BAR CODE

7913627018.

J. S. HERON LTD
5 QUEENS CRESCENT
ST. GEORGES CROSS
GLASGOW 041 332 1883

DESIGN OF DISCRETE TIME CONTROLLERS AND ESTIMATORS

THOMAS JAMES MOIR, BSc

This thesis is submitted to the Council for National Academic Awards in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Collaborating Establishment

British Steel Corporation
Swinden Laboratories
Rotherham

Sheffield City Polytechnic
Department of Electrical &
Electronic Engineering

December 1982

Design of Discrete Time Controllers and Estimators

T J MOIR

ABSTRACT

This thesis considers optimal linear least-squares filtering, smoothing prediction and regulation for discrete-time processes.

A finite interval smoothing filter is derived in the z domain giving a transfer function solution. The resulting time-invariant smoother can be applied to problems where a time varying solution using matrix Riccati equations would diverge if the process is modelled inaccurately.

A self-tuning algorithm is given for the filtering and fixed lag smoothing problems as applied to square multivariable ARMA processes when only the order of the process is assumed known. The dynamics of the process can also be slowly time varying. If the dynamics remain constant and unknown, it is shown how the self-tuning filter or smoother algorithm converges asymptotically to the optimal Wiener solutions.

LQG self-tuning regulation is considered. The LQG algorithms rely on input-output data rather than from the conventional state-space approach employing the Kalman filter. An explicit algorithm is given which is similar to certain pole placement self-tuning regulators, requiring the solution of a diophantine equation. Following this, an implicit algorithm is shown to overcome the problem of solving a diophantine equation by estimating the regulator parameters directly using recursive least squares. The LQG algorithms are shown to be able to cope with processes which are non-minimum phase, open loop unstable and with an unknown time delay.

CONTENTS

ACKNOWLEDGEMENTS

This work has been carried out at the Department of Electrical and Electronic Engineering, Sheffield City Polytechnic, over three years. Throughout this time the members of staff have made my stay enjoyable and worthwhile. I am particularly indebted to my supervisors, Professor M J Grimble (now with the University of Strathclyde) and Mr B S Plumb, for their guidance and encouragement throughout this time. Also worthy of a mention are Dr J Fotakis, Dr J F Barrett and Mr P T K Fung, all formerly of this Department.

Finally, but by no means least, I should like to thank my wife, Lidia, for her endurance and Ann Hughes for swiftly and accurately typing this thesis.

Nomenclature

Unless otherwise stated in the text, the principal nomenclature used in this thesis is as shown below:

$\underline{\Delta}$	defined as
A,B,C,D	Matrices used in the discrete time state space process model.
$A(z^{-1})$, $C(z^{-1})$, $B(z^{-1})$, $D(z^{-1})$	Polynomials or polynomial matrices used in the discrete time AR or CARMA model.
n_x	Order 'x' of polynomial or polynomial matrix.
$\underline{y}(t)$	Process output vector at time t.
$\underline{z}(t)$	Measured output or observation sequence vector at time t.
$\underline{u}(t)$	Control signal at time t.
ℓ	Lag in smoother.
N	Number of sample points in fixed interval.
z	Complex Z-transform variable.
$E\{a\}$	Unconditional expectation of 'a'.
$E\{a\|b\}$	Expectation of 'a' conditional on 'b'.
$\underline{\omega}(t)$	Process noise vector at time t.
$\underline{v}(t)$	Measurement noise vector at time t.
$\underline{\varepsilon}(t)$	Innovations sequence vector at time t.
Q	Process noise covariance matrix.
R	Measurement noise covariance matrix.
R_ε	Innovations covariance matrix.
Q_1	Weighting on error in the cost function.
R_1	Weighting on control in the cost function.
$F(z)$	Return difference matrix.
$W(z)$	Process transfer function matrix.

$\underline{\theta}$	Vector of parameters.
$X(t)$	Vector of inputs and outputs.
K	Kalman gain matrix.
$\mathcal{Z}_1(x(t))$	Single sided Z-transform of the sequence $\{x(t)\} \triangleq \sum_{t=0}^{\infty} x(t) z^{-t}$
$\mathcal{Z}_2(x(t))$	Two-sided Z-transform of the sequence $\{x(t)\} \triangleq \sum_{t=-\infty}^{\infty} x(t) z^{-t}$
$U(t)$	Heaviside unit step function $U(t) = 1$ for all $t \geq 0$, $U(t) = 0$ for $t < 0$.
$\delta(i - j)$	Kronecker delta function $\delta(0) = 1$; $\delta(m) = 0$, $m \neq 0$.

CHAPTER 1

An Introduction to Estimation and Control

1.1 Motivation

This work is concerned with optimal estimation and control with particular reference to industrial applications. It may appear from a careful examination of the existing literature that the work is already well developed [1 - 11]. Closer consideration reveals, however, that although the existing theory is useful for the particular application to which it was applied, it is not always the best for most industrial applications. The development of a 'classical' approach to modern control theory was centred specifically around problems encountered in rocket research. The applications included missile guidance and navigation, well-known missions being Mariner, Apollo and Ranger.

For example, rocket trajectory equations to give a desired orbit are inherently non-linear. To design a control guidance system for the rocket to follow a given trajectory optimally, requires that these equations first be linearised. Then the solution can be formulated into a linear quadratic type of control problem of which the solution is now well-known [3 - 5].

Estimation theory is usually employed in telemetry systems e.g. the radar tracking of aircraft. Linear least-squares filtering or smoothing may be used here [1]. The idea of least-squares prediction first stemmed from the 'anti aircraft fire control problem'. This required the

ability to predict the motion of an aircraft in order to intercept it. Wiener [1] solved the problem for this application intending to have a hardware solution.

A specific advantage of the methods developed in this thesis is that many of the assumptions made in the 'classical' approaches do not have to be made. In particular, the assumption that the model of the process is unknown cannot be made when using the Kalman approach to optimal control or estimation theory. Similarly the Wiener approach to estimation requires at least a knowledge of the spectral density of the corrupted signal. Again this assumption need not be made in this work.

The thesis comprises approximately two thirds estimation theory and one third control theory. As is appropriate at this time, it concentrates only on the discrete-time formulation of the problem, i.e. a solution which will be implemented digitally.

1.2 Linear Least-Squares Estimation

The underlying theme in linear estimation theory is the extraction of a discrete signal vector $\underline{y}(k)$ contaminated with stationary zero mean white noise.

The measured (or observed) process $\underline{z}(k)$ is given as

$$\underline{z}(k) = \underline{y}(k) + \underline{v}(k) \tag{1.1}$$

where $\underline{y}(.)$, $\underline{z}(.)$ and $\underline{v}(.)$ are of dimension r. The white noise vector $\underline{v}(.)$ has a known covariance

$$\text{cov}\left[\underline{v}(k),\underline{v}(n)\right] = R\delta(k-n) \tag{1.2}$$

The linear estimation problem is now posed as follows:

Given the stationary measurement sequence (1.1), determine the estimate of the signal $\underline{y}(k)$ given data up to and including time k_f (denoted $\hat{\underline{y}}(k/k_f)$) such that the criterion

$$J1 = E\{(\underline{y}(k) - \hat{\underline{y}}(k|k_f))^T(\underline{y}(k) - \hat{\underline{y}}(k|k_f))\} \tag{1.3}$$

is minimised.

Depending whether $k < k_f$, $k = k_f$ or $k > k_f$ will determine whether $\hat{\underline{y}}(k|k_f)$ is a smoothed, filtered or predicted estimate respectively. Two approaches are traditionally employed to give the solution to the above.

A. Frequency Domain Solutions

The frequency domain solutions to the linear estimation problem is found by the independent work of Wiener [1] (continuous time) and Kolmogorov [2] (discrete time). Wiener's solution to the above problem which was of the scalar case (r = 1) assumed a complete knowledge of all past observations ($k_o = -\infty$) and was achieved using variational arguments.

The result is the so-called continuous time Wiener-Hopf equation, the solution of which yields the impulse response of the optimal estimator. The transfer function solution can also be found from this. The well-known solution involves a Wiener-Hopf spectral factorisation of the spectral density matrix of the measurement process [1]. Later, Wong and Thomas [6] generalised this work to the multivariable case for filtering and prediction in continuous time.

B. State-Space Solutions in the Time Domain

Kalman [3,4] altered the formulation of the problem by modelling the signal process in the more general time-varying observable form :

$$\underline{y}(k) = C(k)\underline{x}(k) \tag{1.4}$$

$$\underline{x}(k+1) = A(k)\underline{x}(k) + D(k)\underline{\omega}(k) \tag{1.5}$$

with $k_o > -\infty$ and where $\underline{x}(k)$ is an n x 1 state vector, $\underline{\omega}(k)$ is a q x 1 process noise vector and C(k), A(k) and D(k) are known matrices, and

$$\text{cov}\left[\underline{\omega}(k),\underline{\omega}(n)\right] = Q\delta(k-n) \tag{1.6}$$

The Q and R matrices are non-negative definite matrices whose true values are assumed to be known. The resultant time varying Kalman filter solution is found iteratively [60,61].

$$\hat{\underline{x}}(k+1|k) = A(k)\hat{\underline{x}}(k|k) \tag{1.7}$$

$$\hat{\underline{x}}(k|k) = \hat{x}(k|k-1) + K(k)(z(k) - C(k)\hat{\underline{x}}(k|k-1)) \tag{1.8}$$

$$\hat{\underline{y}}(k|k) = C(k)\hat{\underline{x}}(k|k) \tag{1.9}$$

The Kalman gain matrix is updated from

$$K(k) = P(k|k-1)C^T(k)R_\varepsilon^{-1}(k) \qquad (1.10)$$

where the innovations covariance matrix is given as

$$R_\varepsilon(k) = C(k)P(k|k-1)C^T(k) + R \qquad (1.11)$$

and the n x n estimation error covariance matrix is found from

$$P(k|k) = (I - K(k)C(k))P(k|k - 1) \qquad (1.12)$$

and

$$P(k + 1|k) = A(k)P(k|k)A^T(k) + D(k)QD^T(k) \qquad (1.13)$$

Numerous other versions of the Kalman filter exist [11].

Although the Kalman filter is derived for the general use of time varying non-stationary processes, it follows that for stationary systems of fixed dynamics the Kalman filter will converge in the steady state to the Wiener case.

Kalman [3,4] has formulated the linear estimation problem in terms of orthogonal projection in Hilbert space. However, this solution was restricted to prediction and filtering. Rauch [7] solved the smoothing problem in state space followed by Meditch [8,9] and others, e.g. Kailath [11] using an innovations approach.

There are various different kinds of smoothing problem. Fixed lag, fixed interval and fixed point are the major ones, and are discussed in Anderson and Moore [61]. A notable paper which combines the various state space fixed

lag smoothing algorithms is by Moore [10].

Following Kalman's work, interests changed from the input-output type of solution to state space solutions for both estimation and control. There are some drawbacks with the widely-used Kalman filter approach. Firstly, an accurate model of the signal process must be found in state-space form. Secondly, the Q and R matrices must be known or estimated apriori. Thirdly, it is known that the Kalman filter can be used for both time-varying and non-stationary processes. These conditions are met in systems like those for rocket guidance where the changing dynamics and noise statistics are known exactly.

In process control applications, it is not likely that a true time-varying model of the process can be found or accurate representations of the varying Q and R matrices. If the model of the signal process is not correct, divergence may occur in the algorithm [58]. If the Q and R matrices are not known exactly, the filter will give only a sub-optimal estimate. These two major problems have been tackled in several papers and in this thesis. One approach to the divergence problems has been studied by using a 'limited memory' Kalman filter, Jazwinski [12] which deletes old data and thus does not accumulate any major numerical errors. This problem is solved in the frequency domain for a smoothing problem in Chapter 2 of this thesis, and is shown to relate to a Wiener type of smoother and to the recent work of Grimble [13,14,15]. This type of solution has certain advantages since it gives

a transfer function solution to the smoothing problem which is very appropriate for implementation in some applications.

Filtering when the process is known but the Q and R matrices are not, has been solved by mainly Mehra [16,17,18] , Alspach and Abiri [19], Bélanger [20] and Carew and Bélanger [21]. These techniques are surveyed by Leondes and Pearson [22]. The problem which has not been solved before is filtering and smoothing in multivariable processes where both the process is unknown and the noise statistics are unknown. Macfarlane [24], Arcasoy [25], Barrett [26] and Shaked [27] give a relationship in the frequency domain for the steady-state matrix Riccati equation of the Kalman filter. This result shows that the spectral density matrix of the measurement process, when Wiener spectral factorised, is related to the return-difference matrix of the stationary Kalman filter. This combines the alternative frequency domain or time domain approaches of the Wiener filter and Kalman filter. A similar result is provided in the polynomial matrix approach of Kucera [28]. The adaptive estimation scheme which embodies the main part of this thesis uses this type of result..

1.2.1 Self-Tuning Estimation Theory

Kalman filtering for process control needs an accurate model of the process and the knowledge of the Q and R noise covariance matrices. In Wiener filtering, the spectral density of the observation process must be known or the transfer function of the signal generating process together with Q and R. These assumptions are not made here, for the first time, as applied to multi-variable processes. Recall that already Kalman filtering where a knowledge of a process model is not assumed to be known has been covered by Anderson and Moore [92,93]. This is achieved by employing the Kalman filter as a time-varying Wiener filter requiring time-varying spectral factorisation [94]. Signal statistics as would be required if using a Wiener filter must be known. This method therefore has the same limitations as the Wiener filter.

The self-tuning approach taken in this work overcomes these problems by assuming that the signal process model and the observation process model are unknown but can each be modelled as an autoregressive moving average (ARMA) time series of known order. If the order is unknown, it may be estimated beforehand [29,59]. The parameters of the ARMA model can then be estimated using, for example, extended recursive least squares or the real time maximum-likelihood method [56]. The measurement noise covariance can also be estimated simultaneously, assuming that it is white. The filtered, predicted or

smoothed signal estimates can then be found based on this information. The process parameters can either be constant or slowly time-varying. If constant, the self-tuning estimation scheme will converge to a Wiener type solution but the burden of spectral factorisation is omitted, as is the separation of causal from anti-causal modes.

The first adaptive estimation scheme based on the identity of Astrom [34] was by Wittenmark [30]. This was a self-tuning predictor for scalar processes. This has been generalised to the multivariable case by Tanttu [31] and applied to forecasting by Sym and Wellstead [57].

