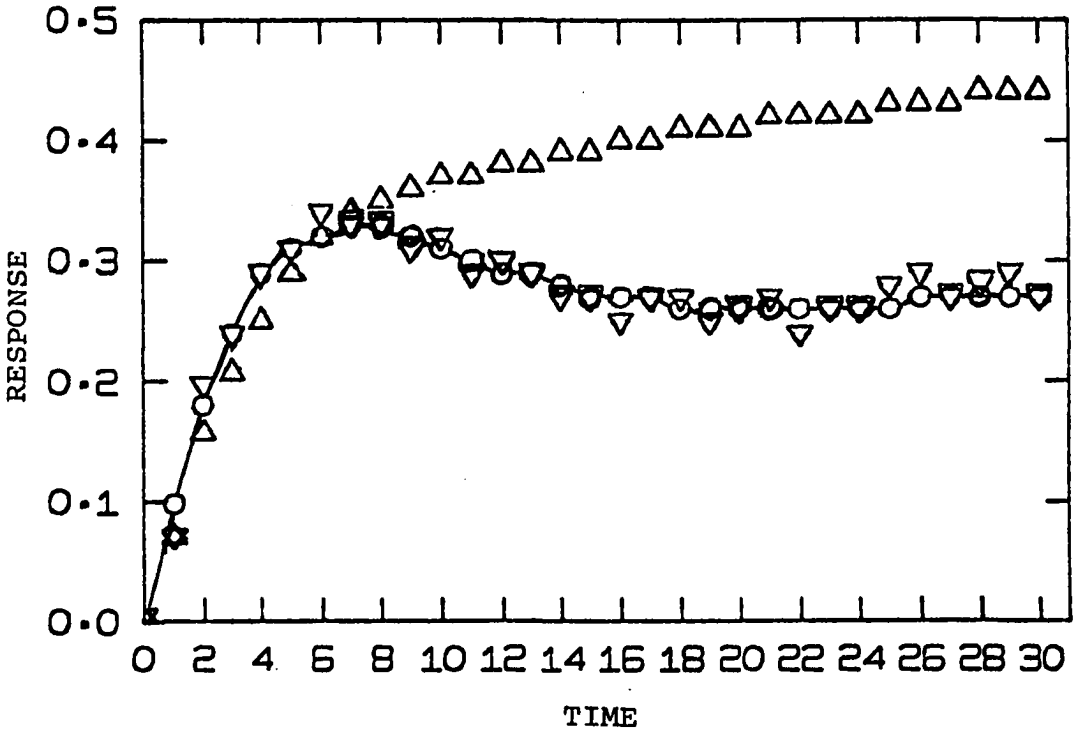
Figure 4.22. Adaptive Estimation of Observation Noise Variance, R_{33}.

developed and used by Wells (34) to demonstrate the applicability of the extended Kalman filter alone for estimating the process state and model coefficients. The system considered is an adiabatic stirred reactor which is simulated by a fourth order mathematical model. The details of this model are presented in Appendix G. The difference between the model used by Wells and the model considered here is that measurements for all four states are assumed to be available; whereas, Wells assumed only three of the states (temperatures) were measured and the fourth (concentration) was then estimated using the extended Kalman filter.

The reactor is assumed to be at a controlled steady state initially. For a plant disturbance of 10 percent in the feed concentration at zero time, i.e.,

$$x^T(0) = [0 \quad 0 \quad 0 \quad 0.1]$$

the simulated response is shown in Figures 4.23 through 4.26. The system was maintained at the final equilibrium by using steady state control. Shown in these same figures are the responses that simulate the observations where the standard deviation of the instruments is 1 percent of the normalized variables. The third curve in these figures is the estimate of the state vector using the extended Kalman filter and assuming that the model of the reactor is completely known with no plant noise present. However, as seen previously for the simpler systems, when this assumption is incorrect



$x_1(k)$ = O

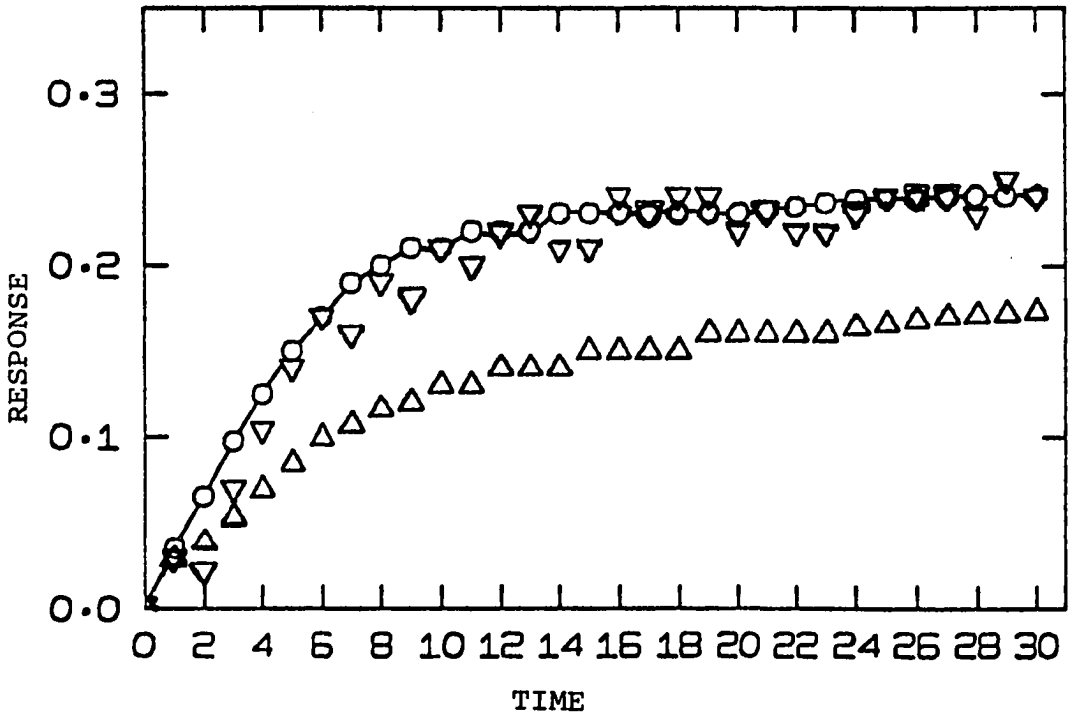
$z_1(k)$ = ▽

$\hat{x}_1(k|k)$ = △

Actual: $Q(k) = \begin{bmatrix} 10^{-4} & 0 & 0 & 0 \\ 0 & 10^{-4} & 0 & 0 \\ 0 & 0 & 10^{-4} & 0 \\ 0 & 0 & 0 & 10^{-4} \end{bmatrix}$; $R(k) = \begin{bmatrix} 10^{-4} & 0 & 0 & 0 \\ 0 & 10^{-4} & 0 & 0 \\ 0 & 0 & 10^{-4} & 0 \\ 0 & 0 & 0 & 10^{-4} \end{bmatrix}$

Assumed: $Q(k) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$; $R(k) = \begin{bmatrix} 10^{-4} & 0 & 0 & 0 \\ 0 & 10^{-4} & 0 & 0 \\ 0 & 0 & 10^{-4} & 0 \\ 0 & 0 & 0 & 10^{-4} \end{bmatrix}$

Figure 4.23. Response, x_1, for Fourth Order Model with Model Error.

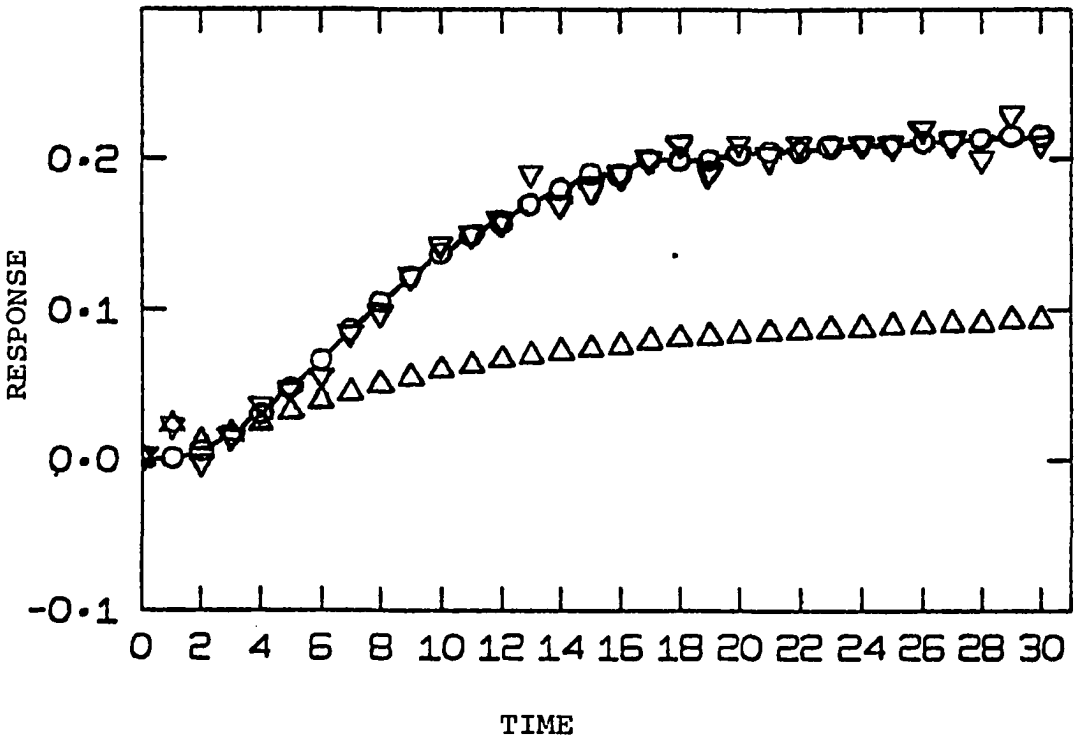


$x_2(k)$ = O

$z_2(k)$ = ▽

$\hat{x}_2(k|k)$ = △

Figure 4.24. Response, x_2, for Fourth Order Model with Model Error.

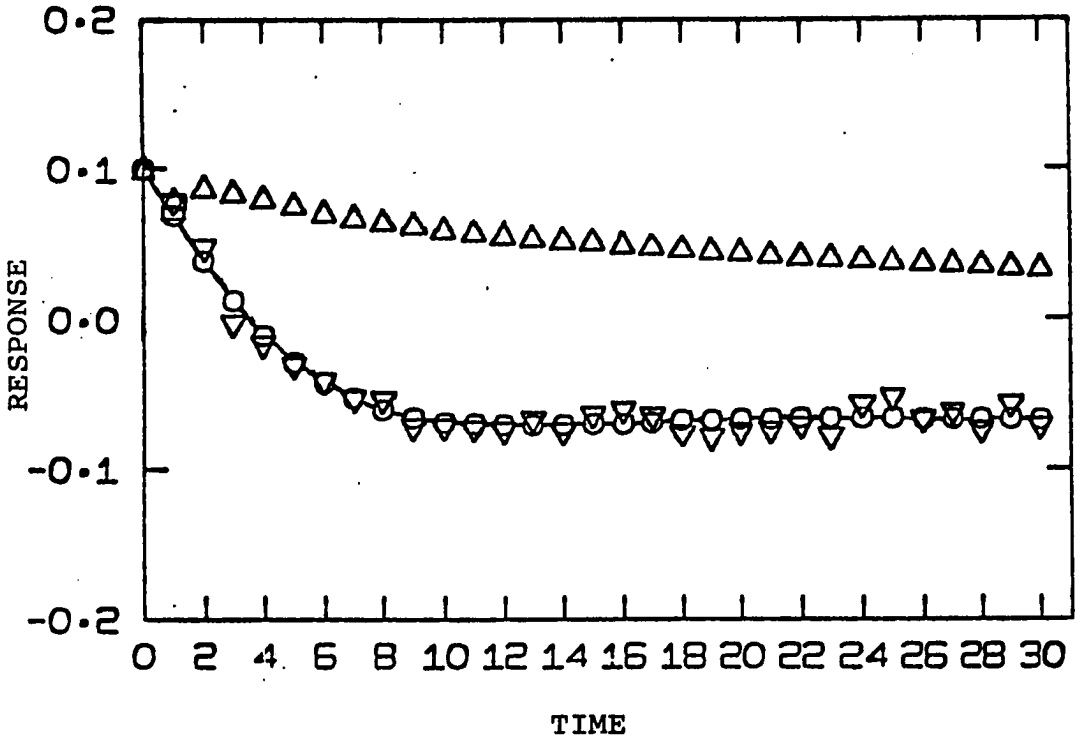


$x_3(k)$ = O

$z_3(k)$ = ∇

$\hat{x}_3(k|k)$ = Δ

Figure 4.25. Response, x_3, for Fourth Order Model with Model Error.



$x_4(k)$ = O

$z_4(k)$ = ▽

$\hat{x}_4(k|k)$ = △

Figure 4.26. Response, x_4, for Fourth Order Model with Model Error.

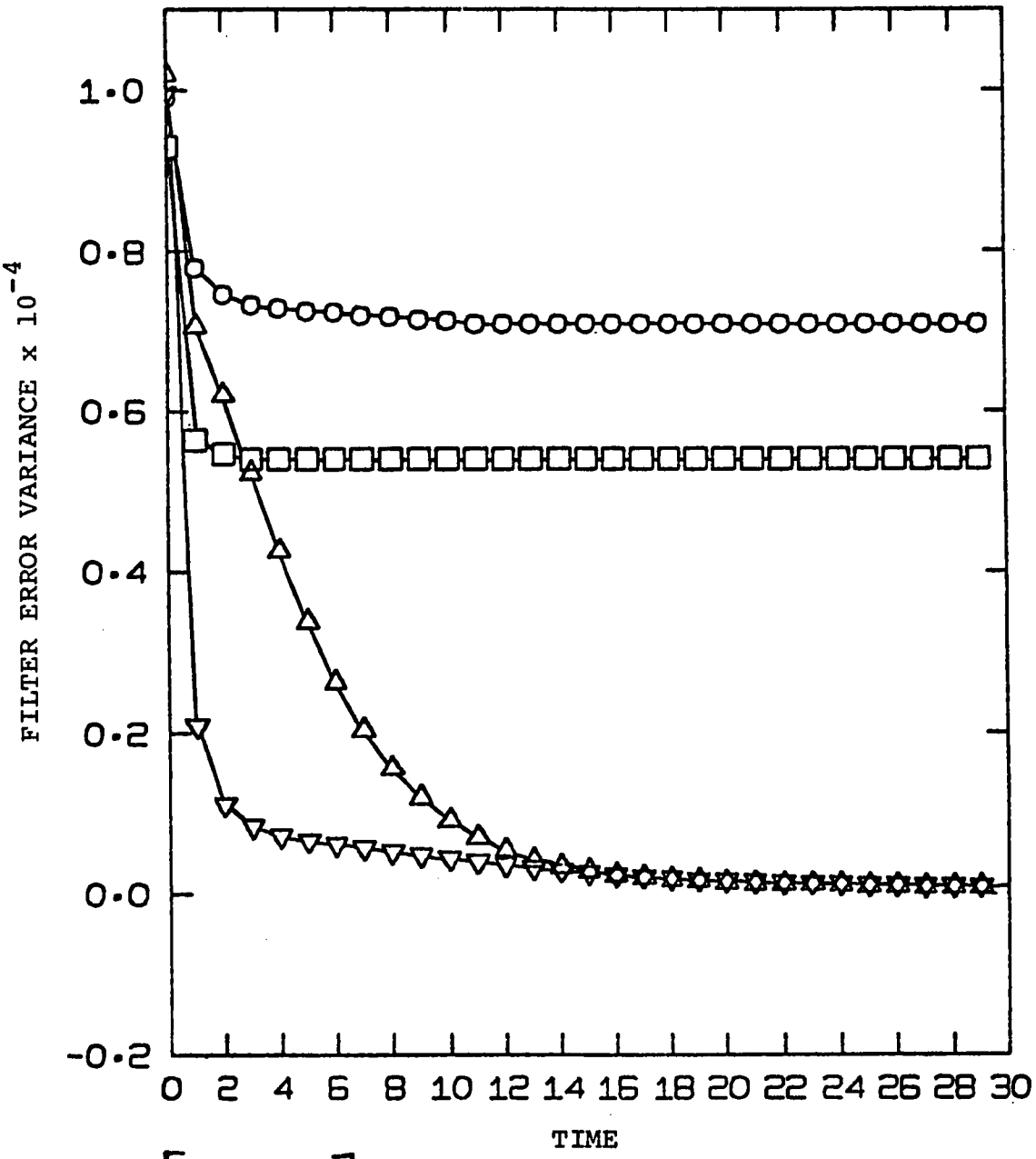
and a random input (e.g., plant noise) does exist, the filter diverges.