A self-tuning filter/predictor for multivariable processes was given by Ledwich and Moore [32], of which Wittenmark's work [30] can be derived as a special case. Hagander and Wittenmark [33] solve the self-tuning fixed-lag smoothing problem for scalar processes. Chapter 3 of this thesis will generalise this work for the multivariable case and show how all three multivariable (filtering, smoothing and prediction) self-tuning estimators may be found from one original frequency domain estimator. This gives a more unified theory. When the process and noise statistics are assumed to be known, the estimators can be compared with the work of other authors e.g. Shaked [27], Astrom [34] and Tanttu [31]. The convergence of the self-tuning filter and smoother will also be shown by using the asymptotic analysis of Ljung [35,36,37]. Thus in this thesis, it is possible to achieve optimal filtering, smoothing or

prediction for multivariable processes without the classical restrictions imposed of a knowledge of signal model or of noise covariances as taken in Kalman theory. Similarly, the assumption of a known spectral density of the observed process is not made.

1.3 Linear Least-Squares Stochastic Control

A problem often solved in this area is Linear Quadratic Gaussian (LQG) optimal control. The problem formulated here considers either making stationary control around a fixed set point value (the regulator problem) or to follow a time-varying reference signal (the servo problem). If the process is observed in the form of (1.1) then $\underline{y}(k)$ is assumed to be the output of a known linear system driven by stationary white noise plus a control signal vector $\underline{u}(k)$ to achieve a feedback control action. The problem now becomes one of deriving the control action $\underline{u}(k)$ such that the performance criterion

$$J2 = E\{\underline{y}^T(k)Q_1\underline{y}(k) + \underline{u}^T(k)R_1\underline{u}(k)\} \qquad (1.14)$$

is minimised,

where the weighting matrices are defined as follows:
Q_1 is a constant symmetrical semi-positive definite matrix,
and
R_1 is a constant symmetrical positive definite matrix.

Kalman [5] solved this problem in the time domain using a state-space approach and Riccati equations. The solution requires a Kalman filter and a time-varying feedback gain control matrix. This solution has similar advantages and disadvantages in application as discussed for the Kalman filter in 1.2. The first frequency domain solution was found by Newton et al [38], but this cannot be used for unstable processes.

Astrom [34] uses so-called minimum variance control assuming no

weighting on control and no measurement noise. The resultant solution gives an unstable control law for non-minimum phase systems and may require a large control effort. Although most continuous time systems may be minimum phase the discrete-time equivalent system may well be non-minimum phase due to zeros which drift out of the unit circle at certain sampling rates. Therefore consideration of the non-minimum phase problem is justified. A more general form of the minimum variance controller is found by Clarke and Hasting-James [39] under the assumption of weighting on the control as well as the output. A minimum variance controller for non-minimum phase systems has been developed by Peterka [40] and Astrom and Wittenmark [41] but these may require a large control effort due to the omission of control weighting. A control-weighted minimum variance controller for non-minimum phase systems has been introduced recently by Grimble [42]. This controller is similar to the Clarke-Hasting-James controller but reduces to the controller of Peterka when the control weighting is zero. It also cannot cope with the problem of measurement noise.

The LQG solution in the frequency domain which minimises (1.14) overcomes the previously discussed difficulties. The solution has been derived by Shaked [43] and by Grimble [44] in z transfer-function format, and by Kucera [45] using polynomial algebra. The LQG solution can deal with open-loop unstable and/or non-minimum phase systems. Measurement noise (usually white) can be

allowed for. Disadvantages are in the complexity of solution. The Kalman time-domain approach requires the iteration of two matrix Riccati equations, one for the filter and one for the control. Solution in the frequency domain is by one of two approaches. The first requires the solution of one or two diophantine equations and two polynomial spectral factorisations when using Kucera's work. The second requires two spectral factorisations and the removal of causal from anti-causal modes if using the Shaked or Grimble approach (op cit). The following section will show how some of these problems can be overcome by using on-line identification in a self-tuning framework.

1.3.1 Self-Tuning Control

Self-tuning control theory was developed in parallel with self-tuning estimation theory, for the same underlying reasons. Models of processes are normally inaccurately known and sometimes the dynamics may even vary with time. Self-tuning control can cope with this problem provided the parameters which define the internal dynamics vary slowly with time about a nominal value. The first practical self-tuner was only for regulation and was based on the minimum variance controller of Astrom (op cit). This was achieved by Astrom and Wittenmark [46] and had the same disadvantages as the minimum variance controller. The addition of control weighting and set point following (Clarke and Gawthrop [48,49]) was achieved by using the controller of Clarke and Hasting-James (op cit). Wellstead et al [50] and [51] , use pole assignment for self-tuning. The advantages of pole placement are that it can be used effectively on non-minimum phase and unstable processes. The disadvantage of this explicit approach is that a diophantine equation has to be solved on-line at each sampling instant to determine the controller parameters. This can be time-consuming if used on real-time systems. Other self-tuners developed have included a minimum variance self-tuning regulator for non-minimum phase systems (Astrom and Wittenmark [41]) and a control weighted self-tuning controller for non-minimum phase systems (Grimble [52]).

More recently a state space approach to pole-assignment self-tuning control has been introduced by Tsay and

Shieh [53] and Warwick [54]. These algorithms have the same advantages as the work of Wellstead et al (op cit).

The LQG self-tuning control problem has already been solved by Lam [55]. The author employs a steady-state Kalman filter and a continuously updated feedback gain matrix. The advantages of the LQG methods have already been discussed. In this thesis, the work will be extended to enable LQG self-tuning from input-output rather than state-space considerations, by using the work of Kucera (op cit) in polynomial equation form. Recursive least squares will be shown to overcome the problem of solving a diophantine equation by deriving an implicit LQG self-tuner. Moreover, non-minimum phase and unstable systems can be dealt with easily. For non-minimum phase systems, no separation of minimum from non-minimum phase terms is required. The separation idea has been used previously by Astrom and Wittenmark [41] and Grimble [52].

CHAPTER 2

Transfer Function Solution to a Multivariable Finite Interval Smoothing Problem

2.1 Introduction

When using linear least squares estimation techniques in applications where the available data is finite, there are two approaches. Consider first the example of least squares filtering. A Wiener frequency domain filter or a time-domain Kalman filter can be used. The Wiener filter is derived under the assumption that an infinite amount of past observed data is available ($k_0 = -\infty$). This assumption has to be made in order that the resulting frequency domain filter is time-invariant. For this situation, the Kalman filter approach only assumes the availability of a finite amount of observed data ($k_0 > -\infty$) and gives a smaller estimation error than that of the Wiener filter. The underlying problem faced here is that the Kalman filter is time varying and cannot be realised in transfer function form for the finite data problem and the Wiener filter is not intended for use in applications where there is only a finite amount of data available and is inferior. Grimble [13,14] has derived new time-invariant filters for finite data records. The transfer function solution gives an optimal filtered state estimate at the end of the finite time interval. This estimate is equivalent to the optimal time-varying Kalman filter estimate and is superior in a least-squares sense to either that of a Wiener filter or steady-state Kalman filter. The finite-time prediction problem was solved first by Zadeh and Ragazzini [66] and enabled the

optimal impulse response of the predictor to be found. The impulse response of the predictor vanishes outside of the specified time interval and thus the predictor has a finite or limited memory.

Limited memory filters are well-known [9], [67-72]. Terminologies vary from 'Limited memory', 'Fixed memory' to 'moving window' and 'sliding memory'. The latter terminology is used mainly in economic systems. The technique involves the processing of the most recent N measurements of data, where the 'window' or 'memory' length N is fixed a priori depending upon the application. At each sampling interval when a new data point is included, an old data point is dropped. This method of discarding old data allows it to be employed in situations where an expanding memory filter (or Kalman filter) would diverge due to erroneous data [12].

The solution to this filtering problem in references [12] and [72] employ two expanding memory filters, one used for data addition and one used for data deletion. The advantages and limitations of these filtering schemes are summarised in [70]. Bierman, in [71] concentrates on achieving greater numerical stability by using a square root covariance technique. Blum [69] uses orthogonal polynomials in a recursive scheme in order to fit a polynomial to the N observed data points. This enables past, present or future values of the data or its derivatives to be found.

Recently a time-invariant smoothing filter has been proposed for finite data records [15]. The problem was solved in the frequency domain by posing an equivalent infinite-time ($t_0 = -\infty$) problem. It was shown how the solution to the infinite-time smoothing problem also gave rise to the solution to the finite-time problem of interest. The equivalence of the two estimators depended upon the choice of the estimation error criterion and the problem descriptions.

In this chapter a discrete-time version of the finite-time smoother is derived [62]. The driving process and measurement noise sequences are assumed to be stationary. The smoother is required to give an optimal estimate of the state vector of the process ($\hat{\underline{x}}(K1|N)$) at time K1, given the observations in [0,N]. This fixed interval problem normally has a solution involving both causal and anticausal filters but cannot be used for on-line smoothing [73-75]. The advantage of the proposed smoother is that it may be used for on-line smoothing by utilising either the weighting sequence or z transfer function forms.

The optimal time-invariant smoother is obtained by solving a discrete-time Wiener-Hopf equation. This equation is expanded and is transformed into the frequency domain by taking the two-sided z transform. This provides a transfer function solution to the smoothing problem from which the weighting sequence is easily found. It is shown that the transfer function solution involves z^{-N} terms

which can be neglected if the smoother is to be used within the fixed interval $[0,N]$. For implementation beyond the finite interval, the z^{-N} terms must be included in the z-transfer function version of the smoother. However, in the weighting sequence form this sequence need only be set to zero outside the interval. The resulting smoother is of the fixed-lag variety but with limited memory.

2.2 System Description

The linear, time-invariant system can be expressed in the discrete time state-space format

$$\underline{x}(k+1) = A\underline{x}(k) + D\underline{\omega}(k) + \underline{u}(k), \quad k = 0, 1, 2 \ldots. \tag{2.1}$$

$$\underline{y}(k) = C\underline{x}(k), \ \underline{x}(k)\varepsilon R^n, \ \underline{y}(k)\varepsilon R^n, \ \underline{\omega}(k)\varepsilon R^q \tag{2.2}$$

The signal $\underline{y}(k)$ is assumed to be corrupted with noise $\underline{v}(k)$, resulting in an observation vector

$$\underline{z}(k) = \underline{y}(k) + \underline{v}(k) \tag{2.3}$$

The Gaussian white noise sequences $\omega(k)$ and $\underline{v}(k)$ are assumed to be zero mean with covariances:

$$\text{COV}\left[\underline{\omega}(k), \ \omega(n)\right] = Q\delta(k-n) \tag{2.4}$$

$$\text{COV}\left[\underline{v}(k), \ \underline{v}(n)\right] = R\delta(k-n) \tag{2.5}$$

$$\text{COV}\left[\underline{\omega}(k), \ \underline{v}(n)\right] = G\delta(k-n) \tag{2.6}$$

The state, process noise and measurement noise vectors have dimensions n, q and m respectively.

In the following, the assumption is made that the initial state vector has a known mean $\underline{m}_0$ and variance Σ_0. The finite interval problem of interest cannot be solved easily by considering only the interval $[0,N]$, where N is the interval length. It is therefore assumed that the smoother has been in operation from time $k_0 = -\infty$ but at time zero the system receives an input which creates new "initial" conditions. In this manner, the invariant problem of interest is embedded within an infinite time problem [13].

In order to initialize the smoother, it is assumed that an impulse of magnitude $\underline{x}_1' = A\underline{x}_0'$ is applied to the system at time zero. That is

$$\underline{u}(k) = \delta(k)\, \underline{x}_1' \tag{2.7}$$

where

$$E\left[\underline{x}_0'\right] \triangleq \underline{m}_0 \tag{2.8}$$

and

$$E\left[\underline{x}_0' \underline{x}_0'^{T}\right] \triangleq \Sigma_0 \geq 0 \tag{2.9}$$

$$E\left[\underline{x}_0' \underline{\omega}^{T}(k)\right] = 0 \tag{2.10}$$

$$E\left[\underline{x}_0' \underline{v}^{T}(k)\right] = 0 \tag{2.11}$$

Define the resolvant matrix

$$\Phi(z) \triangleq (zI - A)^{-1} \tag{2.12}$$

and the transfer function matrix

$$W(z)\ C\Phi(z)D = \mathcal{Z}_1(w(k)) \tag{2.13}$$

where

$$\left.\begin{aligned} w(k) &\triangleq C\Phi(k)D \quad \forall\ k \geq 0 \\ &\triangleq 0 \qquad\ \forall\ k \leq 0 \end{aligned}\right\} \tag{2.14}$$

The state trajectory of the system in convolution summation form is given as:

$$\underline{x}(k) = \sum_{i=-\infty}^{k-1} \phi(k-i-1)(D\underline{\omega}(i) + \delta(i)\underline{x}_1') \tag{2.15}$$

where $\Phi(k) = \phi(k-1)U(k-1) = A^{k-1}U(k-1)$ and where $U(k)$ is the Heaviside Unit Step function. Equation (2.15) becomes

$$\underline{x}(k) = \Phi(k)\underline{x}_1' + \sum_{i=-\infty}^{k-1} \Phi(k-i)D\underline{\omega}(i) \quad \forall\ k > 0 \tag{2.16}$$

from which the state vector at time zero is

$$\underline{x}(0) = \sum_{i=-\infty}^{-1} \Phi(-i)D\underline{\omega}(i) \qquad (2.17)$$

Now

$$\bar{\underline{\omega}}(i) = 0$$

where the bar denotes the expected value and equation (2.16) can be written using (2.8) as

$$\bar{\underline{x}}(k) = \Phi(k)A\underline{m}_0 = \phi(k)\underline{m}_0 \quad \forall\ k > 0 \qquad (2.18)$$

$$= 0 \qquad \forall\ k \leq 0 \qquad (2.19)$$

Thus the delta function input signal establishes the specified mean value for the initial state vector [14].

Let the total state vector at time k = 0 be defined as

$$\underline{x}_0 = \underline{x}_0' + \underline{x}(0) \qquad (2.20)$$

then for time $k \geq 0$ equation (2.16) becomes:

$$\underline{x}(k) = \phi(k)\underline{x}_0 + \sum_{i=0}^{k-1} \Phi(k-i)D\underline{\omega}(i) \qquad (2.21)$$

Equation (2.21) follows directly from substitution for $\underline{x}_0'$ into equation (2.16) and by using $\underline{x}(0)$ from equation (2.17).

2.3 The Optimal Finite Interval Smoothing Filter

Let $\hat{h}(k;N,\ell)$, $k = 0,1 \ldots$ be the weighting sequence of the optimal smoother. The weighting is a function of the interval length N, the smoothing time lag ℓ and the time index k.