Figures 4.27 and 4.28 illustrate the mechanics of filter divergence. The filter error covariance matrix goes to zero and causes the estimate to diverge when the model is incorrectly assumed. The responses of the diagonal elements of the covariance matrix are also plotted to show how the initial values of the state variances were decreased when the model was correctly assumed to include plant noise.

2. Adaptive Estimation

The subroutines used to implement the adaptive filter derived in this work are listed in Appendix H. The Sage, Husa adaptive filter algorithm is available elsewhere (27). For both algorithms, it was found that the estimate of the state was very poor until the estimates of the noise covariance matrices stabilized and converged near the true value.

Even for this four dimensional model, the adaptive estimator developed by this work converges rapidly to an estimate of the covariance matrix much more rapidly than that of Sage and Husa. The Sage, Husa estimator oscillates with a large amplitude initially, then converges to an order of magnitude estimate, and finally after a large number of time increments (30 to 100) is close to the actual covariance. For the 30 estimation intervals shown in Figures 4.29 through 4.32, the estimator for Q developed in this work is always



$$P(0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times 10^{-4}$$

$P_{11}(k|k)$ = ○ , correct model

$P_{22}(k|k)$ = □ , correct model

$P_{11}(k|k)$ = △ , incorrect model

$P_{22}(k|k)$ = ▽ , incorrect model

Figure 4.27. Filtered Error Covariance Matrix for Fourth Order Model, P_{11}, P_{22}.

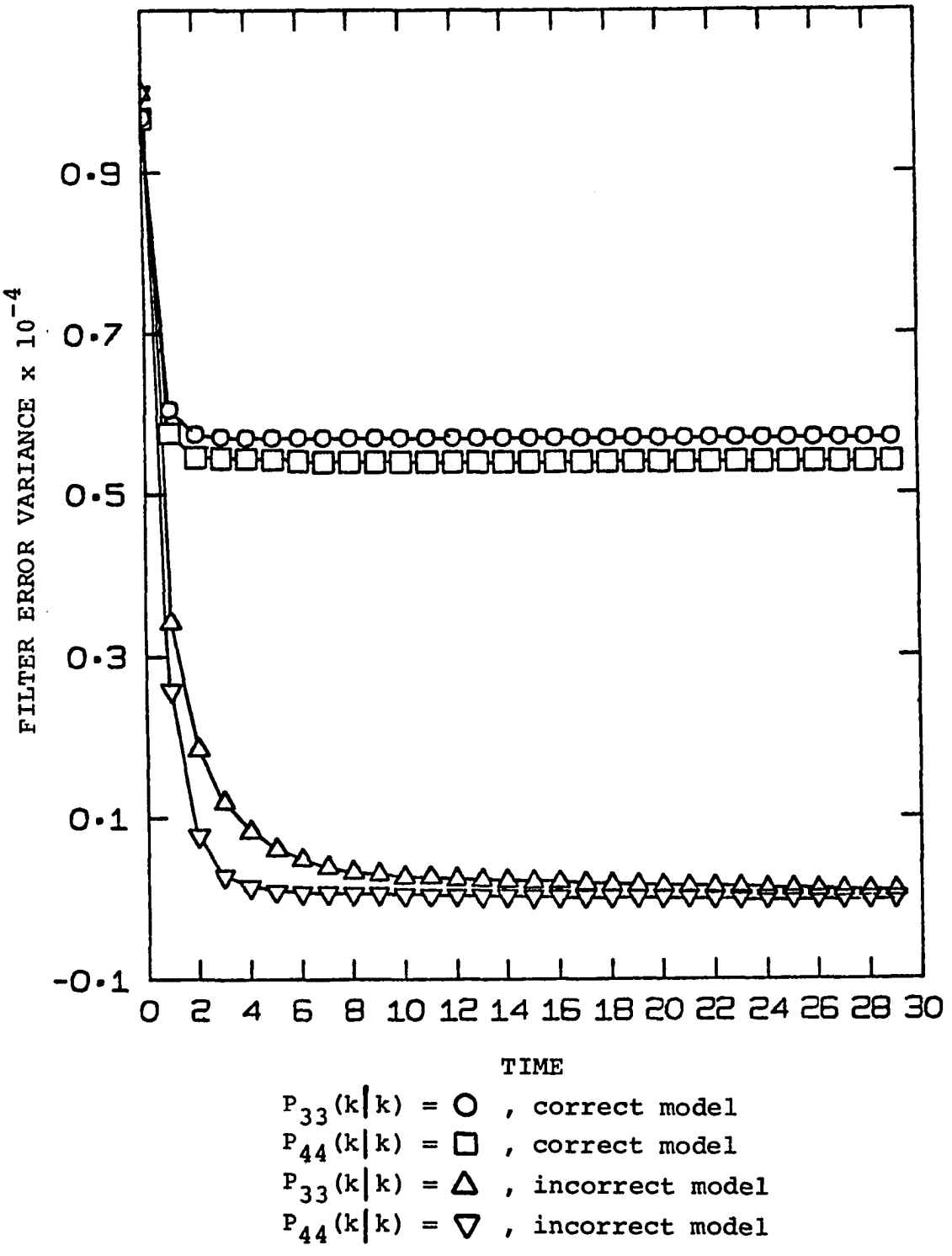
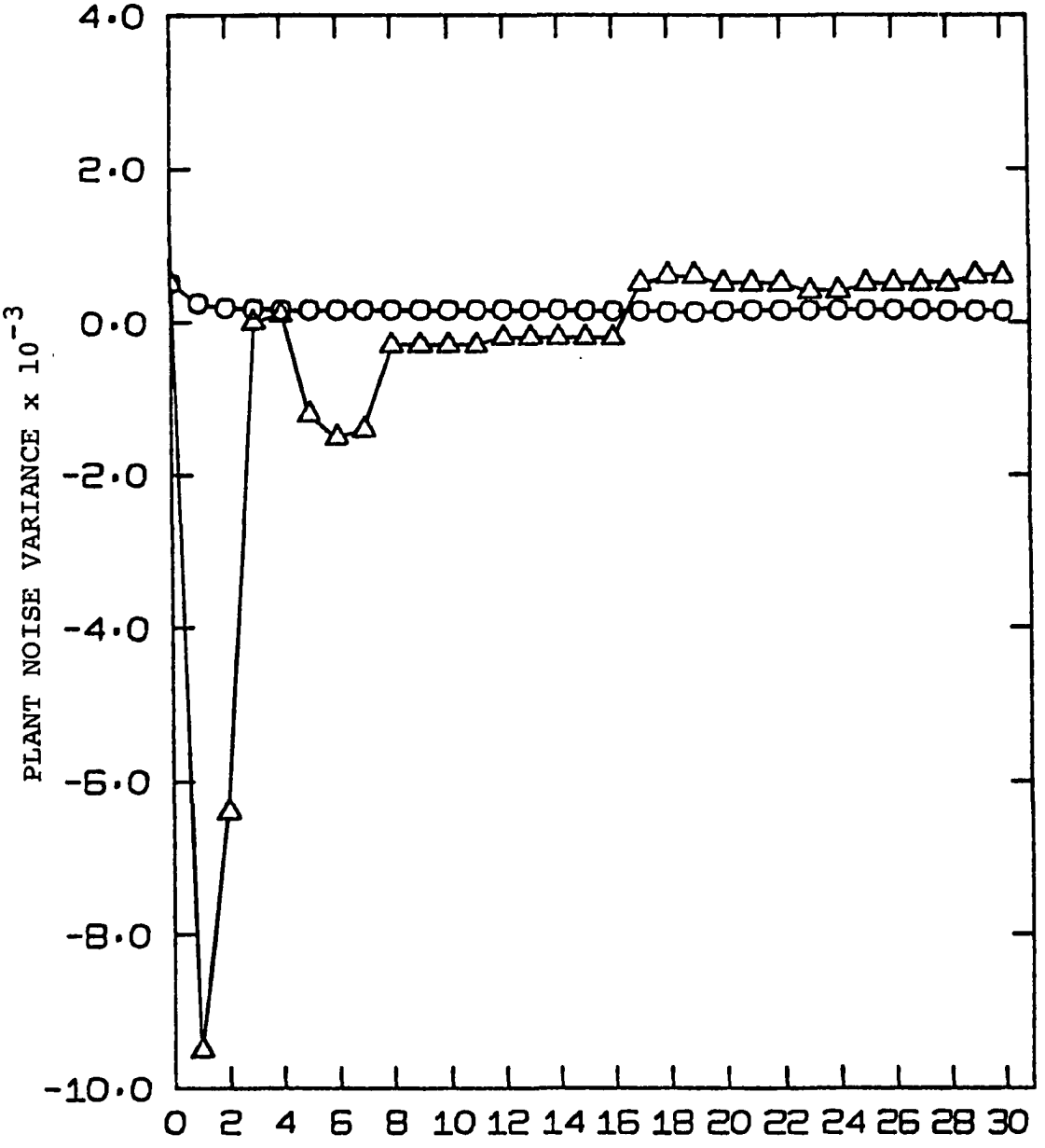


Figure 4.28. Filtered Error Covariance Matrix for Fourth Order Model, P_{33}, P_{44}.



$\hat{Q}_{11}(k|k)$ = O , estimated (this work)

$\hat{Q}_{11}(k|k)$ = Δ , estimated (Sage, Husa)

Figure 4.29. Adaptive Estimation of Plant Noise Variance, Q_{11}.

(Figure continued on page following.)

(Figure 4.29, continued.)

$$Q(k) = \begin{bmatrix} 10^{-4} & 0 & 0 & 0 \\ 0 & 10^{-4} & 0 & 0 \\ 0 & 0 & 10^{-4} & 0 \\ 0 & 0 & 0 & 10^{-4} \end{bmatrix} \qquad R(k) = \begin{bmatrix} 10^{-4} & 0 & 0 & 0 \\ 0 & 10^{-4} & 0 & 0 \\ 0 & 0 & 10^{-4} & 0 \\ 0 & 0 & 0 & 10^{-4} \end{bmatrix}$$

$$\hat{Q}(0|0) = \begin{bmatrix} 5x10^{-4} & 0 & 0 & 0 \\ 0 & 5x10^{-4} & 0 & 0 \\ 0 & 0 & 5x10^{-4} & 0 \\ 0 & 0 & 0 & 5x10^{-4} \end{bmatrix}$$

$$\hat{R}(0|-1) = \begin{bmatrix} 5x10^{-4} & 0 & 0 & 0 \\ 0 & 5x10^{-4} & 0 & 0 \\ 0 & 0 & 5x10^{-4} & 0 \\ 0 & 0 & 0 & 5x10^{-4} \end{bmatrix}$$

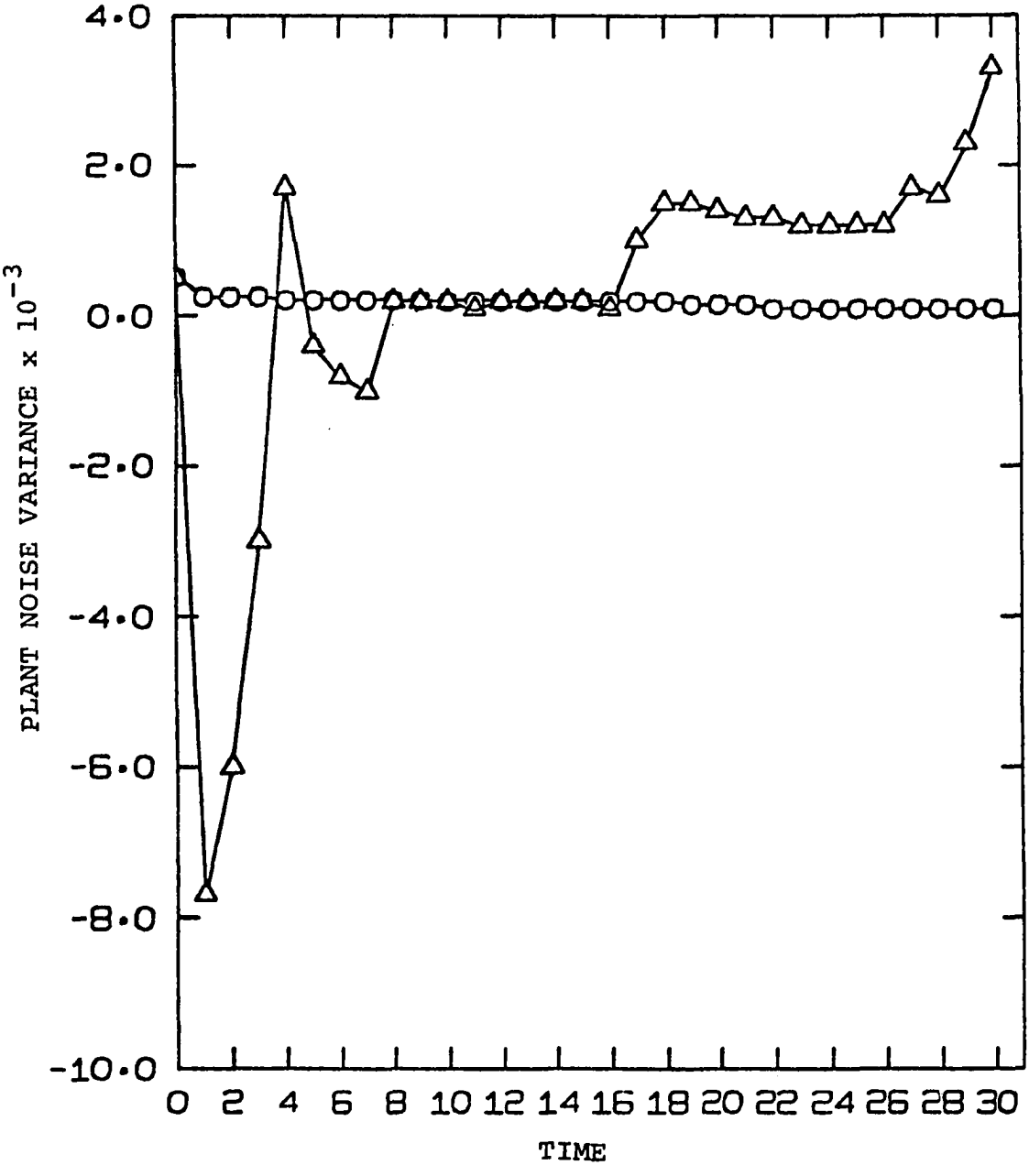


Figure 4.30. Adaptive Estimation of Plant Noise Variance, Q_{22}.