If the ideal output from the smoother at time k is defined to be $\underline{x}_i(k)$, then the estimation error can be defined for all k as

$$\tilde{\underline{x}}(k) \triangleq \underline{x}_i(k) - \hat{\underline{x}}(k) \qquad (2.22)$$

By subtracting signals prior to time zero, the ideal output of the smoother considers the state component due to plant signals at times greater than zero only. That is

$$\underline{x}_i(K1) = \underline{x}(K1) - \psi(N;\ell)\underline{x}(0) \qquad (2.23)$$

and $K1 = N - \ell$ denotes the time at which the state estimate is required. The $\psi(N;\ell)$ matrix is included to optimize the zero input response of the smoother. The estimation error for $k = K1$ can be thought to involve only system inputs for times greater than zero and the vector $\underline{x}_0'$ [13]. Since the vector $\underline{x}_0'$ was assigned the statistics $(\underline{m}_0, \Sigma_0)$ when defining the finite-time problem, it follows that by solving the given infinite-time problem, the desired finite-time problem solution emerges. A second consequence is that the impulse response of the optimal smoother is zero at the end of the interval, giving $\hat{h}(k;N,\ell) = 0$, $\forall\ k \geq N$.

The smoothing problem can now be defined as follows. Given the observation sequence $\underline{z}(k)$ over $k = 0, 1 \ldots N$,

determine the optimal estimate of the state trajectory $\underline{\hat{x}}(K1|N)$ such that the following error criterion is minimised.

$$J = E\left[\underline{\tilde{x}}(K1)^T\, \underline{\tilde{x}}(K1) \,\middle|\, \underline{z}(0) \ \ldots\ \underline{z}(N)\right] \tag{2.24}$$

Note that for this particular problem, the optimal state estimate is a fixed interval smooth estimate of the state at time K1 given observations in [0,N]. Estimates taken at other times within the interval are sub-optimal. Outside the interval the smoother acts as an optimal fixed-lag smoother with lag ℓ, using the optimal impulse sequence as a "sliding memory". Within the first interval $k \leq N$, for an unbiased smoother state estimate:

$$E\left[\underline{\hat{x}}(K1|N)\right] = E\left[\underline{x}(K1)\right] \tag{2.25}$$

and outside the interval, $k > N$

$$E\left[\underline{\hat{x}}(k-\ell|k)\right] = E\left[\underline{x}(k-\ell)\right] \tag{2.26}$$

Theorem 2.1 Time Invariant Smoothing Filter

The optimal transfer function matrix of the discrete time smoother which minimizes equation (2.24) in a least squares sense is given in z-transfer function form as:

$$\hat{H}(z;N,\ell) = n_{11}(z,\ell)\, \Delta^{-1}(z) + \psi(N;\ell)\left[\bar{n}_{12}(z,N) + n_2(z,N)\right]\Delta^{-1}(z) \tag{2.27}$$

where

$$n_{11}(z,\ell) \triangleq \left\{\Phi(z)D(G + QW^T(z^{-1}))\, \Delta^{-T}(z^{-1})z^{-\ell}\right\}_+ \tag{2.28}$$

$$n_{12}(z,N) \triangleq -\, n_{11}(z,N) \tag{2.29}$$

$$n_2(z,N) \triangleq \Sigma_0 \left\{M^T(z^{-1})z^{-N}\right\}_+ \tag{2.30}$$

$$M(z) \triangleq \Delta^{-1}(z)C\Phi(z)A \tag{2.31}$$

If a signal over all time has a 2 sided transform $\{f_{(z)}\}$ then the transform of the positive-time contribution is denoted by $\{f(z)\}_+$.

$\Delta(z)$ is obtained from generalised spectral factorisation, Shaked [27]

$$W(z)QW^T(z^{-1}) + R + W(z)G + G^TW^T(z^{-1})$$
$$= \Delta(z)\,\Delta^T(z^{-1}) \qquad (2.32)$$

The $\psi(N;\ell)$ matrix may be calculated using:

$$\psi(N;\ell) = \left[\phi(N-\ell) - I_{11}(N;\ell)\right]$$
$$\times \left[I_n + I_{12}(N) + \Sigma_0 S(N)\right]^{-1} \qquad (2.33)$$

with $N-\ell = K1$ (2.34)

and

$$I_{11}(N,\ell) \triangleq \sum_{k=0}^{n-1} n_{11}(k,\ell)M(N-k) \qquad (2.35)$$

$$I_{12}(N) \triangleq \sum_{k=0}^{N-1} n_{12}(k,N)M(N-k) = -I_{11}(N,N) \qquad (2.36)$$

$$S(N) \triangleq \sum_{k=0}^{N-1} M^T(N-k)M(N-k) \qquad (2.37)$$

Proof of Theorem 2.1

The proof of Theorem 2.1 is given in Appendix 1

2.4 Other Forms of the Optimal Smoother

Several cases of special interest now emerge from equation (2.27) as follows.

Special Case 1

Consider
$$\lim_{N\to\infty} \hat{H}(z;N,\ell) \to n_{11}(z,\ell)\ \Delta^{-1}(z) \qquad (2.38)$$

This is the steady-state or infinite time fixed-lag smoother as expressed in transfer function form in the z-domain. The smoothing time delay is given by ℓ sampling intervals.

By multiplying by the C matrix, equation (2.38) becomes

$$\lim_{N\to\infty} \hat{H}'(z;N,\ell) \to \left\{(W(z)G + W(z)QW^T(z^{-1})) \times \Delta^{-T}(z^{-1})\, z^{-\ell}\right\}_+ \Delta^{-1}(z) \qquad (2.39)$$

This equation represents an infinite-time fixed lag smoother which can be used to estimate a signal $\underline{y}(k)$ from an observation message $\underline{z}(k)$ when the state estimate is not required.

Special Case 2

If $\ell = 0$ then

$$\lim_{N\to\infty} \hat{H}(z;N,0) \to n_{11}(z,0)\ \Delta^{-1}(z) \qquad (2.40)$$

or multiplying by the C matrix

$$\lim_{N\to\infty} \hat{H}'(z;N,0) \to \left\{(W(z)G + W(z)QW^T(z^{-1})) \times \Delta^{-T}(z^{-1})\right\}_+ \Delta^{-1}(z) \qquad (2.41)$$

Equations (2.40) and (2.41) represent a transfer function solution to the discrete-time stationary Kalman-Bucy filtering problem when dealing with state or signal estimation respectively. The two above cases can also be termed infinite interval or Wiener type estimators. Case 2 has been studied by Shaked [27] and in continuous time by Barrett [26].

Special Case 3

Consider

$$\hat{H}(z;N,o) = n_{11}(z,0)\Delta(z)^{-1} + \psi(N;0) \times [n_{12}(z,N) + n_2(z,N)]\Delta^{-1}(z) \qquad (2.42)$$

Equation (2.42) is a finite interval, time-invariant filter. This filter has been investigated both in discrete-time and continuous-time by Grimble [13] and [14]. The filter has the property that, i.e. $\hat{\Lambda}$ will obtain an optimal state estimate at the end of an interval of length N sampling intervals. Grimble [13] has shown that the error covariance of the state is smaller than that of a Wiener filter and equivalent to that of a time-varying Kalman filter when estimating the state at the end of the interval. The filter has similar advantages to those of the finite interval smoother and can be implemented in either transfer function or weighting sequence forms.

Special Case 4

$$\hat{H}(z;N,N) = n_{11}(z,N)\Delta(z)^{-1} + \psi(N;N) \times [n_{12}(z,N) + n_2(z,N)]\Delta^{-1}(z) \qquad (2.43)$$

Equation (2.43) represents an initial condition estimator

which is again time invariant. In the time domain, Lévy et al [73] have shown how the continuous-time equivalent of equation (2.43) could be found by solving a time-reversed, truncated Wiener-Hopf equation.

The smoother of Lévy et al is a combination of causal and anticausal filters and cannot be used for on-line smoothing. This is a standard solution to the fixed interval smoothing problem, e.g. Fraser and Potter [75].

EXAMPLE 2.1

Finite Interval Smoother

Consider the following scalar signal generating system:

$$x(k+1) = ax(k) + \omega(k)$$

$$z(k) = cx(k) + v(k)$$

for $k \geq 0$ and $|a| \leq 1$.

The independent noise sequences $\omega(k)$ and $v(k)$ are assumed to be zero mean with covariances q and r respectively. The system is assumed to have been in operation from time $k_0 = -\infty$. At time zero, the initial state vector is assumed to have covariance $\Sigma_0 = \sigma_0^2$.

Given the observation sequence $z(k)$, $k = 0,1 \ldots N$ an optimal estimate of the state at time K1 is required from the finite-time smoother at time N. For observations $k > N$ the optimal smoother is required to act as a fixed-lag smoother with smoothing delay $\ell > 0$.

The system resolvant and transfer function are found as:

$$W(z) = C\Phi(z)D = c/(z-a)$$

with

$$\Phi(z) = (zI-A)^{-1} = 1/(z-a)$$

and

$$\Phi(k) = a^{k-1}U(k-1)$$

The spectral factors are:

$$W(z)QW^T(z^{-1}) + R = \Delta(z)\Delta^T(z^{-1})$$

giving

$$\Delta(z) = \left(\frac{ra}{\alpha}\right)^{\frac{1}{2}} (z - \alpha)/(z - a)$$

with

$$\alpha = \xi \pm (\xi^2 - 1)^{\frac{1}{2}} \quad ; \quad |\alpha| < 1$$

and

$$\xi = \left[\underline{q}(c)^2/ra + (a^2 + 1)/a\right]/2$$

$$M(z) \triangleq \Delta^{-1}(z)C\Phi(z)$$

$$= \left(\frac{\alpha}{ra}\right)^{\frac{1}{2}} ca/(z - \alpha)$$

and $n_{11}(z,\ell)$ follows from:

$$n_{11}(z,\ell) = \left\{\Phi(z)DQW^T(z^{-1})\Delta^{-T}(z^{-1})z^{-\ell}\right\}_+$$

$$= \beta\left[\frac{z^{1-\ell}}{z-a} + \frac{\alpha z^{1-\ell}}{1-\alpha z} - \frac{\alpha^{1+\ell}z}{1-\alpha z}\right]$$

where $\beta = qc\left(\frac{\alpha}{ar}\right)^{\frac{1}{2}}/(1-a\alpha)$ and $n_{12}(z,N)$ is found from

$$n_{12}(z,N) = -n_{11}(z,N)$$

The remaining terms are found as follows:

$$n_2(z,N) \triangleq \Sigma_0\left\{M^T(z^{-1})z^{-N}\right\}_+$$

which from reference [13] becomes

$$n_2(z,N) = \Sigma_0 \mathcal{Z}_1\left[M^T(N-k)(U(k) - U(k-N))\right]$$

$$= \sigma_0{}^2\left(\frac{\alpha}{ar}\right)^{\frac{1}{2}} ca\left[z^{-(N-1)} - \alpha^N z\right]/(1-\alpha z)$$

and by inverse z transforming

$$n_2(k,N) = \sigma_0{}^2\left(\frac{\alpha}{ar}\right)^{\frac{1}{2}} ca\alpha^{N-k-1}\left[U(k) - U(k-N)\right]$$

$$n_{11}(k,\ell) = \beta\left[(a^{k-\ell} - \alpha^{\ell-k})U(k-\ell) + a^{\ell-k}U(k)\right]$$

$$n_{12}(k,N) = -\beta\left[(a^{k-N} - \alpha^{N-k})U(k-N) + \alpha^{N-k}U(k)\right]$$

$$M(k) = \left(\frac{\alpha}{ar}\right)^{\frac{1}{2}} ca\alpha^{k-1}U(k-1)$$

The element $\psi(N;\ell)$ is determined from:

$$\psi(N;\ell) = (\phi(K1) - I_{11}(N;\ell)(1 + I_{12}(N) + \Sigma_0 S(N))^{-1}$$

where $\phi(K1) = a^{N-\ell}$

$$I_{11}(N;\ell) = \beta c a \left(\frac{\alpha}{ar}\right)^{\frac{1}{2}} \left[\left(\frac{a^{N-\ell} - \alpha^{N-\ell}}{a - \alpha}\right) + \left(\frac{\alpha^{N-\ell+1} - \alpha^{1+N+\ell}}{1 - \alpha^2}\right)\right]$$

$$I_{12}(N) = -I_{11}(N;N)$$

$$= -\beta c a \left(\frac{\alpha}{ar}\right)^{\frac{1}{2}} \frac{\alpha}{1 - \alpha^2} (1 - \alpha^{2N})$$

and

$$\Sigma_0 S(N) = (\sigma_0 c)^2 \alpha a (1 - \alpha^{2N})/(r(1 - \alpha^2))$$

The transfer function of the time-invariant smoother in the z-domain can now be obtained by using the result of Theorem 2.1.

$$\hat{H}(z;N,\ell) = n_{11}(z,\ell)\Delta^{-1}(z) + \psi(N;\ell) \times [n_{12}(z,N) + n_2(z,N)]\Delta^{-1}(z)$$

Hence using the previous results

$$\hat{H}(z;N,\ell) = \beta\left(\frac{\alpha}{ar}\right)^{\frac{1}{2}} \left\{\frac{(1-a\alpha)z^{1-\ell}}{(1-\alpha z)(z-\alpha)} - \frac{z(z-a)\alpha^{1+\ell}}{(1-\alpha z)(z-\alpha)}\right\}$$

$$+ \psi(N;\ell)\left\{\beta\left(\frac{\alpha}{ar}\right)^{\frac{1}{2}} - \frac{\sigma_0^2 c}{r}\right\}\frac{z(z-a)\alpha^{N+1}}{(1-\alpha z)(z-\alpha)}$$

$$-\psi(N;\ell)\left\{\left(\frac{\alpha}{ar}\right)^{\frac{1}{2}}\beta z^{1-N} + \left(\left(\frac{\alpha}{ar}\right)^{\frac{1}{2}}\beta - \frac{\sigma_0^2 c}{r}\right) \times \frac{(z-a)\alpha z^{1-N}}{(1-\alpha z)(z-\alpha)}\right\}$$

Alternatively, the weighting sequence for the smoother may be implemented as

$$\hat{h}(k;N,\ell) = \beta\left(\frac{\alpha}{ar}\right)^{\frac{1}{2}}\left(\frac{\alpha^{k-\ell} - \alpha^{\ell-k}}{1-\alpha^2}\right)(1-a\alpha)U(k-\ell)$$

$$+ \left\{\psi(N;\ell)\left[\beta\left(\frac{\alpha}{ar}\right)^{\frac{1}{2}} - \frac{\sigma_0^2 c}{r}\right]\right.$$

$$\left(\frac{\alpha^{1+N+k}(\alpha-a) - (1-a\alpha)\alpha^{N-k}}{1-\alpha^2}\right)$$

$$\left. + \beta\left(\frac{\alpha}{ar}\right)^{\frac{1}{2}}\left(\frac{(1-a\alpha)\alpha^{\ell-k} - (\alpha-a)\alpha^{1+\ell+k}}{1-\alpha^2}\right)\right\} U(k)$$

$$- \psi(N;\ell)\left[\beta\left(\frac{\alpha}{ar}\right)^{\frac{1}{2}}\alpha^{k-N} + \left(\frac{\sigma_0^2 c}{r} - \beta\left(\frac{\alpha}{ar}\right)^{\frac{1}{2}}\right)\right.$$

$$\left.\times\left(\frac{(1-a\alpha)\alpha^{N-k} + \alpha^{k-N+1}(1-\alpha-\alpha^2(1-a))}{1-\alpha^2}\right)\right] U(k-N)$$

By direct calculation it may be shown that:

$$\hat{h}(k;N,\ell) = 0 \quad \forall\ k > N$$

Because of the delay terms in the optimal smoother, the weighting sequence is the more attractive method of implementation. In practice, the terms in the weighting sequence for times $k \geq N$ may be neglected and the weighting sequence can be simply set to zero for these times.