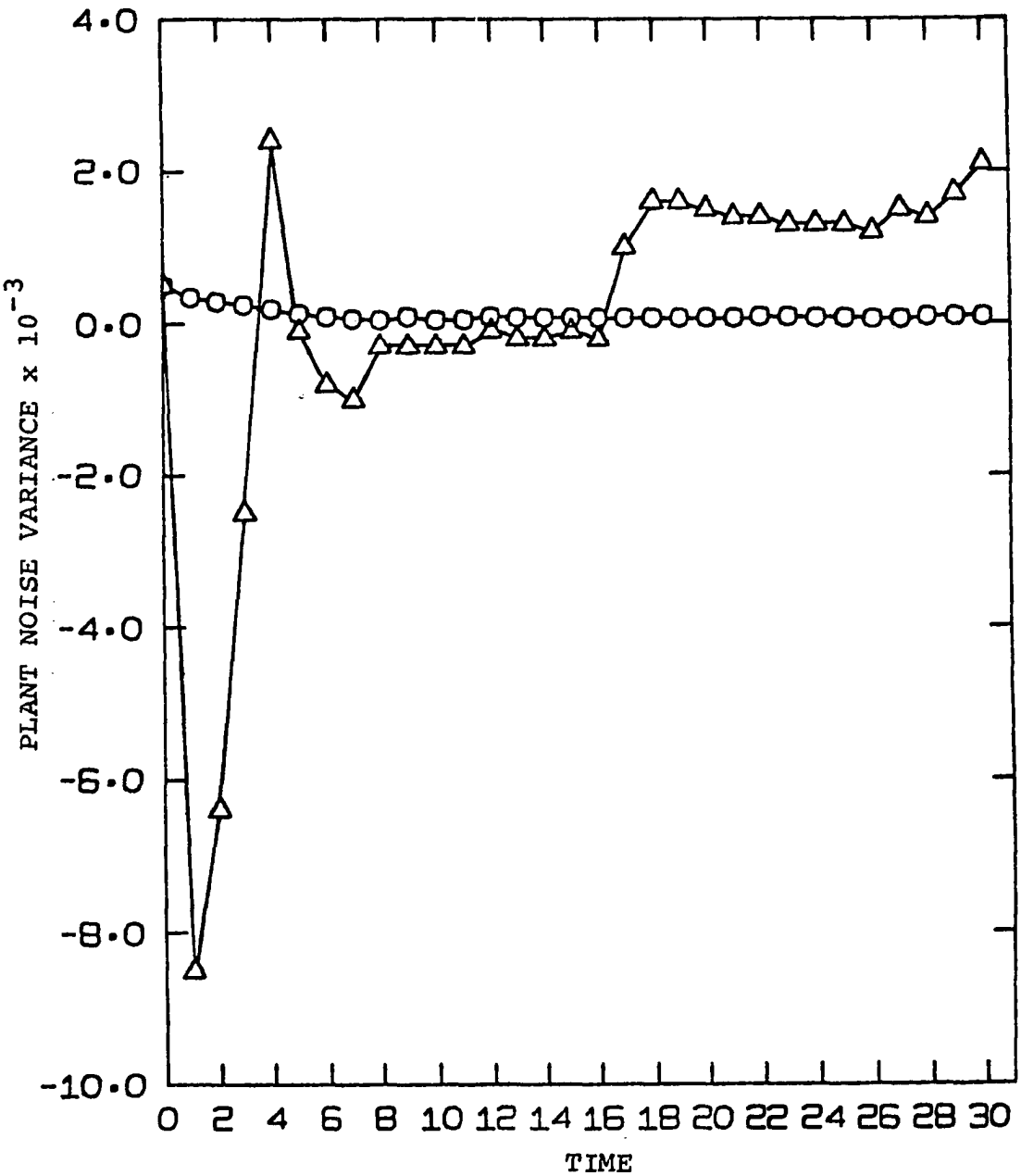


Figure 4.31. Adaptive Estimation of Plant Noise Variance, Q_{33}.

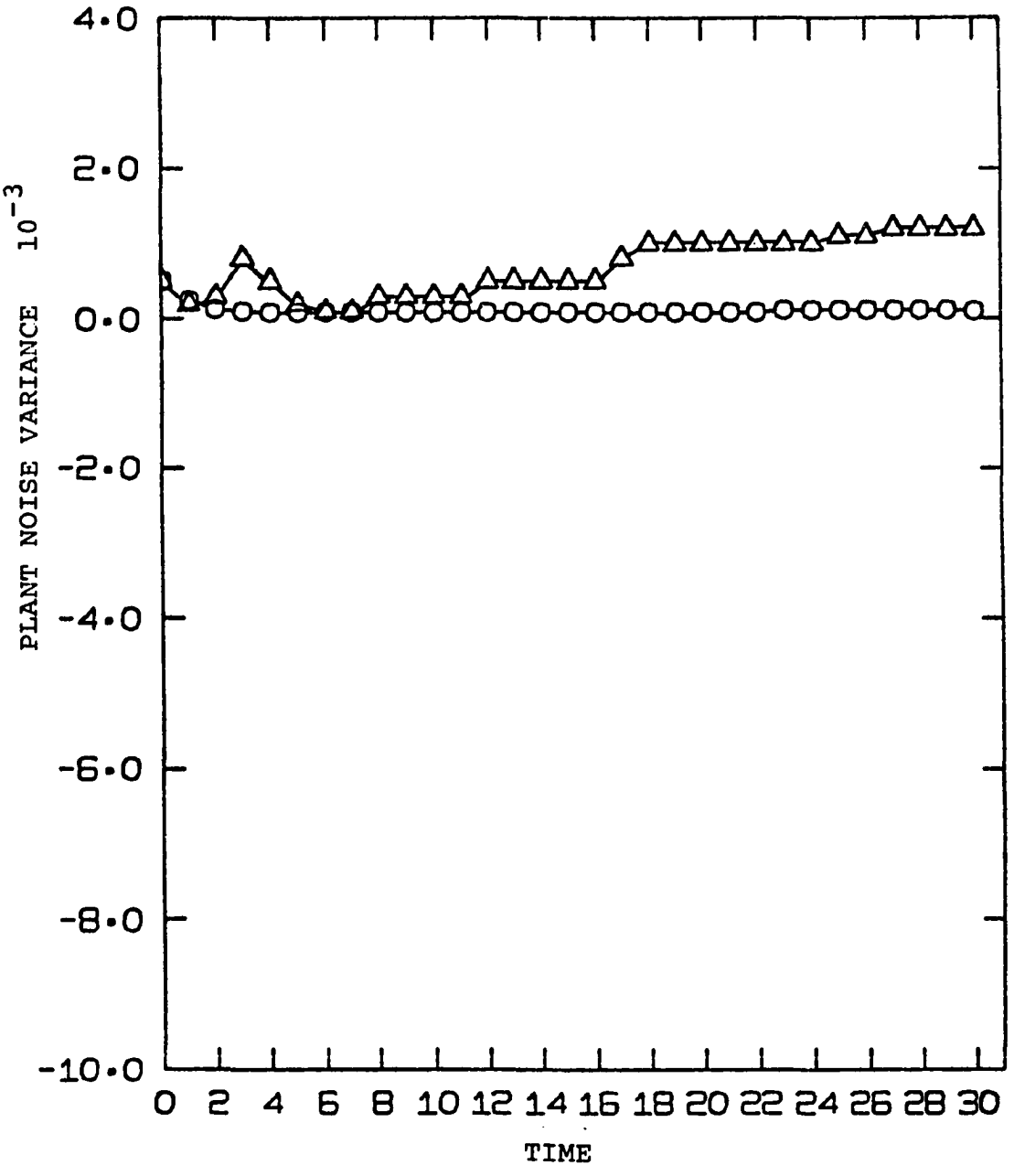


Figure 4.32. Adaptive Estimation of Plant Noise Variance, Q_{44}.

closer to the true value of Q than the Sage, Husa estimator. For a large number of estimation increments, the Sage, Husa estimator was found to converge more rapidly than the method proposed herein.

Figures 4.33 through 4.36 illustrate adaptive estimation of the diagonal elements of the R matrix. Again the estimator of this work converges rapidly to the region of the true value of R and stabilizes with slow convergence to the final estimate of R. The Sage, Husa R estimator oscillates with large amplitude for the first 6 time increments, then reduces the oscillatory behavior and converges slowly to the true value of R. At the end of the 30 time intervals shown, the Sage, Husa estimate of R was 2 to 5 times larger than the estimate from the proposed method, which was about twice as large as the true value of R.

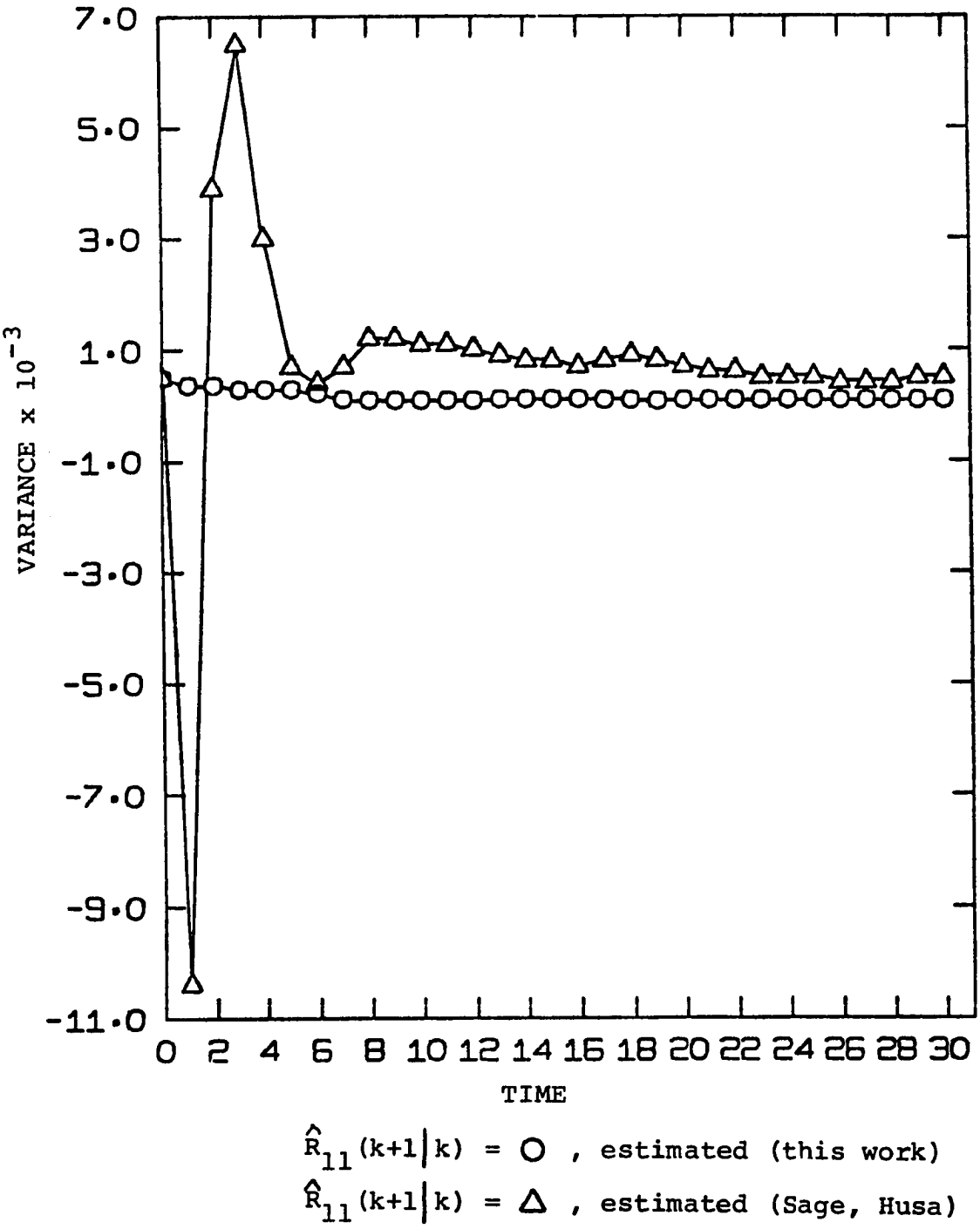


Figure 4.33. Adaptive Estimation of Observation Noise Variance, R_{11}.

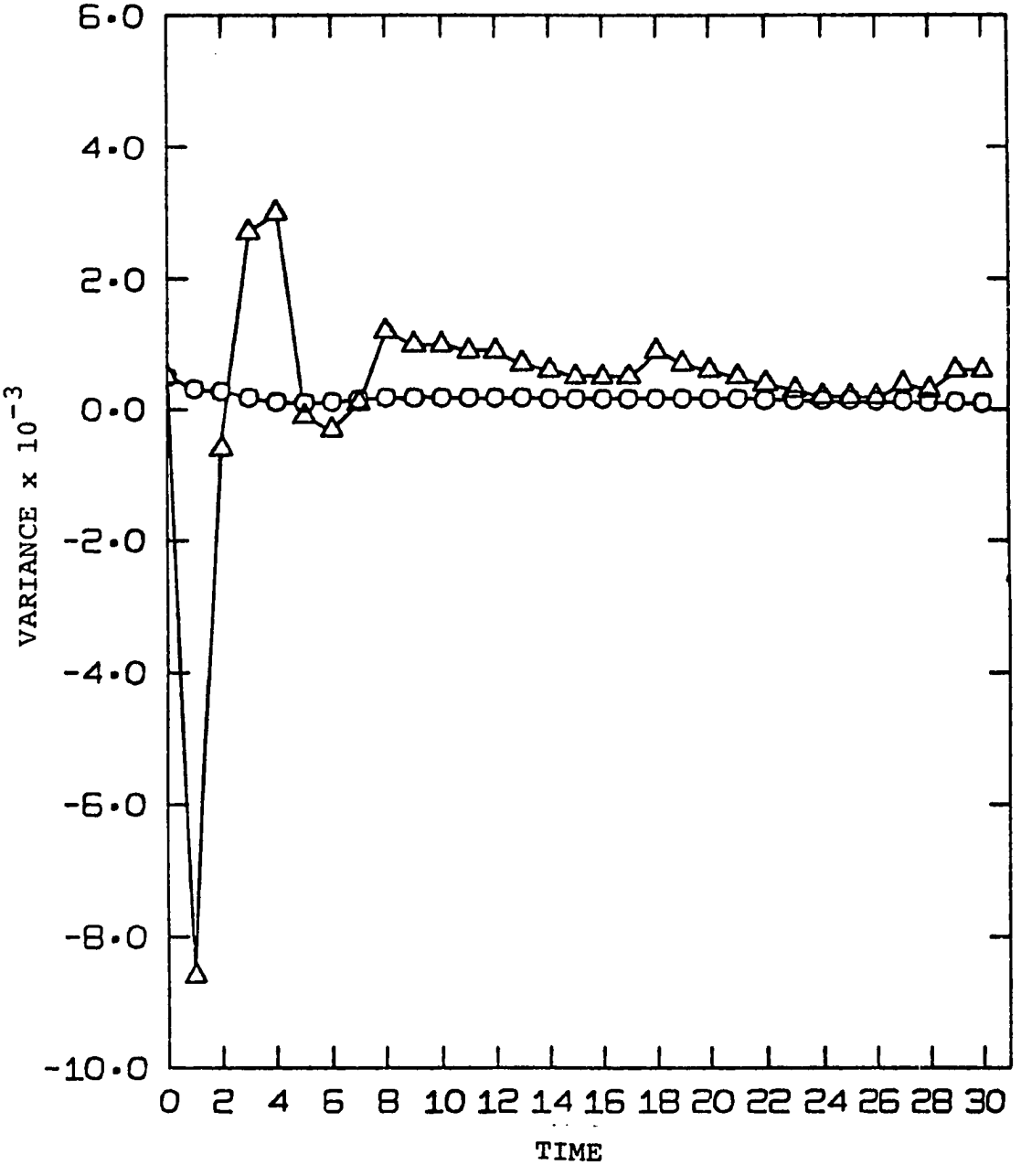


Figure 4.34. Adaptive Estimation of Observation Noise Variance, R_{22}.

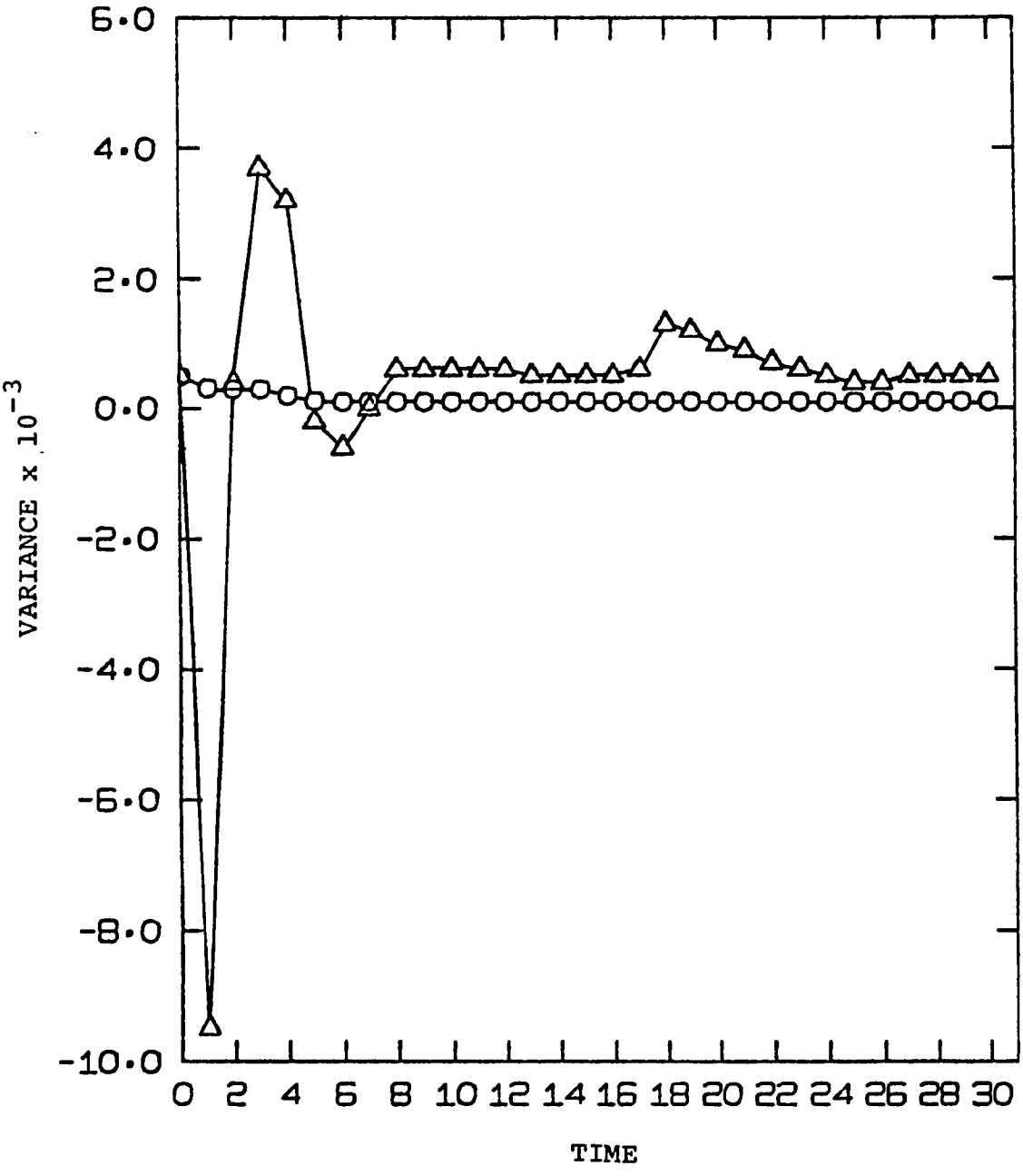


Figure 4.35. Adaptive Estimation of Observation Noise Variance, R_{33}.

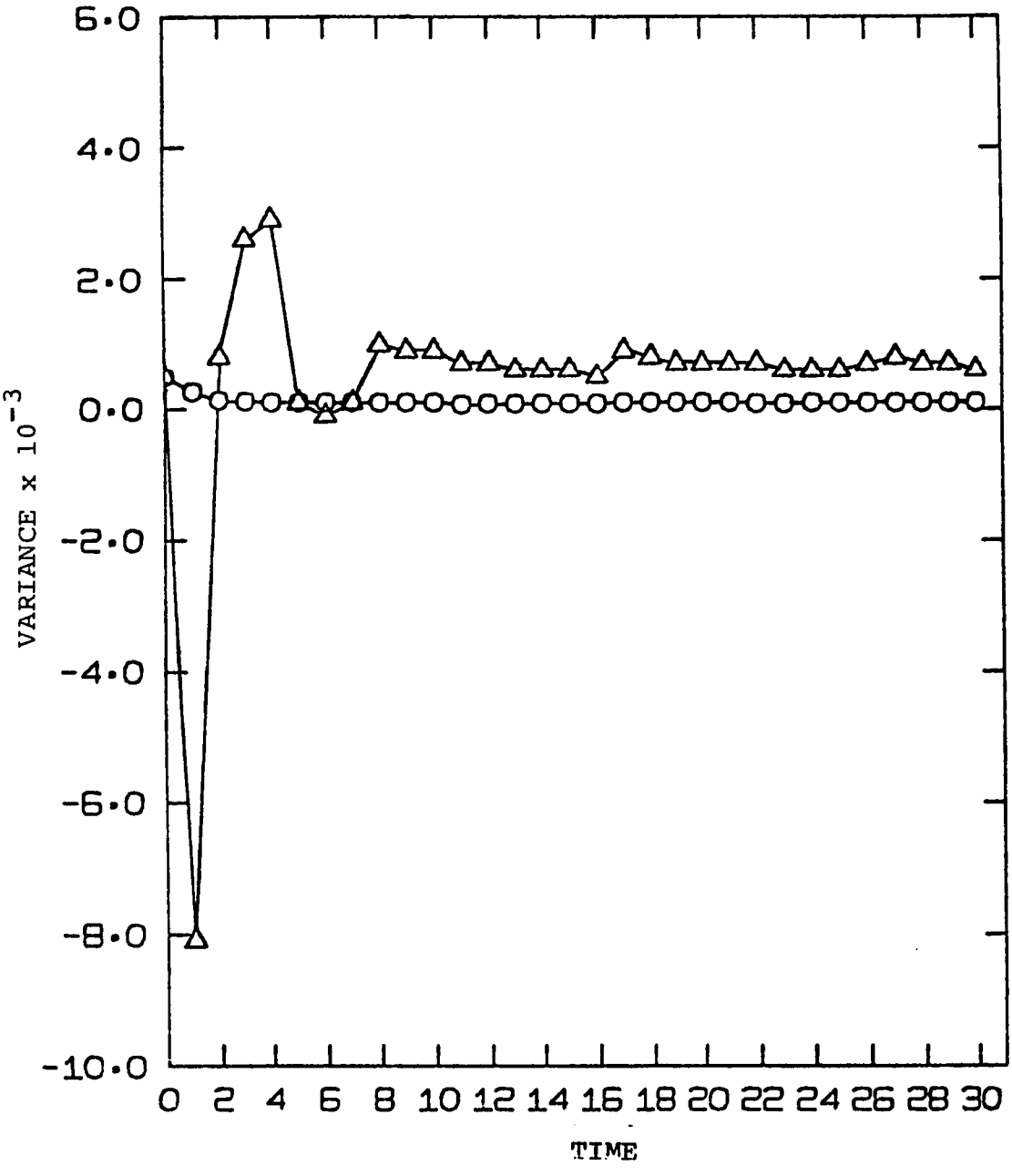


Figure 4.36. Adaptive Estimation of Observation Noise Variance, R_{44}.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

The results of this work indicate the feasibility of applying the adaptive filter to chemical processes. An algorithm for adaptive estimation of the state as well as the plant and observation noise covariance matrices has been developed. Examples of its use indicate acceptable performance of the method.

The limiting cases for zero plant noise covariance and zero observation noise covariance were considered to show the effects on the value of the optimal weighting matrix and how the model dynamics or observational data might be computationally ignored. The use of redundant measurements also was considered.

Incomplete models commonly occur for two reasons. Either the modeller unintentionally overlooks significant components of the model, or in trying to maintain the model as simple as possible, components are intentionally dropped. In either case, consequent modelling error was shown to cause filter divergence with subsequent deterioration of the estimate. In order to overcome this problem, an algorithm was developed to identify the random components which have been

ignored or overlooked in the model for the plant and observational process.

Finally, a comparison was made with a previously published adaptive filter of Sage and Husa (23). The filter developed in this work generally responded more rapidly with less oscillation. It was found that an "exponential smoothing filter" could be used in addition to the algorithm to reduce the noise in the estimate. Reduction of oscillatory or noisy estimates of the state is desirable when the estimate is used as input to a process controller. Although an iterative technique (Method of Wegstein) was needed to solve for estimates of the noise covariance matrices, this can be used to advantage at each estimation interval in order to refine the estimates. The refinement necessitates additional computational time but may be justified when the state estimate is destroyed while the covariance estimate is converging to the true value too slowly.

Several extensions of this work can be recommended. The estimates of the noise covariance matrices should be solved for explicitly. This might satisfactorily be done as Sage and Husa have made simplifying assumptions which led to a suboptimal, but explicit algorithm. Alternatively, the objective function could be optimized by a direct computational method such as a gradient technique.

A measure of the reliability of the estimates of Q and R could be developed. This may take a form such as the trace

of the error covariance matrix for the estimates of Q and R.

Random components of a stochastic model which are non-Gaussian should be considered, for example, the lognormal distribution which is basically as fundamental as the normal distribution. Also, the filter could be applied to the case for nonstationary noise covariance matrices, for "colored" noise inputs, or for the case where plant and measurement noise inputs are correlated. The latter may arise, for example, when plant noise is observed. The correlation becomes apparent by augmenting the measurement noise vector with the plant noise vector. "Colored" noise could be handled by adjoining to the state equations those equations for "linear signal filters" used to describe the "colored" noise generated from white noise inputs.

In this work, the discrete form of the filter was derived for purposes of implementation on the digital computer. But the properties of the filter could be more thoroughly evaluated for the continuous form of the filter. A nonrigorous limiting technique might be used to obtain the continuous case by allowing the samples to become dense such that as $k \to \infty$, then $k\Delta \to t$ and

$$F[x(t),t] \triangleq \lim_{k\to\infty} \left(f[x(k),k] - x(k)\right)/\Delta$$

$$\rho \triangleq \lim_{k\to\infty} \Delta R(k\Delta)$$

$$H(t) \triangleq \lim_{k\to\infty} H(k\Delta), \qquad \text{etc.}$$

Finally, the proposed technique could be extended to the smoothing problem using the density function p[Z(k+j) x(k), R(k), Q(k)] instead of p[Z(k) x(k), R(k), Q(k)].

Potential applications of adaptive filtering are diverse. A general application and perhaps the most obvious is identification of random inputs to an industrial process. A less obvious application of the extended Kalman filter is estimation of coefficients as well as the state property itself in physical property correlations. The sequential nature of the algorithm particularly is of value for correlations being fit to large data sets, and is much preferred for purposes of computer storage space economy. The adaptive estimation of random components might be of use where the correlation parameters have a random component, for example, the dependence of a heat transfer correlation on the random formation of nucleation sites during the boiling phenomenon.

BIBLIOGRAPHY

1. Aitchison, J., and Brown, J. A. C. The Lognormal Distribution. London, England: Cambridge University Press, 1963.

2. Aoki, M. Optimization of Stochastic Systems. New York: Academic Press, 1967.

3. Athens, M.; Wishner, R. P.; and Bertolini, A. "Suboptimal State Estimation for Continuous Time Nonlinear Systems from Discrete Noisy Measurements." IEEE Trans, Automatic Control, Vol. AC-13 (1968):504-514.

4. Bucy, R. S., and Joseph, P. D. Filtering for Stochastic Processes with Applications to Guidance. New York: Interscience Publishers, 1968.

5. Cox, H. "Sequential Minimax Estimation." IEEE Trans., Automatic Control, Vol. AC-11 (1966):323-324.

6. Franks, R. G. E. Modelling and Simulation in Chemical Engineering. New York: Wiley-Interscience, 1972.

7. Gauss, K. F. Theory of the Motion of the Heavenly Bodies Moving About the Sun in Conic Sections. A translation of Theoria Motus by C. H. Davis. New York: Dover Publications, Inc., 1963.

8. Heffes, J. "The Effect of Erroneous Models on the Kalman Filter Response." IEEE Trans., Automatic Control, Vol. AC-11 (1966):541-543.

9. IBM Application Program, System/360 Scientific Subroutine Package Version II Programmers' Manual. H20-0205-2, IBM Corp., White Plains, New York, 1967.

10. IEEE Trans., Automatic Control. "Special Issue on Linear-Quadratic-Gaussian Problem." Vol. AC-16, 6 (December 1971):527-869.

11. Jazwinski, A. H. "Adaptive Filtering." Automatica, Vol. 5 (1969):475-485.

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

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

14. Kalman, R. E., and Bucy, R. S. "New Results in Linear Filtering and Prediction Theory." Trans. ASME, J. Basic Engr., Series D, Vo. 83 (1961):95-108.

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

16. Larson, R. E. "Report on Estimation and Identification." Report for Bonneville Power Administration, Portland, Oregon, November 1968.

17. Lee, R. C. K. Optimal Estimation, Identification, and Control. Cambridge, Mass: M.I.T. Press, 1964.

18. Magill, D. T. "Optimal Adaptive Estimation of Sampled Stochastic Processes." IEEE Trans., Automatic Control, Vol. AC-10 (1965):434-439.

19. Magill, D. T. "Optimal Adaptive Estimation of Samples Stochastic Processes." Systems Theory Lab, SEL, Stanford University Technical Report 6302-3, Stanford, California, December 1963.

20. Meditch, J. S. "Sequential Estimation for Discrete Time Nonlinear Systems." Information Sciences Report No. 26, Boeing Scientific Research Lab, April 1969.

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

22. Mehra, R. K. "On the Identification of Variances and Adaptive Kalman Filtering." JACC, Boulder, Colorado, (1969):494-503.

23. Mollenkamp, R. A.; Smith, C. L.; and Corripio, A. B. "On-Line Model Identification and Control Using the Kalman Filter." JACC, St. Louis, Missouri (1971): 751-759.

24. Nishimura, T. "On the A Priori Information in Sequential Estimation Problems." IEEE Trans., Automatic Control, Vol. AC-11 (1966):197-204.

25. Perry, J. H., ed. Chemical Engineers' Handbook. New York: McGraw-Hill Book Company, 1963.

26. Rauch, H. E.; Tung, F.; and Striebel, C. T. "Maximum Likelihood Estimates of Linear Dynamic Systems." AIAA J., Vol. 3, 8 (1965):1445-1450.

27. Sage, A. P., and Husa, G. W. "Adaptive Filtering with Unknown Prior Statistics." JACC, Boulder, Colorado (1969):760-769.

28. Sage, A. P., and Melsa, J. L. Estimation Theory with Applications to Communications and Control. New York: McGraw-Hill Book Company, 1971.

29. Sengbush, R. L., and Lainiotis, D. G. "Simplified Parameter Quantization Procedure for Adaptive Estimation." IEEE Trans., Automatic Control, Vol. AC-14 (1969):424-425.

30. Shellenbarger, J. C. "A Multivariance Learning Technique for Improved Dynamic System Performance." Proc. Nat. Elect. Conf.,(1967):146-151.

31. Shellenbarger, J. C. "Estimation of Covariance Parameters for an Adaptive Kalman Fitler." Proc. Nat. Elect. Conf. (1966):698-702.

32. Smith, G. L. "Sequential Estimation of Observation Error Variances in a Trajectory Estimation Problem." AIAA J., Vol. 5 (1967):1964-1970.

33. Wegstein, J. H. "Accelerating Convergence of Iterative Processes." Communications of the ACM, 1, 9 (1958).

34. Wells, C. H. "Application of Modern Estimation and Identification Techniques to Chemical Processes." AIChE J., Vol. 17, 4 (1971):966-973.

APPENDICES

APPENDIX A

MATRIX INVERSION LEMMA

A.1 The Lemma

A is an n-square matrix; B and C are n x m matrices. If A and the two matrices $(A + BC^T)$ and $(I + C^TA^{-1}B)$ are nonsingular, then a matrix indentity may be obtained,

$$(A + BC^T)^{-1} \equiv A^{-1} - A^{-1}B(I + C^TA^{-1}B)^{-1}C^TA^{-1}$$

A derivation is available in Sage and Melsa (28).

A.2 A Special Case of Interest

Let $A \triangleq P^{-1}$, $B \triangleq h_x^TR^{-1}$ and $C^T \triangleq h_x$. Substituting for A, B and C^T

$$\begin{aligned}(P^{-1} + h_x^TR^{-1}h_x)^{-1} &\equiv P - Ph_x^TR^{-1}(I + h_xPh_x^TR^{-1})^{-1}h_xP \\ &= P - Ph_x^T(R + h_xPh_x^T)^{-1}h_xP\end{aligned}$$

where P and R are symmetric and positive definite.