Note that the weighting sequence for the optimal smoother also yields the following

$\hat{h}(k;\infty,\ell) \rightarrow$ infinite time smoother weighting sequence

$\hat{h}(k;\infty,0) \rightarrow$ Wiener or steady state Kalman filter

$\hat{h}(k,N,N) \rightarrow$ initial condition estimator weighting sequence

$\hat{h}(k,N,0) \rightarrow$ finite time filter weighting sequence

The finite interval smoother was simulated for the process of this example with the following parameters selected

$$a = 0.99,\ c = 0.1,\ q = 10,\ r = 1,\ \sigma_0^2 = 10$$

The interval length was chosen as 10 sampling increments ($N = 10$) with a smoothing lag of $\ell = 5$. The observation sequence for such parameters, together with the state and smoothed state estimate are shown in Figure 2.2. Clearly, the state estimate is acceptable. There is a delay of 5 steps between the state and the corresponding estimate and this is unavoidable. The impulse response for the time invariant smoother is shown in Figure 2.1 and takes the form of a rising and falling curve which has a peak value occurring at $k = \ell$. It is important to note that for times $k > N$, the smoothing filter has been used as a fixed memory, fixed lag smoother, using the 'moving window' of filter information present in the optimal impulse sequence. Figure 2.1 shows the impulse response to be zero for $k \geq N$ resulting in an observation length at any time k of only N, the observation length being $[\underline{k}-N, \underline{k}]$. Any observations prior to the interval $[\underline{k}-N, \underline{k}]$ is set to zero, thus discarding old data.

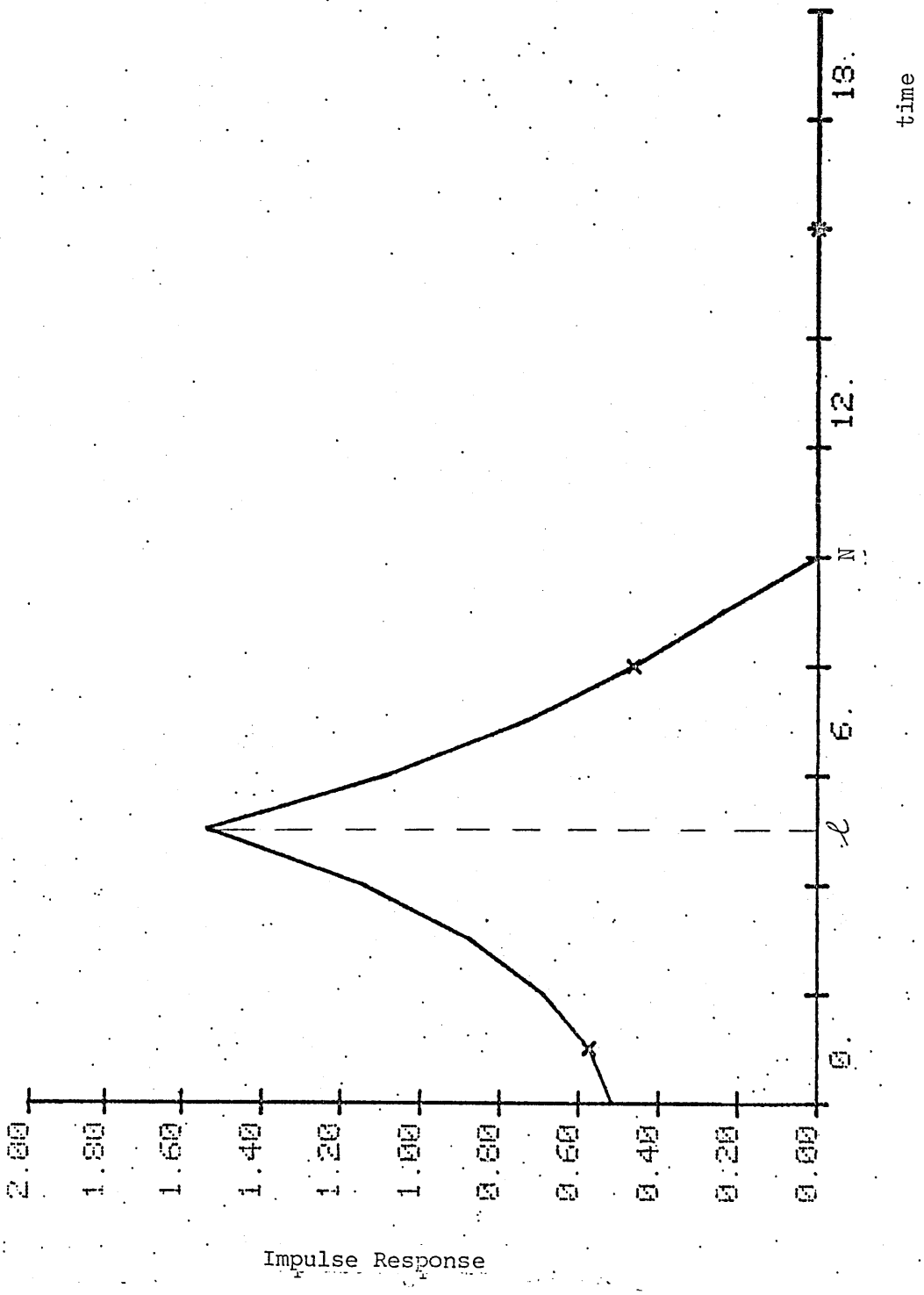


Figure 2.1 Impulse Response Sequence

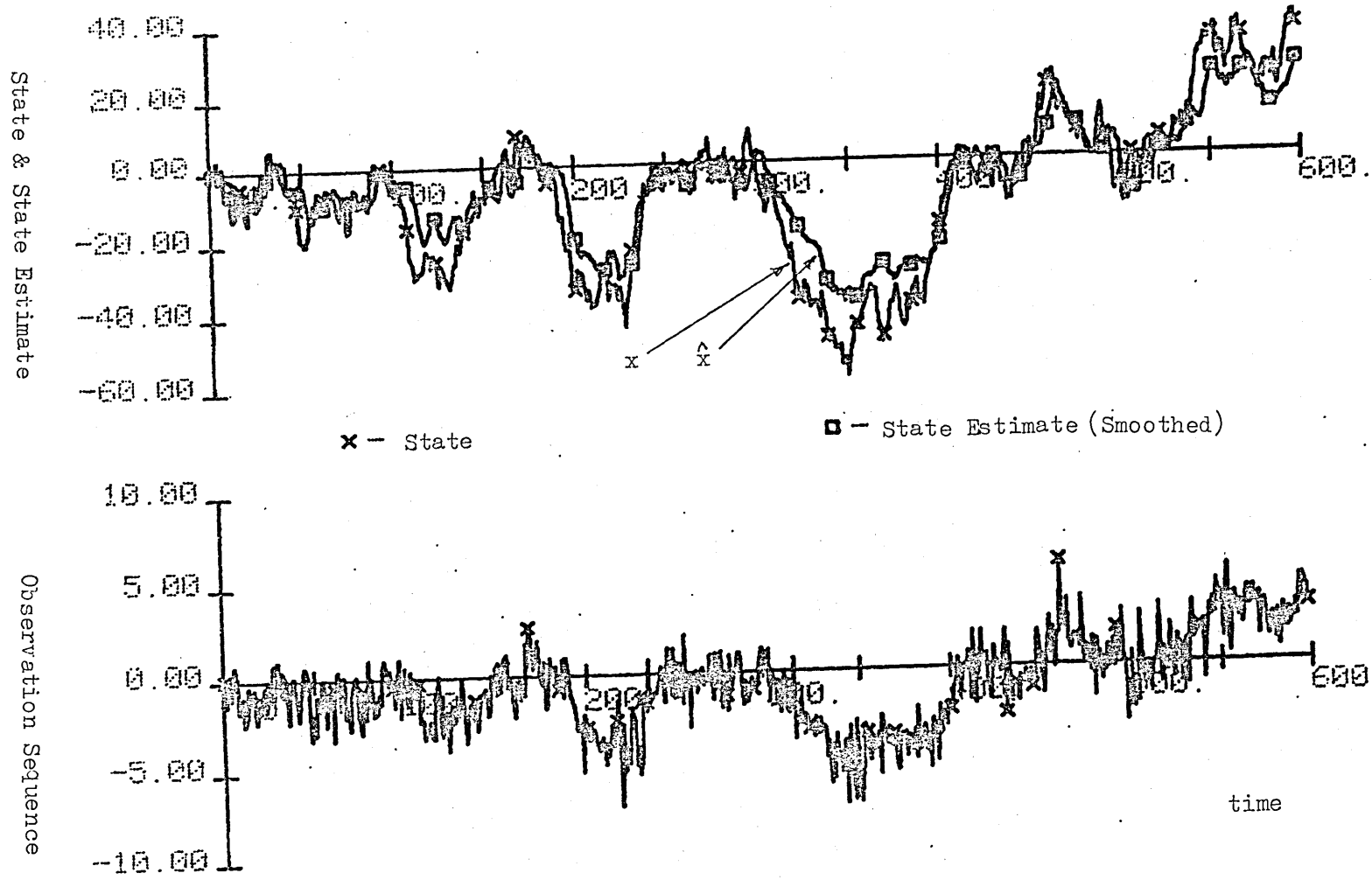


Figure 2.2 State, State Estimate & Observation

2.5 Further Simplifications to Smoother

The section will give a simplified algorithm for calculating the optimal smoother. The algorithm is found by manipulation of the steady-state matrix Riccati equation and thus links the steady-state time domain approach with the frequency domain. Further insight into the solution can then be seen. The assumption is made throughout that the process and measurement noises are uncorrelated ($G = 0$) and that all terms of z^{-N} can be ignored since these occur outside the finite interval. Smoothing outside the interval can still be achieved by setting the weighting sequence to zero at the end of the interval (op cit).

The steady-state discrete-time observer Riccati equation given by [60]:

$$P = APA^T - AKCPA^T + DQD^T \tag{2.44}$$

where K is the steady-state Kalman gain matrix

$$K = PC^T R_\varepsilon^{-1} \tag{2.45}$$

The covariance matrix for the innovations is given as (Sage and Melsa [60])

$$R_\varepsilon = R + CPC^T \tag{2.46}$$

P is the positive definite symmetric filtering error-covariance matrix. The relationship between the return-difference matrix for the steady-state Kalman filter and the spectral factor of the observation is (Arcasoy [27])

$$F(z)R_\varepsilon F^T(z^{-1}) = \Delta(z)\Delta^T(z^{-1}) \tag{2.47}$$

where the return difference matrix of the Kalman filter is

$$F(z) = C\Phi(z)AK + I \tag{2.48}$$

From (2.47)

$$\Delta(z) = F(z)R_\varepsilon^{\frac{1}{2}} \tag{2.49}$$

and

$$\Delta^T(z^{-1}) = R_\varepsilon^{\frac{T}{2}}F^T(z^{-1}) \tag{2.50}$$

From (2.28)

$$n_{11}(z,\ell)\underline{\Delta}\{\Phi(z)QW^T(z^{-1})\Delta^{-T}(z^{-1})z^{-\ell}\}_+ \tag{2.51}$$

Now it is possible to manipulate equation (2.44) by the addition and subtraction of various terms as demonstrated in Appendix 4.

$$\begin{aligned}\{\Phi(z)QW^T(z^{-1})\Delta^{-T}(z^{-1})z^{-\ell}\}_+ &= \{T(z^{-1})z^{-\ell} + \Phi(z)AKR_\varepsilon^{\frac{1}{2}}z^{-\ell}\}_+ \\ &= \{T(z^{-1})z^{-\ell}\}_+ + \Phi(z)AKR_\varepsilon^{\frac{1}{2}}z^{-\ell}\end{aligned} \tag{2.52}$$

where

$$T(z^{-1}) = P\left[\underline{M}^T(z^{-1}) + C^T\Delta^{-T}(z^{-1})\right] \tag{2.53}$$

$$= P\left[\underline{A}^T\Phi^T(z^{-1}) + I\right]C^T\Delta^{-T}(z^{-1}) \tag{2.54}$$

and M(z) is defined as in (2.31).