A.3 An Alternate Expression

Show: $(P - Wh_xP)h_x^TR^{-1} = W$ using the definition $W \triangleq Ph_x(R + h_xPh_x^T)^{-1}$.

Proof: Reduce the above definition of W in the following successive steps.

$$W(R + h_x Ph_x^T) \equiv Ph_x$$

$$W(I + h_x Ph_x^T R^{-1}) = Ph_x R^{-1}$$

$$W = Ph_x R^{-1} - Wh_x Ph_x^T R^{-1}$$

$\therefore$ $W = (P - Wh_x P)\, h_x^T R^{-1}$ which is an alternate expression for W.

APPENDIX B

VECTOR AND MATRIX OPERATIONS

B.1 Vector Differentiation

Let x be an n-vector; z is an m-vector; and M is an m-square matrix.

1. Prove: $\partial(x^T Mz)/\partial x = Mz$

Proof:

$$x^T \cdot M \equiv \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \cdot \begin{bmatrix} M_{11} & \cdots & M_{1m} \\ \vdots & & \vdots \\ M_{n1} & \cdots & M_{nm} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_i^n x_i M_{i1} & \cdots & \sum_i^n x_i M_{im} \end{bmatrix}$$

$$\equiv \left[\sum_i^n x_i M_{ij} \right], \quad j = 1 \ldots n$$

$$\text{and } x^T \cdot M \cdot z = \begin{bmatrix} \sum_i^n x_i M_{i1} & \cdots & \sum_i^n x_i M_{im} \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}$$

$$= z_1 \sum_i^n x_i M_{i1} + \ldots z_m \sum_i^n x_i M_{im}$$

$$\equiv \left[\sum_j^m z_j \sum_i^n x_i M_{ij} \right]$$

$$\frac{\partial\, x^T \cdot M \cdot z}{\partial x} = \begin{bmatrix} \partial\, x^T Mz\, /\partial x_1 \\ \vdots \\ \partial\, x^T Mz\, /\partial x_n \end{bmatrix}$$

$$\equiv \begin{bmatrix} \partial \sum_j z_j \sum_i x_i M_{ij}\, /\partial x_1 \\ \vdots \\ \partial \sum_j z_j \sum_i x_i M_{ij}\, /\partial x_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum_j z_j M_{ij} \\ \vdots \\ \sum_i z_j M_{nj} \end{bmatrix}, \text{ where } \frac{\partial x_i}{\partial x_j} = \delta_{ij}$$

$$= \begin{bmatrix} M_{11} & \cdots & M_{1m} \\ \vdots & & \vdots \\ M_{n1} & & M_{nm} \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}$$

$$= M \cdot z$$

2. Prove: $\partial\left[x^T Mx\right]/\partial x = 2Mx$

Proof: $x^T Mx = \sum_j x_j \sum_i x_i M_{ij}$ from previous proof.

and $\frac{\partial}{\partial x} \triangleq \begin{bmatrix} \partial/\partial x_i \\ \vdots \\ \partial/\partial x_n \end{bmatrix}$

expanded,

$$x^T Mx = \sum_j x_j (x_1 M_{1j} + \cdots x_n M_{nj})$$

$$= x_1(x_1 M_{11} + \cdots x_n M_{n1}) + x_2(x_1 M_{12} + \cdots x_n M_{n2}) + \cdots x_n(x_1 M_{1n} + \cdots x_n M_{nn})$$

$$\therefore \begin{matrix} \partial x^T Mx/\partial x_1 \\ \vdots \\ \partial x^T Mx/\partial x_n \end{matrix} = \begin{matrix} (x_1 M_{11} + \ldots x_n M_{n1}) + x_1 M_{11} + x_2 M_{12} + \ldots x_n M_{1n} \\ \vdots \qquad\qquad\qquad\qquad \vdots \qquad\qquad \vdots \\ x_1 M_{n1} + x_2 M_{n2} + \ldots x_1 M_{1n} + \ldots x_n M_{nn} + x_n M_{nn} \end{matrix}$$

$$\frac{\partial x^T Mx}{\partial x} = \begin{bmatrix} \sum_i^n x_i M_{i1} + \sum_i^n x_i M_{1i} \\ \vdots \\ \sum_i^n x_i M_{in} + \sum_i^n x_i M_{ni} \end{bmatrix}$$

$$= \begin{bmatrix} M_{11} & \cdots & M_{n1} \\ \vdots & & \vdots \\ M_{1n} & \cdots & M_{nn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} M_{11} & \cdots & M_{1n} \\ \vdots & & \vdots \\ M_{n1} & \cdots & M_{nn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= M^T \cdot x + M \cdot x$$

$$= \left[M^T + M\right] \cdot x$$

$$= 2Mx \text{ if } M^T = M$$

3. Prove: $\partial z^T Mx/\partial x = M^T z$

Proof: From previous proof, $z^T M = \sum_i^m z_i M_{ij}$; $j = 1,\ldots n$

Similarly, $z^T Mx = \sum_j^n x_j \sum_i^m z_i M_{ij}$

$$\therefore \quad \frac{\partial z^T Mx}{\partial x} \triangleq \begin{bmatrix} \partial/\partial x_1 \\ \vdots \\ \partial/\partial x_n \end{bmatrix} z^T Mx$$

$$= \begin{bmatrix} \sum_i^m z_i M_{i1} \\ \vdots \\ \sum_i^m z_i M_{in} \end{bmatrix}$$

$$= \begin{bmatrix} M_{11} & \cdots & M_{m1} \\ \vdots & & \vdots \\ M_{1n} & \cdots & M_{mn} \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}$$

$$= M^T \cdot z$$

B.2 Vector Gradient Identities

Let $h(x)$ and z be m-vectors; x is an n-vector; M is an $m \times m$ matrix; and h_x^T is an $n \times m$ matrix. Also $H_{i\ell} \overset{\Delta}{=} \partial h_i / \partial x_\ell$

1. Prove: $\partial \left[h^T(x) Mz \right] / \partial x = h_x^T(x) Mz$

Proof: $h^T Mz = \sum_j^m z_j \sum_i^m h_i M_{ij}$

$$\therefore \quad \frac{\partial h^T Mz}{\partial x} = \begin{bmatrix} \vdots \\ \sum_j^m z_j \sum_i^m \frac{\partial h_i}{\partial x_\ell} M_{ij} \\ \vdots \end{bmatrix}, \quad \ell = 1 \ldots n$$

$$= \begin{bmatrix} \vdots \\ \sum_j^m z_j \sum_i^m H_{i\ell} M_{ij} \\ \vdots \end{bmatrix}$$

$$= \begin{bmatrix} H_1^T Mz \\ \vdots \\ H_n^T Mz \end{bmatrix}$$

$$= H^T Mz$$

$$= h_x^T Mz$$

2. Prove: $\partial \left[h^T(x) Mh(x) \right] / \partial x = 2 h_x^T(x) Mh(x)$

Proof: $h^T Mh = \sum_j^m h_j \sum_i^m h_i M_{ij}$

$$\frac{\partial h^T M h}{\partial x} = \begin{bmatrix} \vdots \\ \sum_j H_{j\ell} \sum_i h_i M_{ij} + \sum_i h_j \sum_i H_{i\ell} M_{ij} \\ \vdots \end{bmatrix}$$

$$= \begin{bmatrix} \vdots \\ H_\ell^T M^T h + H_\ell^T M h \\ \vdots \end{bmatrix}, \quad \ell = 1 \ldots n$$

$$= h_x^T M^T h + h_x^T M h$$

$$\equiv 2 h_x^T M h \text{ if } M^T = M$$

3. Prove: $\partial \left[z^T M h(x) \right] / \partial x = h_x^T(x) M^T z$

Proof: $z^T M h = \sum_j h_j \sum_i z_i M_{ij}$

$$\frac{\partial z^T M h}{\partial x} = \begin{bmatrix} \vdots \\ \sum_j \frac{\partial h_j}{\partial x_\ell} \sum_i z_i M_{ij} \\ \vdots \end{bmatrix}, \quad \ell = 1 \ldots n$$

$$= \begin{bmatrix} \vdots \\ \sum_j H_{j\ell} \sum_i z_i M_{ij} \\ \vdots \end{bmatrix}$$

$$= \begin{bmatrix} \vdots \\ H_\ell^T M^T z \\ \vdots \end{bmatrix}$$

$$= h_x^T M^T z$$

B.3 Matrix Gradient Identities

1. Prove: $\partial |A| / \partial A = |A| A^{T^{-1}}$

Matrix A is n x n. $A_{ij} \triangleq$ cofactor of a_{ij} in A matrix.

Proof: Consider differentiation with respect to one element a_{ij}.

The determinant $|A|$ may be expanded by a column of elements as

$$|A| = \sum_{i}^{n} a_{ik}A_{ik} \text{ for any kth column.}$$

For differentiation choose element a_{ij}, then expand the determinant $|A|$ about the column containing this element,

$$\frac{\partial |A|}{\partial a_{ij}} = \frac{\partial}{\partial a_{ij}} \sum_{i=1}^{n} a_{ij}A_{ij}$$

$$= \sum_{i=1}^{n} \frac{\partial a_{ij}}{\partial a_{ij}} A_{ij}$$

$$= \sum_{i=1}^{n} \delta(i-j)\ A_{ij}$$

$$= A_{ij}, \text{ where } \delta(i-j) \overset{\Delta}{=} \text{Kronecker delta}$$

$$\Rightarrow \frac{\partial |A|}{\partial A} = (\text{adjoint } A)^T$$

By definition $A^{-1} \overset{\Delta}{=} \frac{\text{adjoint } A}{|A|}$

$$\therefore \frac{\partial |A|}{\partial A} = |A| \cdot A^{T^{-1}}$$

2. Prove: $\partial|A+B|/\partial A = |A+B|\,(A+B)^{T^{-1}}$

A and B are n x n matrices.

Proof: $|A + B| \overset{\Delta}{=} |C|$

$$\overset{\Delta}{=} \sum_{i=1}^{n} c_{ik}C_{ik} \text{ for any kth column}$$

$$\frac{\partial |C|}{\partial a_{ij}} = \sum_{i=1}^{n} \frac{\partial (a_{ij} + b_{ij})}{\partial a_{ij}} C_{ij}$$

$$= \sum_{i=1}^{n} \delta(i-j)\, C_{ij}$$

$$= C_{ij}$$

$$\therefore \quad \frac{\partial |C|}{\partial A} = (\text{adjoint } C)^T$$

also, $C^{-1} \triangleq \dfrac{\text{adjoint } C}{|C|}$

$$\therefore \frac{\partial |A + B|}{\partial A} = |A + B| (A + B)^{T^{-1}}$$

3. Prove: $\partial (x^T A^{-1} z)/\partial A = -A^{T^{-1}} x z^T A^{T^{-1}}$

Matrix A is n x n; x is an n-vector; z is an m-vector.

Proof: $\dfrac{\partial (x^T A^{-1} z)}{\partial a_{ij}} = x^T \left(\dfrac{\partial A^{-1}}{\partial a_{ij}}\right) z$

where differentiation is carried out with respect to a scalor, of which only A is a function.

(1) $\partial A^{-1}/\partial a_{ij}$ may be obtained as follows:

Since $\dfrac{\partial (AA^{-1})}{\partial a_{ij}} = \dfrac{\partial A}{\partial a_{ij}} A^{-1} + A \dfrac{\partial A^{-1}}{\partial a_{ij}}$

$$= \frac{\partial I}{\partial a_{ij}} = 0$$

Then $A \dfrac{\partial A^{-1}}{\partial a_{ij}} = \dfrac{-\partial A}{\partial a_{ij}} A^{-1}$

or $\dfrac{\partial A^{-1}}{\partial a_{ij}} = -A^{-1} \dfrac{\partial A}{\partial a_{ij}} A^{-1}$

where $\partial A/\partial a_{ij}$ is a matrix of zeroes, except for the ij^{th} element which is 1.

$$\therefore \frac{\partial x^T A^{-1} z}{\partial a_{ij}} = -x^T A^{-1} \frac{\partial A}{\partial a_{ij}} A^{-1} z$$

(2) Rename for notational convenience $B \triangleq A^{-1}$

$$\therefore -x^T A^{-1} \frac{\partial A}{\partial a_{ij}} A^{-1} z \equiv -x^T B \frac{\partial A}{\partial a_{ij}} Bz$$

using $x^T B \equiv \left[\ldots \sum_i^n x_i b_{ij} \ldots \right]$, $j = 1 \ldots n$

and $$Bz \equiv \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$$

$$= \begin{bmatrix} \vdots \\ \sum_k^n b_{\ell k} z_k \\ \vdots \end{bmatrix}, \quad \ell = 1 \ldots n$$

then $$\frac{\partial A}{\partial a_{ij}} Bz = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & 1_{ij} & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \cdot \begin{bmatrix} \sum_k^n b_{\ell k} z_k \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ \vdots \\ \sum_k^n b_{jk} z_k \\ \vdots \\ 0 \end{bmatrix}$$ all elements are zero, except row j.

$$\therefore -x^T B \frac{\partial A}{\partial a_{ij}} Bz = -\left[\ldots \sum_\ell^n x_\ell b_{\ell i} \ldots \right] \cdot \begin{bmatrix} 0 \\ \vdots \\ \sum_k^m b_{jk} z_k \\ \vdots \\ 0 \end{bmatrix}$$

$$= - \sum_{\ell}^{n} x_\ell b_{\ell i} \sum_{k}^{m} b_{jk} z_k$$

is the only nonzero product.

(3) Using, $\begin{bmatrix} \sum_{\ell}^{n} x_\ell b_{\ell i} \end{bmatrix} \equiv \begin{bmatrix} b_{11} & \cdots & b_{n1} \\ \vdots & & \vdots \\ b_{1n} & \cdots & b_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \equiv B_x^T$

and $\begin{bmatrix} \cdots & \sum_{k}^{n} b_{jk} z_k & \cdots \end{bmatrix} \equiv \begin{bmatrix} z_1 & \cdots & z_n \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{n1} \\ \vdots & & \vdots \\ b_{1n} & \cdots & b_{nn} \end{bmatrix} \equiv z^T B^T$

then $-x^T B \frac{\partial A}{\partial a_{ij}} Bz$, or $-x^T A^{-1} \frac{\partial A}{\partial a_{ij}} A^{-1} z$, is the ij^{th} element of the product $-B^T x z^T B^T$, or $-A^{-1^T} x z^T A^{-1^T}$.

4. Prove: $\partial(x^T(A+B)^{-1}z)/\partial A = -(A+B)^{T^{-1}} x z^T (A+B)^{T^{-1}}$

Proof: $\frac{\partial(x^T(A+B)^{-1}z)}{\partial a_{ij}} = x^T \frac{\partial(A+B)^{-1}z}{\partial a_{ij}}$

(1) Similar to previous proof,

$$\frac{\partial(A+B)^{-1}}{\partial a_{ij}} = - (A+B)^{-1} \frac{\partial A}{\partial a_{ij}} (A+B)^{-1}$$

$$\therefore \frac{\partial(x^T(A+B)^{-1}z)}{\partial a_{ij}} = -x^T(A+B)^{-1} \frac{\partial A}{\partial a_{ij}} (A+B)^{-1}z$$

(2) Similar to previous proof,

$$\therefore -x^T(A+B)^{-1} \frac{\partial A}{\partial a_{ij}} (A+B)^{-1}z = - \sum_{\ell}^{n} x_\ell c_{\ell i} \sum_{k}^{n} c_{mk} z_k$$

where $C \triangleq (A + B)^{-1}$

(3) Similar to previous proof,

$-x^T(A+B)^{-1} \frac{\partial A}{\partial a_{ij}} (A+B)^{-1}z$ is the ij^{th} element of the product $-(A+B)^{-1^T} xz^T (A+B)^{-1^T}$.

B.4 A Useful Identity

Prove: $\frac{\partial J}{\partial A} \equiv \frac{\partial}{\partial A} \{|B+A|^{-1/2} \operatorname{Exp}[-\frac{1}{2} x^T (B+A)^{-1} z]\}$

$$= -\frac{1}{2} |B+A|^{-1/2} \operatorname{Exp}[-\frac{1}{2} x^T (B+A)^{-1} z]$$
$$\cdot [(B+A)^{T^{-1}} - (B+A)^{T^{-1}} xz^T (B+A)^{T^{-1}}]$$

Proof: $\frac{\partial J}{\partial A} = -\frac{1}{2} |B+A|^{-3/2} \frac{\partial |B+A|}{\partial A} \operatorname{Exp}[-\frac{1}{2} x^T (B+A)^{-1} z]$

$$- \frac{1}{2} |B+A|^{-1/2} \operatorname{Exp}[-\frac{1}{2} x^T (B+A)^{-1} z] \cdot \frac{\partial x^T (B+A)^{-1} z}{\partial A}$$

from previous gradient matrix identities,

$$\frac{\partial J}{\partial A} = -\frac{1}{2} |B+A|^{-3/2} |B+A| (B+A)^{T^{-1}} [\operatorname{Exp} -\frac{1}{2} x^T (B+A)^{-1} z]$$
$$+ \frac{1}{2} |B+A|^{-1/2} \operatorname{Exp}[-\frac{1}{2} x^T (B+A)^{-1} z] (B+A)^{T^{-1}} xz^T (B+A)^{T^{-1}}$$

reducing,

$$\frac{\partial J}{\partial A} = -\frac{1}{2} |B+A|^{-1/2} \operatorname{Exp}[-\frac{1}{2} x^T (B+A)^{-1} z] \cdot [(B+A)^{T^{-1}}$$
$$- (B+A)^{T^{-1}} xz (B+A)^{T^{-1}}].$$

APPENDIX C

NECESSARY AND SUFFICIENT CONDITIONS

C.1 Necessary Conditions of Optimization

The necessary condition for minimization is required for solution of the optimal adaptive estimation problem. This condition is expressed as

$$\frac{\partial L}{\partial S} = 0 \; ; \; S \triangleq \{x(k), Q, R\}$$

where the cost function L is obtained in Appendix E. The requisite vector and matrix gradients may be obtained using the identities in Appendix B.

1. $\partial L/\partial x(k) = 0$

Considering only those terms in L which are a function of $x(k)$,

$$0 = \frac{\partial}{\partial x(k)} \left\{ \text{Exp}\left[-\frac{1}{2}\{z(k) - h[x(k)] - r\} R^{-1} \{z(k) - h[x(k)] - r\}^T \right.\right.$$
$$\left.\left. - \frac{1}{2}\{x(k) - \hat{x}(k|k-1)\} P^{-1}(k|k-1) \{x(k) - \hat{x}(k|k-1)\}^T \right] \right\}$$

$$0 = \frac{\partial}{\partial x(k)} \{-h^T[x(k)]R^{-1}z(k) - z^T(k)R^{-1}h[x(k)]$$
$$+ h^T[x(k)]R^{-1}h[x(k)] + r^TR^{-1}h[x(k)] + h^T[x(k)]R^{-1}r\}$$
$$+ \frac{\partial}{\partial x(k)} \{x^T(k)P^{-1}(k|k-1)x(k) - \hat{x}^T(k|k-1)P^{-1}(k|k-1)x(k)$$
$$- x^T(k)P^{-1}(k|k-1)\hat{x}(k|k-1)\}$$

where terms which are not a function of x(k) have not been retained.

Differentiation may be completed using the vector gradient identities of Appendix B.

$$0 = [-h_x^TR^{-1}z(k) - h_x^TR^{-1}z(k) + 2h_x^TR^{-1}h + h_x^TR^{-1}r + h_x^TR^{-1}r]$$
$$+ [2P^{-1}(k|k-1)x(k) - P^{-1}(k|k-1)\hat{x}(k|k-1)$$
$$- P^{-1}(k|k-1)\hat{x}(k|k-1)]$$

where h and h_x are evaluated at $\hat{x}(k|k-1)$. From the series approximation for h[x(k)],

$$h_x[x(k)] = h_x[\hat{x}(k|k-1)]$$

After substitution and utilization of the symmetry of R and P(k|k-1) the final expression may be rearranged for x(k).

$$x(k) = [h_x^TR^{-1}h_x + P^{-1}(k|k-1)]^{-1} [P^{-1}(k|k-1)\hat{x}(k|k-1)$$
$$+ h_x^TR^{-1}\{z(k) - h + h_x\hat{x}(k|k-1) - r\}]$$

2. $\partial L/\partial R = 0$

$$0 = \frac{\partial}{\partial R}\left\{|R|^{-1/2}\ |h_x[\hat{x}(k|k-1)]P(k|k-1)h_x^T[\hat{x}(k|k-1)] + R|^{-1/2}\right.$$
$$\cdot \mathrm{Exp}\left[-\tfrac{1}{2}\{z(k) - h[x(k)] - r\}^T R^{-1}\{z(k) - h[x(k)] - r\}\right.$$
$$- \tfrac{1}{2}\{z(k) - h[\hat{x}(k|k-1)] - r\}^T$$
$$\cdot \{h_x[\hat{x}(k|k-1)]P(k|k-1)h_x^T[\hat{x}(k|k-1)] + R\}^{-1}$$
$$\left.\left.\cdot \{z(k) - h[\hat{x}(k|k-1)] - r\}\right]\right\}$$

where terms not a function of R have been dropped. By the chain rule of differentiation,

$$0 = \mathrm{Exp}\ (...)\{|h_x P h_x^T + R|^{-1/2}\ \frac{\partial |R|^{-1/2}}{\partial R}$$
$$+ |R|^{-1/2}\ \frac{\partial}{\partial R}\ |h_x P h_x^T + R|^{-1/2} + |R|^{-1/2}$$
$$\cdot |h_x P h_x^T + R|^{-1/2}\ \frac{\partial}{\partial R}\ [-\tfrac{1}{2}\|z - h - r\|^2_{R^{-1}}$$
$$- \tfrac{1}{2}\|z - \hat{h} - r\|^2_{Y^{-1}}]\}$$

For notational convenience, the argument of Exp is not repeated explicitly since there is no change between steps. The subscripts have been deleted also, but with no loss of clarity however. Finally, $\hat{h}$ merely denotes h evaluated at $\hat{x}(k|k-1)$.

Differentiation is completed using matrix gradient identities in Appendix B.

$$0 = -\frac{1}{2}\,\mathrm{Exp}(\ldots)\{|Y|^{-1/2}\ |R|R^{-1}\ |R|^{-3/2}$$
$$+\ |R|^{-1/2}\ |Y|^{-3/2}\ |Y|(Y)^{-1}$$
$$+\ |R|^{-1/2}\ |Y|^{-1/2}[-R^{-1}(z-h-r)(z-h-r)^{T}R^{-1}$$
$$-\ (Y)^{-1}(z-\hat{h}-r)(z-\hat{h}-r)^{T}(Y)^{-1}]\}$$

The matrix $Y(k|k-1)$ has previously been derived as $\{h_x[x(k|k-1)]P(k|k-1)h_x^T[x(k|k-1)] + R\}$. The abbreviated notation is used here however. Also, the symmetry of R and Y has been utilized. Factoring out the nontrivial value of R which satisfies this equation,

$$0 = R^{-1} + Y^{-1} - R^{-1}BR^{-1} - Y^{-1}CY^{-1}$$

where Y implicitly depends on R as previously derived. B and C are defined for notation convenience only as

$$B \triangleq (z - h - r)(z - h - r)^{T}$$
$$C \triangleq (z - \hat{h} - r)(z - \hat{h} - r)^{T}$$

3. $\partial L/\partial Q = 0$

$$0 = \frac{\partial}{\partial Q}\Big\{|f_x[\hat{x}(k-1|k-1)]P(k-1|k-1)f_x^T[\hat{x}(k-1|k-1)] + Q|^{-1/2}$$
$$\cdot\ |h_x[\hat{x}(k|k-1)]P(k|k-1)h_x^T[\hat{x}(k|k-1)] + R|^{-1/2}$$
$$\cdot\ \mathrm{Exp}\Big[-\frac{1}{2}\{x(k)-f[\hat{x}(k-1|k-1)]-q\}^{T}\{f_x[\hat{x}(k-1|k-1)]$$
$$\cdot\ P(k-1|k-1)f_x^T[\hat{x}(k-1|k-1)] + Q\}^{-1}$$
$$\cdot\ \{x(k)-f[\hat{x}(k-1|k-1)]-q\} - \frac{1}{2}\{z(k)-h[\hat{x}(k|k-1)]-r\}^{T}$$
$$\cdot\ \{h_x[\hat{x}(k|k-1)]P(k|k-1)h_x^T[\hat{x}(k|k-1)]+R\}^{-1}$$
$$\{z(k) - h[\hat{x}(k|k-1)] - r\}^{T}\Big]\Big\}$$

By the chain rule of differentiation,

$$
\begin{aligned}
0 = \mathrm{Exp}(\ldots)\{ & |P(k|k-1)|^{-1/2} \, |Y(k|k-1)|^{-1/2} \\
& \cdot \frac{\partial}{\partial Q} \left[-\frac{1}{2} \| x(k) - \hat{x}(k|k-1) \|^{2}_{P^{-1}(k|k-1)} \right. \\
& \qquad \left. - \frac{1}{2} \| z(k) - \hat{z}(k|k-1) \|^{2}_{Y^{-1}(k|k-1)} \right] \\
& + |P(k|k-1)|^{-1/2} \, \frac{\partial |Y(k|k-1)|^{-1/2}}{\partial Q} \\
& + |Y(k|k-1)|^{-1/2} \, \frac{\partial}{\partial Q} \, |P(k|k-1)|^{-1/2} \}
\end{aligned}
$$

where we have previously defined

$$
\begin{aligned}
P(k|k-1) &= f_x[\hat{x}(k-1|k-1)]P(k-1|k-1)f_x^T[\hat{x}(k-1|k-1)] + Q \\
Y(k|k-1) &= h_x[\hat{x}(k|k-1)]P(k|k-1)h_x^T[\hat{x}(k|k-1)] + R \\
\hat{x}(k|k-1) &= f[\hat{x}(k-1|k-1)] + q \\
\hat{z}(k|k-1) &= h[\hat{x}(k|k-1)] + r
\end{aligned}
$$

For notational brevity, the argument of Exp is deleted.

The following two steps are carried out by utilizing matrix gradient identities from Appendix B; the symmetry of $P(k|k-1)$ and $Y(k|k-1)$ is noted; and the exponential term is factored out.

$$0 = -\tfrac{1}{2}\,\mathrm{Exp}(\ldots)\Big\{|P(k|k-1)|^{-1/2}\;|Y(k|k-1)|^{-1/2}$$

$$\cdot\;\big[-P^{-1}(k|k-1)\,[x(k)-\hat{x}(k|k-1)]\,[x(k)-\hat{x}(k|k-1)]^{T}P^{-1}(k|k-1)$$

$$-\;h_x^{T}Y^{-1}(k|k-1)\,[z(k)-\hat{z}(k|k-1)]\,[z(k)-\hat{z}(k|k-1)]^{T}$$

$$\cdot\;Y^{-1}(k|k-1)h_x\big] + |P(k|k-1)|^{-1/2}\;|Y(k|k-1)|^{-3/2}h_x^{T}$$

$$\cdot\;Y^{-1}(k|k-1)h_x\;Y(k|k-1)$$

$$+\;|Y(k|k-1)|^{-1/2}|P(k|k-1)|^{-3/2}|P(k|k-1)|P^{-1}(k|k-1)\Big\}$$

$$\therefore\quad 0 = -P^{-1}AP^{-1} - h_x^{T}Y^{-1}DY^{-1}h_x + h_x^{T}Y^{-1}h_x + P^{-1}$$

where P and Y implicitly depend upon Q. For notational convenience,

$$P \equiv P(k|k-1) \quad \text{and} \quad Y \equiv Y(k|k-1)$$

$$A \triangleq [x(k) - \hat{x}(k|k-1)]\,[x(k) - \hat{x}(k|k-1)]^{T}$$

$$D \triangleq [z(k) - \hat{z}(k|k-1)]\,[z(k) - \hat{z}(k|k-1)]^{T}$$

C.2 Sufficient Conditions for Optimization

The sufficient condition for a maximum of the objective function L is $\partial^2 L/\partial x^2(k) < 0$. Therefore, from part 1 of this Appendix, $\partial^2 L/\partial x^2(k)$ is less than zero if

$$\frac{\partial}{\partial x(k)}\,\{h_x^{T}R^{-1}[z(k) - h - r] - P^{-1}(k\;k-1)x(k)\} < 0$$

Equivalently, $P^{-1}(k|k-1) - g_x[x(k),R,z(k)]$ is positive definite where $g_x[x(k),R,z(k)]$ is the n x n Jacobian matrix of $h_x^{T}R^{-1}[z(k)-h-r]$ evaluated at $\hat{x}(k|k)$ for the sufficient condition for optimal filtering.

APPENDIX D

PROBABILITY DENSITY FUNCTION IDENTITIES

D.1 $p[x(k),Q,R|Z(k)]$

Show

$$p[x(k),Q,R|Z(k)] \equiv p[z(k)|x(k),Q,R] \cdot p[x(k),Q,R|Z(k-1)]p[Z(k-1)]/p[Z(k)]$$

By the definition of the conditional density

$$p(\alpha|\beta) = p(\alpha,\beta)/p(\beta)$$

where $p(\alpha,\beta)$ is the joint probability density function of random vectors α and β; $p(\beta)$ is the marginal density function of β; and $p(\alpha|\beta)$ is the conditional probability density function of α given β.

Proof:

(1) Using the above definition,

$$p[x(k),Q,R|Z(k)] = p[x(k),Q,R,Z(k)]/p[Z(k)]$$

(2) Develop an intermediate result by using the definition

$$p[z(k)|x(k),Z(k-1)] \equiv p[z(k)|x(k),z(k-1),z(k-2)...z(1)]$$

where $Z(k) = \{z(i),\ i = 1,2...k\}$. Applying the definition

$$p[z(k)|x(k),Z(k-1)] = p[z(k),x(k),z(k-1)\ldots z(1)] \div p[x(k),z(k-1)\ldots z(1)]$$

Rearranging,

$$p[x(k),z(k),z(k-1)\ldots z(1)] = p[z(k)|x(k),z(k-1)\ldots z(1)] \cdot p[x(k),z(k-1)\ldots z(1)]$$

$$\text{Or,}\quad p[x(k),Z(k)] = p[z(k)|x(k),Z(k-1)]p[x(k),Z(k-1)] \equiv p[z(k)|x(k)]p[x(k),Z(k-1)]$$

since knowledge of $x_{(k)}$supercedes the need for Z(k-1).

$$\therefore\ p[x(k),Q,R,Z(k)] = p[z(k)|x(k),Q,R]p[x(k),Q,R,Z(k-1)]$$

(3) Again applying the definition,

$$p[x(k),Q,R,Z(k-1)] = p[x(k),Q,R|Z(k-1)]p[Z(k-1)]$$

(4) Consecutive application of parts 1 through 3 yields the desired equality.

D.2 $p[x(k),Q,R|Z(k-1)]$

From the property of independent vectors

$$p(\alpha,\beta) = p(\alpha)\ p(\beta)$$

where α and β are statistically independent random vectors.

It follows for x(k), Q, R mutually independent,

$$p[x(k),Q,R|Z(k-1)] = p[x(k)|Z(k-1)]p[Q|Z(k-1)]p[R|Z(k-1)] = p[x(k)|Z(k-1)]p(Q)p(R)$$

where the true moments, Q, R are independent of Z(k-1).