Now define the time function matrix T(k)

$$T(k) = \underset{\sim}{\mathcal{Z}}_1^{-1}(T(z)) \tag{2.55}$$

where $\underset{\sim}{\mathcal{Z}}_1^{-1}$ denotes inverse z transforming. Equation (2.52) then simplifies to

$$\begin{aligned}n_{11}(z,\ell) &= \underset{\sim}{\mathcal{Z}}_1\left[\underline{T}(\ell-k)(U(k) - U(k-\ell))\right] \\ &\qquad + \Phi(z)AKR_\varepsilon^{\frac{1}{2}}z^{-\ell}\end{aligned} \tag{2.56}$$

$$= \underset{\sim}{\mathcal{Z}}_1\left[\underline{T}(\ell-k)U(\ell-k)\right] + \Phi(z)AKR_\varepsilon^{\frac{1}{2}}z^{-\ell} \tag{2.57}$$

The filtering case when $\ell = 0$ is of importance since it is a special case of the smoother. Under these conditions it is simpler to determine $n_{11}(z,0)$ by the z^0 coefficient of the matrix Taylor series expansion of $T(z^{-1})$. By

defining

$$\{T(z^{-1})z^{-\ell}\}_+ \triangleq [T_0 + T_1 z + T_2 z^2 + \ldots + T_\ell z^\ell] z^{-\ell} \tag{2.58}$$

Then when $\ell = 0$

$$T_0 \triangleq \{T(z^{-1})\}_+ \tag{2.59}$$

giving(Chen [76])

$$T_0 = \lim_{z\to 0} T(z^{-1}) \tag{2.60}$$

$$= \lim_{z\to 0}\{PM^T(z^{-1}) + PC^{T_\Delta - T}(z^{-1})\}$$

Now

$$\lim_{z\to 0} PM^T(z^{-1}) \to 0 \tag{2.61}$$

and similarly

$$\lim_{z\to 0} PC^{T_\Delta - T}(z^{-1})$$

using (2.50)

$$= \lim_{z\to 0} PC^{TF-T}(z^{-1})R_\varepsilon^{-\frac{1}{2}}$$

$$= PC^T R_\varepsilon^{-\frac{1}{2}} \tag{2.62}$$

The expression for $n_{11}(z,0)$ now becomes

$$n_{11}(z,0) = T_0 + \Phi(z)AKR_\varepsilon^{\frac{1}{2}} \tag{2.63}$$

where

$$T_0 = PC^T R_\varepsilon^{-\frac{1}{2}} \tag{2.64}$$

The other terms of the optimal finite interval smoother also hold for the filtering case and follow

$$n_{12}(z,N) = -n_{11}(z,N)$$

$$= -\mathcal{Z}_1(T(N-k)) \tag{2.65}$$

(Ignoring the z^{-N} terms which occur outside the interval N)

Substitute the initial error covariance matrix $P_0 = \Sigma_0$, then

$$n_2(z,N) = \mathcal{Z}_1(P_0 M^T(N-k)(U(k) - U(k-N))) \qquad (2.66)$$

Define

$$n_{22}(z,N) = n_{12}(z,N) + n_2(z,N) \qquad (2.67)$$

$$= \mathcal{Z}_1(L(N-k)) \qquad (2.68)$$

where

$$L(N-k) = P_0 M^T(N-k) - T(N-k) \qquad (2.69)$$

Hence the transfer function matrix for the optimal smoother is greatly simplified, and becomes

$$\hat{H}(z;N,\ell) = [n_{11}(z,\ell) + \psi(N;\ell)n_{22}(z,N)]\Delta(z)^{-1} \qquad (2.70)$$

and $\psi(N;\ell)$ is also greatly simplified since

$$I_{11}(N,\ell) = \sum_{k=0}^{N-1} n_{11}(k,\ell)M(N-k)$$

$$= \sum_{k=0}^{N-1} T(\ell-k)M(N-k)$$

$$+ \sum_{k=\ell}^{N-1} [\phi(k-\ell)AKR_\varepsilon^{\frac{1}{2}} - T(\ell-k)]M(N-k) \qquad (2.71)$$

Put $I_{22}(N) = I_{12}(N) + P_0 S(N)$ (2.72)

and $I_{22}(N)$ can be simplified to be

$$I_{22}(N) = \sum_{k=0}^{N-1} L(N-k)M(N-k) \qquad (2.73)$$

and hence the $\psi(n;\ell)$ matrix becomes

$$\psi(N;\ell) = (\phi(N-\ell) - I_{11}(N,\ell))(I_n + I_{22}(N))^{-1} \qquad (2.74)$$

These simplifications lead to the following algorithm.

Algorithm 2.1 Calculation of Finite Interval Smoother

1. Compute the steady-state innovation covariance matrix $R_\varepsilon = R + CPC^T$. Hence find the spectral factor $\Delta(z)$ from:

$$\Delta(z) = (I + C\Phi(z)AK)R_\varepsilon^{\frac{1}{2}}$$

where P and K are found from the steady-state Kalman filter Riccati equation.

2. Compute $M(z) \triangleq \Delta^{-1}(z)C\Phi(z)A$

and hence find the time function M(N-k) by inverse z transforming.

3. Compute $T(z^{-1})$ from $P[\underline{M}^T(z^{-1}) + C^T\Delta^{-T}(z^{-1})]$ or

$P[\underline{A}^T\Phi^T(z^{-1}) + I]C^T\Delta^{-T}(z^{-1})$

and hence find T(z) and the time function T(k).

4. If $\ell = 0$, then $n_{11}(z,0) = T_0 + \Phi(z)AKR_\varepsilon^{\frac{1}{2}}$ else find

$n_{11}(z,\ell) \triangleq \mathcal{Z}_1\{T(\ell-k)(U(k) - U(k-\ell)\} + \Phi(z)AKR_\varepsilon^{\frac{1}{2}}z^{-\ell}$

5. Find $L(N-k) \triangleq P_0M^T(N-k) - T(N-k)$

6. Compute $n_{22}(z,N) = \mathcal{Z}_1(L(N-k))$

and $I_{22}(N) \triangleq \sum_{k=0}^{N-1} L(N-k)M(N-k)$

7. Compute $I_{11}(N,\ell) \triangleq \sum_{k=0}^{N-1} T(\ell-k)M(N-k)$

$$+ \sum_{k=\ell}^{N-1} [\Phi(k-\ell)AKR_\varepsilon^{\frac{1}{2}} - T(\ell-k)]M(N-k)$$

8. Calculate $\psi(N;\ell) = (\phi(N-\ell) - I_{11}(N,\ell))(I_n + I_{22}(N,k))^{-1}$

9. Find $\hat{H}(z;N,\ell) = [n_{11}(z,\ell) + \psi(N;\ell)n_{22}(z,N)]\Delta^{-1}(z)$

EXAMPLE 2.2

Consider the linear discrete-time scalar system given by

$$x(k+1) = x(k) + w(k)$$

$$y(k) = x(k) + v(k)$$

with process and measurement noise covariances q and r respectively. It is required to find the transfer function of the optimal finite memory smoother with memory length N and smoothing time lag $\ell > 0$.

It is known that

$$\Phi(z) = 1/(z-1) \text{ and } W(z) = 1/(z-1)$$

The solution is found by following the Riccati based algorithm 2.1.

Step 1

$$R_\varepsilon = R + CPC^T \quad \text{hence} \quad R_\varepsilon = r + P$$

$$\Delta(z) = (I + C\Phi(z)AK)R_\varepsilon^{\frac{1}{2}}$$

$$= R_\varepsilon^{\frac{1}{2}} \left(\frac{z - \alpha}{z - 1}\right) \qquad \alpha = r/(P + r)$$

Where the steady-state Kalman filter gain matrix is found to be:

$$K = P/(r+P)$$ and P is found from the steady-state Riccati equation

$$P = P - P^2/(r+P) + q$$

to be

$$P = q/2 + \{q^2 + 4q\ r\}^{\frac{1}{2}}/2$$

Step 2

$$M(z) \triangleq \Delta(z)^{-1}C\Phi(z)A$$

$$= R_\varepsilon^{-\frac{1}{2}}/(z-\alpha)$$

$$M(N-k) = R_\varepsilon^{-\frac{1}{2}}\ \alpha^{N-k-1}$$

Step 3

$$T(z^{-1}) = P\left[M^T(z^{-1}) + C^T\Delta^{-T}(z^{-1})\right]$$

$$= P\, R_\varepsilon^{-\frac{1}{2}}\, z^{-1}/(z^{-1}-\alpha)$$

and $T(z) = P\, R_\varepsilon^{-\frac{1}{2}} z/(z-\alpha)$

giving $T(\ell-k) = P\, R_\varepsilon^{-\frac{1}{2}}\, \alpha^{\ell-k}$

Step 4

$$n_{11}(z,\ell) = \mathcal{Z}_1\{T(\ell-k)(U(k) - U(k-\ell))\} + \Phi(z)AKR_\varepsilon^{\frac{1}{2}} z^{-\ell}$$

$$= PR_\varepsilon^{-\frac{1}{2}}\left[\frac{z^{-\ell}}{z-1} - \frac{z^{-\ell}}{\alpha z-1} + \alpha^{1+\ell}\,\frac{z}{\alpha z-1}\right]$$

Step 5

$$L(N-k) = P_0 M^T(N-k) - T(N-k)$$

$$= R_\varepsilon^{-\frac{1}{2}}(P_0\alpha^{N-k-1} - P\alpha^{N-k})$$

Step 6

$$n_{22}(z,N) = \mathcal{Z}_1(L(N-k))$$

$$= R_\varepsilon^{-\frac{1}{2}}\left(\frac{P_0\alpha^N z}{\alpha z-1} - \frac{P\alpha^{N+1} z}{\alpha z-1}\right)$$

$$I_{22}(N) = \sum_{k=0}^{N-1} L(N-k)M(N-k)$$

$$= R_\varepsilon^{-1}(P_0 - \alpha P)\left(\frac{1-\alpha^{2N}}{1-\alpha^2}\right)$$

Step 7

$$I_{11}(N,\ell) = \sum_{k=0}^{N-1} T(\ell-k)M(N-k)$$

$$+ \sum_{k=\ell}^{N-1}\left[\Phi(k-\ell)AKR_\varepsilon^{\frac{1}{2}} - T(\ell-k)\right]M(N-k)$$

$$= \left(\frac{P}{P+r}\right)\left[\frac{1-\alpha^{N-\ell}}{1-\alpha} + \alpha^{N+1}\left(\frac{\alpha^{-\ell}-\alpha^{\ell}}{1-\alpha^2}\right)\right]$$

$$= 1 - \alpha^{N-\ell} + \alpha^{N+1}(\alpha^{-\ell} - \alpha^{\ell})/(1+\alpha)$$

Step 8

Hence $\psi(N:\ell)$ is found by substitution into

$$\psi(N;\ell) = (\phi(N-\ell) - I_{11}(N,\ell))(I + I_{22}(N))^{-1}$$

Step 9

$$H(z;N,\ell) = \left[n_{11}(z,\ell) + \psi(N;\ell)n_{22}(z,N)\right]\Delta^{-1}(z)$$

$$= \left(\frac{P}{P+r}\right)\left\{\frac{z(z-1)}{(\alpha z-1)(z-\alpha)}\left(\frac{r}{P+r}\right)^{1+\ell}\right.$$

$$\left. - \left(\frac{P}{P+r}\right)\frac{z^{1-\ell}}{(\alpha z-1)(z-\alpha)}\right\}$$

$$+ \psi(N;\ell)\left(\frac{P_0}{r} - \frac{P}{P+r}\right)\left(\frac{r}{P+r}\right)^{N+1}\frac{z(z-1)}{(\alpha z-1)(z-\alpha)}$$

Note the exclusion of any terms of z^{-N}.

These terms occur outside the interval length of N, and thus are excluded in this algorithm.

2.6 Summary

A finite interval, time invariant smoothing filter has been derived. The smoother can either be implemented in Z transfer function or weighting sequence form. For implementation of the smoothing filter beyond the finite interval N, the smoother has a fixed memory length. Implementation is achieved for these times by sliding a 'window' of the N most recent observations and weighting these observations with the optimal smoother weighting sequence. In this way, divergence problems associated with erroneous data can be avoided. Four special cases emerge from the solution to the finite-interval smoother. They have been shown to be equivalent to related work in this area.

By manipulation of the steady-state matrix Riccati equation, a simplified algorithm has been found for computing the optimal smoother. Some of the steps of the algorithm are easily obtainable by using standard Riccati equation subroutine to avoid spectral factorisation. The remaining terms may be more difficult to compute since they require z-transforming and inverse z-transforming. However, by using this kind of algorithm the problem of removing causal from anti-causal terms, which is manifest in the original solution, is overcome and further insight into the link between input-output and state-space Riccati equations is seen.

CHAPTER 3

Multivariable Self-Tuning Filtering, Smoothing and Prediction

3.1 Introduction

The optimal filtering, smoothing and prediction of multivariable processes traditionally employs a knowledge of signal statistics or signal model dynamics. This approach is taken by Wiener [1] and Kalman [3,4] in their respective frequency and time domain approaches (op cit). It is assumed using these approaches that signal statistics or signal model dynamics are found off-line from calculations using experimental data. This chapter considers the more difficult case for many real applications where these quantities are unknown and leads to a solution in which they need not be identified off-line. In addition, the parameters of the signal model may be slowly time varying.

A z transfer function solution to a finite interval smoothing problem has been derived in Chapter 2 of this thesis. It is shown how the first terms of the smoother solution relates to the infinite interval or Wiener type of smoother or filter.

Hagander and Wittenmark [33] derived a computationally attractive implicit (without explicit spectral factorisation calculations), fixed-lag, self-tuning solution to the smoothing problem for scalar ARMA processes. The explicit form of their result is equivalent to the aforementioned infinite interval frequency domain solution. Hagander and Wittenmark [33] employ a Bayesian approach

and innovations signal model in the solution of the fixed lag smoothing problem. The innovation model parameters are estimated on line using extended recursive least squares, or the real-time maximum likelihood method. The assumption is made that the order of the ARMA process is known or can be identified beforehand.

In this chapter a multivariable version of the problem in [33] is discussed using a z-domain approach and polynomial matrix algebra [45]. An explicit solution for the optimal smoothing, filtering and prediction problems is presented in Section 3.3. This solution assumes the process and noise statistics are given. No such a priori information is assumed in the self-tuning Section 3.4. The parameters of the innovations signal model are estimated using a technique due to Panuska [77] or extended recursive least squares. The optimal estimators are then implemented directly from these results. Self-tuning filters prove valuable in many industrial filtering applications [78].

3.2 Discrete-Time Process Description

The discrete-time process will be represented in both state space and vector difference equation forms. This will enable the relationship between the Kalman and the z-domain results to be discussed.

(A) State Space Model

Consider a signal $\underline{y}(k)$ corrupted by a white noise sequence $\underline{v}(k)$ and observed as:

$$\underline{z}(k) = \underline{y}(k) + \underline{v}(k) \tag{3.1}$$

The state $\underline{x}(k)$ is generated from the following stable, discrete-time Markov process:

$$\underline{x}(k+1) = A\underline{x}(k) + D\underline{\omega}(k) \tag{3.2}$$

and

$$\underline{y}(k) = C\underline{x}(k),\ \underline{x}(k)\varepsilon R^n,\ \underline{y}(k)\varepsilon R^r,\ \underline{\omega}(k)\varepsilon R^q \tag{3.3}$$

where by assumption $r = q$. The Gaussian white noise sequences $\underline{\omega}$ and $\underline{v}$ are assumed to be zero mean with covariances

$$\text{cov}\left[\underline{\omega}(k),\ \underline{\omega}(n)\right] = Q\delta(k-n) \tag{3.4}$$

$$\text{cov}\left[\underline{v}(k),\ \underline{v}(n)\right] = R\delta(k-n) \tag{3.5}$$

$$\text{cov}\left[\underline{\omega}(k),\ \underline{v}(k)\right] = 0 \tag{3.6}$$

The matrices Q and R are assumed to be symmetric and positive definite and the system is assumed to be in operation from time $-\infty$.

The input-output transfer-function relating the signal and the driving noise is given as:

$$W(z) = C(zI - A)^{-1}D \tag{3.7}$$

and

$$W(k) \triangleq C\Phi(k)D, \text{ for } k > 0 \tag{3.8}$$

$$\triangleq 0 \qquad \text{for } k \leq 0$$

where $\Phi(z) = (zI - A)^{-1}$ and $\Phi(k) = A^{k-1}U(k-1)$. Here $U(k) \triangleq 1$ for $k \geq 0$ and $U(k) \triangleq 0$ for $k < 0$.

The signal can be calculated from the convolution summation:

$$\underline{y}(k) = \sum_{i=-\infty}^{k-1} W(k-i)\underline{\omega}(i) \tag{3.9}$$

The observations signal can be represented similarly:

$$\underline{z}(k) = \sum_{i=-\infty}^{k} F(k-i)\underline{\varepsilon}(i) \tag{3.10}$$

where $\{\varepsilon(i)\}$ is a white zero mean sequence with covariance:

$$\text{cov}[\underline{\varepsilon}(i)\varepsilon(n)] = R_\varepsilon \delta(i-n) \tag{3.11}$$

and $\delta(i) = 1$ for $i = 0$ and $\delta(i) = 0$ for $i \neq 0$. The signal $\underline{\varepsilon}$ must, through (3.10) generate the same spectrum as the observations signal, so that

$$F(z)R_\varepsilon F^T(z^{-1}) = W(z)QW(z)^T + R \tag{3.12}$$

where

$$F(z) \triangleq \mathcal{Z}_1(F(k)) \tag{3.13}$$

It is well-known (Anderson and Moore [61]) that a Kalman filter can be used to construct the observations signal $\underline{z}(\cdot)$ given the white noise innovations signal. Thus, $\underline{\varepsilon}(\cdot)$ can be identified with the innovations (Gevers and Kailath [79]) and $F(z)$ represents the return difference matrix for the filter (Arcasoy [25]). The covariance matrix for the innovations (Sage and Melsa [60]) is given as $R_\varepsilon = R + CPC^T$. Note that,

$$F(z) = I_r + C\Phi(z)AK \tag{3.14}$$

and the Kalman gain matrix (Sage and Melsa [60]):

$$K = PC^T R_\varepsilon^{-1} \tag{3.15}$$

where P represents the positive definite symmetric matrix which satisfies the steady state Riccati equation:

$$P = APA^T - APC^T R_\varepsilon^{-1} CPA^T + DQD^T \qquad (3.16)$$

The return-difference relationship (3.12) follows by algebraic manipulation of the Riccati equation.

(B) Vector Difference Equation Model

The vector difference equation (VDE) model or multivariable autoregressive moving average (ARMA) model for the system becomes

$$A(z^{-1})\underline{y}(k) = C(z^{-1})\underline{\omega}(k) \qquad (3.17)$$

where the zeros of det(A(x)) and det(C(x)) are assumed to be strictly outside the unit disc in the x-plane. Note that $A(z^{-1})$ and $C(z^{-1})$ should not be confused with the matrices occurring in the state equations since the arguments in z^{-1} will be retained. Unfortunately the notation for both system descriptions is now standard. The equivalent innovations signal representation becomes [80,81] :

$$A(z^{-1})\underline{z}(k) = D(z^{-1})\underline{\varepsilon}(k) \qquad (3.18)$$

The polynomial matrices $A(z^{-1})$, $C(z^{-1})$ and $D(z^{-1})$ can be represented as:

$$A(z^{-1}) = I_r + A_1 z^{-1} + A_2 z^{-2} + \ldots + A_{n_a} z^{-n_a} \qquad (3.19a)$$

$$C(z^{-1}) = C_1 z^{-1} + C_2 z^{-2} + \ldots + C_{n_c} z^{-n_c} \qquad (3.19b)$$

$$D(z^{-1}) = I_r + D_1 z^{-1} + D_2 z^{-2} + \ldots D_{n_d} z^{-n_d} \qquad (3.19c)$$

The matrix $A(z^{-1})$ will be assumed regular (that is A_{n_a} is assumed to be non-singular). The zeros of det(D(x)) and thence of D(x) are required to lie strictly outside the unit disc in the x-plane. That is, $D(z^{-1})^{-1}$ is required to be an asymptotically stable transfer function when performing the spectral factorisation.

The equivalent VDE spectral factorisation result follows immediately as:

$$D(z^{-1})R_\varepsilon D^T(z) = C(z^{-1})QC^T(z) + A(z^{-1})RA^T(z) \qquad (3.20a)$$

and from (3.12) the relationship between $D(z^{-1})$ and F(z)

can be established as:

$$D(z^{-1}) = A(z^{-1})F(z) \qquad (3.20b)$$

To simplify the following discussions, the usual case where $n_c \leq n_a \Rightarrow n_d = n_a$ will be assumed to hold. If (3.20a) is multiplied by z^{n_a} and the limit as $z \to 0$ is taken then

$$D_{n_d} R_\varepsilon = A_{n_a} R$$

and with the assumption that $A(z^{-1})$ is regular then

$$R = A_{n_a}^{-1} D_{n_d} R_\varepsilon$$

Note that the above assumption is only necessary to allow R to be calculated and it follows that the assumption can be relaxed if R is known.

3.3 Frequency Domain Solution to the Linear Estimation Problems

The optimal linear estimator to minimise the performance criterion:

$$J = E\{(\underline{y}(k - \ell) - \hat{\underline{y}}(k - \ell|k))^T(\underline{y}(k - \ell) - \hat{\underline{y}}(k - \ell|k))\} \quad (3.21)$$

for the system described in Section 2 and given the observations $\{\underline{z}(i)\}$, $i = k, k - 1, k - 2, \ldots$ can be derived in the z-domain and can be calculated using the following results.

Theorem 3.1 Optimal Linear Estimator

The optimal steady state filter ($\ell = 0$), smoother ($\ell > 0$) or predictor ($\ell < 0$) can be calculated using:

$$H(z;\ell) = \{W(z)QW^T(z^{-1})F^T(z^{-1})^{-1}z^{-\ell}\}_+ R_\varepsilon^{-1}F(z)^{-1} \quad (3.22)$$

The notation $\{f(z^{-1})\}_+$ is used to denote the transform of the positive time terms in the expansion of $f(z^{-1})$.

Proof The proof follows from the results in Grimble [13] ($\ell \leq 0$) and Moir and Grimble [62], ($\ell > 0$), and is given in Appendix 5.

The time invariant estimator can be implemented in the weighting sequence form:

$$\hat{\underline{y}}(k - \ell|k) = \sum_{i=-\infty}^{k} H(k - i;\ell)\underline{z}(i)$$

where $H(k;\ell)$ is the impulse response sequence of the optimal estimator, and $\hat{\underline{y}}(k - \ell|k)$ denotes the least squares estimate of the signal at time $k - \ell$ given observations up to and including time k. The estimator transfer function matrix:

$$H(z;\ell) \triangleq \mathcal{Z}_1(H(k;\ell))$$

The expression for the transfer function matrix for the estimator can be simplified using (3.12):

$$\{W(z)QW^T(z^{-1})F^T(z^{-1})^{-1}z^{-\ell}\}_+$$
$$= \{(F(z)R_\varepsilon - RF^T(z^{-1})^{-1})z^{-\ell}\}_+ \qquad (3.23)$$

(A) Filtering and Smoothing

The results will now be specialised to the filtering and smoothing problems ($\ell \geq 0$). Recall that $F(z)$ can be expanded as a convergent matrix series in z^{-1}:

$$F(z) = I_r + F_1 z^{-1} + F_2 z^{-2} + \ldots \qquad (3.24)$$

thence from (3.23):

$$\{\cdot\}_+ R_\varepsilon^{-1} F(z)^{-1} = I_r z^{-\ell} - R\{F^T(z^{-1})^{-1} z^{-\ell}\}_+ \times R_\varepsilon^{-1} F(z)^{-1} \qquad (3.25)$$

and the final term in braces can be expanded as a Taylor series up to ℓ terms:

$$\{F^T(z^{-1})^{-1} z^{-\ell}\}_+ = (I_r + P_1 z + P_2 z^2 + \ldots + P_\ell z^\ell) z^{-\ell}$$

Thus, let $P_\ell(z) z^{-\ell} \triangleq \{F^T(z^{-1})^{-1} z^{-\ell}\}_+ = \{A(z)^T D^T(z)^{-1} z^{-\ell}\}_+$ so that

$$P_\ell(z) = I_r + P_1 z + P_2 z^2 + \ldots + P_\ell z^\ell \qquad (3.26)$$

then collecting the above results, for $\ell \geq 0$, the transfer function solution becomes

$$H(z;\ell) = (I_r - R P_\ell(z) R_\varepsilon^{-1} F(z)^{-1}) z^{-\ell} \qquad (3.27)$$

The polynomial matrix $P_\ell(z)$ can be calculated using the algorithm in Appendix 6.

The optimal smoothed estimate can be calculated using:

$$\underline{\hat{y}}(k - \ell | k) = H(z;\ell)\underline{z}(k) = \underline{z}(k - \ell) - R P_\ell(z) R_\varepsilon^{-1} \underline{\varepsilon}(k - \ell) \qquad (3.28)$$

where in this last equation $\underline{z}(k) = F(z)\underline{\varepsilon}(k)$ and

$$\underline{\varepsilon}(k - \ell) = \underline{z}(k - \ell) - \underline{\hat{y}}(k - \ell | k - \ell - 1) \qquad (3.29)$$

Notice that in the filtering case $\ell = 0$ and $P_\ell(z) = I_r$ which leads to some simplifications in (3.28) and (3.27). Also note that the smoothed estimate can be expressed in terms of a filtered estimate and a correction term (Appendix 7).

(B) Prediction

The results for the prediction case are somewhat different from the foregoing. Let $\ell_1 = -\ell > 0$ then from (3.23)

$$\{\cdot\}_+ R_\varepsilon^{-1} F(z)^{-1} = \{F(z)z^{\ell_1}\}_+ F(z)^{-1} \qquad (3.30)$$

$$= \{A(z^{-1})^{-1} D(z^{-1}) z^{\ell_1}\}_+ F(z)^{-1} \qquad (3.31)$$

Introduce $\tilde{D}(z^{-1})$ and $\tilde{A}(z^{-1})$, where $\det(\tilde{A}(z^{-1})) = \det(A(z^{-1}))$ and $\tilde{A}(0) = I_r$ via the relationship

$$D(z^{-1})\tilde{A}(z^{-1}) = A(z^{-1})\tilde{D}(z^{-1}) \qquad (3.32)$$

and let $H(z^{-1})$ and $G(z^{-1})$ satisfy the diophantine equation:

$$H(z^{-1})\tilde{A}(z^{-1}) + z^{-\ell_1} G(z^{-1}) = \tilde{D}(z^{-1}) \qquad (3.33)$$

where the order of $H(z^{-1})$ is $n_h = \ell_1 - 1$. Thence

$$z^{\ell_1} H(z^{-1}) + G(z^{-1})\tilde{A}(z^{-1})^{-1} = A(z^{-1})^{-1} D(z^{-1}) z^{\ell_1}$$

and equation (3.31) may be simplified as:

$$\hat{H}(z;\ell_1) = G(z^{-1})\tilde{A}(z^{-1})^{-1} D(z)^{-1} A(z^{-1})$$

$$= G(z^{-1})\tilde{D}(z^{-1})^{-1}$$

The optimal predicted estimate can be calculated using:

$$\underline{\hat{y}}(k + \ell_1 | k) = G(z^{-1})\tilde{D}(z^{-1})^{-1}\underline{z}(k) \qquad (3.34)$$

This is a multivariable version of the predictor derived by Wittenmark (1974) [30] and presented in a multivariable self-tuning form by Tanttu (1980) [31]. The work has been applied by Sym and Wellstead [57] and by Holst [82]. Thus, the multivariable self-tuning predictor is already well-established. The results here, however, do lead to some unification in the area.

3.4 Self-Tuning Smoothing and Filtering

3.4.1 The Scalar Case

Hagander and Wittenmark [33] considered the difference equation representation of the process of equations (3.18) and (3.19). They assumed that the noise covariances and process parameters were unknown, but that the order of the polynomials were given. The following algorithm was proposed in [33].

Algorithm 3.1 Scalar Case

Step 1: Estimate the parameters A_i, D_i, $i = 1, \ldots, n_a$ in the polynomials A and D using ELS or the RML identification method [56]. The estimated polynomials are denoted $\hat{A}$ and $\hat{D}$, (estimate A and D from the innovations model of equation (3.18).

Step 2: Compute the polynomial $\hat{P}_\ell(z)$ from $\hat{A}(z) = \hat{P}_\ell(z)\hat{D}(z)$ and the smoothing estimate from

$$\underline{\hat{y}}(k - \ell|k) = \underline{z}(k - \ell) - (\hat{D}_{n_d}/\hat{A}_{n_a})\ \hat{P}_\ell(z)\underline{\hat{\varepsilon}}(k - \ell) \quad (3.35)$$

where

$$\underline{\hat{\varepsilon}}(k - \ell) = \underline{z}(k - \ell) - \underline{\hat{y}}(k - \ell|k - \ell - 1) \quad (3.36)$$

The two steps of the algorithm are repeated at each step of time.

3.4.2 The Multivariable Case

This section serves as a natural extension to the work of Hagander and Wittenmark. An explicit polynomial matrix solution to the multivariable fixed-lag smoothing problem was presented in Section 3.3. To determine the expression for the z-domain smoother, prior knowledge of the noise statistics and the plant matrix parameters was assumed. Further, a matrix spectral factorisation (equation 3.12) has to be performed in order to determine the polynomial

matrix $F(z^{-1})$ and the innovations covariance matrix R_ε.

The self-tuning multivariable smoother proposed in the following section alleviates these difficulties in the same way as that of the scalar smoother of Hagander and Wittenmark. The following assumptions are made:

1. The signal generating process is linear, stable and can be expressed in VDE form both from an innovations model (equation 3.18) and a process model (equation 3.17).
2. The degree of the $A(z^{-1})$ polynomial matrix is assumed known and $n_a = n_d$.
3. The dimensions of the observation vector and the process noise vector are equal ($q = r$) that is, the process is assumed to be square.
4. The parameters in the process are assumed to be constant or slowly time varying and the noise sources are assumed to be stationary.

The self-tuning smoother may be constructed by reference to equation (3.28). Clearly, the innovations sequence, $P_\ell(z)$, R and R_ε must be calculated before the estimate may be found. The matrix $P_\ell(z)$ can be obtained using the algorithm in Appendix 6 once the innovations model polynomial matrices $A(z^{-1})$ and $D(z^{-1})$ have been obtained. These matrices and the noise covariance matrices can be estimated using the algorithm due to Panuska [77].