D.3 p[z(k)|Z(k-1)]

Show $p[Z(k)]/p[Z(k-1)] = p[z(k)|Z(k-1)]$. From the definition of the conditional probability density function, p[Z(k)] may be rewritten as

$$p[Z(k)] \equiv p[z(k),Z(k-1)]$$
$$= p[z(k)|Z(k-1)]\ p[Z(k-1)]$$

$$\therefore\ p[Z(k)]/p[Z(k-1)] = p[z(k)|Z(k-1)]p[Z(k-1)]/p[Z(k-1)]$$
$$= p[z(k)|Z(k-1)]$$

where $$Z(k) \triangleq \{z(k),z(k-1)\ldots z(1)\}$$
$$\equiv \{z(k),Z(k-1)\}$$

D.4 Uniform Probability Density Distributions

The following uniform probability densities have been utilized to obtain a solution to the maximization of the likelihood function with respect to the plant input and measurement error statistics.

$$p(Q) = \begin{cases} (Q_{max} - Q_{min})^{-1}, & Q_{min} < Q < Q_{max} \\ 0, & \text{otherwise} \end{cases}$$

$$p(R) = \begin{cases} (R_{max} - R_{min})^{-1}, & R_{min} < R < R_{max} \\ 0, & \text{otherwise} \end{cases}$$

These functions contribute only a pre-exponentail constant to the likelihood function which can be ignored if the above limiting values are not exceeded during estimation.

APPENDIX E

PROBABILITY DENSITY FUNCTION EVALUATION

Some definitions which are utilized in this Appendix E are the following.

$$\hat{x}(k|j) \triangleq E[x(k)|Z(j)]$$

for the Gaussian stochastic sequences $\{x(k),\ k = 0,1,2...\}$ and $\{z(j),\ j = 1,2,...\}$.

$$\|a\|_B^2 \triangleq a^T B a$$

This is the Euclidean norm of a over B.

Assuming $\|x(k) - \hat{x}(k|j)\|$ is sufficiently small for us to represent $f[x(k)]$ and $h[x(k)]$ by the first two terms of a Taylor series expansion about $\hat{x}(k|k)$ and $\hat{x}(k|k-1)$ respectively, then

$$f[x(k)] \cong f[\hat{x}(k|k)] + f_x[\hat{x}(k|k)][x(k) - \hat{x}(k|k)]$$

for $k = 0,1,2...$

and $h[x(k)] \cong h[\hat{x}(k|k-1)] + h_x[\hat{x}(k|k-1)][x(k) - \hat{x}(k|k-1)]$

for $k = 1,2...$

The definition of the estimation error covariance is

$$P(k|j) \triangleq E\{[x(k) - \hat{x}(k|j)][x(k) - \hat{x}(k|j)]^T | Z(j)\}$$
$$\equiv \text{Cov}\,[x(k)|Z(j)]$$

where

$$Z(j) \triangleq \{z(i),\ i = 1,2...j\}$$

Other definitions for the convenience of notation only are

$$\tilde{x}(k|j) \triangleq [x(k) - \hat{x}(k|j)]$$
$$\tilde{v}(k) \triangleq [v(k) - r(k)]$$
$$\tilde{w}(k) \triangleq [w(k) - q(k)]$$
$$\hat{x}(k) \triangleq x(k|k), \text{ for identical double subscripts.}$$

E.1 $p[x(k)|Z(k-1)]$

Assuming a Gaussian density, only the first two conditional moments are required. The conditional mean is defined by

$$E[x(k)|Z(k-1)] \triangleq \hat{x}(k|k-1)$$
$$= E\{f[\hat{x}(k-1)] + f_x[\hat{x}(k-1)][x(k-1) - \hat{x}(k-1)]$$
$$+ w(k-1)|Z(k-1)\}$$
$$= f[\hat{x}(k-1)] + q(k-1)$$

which is the single stage prediction estimate using

$$E[\tilde{x}(k-1)|Z(k-1)] = 0$$
$$E[w(k-1)|Z(k-1)] = q(k-1)$$
$$E\{f[\hat{x}(k-1)]|Z(k-1)\} = f[\hat{x}(k-1)]$$

and $\hat{x}(k-1) \triangleq \hat{x}(k-1|k-1)$ for notational convenience.

The conditional variance is defined by

$$E\left([x(k) - \hat{x}(k|k-1)][x(k) - \hat{x}(k|k-1)]^T | Z(k-1)\right) \triangleq P(k|k-1)$$

$$= E\left(\{f[\hat{x}(k-1)] + f_x[\hat{x}(k-1)][x(k-1) - \hat{x}(k-1)] + w(k-1) - f[\hat{x}(k-1)] - q(k-1)\} \cdot \{f[\hat{x}(k-1)] + f_x[\hat{x}(k-1)][x(k-1) - \hat{x}(k-1)] + w(k-1) - f[\hat{x}(k-1)]\}^T | Z(k-1)\right)$$

$$= f_x[\hat{x}(k-1)]P(k-1)f_x^T[\hat{x}(k-1)] + E\left(\tilde{w}(k-1)\tilde{x}^T(k-1) \cdot f_x^T[\hat{x}(k-1)] | Z(k-1)\right) + E\left(f_x[\hat{x}(k-1)]\tilde{x}(k-1)\tilde{w}^T(k-1) | Z(k-1)\right) + Q(k-1)$$

Assuming the state vector and plant input vector first moments are statistically independent,

$$E\,[w(k-1) - q(k-1)][x(k-1) - \hat{x}(k-1)]^T | Z(k-1)]$$

$$= E[w(k-1)|Z(k-1)]\; E[x^T(k-1)|Z(k-1)] - q(k-1)\; E[x^T(k-1)|Z(k-1)] - E[w(k-1)|Z(k-1)]\hat{x}^T(k-1) + q(k-1)\;\hat{x}^T(k-1)$$

$$= q(k-1)\;\hat{x}^T(k-1) - q(k-1)\;\hat{x}^T(k-1) - q(k-1)\;\hat{x}^T(k-1) + q(k-1)\;\hat{x}^T(k-1)$$

$$= 0$$

Similarly, $E[\tilde{x}(k-1)\;\tilde{w}^T(k-1)|Z(k-1)] = 0$. Therefore, $P(k|k-1) = f_x[\hat{x}(k-1)]P(k-1)f_x^T[\hat{x}(k-1)] + Q(k-1)$ where $Q(k-1) = E[\tilde{w}(k-1)\tilde{w}^T(k-1)|Z(k-1)]$.

Assuming the state vector remains normally distributed, the density function for the conditional random n-vector $x(k)$ given $Z(k-1)$ becomes

$$p[x(k)|Z(k-1)] = (2\pi)^{-n/2}|P(k|k-1)|^{-1/2} \cdot \text{Exp}\left\{-\frac{1}{2}\|x(k) - \hat{x}(k|k-1)\|^2_{P^{-1}(k|k-1)}\right\}$$

E.2 $p[z(k)|x(k),Q(k),R(k)]$

The conditional mean is defined by

$$\begin{aligned} E[z(k)|x(k),Q(k),R(k)] &= E\{h[x(k)] + v(k)|x(k)...\} \\ &= h[x(k)] + E[v(k)|x(k)...] \\ &= h[x(k)] + r(k) \\ &\triangleq \bar{z}(k) \end{aligned}$$

where $h[x(k)]$ is constant for a given value of $x(k)$ and the error vector $v(k)$ is assumed nonzero mean.

The conditional variance is defined by

$$\begin{aligned} &E\{[z(k) - \bar{z}(k)][z(k) - \bar{z}(k)]^T|x(k)...\} \\ &\quad = E\big(\{h[x(k)] + v(k) - h[x(k)] - r(k)\} \\ &\qquad \cdot \{h[x(k)] + v(k) - h[x(k)] - r(k)\}^T|x(k)...\big) \\ &\quad = E\{[v(k) - r(k)][v(k) - r(k)]^T|x(k)...\} \\ &\quad \triangleq R(k) \end{aligned}$$

The Gaussian density function for the conditional m-vector $z(k)$, given $x(k)$, becomes

$$p[z(k)|x(k),Q(k),R(k)] = (2\pi)^{-m/2}|R(k)|^{-1/2} \cdot \text{Exp}\left[-\frac{1}{2}\|z(k) - \bar{z}(k)\|^2_{R^{-1}(k)}\right]$$

E.3 $p[z(k)|Z(k-1)]$

The conditional mean is defined by

$$\begin{aligned}
E[z(k)|Z(k-1)] &= E\{h[\hat{x}(k|k-1)] + h_x[\hat{x}(k|k-1)][x(k) \\
&\qquad - \hat{x}(k|k-1)] + v(k)|Z(k-1)\} \\
&= h[\hat{x}(k|k-1)] + h_x[\hat{x}(k|k-1)]E[\hat{x}(k|k-1)|Z(k-1)] \\
&\quad + E[v(k)|Z(k-1)] \\
&= h[\hat{x}(k|k-1)] + r(k) \\
&= \hat{z}(k|k-1)
\end{aligned}$$

where $h[\hat{x}(k|k-1)]$ and $h_x[\hat{x}(k|k-1)]$ are constant for a given value of $\hat{x}(k|k-1)$, and $r(k) = E[v(k)|Z(k-1)]$.

The conditional variance is defined by

$$\begin{aligned}
&E\{[z(k) - \hat{z}(k|k-1)][z(k) - \hat{z}(k|k-1)]^T|Z(k-1)\} \triangleq Y(k|k-1) \\
&\quad = E\big(\{h[\hat{x}(k|k-1)] + h_x[\hat{x}(k|k-1)][x(k) - \hat{x}(k|k-1)] \\
&\qquad + v(k) - h[\hat{x}(k|k-1)] - r(k)\}\cdot\{h[\hat{x}(k|k-1)] \\
&\qquad + h_x[\hat{x}(k|k-1)][x(k) - \hat{x}(k|k-1)] + v(k) \\
&\qquad - h[\hat{x}(k|k-1)] - r(k)\}^T|Z(k-1)\big) \\
&\quad = h_x[\hat{x}(k|k-1)E\{[x(k) - \hat{x}(k|k-1)][x(k) - \hat{x}(k|k-1)]^T \\
&\qquad |Z(k-1)\} \cdot h_x^T[\hat{x}(k|k-1)] + h_x[\hat{x}(k|k-1)] \\
&\qquad \cdot E[\tilde{x}(k|k-1)\ \tilde{v}^T(k)|Z(k-1)] \\
&\qquad + E[\tilde{v}(k)\ \tilde{x}^T(k|k-1)|Z(k-1)]h_x^T[\hat{x}(k|k-1)] + R(k)
\end{aligned}$$

Assuming the state vector and measurement error vector first moment are statistically independent,

$$E\{[v(k) - r(k)][x(k) - \hat{x}(k|k-1)]^T Z(k-1)\}$$
$$= E[v(k)|Z(k-1)]\ E[x^T(k)|Z(k-1)] - r(k)E[x^T(k)|Z(k-1)]$$
$$- E[v(k)|Z(k-1)]\ \hat{x}^T(k|k-1) + r(k)\ \hat{x}^T(k|k-1)$$
$$= r(k)\ \hat{x}^T(k|k-1) - r(k)\ \hat{x}^T(k|k-1) - r(k)\ \hat{x}^T(k|k-1)$$
$$+ r(k)\ \hat{x}^T(k|k-1)$$
$$= 0$$

Similarly, $E[\tilde{x}(k|k-1)\ \tilde{v}^T(k)|Z(k-1)] = 0$. Therefore, $Y(k|k-1) = h_x[\hat{x}(k|k-1)]P(k|k-1)h_x^T[\hat{x}(k|k-1)] + R(k)$ where $R(k) = E[\tilde{v}(k)\ \tilde{v}^T(k)|Z(k-1)] = R(k)$.

The Gaussian density function for the conditional m-vector $z(k)$, given $Z(k-1)$, becomes

$$p[z(k)|Z(k-1)] = (2\pi)^{-m/2}|Y(k|k-1)|^{-1/2}$$
$$\cdot \text{Exp}\ [-\tfrac{1}{2}\|z(k) - \bar{z}(k|k-1)\|^2_{Y^{-1}(k|k-1)}]$$

E.4 $p[x(k)|Z(k)]$

The consequent probability density function of $x(k)$, given $Z(k)$, which is of interest may be obtained from the results of parts E.1 through E.3.

$$L = p[z(k)|x(k),Q(k),R(k)]p[x(k)|Z(k-1)]/p[z(k)|Z(k-1)]$$
$$= (2\pi)^{-(2m+n)/2}\cdot|R(k)|^{-1/2}\cdot|P(k|k-1)|^{-1/2}\cdot|Y(k|k-1)|^{-1/2}$$
$$\cdot \text{Exp}\ \left(-\tfrac{1}{2}\|z(k) - \bar{z}(k)\|^2_{R^{-1}(k)} - \tfrac{1}{2}\|x(k) - x(k|k-1)\|^2_{P^{-1}(k|k-1)}\right.$$
$$\left. - \tfrac{1}{2}\|z(k) - z(k|k-1)\|^2_{Y^{-1}(k|k-1)}\right)$$

APPENDIX F

KALMAN EXTENDED FILTER ALGORITHM

F.1 Extended Filter

Evaluate $[\partial L/\partial x_{(k)}]_{x_{(k)}} \rightarrow \hat{x}_{(k|k)}$ from Appendix C, and using the matrix inversion lemma of Appendix A replace the first factor to obtain the results

$$\hat{x}_{(k|k)} = (P - Ph_x^T[R + h_xPh_x^T]^{-1} h_xP)\ h_x^TR^{-1}(z_{(k)} - h + h_x\hat{x}_{(k|k-1)} - r_{(k)} + P^{-1}\hat{x}_{(k|k-1)})$$

Substitute the definition of the gain matrix W.

$$W \triangleq Ph_x^T(R + h_xPh_x^T)^{-1}$$

$$\hat{x}_{(k|k)} = (P - Wh_xP)(h_x^TR^{-1}[z_{(k)} - h + h_x\hat{x}_{(k|k-1)} - r_{(k)}]) + P^{-1}\hat{x}_{(k|k-1)}$$

Rearranging this expression and substituting the alternate expression for W from Appendix A, one obtains the result.

$$\hat{x}_{(k|k)} = \hat{x}_{(k|k-1)} + W_{(k)}(z_{(k)} - h[\hat{x}(k|k-1)] - r_{(k)})$$

for nonzero mean $r_{(k)}$, $Q_{(k)}$ and $R_{(k)}$ assumed to be known.

F.2 Extended Filter Error Covariance Matrix

For a dynamic system modeled by the equation

$$x_{(k+1)} = f[x_{(k)}] + w_{(k)}$$

with observations of the system available in the form

$$z_{(k+1)} = h[x_{(k+1)}] + v_{(k\ 1)}$$

the filtered estimate of the state is given by

$$\hat{x}_{(k|k)} = \hat{x}_{(k|k-1)} + W[z_{(k)} - h[x_{(k|k-1)}] - r_{(k)}]$$

where the filter gain matrix, W(k) is given by

$$W(k) \triangleq P_{(k|k-1)}h_x[R_{(k)} + h_x P_{(k|k-1)}h_x^T]^{-1}$$

The estimation error is defined by

$$\begin{aligned}\tilde{x}_{(k|k)} &= x_{(k)} - \hat{x}_{(k|k)} \\ &= f[x_{(k-1)}] + w_{(k-1)} - f[\hat{x}_{(k-1)}] - q_{(k-1)} \\ &\quad - Wh[\hat{x}_{(k|k-1)}] - Wh_x[\hat{x}_{(k|k-1)}]f[x_{(k-1)}] \\ &\quad - Wh_x[\hat{x}_{(k|k-1)}]w_{(k-1)} + Wh_x[\hat{x}_{(k|k-1)}]f[x_{(k-1)}] \\ &\quad + Wh_x[\hat{x}_{(k|k-1)}]q_{(k-1)} - Wr_{(k)} + Wh[\hat{x}_{(k|k-1)}] \\ &\quad + Wr_{(k)}\end{aligned}$$

where substitutions using the model, observation, and prediction equations have been made. This is further easily reduced to the recursive equation for filter error

$$\tilde{x}_{(k|k)} = f_x[\hat{x}_{(k-1)}]\tilde{x}_{(k-1)} + \tilde{w}_{(k-1)} - Wh_x[\hat{x}_{(k|k-1)}] \cdot f_x[\hat{x}_{(k-1)}]\tilde{x}_{(k-1)} - Wh_x[\hat{x}_{(k|k-1)}]\tilde{w}_{(k-1)} - W\tilde{v}_{(k)}$$

The filter error covariance matrix is obtained from the definition.

$$P_{(k|k)} \triangleq \text{Cov}\{\tilde{x}_{(k|k)}, \tilde{x}_{(k|k)}\}$$

$$= E[\tilde{x}_{(k|k)}\tilde{x}^T_{(k|k)}]$$

$$= E\{[(I - Wh_x)\tilde{x}_{(k|k-1)} - W\tilde{v}_{(k)}] \cdot [(I - Wh_x)\tilde{x}_{(k|k-1)} - W\tilde{v}_{(k)}]^T\}$$

where $\tilde{x}_{(k|k-1)} \triangleq x_{(k)} - \hat{x}_{(k|k-1)}$

$$= f[x_{(k-1)}] + w_{(k-1)} - f[\hat{x}_{(k-1)}] - q_{(k-1)}$$

$$= f_x[\hat{x}_{(k-1)}]\tilde{x}_{(k-1)} + \tilde{w}_{(k-1)}$$

Using $P_{(k|k-1)} \triangleq E[\tilde{x}_{(k|k-1)}\tilde{x}^T_{(k|k-1)}]$

and $R_{(k)} \triangleq E[\tilde{v}_{(k)}\tilde{v}^T_{(k)}]$

the covariance matrix equation reduces to

$$P_{(k|k)} = (I - Wh_x)P_{(k|k-1)}(I - Wh_x)^T \quad W\,E[\tilde{v}_{(k)}\tilde{x}^T_{(k|k-1)}](I - Wh_x)^T - (I - Wh_x)E_T[\tilde{x}_{(k|k-1)}\tilde{v}_{(k)}^T]W^T + WR_{(k)}W^T$$

$$= (I - Wh_x)P_{(k|k-1)}(I - Wh_x)^T + WR_{(k)}W^T$$

where $x_{(k|k-1)}$ and $v_{(k)}$ are statistically independent.