Parameter Estimation Procedure

First write the innovations model of (3.10) in the form

$$\underline{z}(k) = \psi(k)\underline{\Theta} + \underline{\varepsilon}(k) \qquad (3.37)$$

with $\psi(k)$ = block diag $\{\eta^T(k), \eta^T(k), \ldots, \eta^T(k)\}$ (3.38)

and

$$\eta^T(k) = [-\underline{z}^T(k-1), \ldots, -\underline{z}^T(k-n_a), \underline{\varepsilon}^T(k-1), \ldots, \underline{\varepsilon}^T(k-n_d)]^T$$

$$\underline{\theta} = [\underline{\xi}_1^T, \underline{\xi}_2^T, \ldots, \underline{\xi}_r^T]^T \tag{3.39}$$

$$\underline{\xi}_1^T \triangleq \text{ith row of } [A_1, \ldots, A_{n_a}; D_1, \ldots, D_{n_d}] \tag{3.40}$$

Equation (3.37) can be interpreted as the observation equation for a state space model of (3.18) with the state vector representing the (unknown) constant parameter vector $\underline{\theta}(k) = \underline{\theta}$. The state equations for the parameter estimation model become:

$$\underline{\theta}(k+1) = \underline{\theta}(k) \tag{3.41}$$

$$\underline{z}(k) = \psi(k)\underline{\theta}(k) + \underline{\varepsilon}(k) \tag{3.42}$$

The extended Kalman filter algorithm proposed by Panuska [77] may be employed to estimate $\underline{\theta}$ and the covariance matrices R and R_ε.

Algorithm 3.2 Multivariable Parameter Estimation

Initialise $\hat{\underline{\theta}}(0) = \underline{\theta}^o$, $\Lambda(0) = \Lambda^o$, $\hat{R}_\varepsilon(0) = R_\varepsilon^o$

$$\hat{\underline{\theta}}(k+1|k) = \hat{\underline{\theta}}(k|k-1) + \Gamma(k)(\underline{z}(k) - \hat{\psi}(k)\hat{\underline{\theta}}(k|k-1)) \tag{3.43}$$

$$\Gamma(k) = \Lambda(k)\hat{G}^T(k)(\hat{G}(k)\Lambda(k)\hat{G}^T(k) + \hat{R}_\varepsilon(k+1))^{-1} \tag{3.44}$$

$$\Lambda(k+1) = \Lambda(k) - \Gamma(k)(\hat{G}(k)\Lambda(k)\hat{G}^T(k) + \hat{R}_\varepsilon(k+1))\Gamma^T(k) \tag{3.45}$$

$$\hat{R}_\varepsilon(k+1) = \hat{R}_\varepsilon(k) + \frac{1}{k+1}(\hat{\underline{\varepsilon}}(k)\hat{\varepsilon}^T(k) - \hat{R}_\varepsilon(k)) \tag{3.46}$$

where

$$\hat{\underline{\varepsilon}}(k) = \underline{z}(k) - \hat{\psi}(k)\hat{\underline{\theta}}(k|k-1) \tag{3.47}$$

and

$$\hat{G}(k) = \frac{\partial(\psi(k)\underline{\hat{\theta}}(k|k-1))}{\partial\underline{\hat{\theta}}(k|k-1)} \tag{3.48}$$

The measurement noise variance may be estimated from

$$\hat{R}(k) = \hat{A}_{n_a}^{-1}\hat{D}_{n_d}\hat{R}_{\varepsilon}(k) \tag{3.49}$$

Now denote $S = R_{\varepsilon} R^{-1}$ and let

$$\hat{S}(k) = \hat{A}_{n_a}^{-1}\hat{D}_{n_d} \tag{3.50}$$

The self-tuning filtering or smoothing algorithm now follows:

Algorithm 3.3 Multivariable Self-Tuning Filter or Smoother

Step 1 : Using the Kalman filter algorithm of Panuska, estimate the parameter vector $\hat{\underline{\theta}}$, and hence construct the matrix coefficients A_i, D_i, $i = 1,2,\ldots, n_a$, $(n_a = n_d)$.

Step 2 : Estimate the innovations vector

$$\hat{\underline{\varepsilon}}(k - \ell) = \underline{z}(k - \ell) - \hat{\psi}(k - \ell)\hat{\underline{\theta}}(k - \ell | k - \ell - 1) \quad (3.51)$$

For $\ell \neq 0$ compute the variance $\hat{R}_\varepsilon(k - \ell)$.

Step 3 : If $\ell = 0$, compute $\hat{S}(k) = \hat{A}_{n_a}^{-1}\hat{D}_{n_d}$ else compute

$$\hat{R}(k - \ell) = \hat{A}_{n_a}^{-1}\hat{D}_{n_d}\hat{R}_\varepsilon(k - \ell)$$

Step 4 : If $\ell = 0$ go to Step 5 else compute $P_\ell(z)$ from $A^T(z)\ D^T(z)^{-1}$ using the algorithm in Appendix 6.

Step 5 : For $\ell = 0$ compute the filtered estimate using

$$\hat{\underline{y}}(k|k) = \underline{z}(k) - \hat{S}(k)\hat{\underline{\varepsilon}}(k) \quad (3.52)$$

else compute the ℓ step smoothed estimate from

$$\hat{\underline{y}}(k - \ell | k) = \underline{z}(k - \ell) - \hat{R}(k - \ell)\hat{P}_\ell(z) \times \hat{R}_\varepsilon(k - \ell)^{-1}\hat{\underline{\varepsilon}}(k - \ell) \quad (3.53)$$

The above steps are repeated at each sampling instant.

Remarks

1. The system matrices are identified directly from the innovations model for the process and this avoids any spectral factorisation stage. The algorithm can therefore be classed as an implicit rather than an explicit self-tuning estimator.

2. The implicit smoothing filter may be compared with the work of Hagander and Wittenmark since for the scalar case:

$$D_{n_d}/A_{n_a} = R/R_\varepsilon$$

and hence from (3.48)

$$\hat{\underline{y}}(k - \ell|k) = \underline{z}(k - \ell) - (\hat{D}_{n_d}/\hat{A}_{n_a})\hat{P}_\ell(z)\hat{\underline{\varepsilon}}(k - \ell)$$

with $\hat{P}_\ell(z)$ calculated using $\hat{A}(z)/\hat{D}(z)$.

3. For the scalar case above, the noise variances R and R_ε may be found together as a ratio and can be employed directly in the smoother/filter equation. Under multivariable conditions, only in the filtering case can RR_ε^{-1} be used in the solution directly ($S = RR_\varepsilon^{-1}$).

4. The observation matrix $\hat{\psi}(k)$ is dependent on the previously estimated innovations. The most recent estimate of the plant parameters is used to calculate this approximation to the innovations:

$$\hat{\underline{\varepsilon}}(k - i) = \underline{z}(k - i) - \hat{\psi}(k - i)\hat{\underline{\theta}}(k|k - 1)$$

for $i = 1,2,\ldots, n_d$. The dependence of the observation matrix on the parameter $\hat{\underline{\theta}}$ gives a non-linear estimation problem which is solveable using the extended Kalman filter.

5. If the computed innovations (residuals) in the information matrix $\psi(k)$ are assumed to be independent of the estimated parameter vector $\theta(k)$ then it is clear that:

$$\hat{G}(k) = \frac{\partial(\hat{\psi}(k)\hat{\underline{\theta}}(k|k - 1))}{\partial\hat{\underline{\theta}}(k|k - 1)} = \hat{\psi}(k)$$

6. The parameter estimation algorithm need not be multivariable. For filtering problems the matrices A_{n_a} and D_{n_d} can be estimated by calling a standard SISO extended recursive least squares algorithm r times. For smoothing problems, the innovations variance matrix R_ε may be estimated via (3.46) and an extended least squares

algorithm can be called r times to estimate $A(z^{-1})$ and $D(z^{-1})$. There are clearly similarities with the work of Mehra [16-18] on the identification of dynamic systems.

7. Equation (3.46) is merely a recursive form of the more familiar equation :

$$\hat{R}_\varepsilon(k) = \frac{1}{k} \sum_{i=0}^{k-1} \hat{\underline{\varepsilon}}(i)\, \hat{\underline{\varepsilon}}^T(i)$$

which is asymptotically unbiased.

3.4.3 Numerical Considerations

To avoid the matrix $\Lambda(k+1)$ becoming singular (3.45) may be replaced by

$$\Lambda(k+1)^{-1} = \Lambda(k)^{-1} + \hat{G}^T(k)\hat{R}_\varepsilon^{-1}(k+1)\hat{G}(k) + \delta I \quad (3.54)$$

for some small real scalar $\delta > 0$. In the same spirit, to avoid A_{n_a} becoming singular, (3.49) may be replaced by:

$$\hat{R}(k) = (\hat{A}_{n_a} + \delta_1 I)^{-1}\hat{D}_{n_d}\hat{R}_\varepsilon(k) \quad (3.55)$$

for some small real scalar $\delta_1 > 0$. Under the given assumptions the problem with singularity should not arise after the first few iterations of the parameter estimation algorithm.

3.4.4 Convergence Analysis

The following discussion of the convergence of the self-tuning smoother or filter is based upon the work of Panuska [77] who used the asymptotic convergence analysis of Ljung [35-37] . It will be shown that the parameter estimates generated by the algorithm converge with probability one (w.p.1.).

Assume that all absolute moments of the sequence $\{\underline{\omega}(k)\}$ exist and are bounded. Recall that the system (3.1), (3.2) was assumed to be stable and assume that the equation generating $\hat{\underline{\varepsilon}}(t)$ in (3.47) be stable. Note that the regularity conditions described by Ljung [36] can be satisfied. Thus, to apply the theory of [36] first obtain an alternative expression for the Kalman gain $\Gamma(k)$ in (3.43). Write (3.43) and (3.54) in terms of

$$P(k) = (k\Lambda(k))^{-1} \tag{3.56}$$

Noting from (3.44) and (3.54):

$$\Gamma(k) = (\Lambda(k)^{-1} + \hat{G}^T(k)\hat{R}_\varepsilon(k+1)^{-1}\hat{G}(k))^{-1}\hat{G}^T(k)\hat{R}_\varepsilon(k+1)^{-1}$$

$$\Lambda(k+1)^{-1} = \Lambda(k)^{-1} + \hat{G}^T(k)\hat{R}_\varepsilon(k+1)^{-1}\hat{G}(k)$$

thus

$$\hat{\underline{\theta}}(k+1|k) = \hat{\underline{\theta}}(k|k-1) + \frac{P(k+1)^{-1}}{k+1}\hat{G}^T(k)\hat{R}_\varepsilon(k+1)^{-1} \times (\underline{z}(k) - \hat{\psi}(k)\hat{\underline{\theta}}(k|k-1)) \tag{3.57}$$

and

$$P(k+1) = P(k) + \frac{1}{k+1}(\hat{G}^T(k)\hat{R}_\varepsilon(k+1)^{-1}G(k) + \delta I - P(k)) \tag{3.58}$$

The convergence analysis depends upon associating (3.57) and (3.58) with an ordinary differential equation that includes the necessary information about the asymptotic behaviour [36]. This equation will now be defined in terms of stationary processes resulting from the algorithm when the parameter estimates are held at constant values $\hat{\underline{\theta}}(k) = \underline{\theta}$ and $R_\varepsilon(k) = R_\varepsilon$. Then $\hat{G}(k)$, $\hat{\psi}(k)$ and $\hat{\underline{\varepsilon}}(k)$ would give rise to some $\overline{G}(k;\underline{\theta})$, $\overline{\psi}(k;\underline{\theta})$ and $\overline{\underline{\varepsilon}}(k;\underline{\theta})$ for large k. Taking now the expectation with respect to $\{\underline{\omega}(k)\}$ define:

$$f(\underline{\theta},R_\varepsilon) = E\{\overline{G}^T(k;\underline{\theta})R_\varepsilon^{-1}\ \overline{\underline{\varepsilon}}(k;\underline{\theta})\} \tag{3.59a}$$

$$g(\underline{\theta},R_\varepsilon) = E\{\overline{G}^T(k;\underline{\theta})R_\varepsilon^{-1}\ \overline{G}(k;\underline{\theta})\} \tag{3.59b}$$

$$h(\underline{\theta}) = E\{\overline{\underline{\varepsilon}}(k;\underline{\theta})\overline{\underline{\varepsilon}}(k;\underline{\theta})^T\} \tag{3.59c}$$

It now follows from [36] and by reference to equations (3.57), (3.58) and (3.46) that the behaviour of the estimates can be described by the following coupled ordinary differential equations, as $k \to \infty$:

$$\frac{d\underline{\theta}}{d\tau}(\tau) = P(\tau)^{-1}\ f(\underline{\theta}(\tau),\ R_\varepsilon(\tau)) \tag{3.60a}$$

$$\frac{dP}{d\tau}(\tau) = g(\underline{\theta}(\tau),\ R_\varepsilon(\tau)) + \delta I - P(\tau) \tag{3.60b}$$

$$\frac{d}{d\tau}R_\varepsilon(\tau) = h(\underline{\theta}(\tau)) - R_\varepsilon(\tau) \tag{3.60c}$$

Global stability of (3.60) implies convergence w.p.1. of (3.57), (3.58) and furthermore the possible convergence points of the algorithm are the stable stationary points of the corresponding differential equation (3.60).

The convergence properties of the algorithm may now be summarised.

Theorem 3.2 Parameter Convergence

Consider the stable system described in Section 3.2 and assume the algorithm assures stable generation of $\underline{\varepsilon}(k)$ in (3.47). This can be achieved by allowing only the parameter vectors $\hat{\underline{\theta}}(t)$ for which $\hat{D}(z^{-1})^{-1}$ is stable. Then the estimate $\hat{\underline{\theta}}(k)$ converges w.p.1. to a stationary point of the function:

$$V(\underline{\theta},R_\varepsilon) = E\{\bar{\underline{\varepsilon}}^T(k;\underline{\theta})R_\varepsilon^{-1}\bar{\underline{\varepsilon}}(k;\underline{\theta})\} + \ell n(\det(R_\varepsilon)) \quad (3.61)$$

where

$$\bar{\underline{\varepsilon}}(k;\underline{\theta}) = D_\theta^{-1}(z^{-1})A_\theta(z^{-1})\ \underline{z}(k) \quad (3.62)$$

where A_θ and D_θ are polynomial matrices corresponding to $\underline{\theta}$.