Consider the term

$$E[\tilde{x}_{(k|k-1)}\tilde{v}^T_{(k)}] = E[(x_{(k)} - \hat{x}_{(k|k-1)})(v_{(k)} - r_{(k)})^T]$$

Since $E[\hat{x}_{(k|k-1)}(v_{(k)} - r_{(k)})^T] = \hat{x}_{(k|k-1)}E[(v_{(k)} - r_{(k)})^T] = 0$

and $$E[x_{(k)}v^T_{(k)}] = E[x_{(k)}]\ E[v^T_{(k)}] = E[x_{(k)}]r^T_{(k)}$$

by the statistical independence of $x_{(k)}$ and $v_{(k)}$; also $-E[x_{(k)}r^T_{(k)}] = -E[x_{(k)}]r^T_{(k)}$ finally then $E[\tilde{x}_{(k|k-1)}\tilde{v}^T_{(k)}] = 0$. Similarly for $E[\tilde{v}_{(k)}\tilde{x}^T_{(k|k-1)}] = 0$.

The expression for $P_{(k|k)}$ may be reduced to the final recursive equation by rearranging to

$$\begin{aligned} P_{(k|k)} &= (I - W_{(k)}h_x)P_{(k|k-1)} + W_{(k)}(h_xP_{(k|k-1)}h^T_x + R_{(k)})W^T_{(k)} \\ &\quad - P_{(k|k-1)}h^T_xW^T_{(k)} \\ &= (I - Wh_x)P_{(k|k-1)} \end{aligned}$$

where the alternate expression for the gain matrix W_k has been substituted after the following rearrangement of the expression

$$WR_{(k)} = (P_{(k|k-1)} - Wh_xP_{(k|k-1)})h^T_x$$

then $$(WR_{(k)} + Wh_xP_{(k|k-1)}h^T_x) = P_{(k|k-1)}h^T_x.$$

APPENDIX G

CHEMICAL REACTOR MODEL

Wells (34) previously has developed a mathematical model for an adiabatic stirred reactor. Figure G.1 is a schematic of the physical system. It is assumed that:

1. The exiting solutions have been perfectly mixed inside the reactor and coolant containment.
2. There is no heat loss from the coolant containment to the surroundings.
3. The reactor wall has sufficient capacity to influence system dynamics.
4. The reactor contents undergo a second order irreversible exothermic reaction

$$2A \xrightarrow{k} B$$

A model for the reactor dynamics is obtained from the following four balances.

1. Energy balance on reactor contents

$$\rho V C_p \frac{dT}{dt} = \rho F C_p (T_i - T) + \Delta H k V C_A^2 e^{-E/RT} - hA(T - T_w)$$

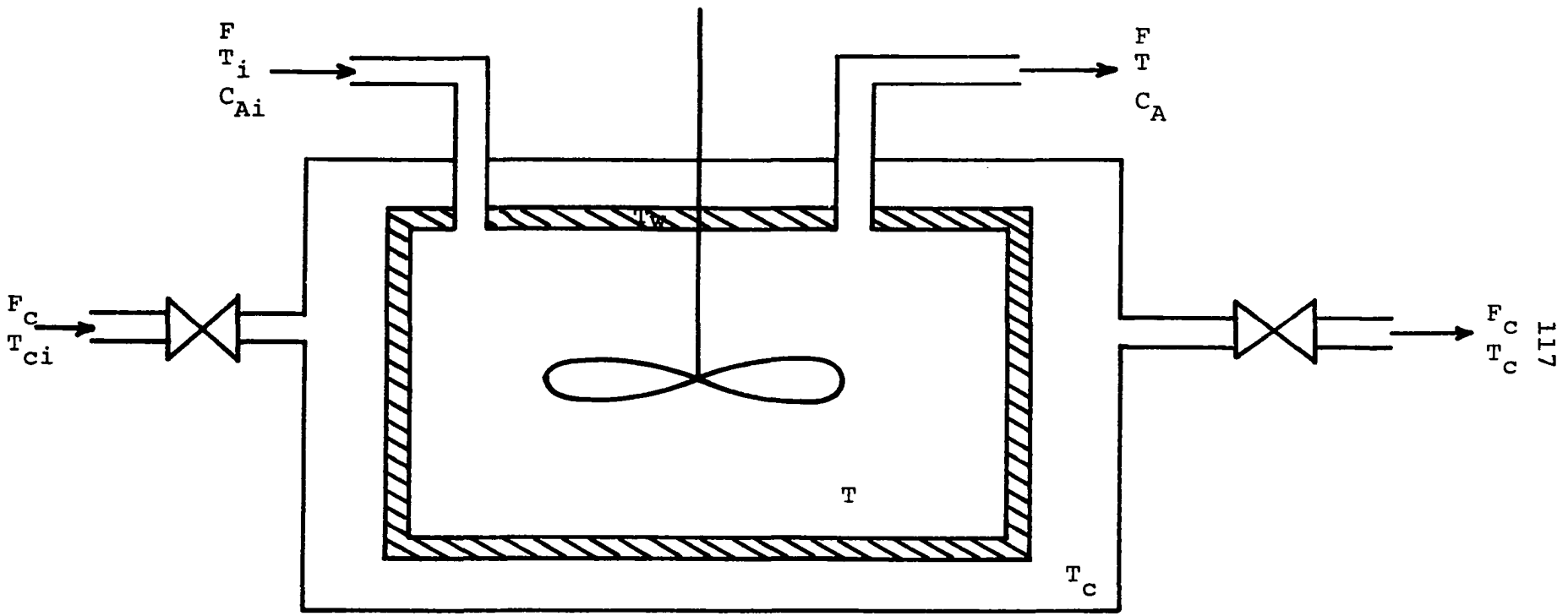


Figure G.1. Stirred Chemical Reactor.

2. Energy balance on reactor wall

$$\rho_w V_w C_{pw} \frac{dT_w}{dt} = hA(T - T_w) - h_w A_w (T_w - T_c)$$

3. Energy balance on coolant

$$\rho_c V_c C_{pc} \frac{dT_c}{dt} = \rho_c F_c C_{pc} (T_{ci} - T_c) + h_w A_w (T_w - T_c)$$

4. Mass balance on reactor contents

$$V \frac{dC_A}{dt} = F(C_{Ai} - C_A) - kVC_A^2 \, e^{-E/RT}$$

For convenience, these four state equations are normalized about an operating point in the temperature-concentration plane T_s, C_s. The resulting equations are

$$\dot{x}_1 = -(c_1 + c_4) x_1 + c_3 (1 + x_4)^2 \exp(k_1 x_1/[1+x_1]) + c_4 x_2 + c_1 u_1$$

$$\dot{x}_2 = -(c_5 + c_6) x_2 + c_5 x_1 + c_6 x_3$$

$$\dot{x}_3 = -(c_7 + c_8) x_3 + c_8 x_2 + c_7 u_2$$

$$\dot{x}_4 = -c_1 x_4 - c_2 (1 + x_4)^2 \exp(k_1 x_1/[1+x_1]) + c_1 u_3$$

which have the vector form $\dot{x} = f(x,u)$. The process parameters for this fourth order model are defined in Tables G.1, G.2 and G.3. To maintain the system at the operating point (T_s, C_s), steady state control functions are used for u_1, u_2 and u_3.

TABLE G.1

PROCESS VARIABLES AND CONSTANTS

Symbol	Definition	Units
A	reactor wall surface area	ft^2
A_w	coolant wall surface area	ft^2
C_A	weight concentration A in reactor	lb/ft^3
C_{Ai}	weight concentration A at inlet	lb/ft^3
C_p	mean heat capacity of reactor contents	Btu/lb-°R
C_{pc}	mean heat capacity of coolant	Btu/lb-°R
C_{pw}	mean heat capacity of wall	Btu/lb-°R
E	activation energy	Btu/mol
F	volumetric flow rate of reactor effluent	ft^3/sec
F_c	volumetric flow rate of coolant	ft^3/sec
ΔH	heat of reaction	Btu/lb
h	film coefficient in reactor	Btu/ft^2-°F-sec
h_w	film coefficient between wall and coolant	Btu/ft^2-°F-sec
k	Arrhenius rate constant	sec^{-1}
R	gas constant	Btu/mol-°R
T	reactor effluent temperature	°R
T_c	coolant temperature	°R
T_i	reactor inlet stream temperature	°R
T_w	wall temperature	°R
t	time	sec
V	reactor volume	ft^3
V_c	coolant containment volume	ft^3
V_w	wall volume	ft^3
C_s	operating point concentration	lb/ft^3

TABLE G.1--Continued

Symbol	Definition	Units
T_s	operating point temperature	°R
ρ	mean density of reactor contents	lb/ft^3
ρ_c	mean density of coolant	lb/ft^3
ρ_w	mean density of wall	lb/ft^3

TABLE G.2
DIMENSIONLESS CONSTANTS

Symbol	Definition	Value
k_1	E/RT_s	1
c_1	$F/(V\beta)$	0.2
c_2	$(k/\beta)C_s^2 \exp(-K_1)$	0.2
c_3	$\Delta Hk \exp(-K_1)C_s^2/(\rho C_p T_s \beta)$	1.0
c_4	$hA/(\rho V C_p)$	0.5
c_5	$hA/(\rho_w V_w C_{pw} \beta)$	0.5
c_6	$h_w A_w/(\rho_w V_w C_{pw} \beta)$	0.5
c_7	$F_c/(V_c \beta)$	0.05
c_8	$h_w A_w/(\rho_c V_c C_{pc} \beta)$	0.5
β	τ/t	1/240
u_1	$-c_3/c_1$	-5.0
u_2	0	0
u_3	c_2/c_1	1.0

TABLE G.3

DIMENSIONLESS VARIABLES

Symbol	Definition
$x_1 = (T-T_s)/T_s$	normalized reactor effluent temperature
$x_2 = (T_w- T_s)/T_s$	normalized wall temperature
$x_3 = (T_c- T_s)/T_s$	normalized coolant temperature
$x_4 = (C_A - c_s)/C_s$	normalized reactor effluent concentration
$\dot{x} = dx/d\tau$	rate of change in normalized state variables

All states are assumed to be measurable; and the measurements are assumed to contain white Gaussian noise. The vector matrix measurement model is

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 000 & & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

The state transition matrix for the linearized system is obtained from the Jacobian matrix for the state equations. The matrix of partial derivatives for these equations is

$$\phi = \begin{bmatrix} f_{11} & f_{12} & 0 & f_{14} \\ f_{21} & f_{22} & f_{23} & 0 \\ 0 & f_{32} & f_{33} & 0 \\ f_{41} & 0 & 0 & f_{44} \end{bmatrix}$$

where $f_{11} = -(c_1 + c_4) + c_3 k_1 B^2 DE$

$f_{12} = c_4$

$f_{14} = 2\, c_3\, BE$

$f_{21} = c_5$

$f_{22} = -(c_5 + c_6)$

$f_{23} = c_6$

$f_{32} = c_8$

$f_{33} = -(c_7 + c_8)$

$f_{41} = -c_2\, k_1 B^2 DE$

$f_{44} = -c_1 - 2c_2 BE$

and $B = (1 + x_4)$

$D = (1 + x_1)^{-2}$

$E = \exp\ (k_1 x_1 / [1 + x_1])$

To complete the model, the initial condition state vector is

$$x(0) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0.1 \end{bmatrix}$$

where the reactor is perturbed from the steady state with a 10 percent disturbance in feed concentration. The prior statistics for the measurement error are given by the measurement noise covariance matrix

$$R = \begin{bmatrix} R_{11} & 0 & 0 & 0 \\ 0 & R_{22} & 0 & 0 \\ 0 & 0 & R_{33} & 0 \\ 0 & 0 & 0 & R_{44} \end{bmatrix}$$

where the noise is assumed to be: (1) zero mean, (2) Gaussian distributed, and (3) uncorrelated. The prior statistics for the plant noise are given by the plant noise covariance matrix

$$Q = \begin{bmatrix} Q_{11} & 0 & 0 & 0 \\ 0 & Q_{22} & 0 & 0 \\ 0 & 0 & Q_{33} & 0 \\ 0 & 0 & 0 & Q_{44} \end{bmatrix}$$

and the same previous three assumptions.

APPENDIX H

COMPUTER PROGRAMS

There are three types of subroutines incorporated into the Computer Program for Adaptive Estimation. These three are as follows:

1. operational--utility programs for matrix operations and information retrieval.
2. user supplied--model specific programs reflecting user's choice of mathematical model and options.
3. functional--filter specific programs which depend upon the particular type of estimation method employed.

H.1 Operational Subroutines

(1)	ADD	matrix addition, A + B = C
(2)	SUB	matrix subtraction, A - B = C
(3)	MULT	matrix multiplication, A x B = C
(4)	INVRS	matrix inversion, $A^{-1} = C$, (sequentially (4)]
(5)	TRANS	matrix transpose, $A^{T} = C$
(6)	SYM	matrix symmetry, $A_{ij} = (A_{ij} + A_{ji})/2$
(7)	IDEN	matrix identity, A = [I]
(8)	SKALR	multiply a matrix by a scalar, b · A = C
(9)	MATPT	print a 1-dimension array as an (n x m)-dimension matrix

(10)	TYM	print current model time
(11)	STR	store a set of vectors esquentially to facilitate printing later
(12)	SUMRY	print summary of states, estimates, and observations as a function of model time
(13)	DIAGV	store diagonal elements of a matrix in a vector to facilitate printing later
(14)	RPTQR	print summary of plant and observation noise and covariances

H.2 User Supplied Subroutines

(1)	C	constants used in mathematical model
(2)	CNTRL	desired control function used in mathematical model
(3)	SYS	mathematical model; non-linear state function for continuous time model
(4)	PHI	partial derivative of continuous time model [Jacobian]
(5)	MEASR	non-linear measurement function; continuous time
(6)	PARTH	partial derivations of non-linear measurement function [Jacobian]
(7)	STATE	numerical integrates the non-linear state functions over given model time interval
(8)	DSKSYS	discretized non-linear state function; discrete time model

(9) DSKPHI discretized linearized state functions; discrete time linearized model

(10) OBSRV generates "experimental observations" from exact measurement model and observational noise

H.3 Functional Subroutines

(1) PREDX predicts estimate of state

(2) FILTX filters estimate of state

(3) PREDP predicts estimate of state covariance

(4) FILTP filters estimate of state covariance

(5) PLTMN estimates plant noise mean

(6) PLTCO estimates plant noise covariance

(7) OBSMN estimates observation noise mean

(8) OBSCO estimates observation noise covariance

(9) RESID estimates observation residual

(10) WAIT estimates optimal filter gain

(11) SUCES implements Wegstein successive approximation method