The asymptotic properties of the algorithm 3.2 are described by the differential equation (3.60). From (3.59a), (3.47) and (3.61) obtain:

$$\frac{d}{d\underline{\theta}}\ V(\underline{\theta},R_\varepsilon) = -2f(\underline{\theta},R_\varepsilon)$$

and choose (3.61) as a Lyapunov function so that:

$$\begin{aligned}
\frac{d}{d\tau} &= \frac{dV}{d\tau}(\underline{\theta}(\tau),R_\varepsilon(\tau)) \\
&= \frac{d}{d\tau}\ \big[\bar{E}\{\bar{\underline{\varepsilon}}^T(k;\underline{\theta}(\tau))R_\varepsilon(\tau)^{-1}\bar{\underline{\varepsilon}}(k;\underline{\theta}(\tau))\} \\
&\qquad + \ell n(\det(R_\varepsilon(\tau)))\big] \\
&= E\Big(\frac{d}{d\underline{\theta}}\ \bar{\underline{\varepsilon}}^T(k;\underline{\theta})R_\varepsilon^{-1}\bar{\underline{\varepsilon}}(k;\underline{\theta})\Big)^T P(\tau)^{-1}f(\underline{\theta},R_\varepsilon) \\
&\quad - E\ \bar{\underline{\varepsilon}}^T(k;\underline{\theta})R_\varepsilon^{-1}(h(\underline{\theta}) - R_\varepsilon)R_\varepsilon^{-1}\bar{\underline{\varepsilon}}(k;\underline{\theta}) \\
&\quad + \mathrm{trace}\{R_\varepsilon^{-1}(h(\underline{\theta}) - R_\varepsilon)\} \\
&= -2f^T(\underline{\theta},R_\varepsilon)P(\tau)^{-1}f(\underline{\theta},R_\varepsilon) \\
&\quad - \mathrm{trace}\{R_\varepsilon^{-1}(h(\underline{\theta})-R_\varepsilon)R_\varepsilon^{-1}(h(\underline{\theta})-R_\varepsilon)\} \leq 0 \quad (3.63)
\end{aligned}$$

For the differential equation (3.60a,b,c) to be globally stable and thence for (3.57) to converge w.p.1., the derivative of the Lyapunov function must be negative semi-definite. The convergence points of the algorithm are determined by the stable stationary points of the differential equation [36]. These follow from (3.63) as: $\underline{\theta} = \underline{\theta}^*$ and $R_\varepsilon = R_\varepsilon^*$ where $\underline{\theta}^*$, R_ε^* satisfy $f(\underline{\theta}^*, R_\varepsilon^*) = 0$ and $R_\varepsilon^* = h(\underline{\theta}^*) .= E\{\underline{\bar{\varepsilon}}(k;\underline{\theta}^*)\ \underline{\bar{\varepsilon}}(k;\theta^*)^T\}$. These are the stationary points of (3.61) which concludes the proof.

If R_ε is known apriori the algorithm 3.2 will asymptotically minimise the usual Kalman filter minimum variance criterion:

$$V = E\{\underline{\bar{\varepsilon}}^T R_\varepsilon^{-1} \underline{\bar{\varepsilon}}\},$$

however, this is not the usual situation. For scalar systems for which $D(z^{-1})$ satisfies a positive real property, a stronger result than Theorem 3.2 can be obtained.

Theorem 3.3 Convergence for the Scalar Case

Consider the system (scalar) described in Section 3.2 where $A(z^{-1})$ is stable and $A(z^{-1})$ and $D(z^{-1})$ have no common factors. Assume that the polynomial $D^{-1}(z^{-1})-1/2$ is positive real (i.e. $Re\{D^{-1}(e^{j\omega}) - 1/2\}>0$, $-\pi<\omega\leq\pi$) and $D(z^{-1})$ is maintained stable, then the parameter estimates generated by the algorithm 3.2 converge to their true values with probability one, i.e. $\hat{\underline{\theta}}(k) \to \underline{\theta}$ and $\hat{R}_\varepsilon(k) \to R_\varepsilon$ for $k\to\infty$.

Proof The result follows from Theorem 5.1 of Panuska [77] with minor changes.

Convergence of the Smoother or Filter

The smoother of filter transfer function may be written in parametric form as:

$$H(z;\ \ell;\ \underline{\theta},\ R_\varepsilon)$$
$$= (I_r - R(\underline{\theta},R_\varepsilon)P_\ell(z;\underline{\theta})R^{-1}F^{-1}(z;\underline{\theta}))z^{-\ell} \qquad (3.64)$$

Note also that the feature which ensures $D(z^{-1})$ remains stable in the estimation algorithm also ensures that the smoother or filter is stable (see the return difference relationship (3.20b)). The above results may now be collected in the following theorem.

Theorem 3.4 Smoother or Filter Convergence

The self-tuning smoother or filter of Algorithm 3.3 converges with probability one to the estimator given by $H(z;\ \ell,\ \underline{\theta}^*,\ R_\varepsilon^*)$. Under the conditions of Theorem 3.2 the estimator converges with probability one to the minimum variance estimator (this also applies in the multivariable problem when $D(z^{-1}) = I_r$, [77].)

The representation of the multivariable system is not unique but the optimal estimator depends only upon the input-output transfer function matrix and thus the non-uniqueness of $\hat{A}(z^{-1})$ and $\hat{D}(z^{-1})$ should not cause a difficulty in calculating the estimator. The identification problem can be simplified by representing the system in a canonical form [95-98]. An alternative approach to the analysis of the convergence problem would be via the work of Goodwin, Ramadge and Caines[85] or Solo [87].

3.5 Illustrative Examples

To illustrate the operation of the multivariable self-tuning filter and smoother, three examples have been chosen.

Example 3.1

The process is two-input two-output in the VDE (or ARMA) form:

$$A(z^{-1})\underline{y}(k) = C(z^{-1})\underline{\omega}(k)$$

with

$$A(z^{-1}) = I_2 + A_1 z^{-1}$$

$$C(z^{-1}) = C_1 z^{-1}$$

where

$$A_1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 0.9 & 0.4 \\ 0.4 & -0.9 \end{bmatrix}$$

$$C_1 = -\begin{bmatrix} 0.2 & 0.9 \\ 0.9 & 0.8 \end{bmatrix}$$

The signal vector is corrupted by white noise $\underline{v}$ and the white noise vectors $\underline{\omega}(k)$ and $\underline{v}(k)$ have the variances:

$$Q = \text{diag}\{0.01, 0.01\} \text{ and } R = \text{diag}\{1.0, 1.0\}$$

Some idea of the signal to noise ratio for this example can be gleaned from figures 3.1 and 3.2 which show the signal, noise and signal + noise for outputs (channels) 1 and 2. The order of the above process ($n_a = 1$) was assumed to be known and the signal estimates for each channel were found using either the self-tuning filter ($\ell = 0$) or the self-tuning smoother ($\ell = 10$).

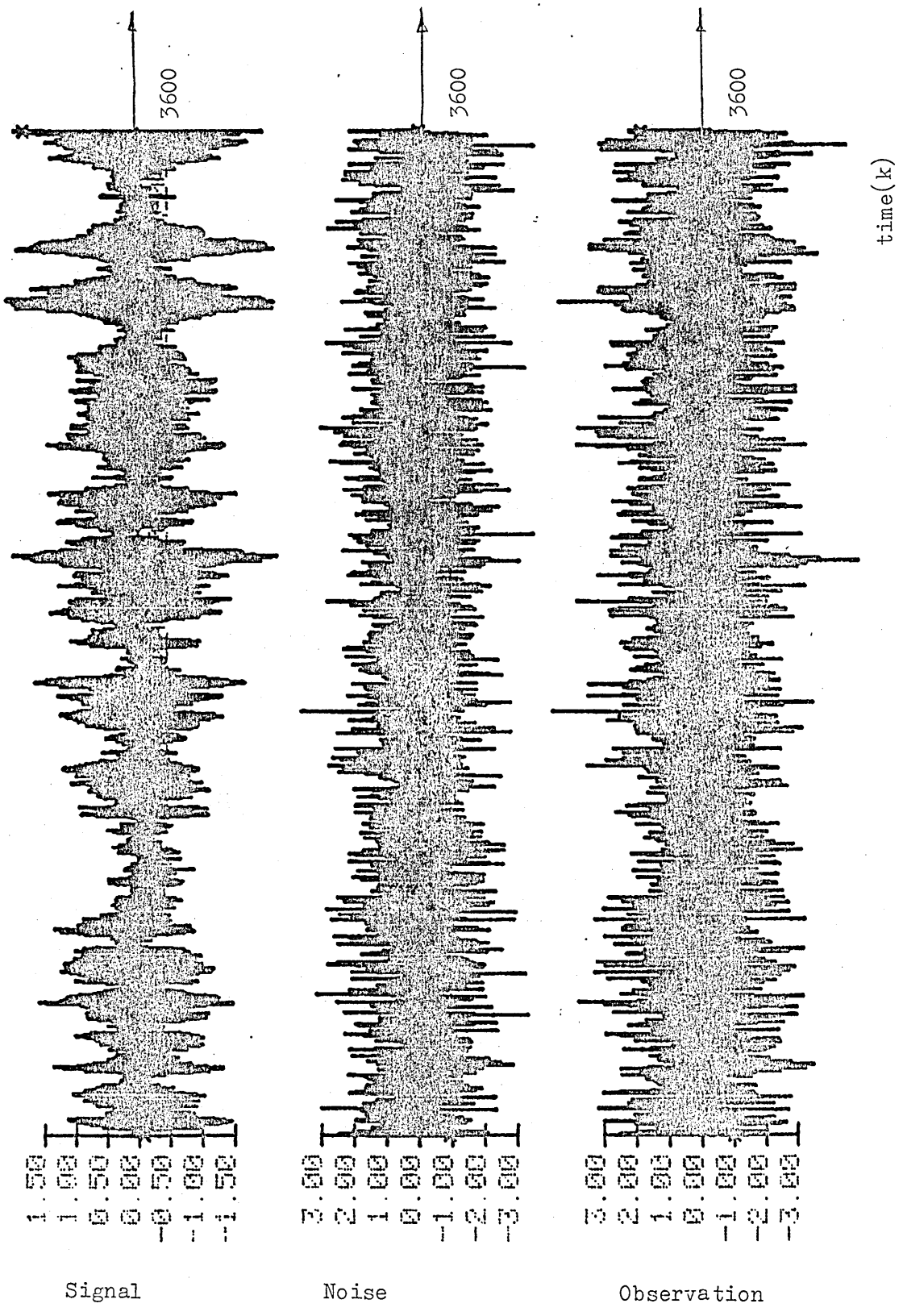


Figure 3.1 Signal, Noise & Observation for Channel 1

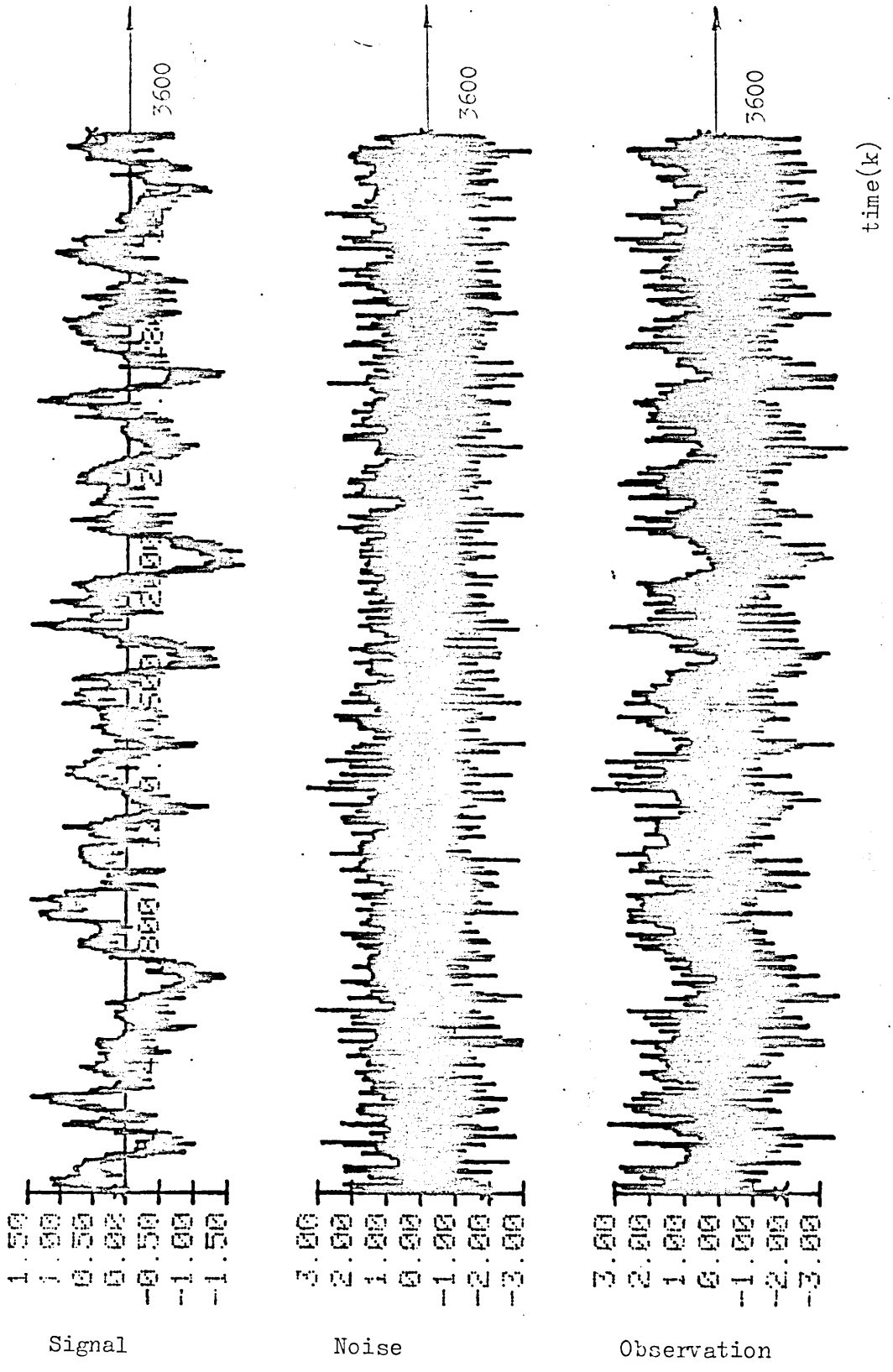


Figure 3.2 Signal, Noise & Observation for Channel 2

For comparison purposes, the optimal Wiener filter or smoothing filter estimates were also calculated based upon the actual plant model. The process was modelled using (3.18) and the spectral factorisation equation (3.20) was evaluated as:

$$D(z^{-1})R_\varepsilon D^T(z)$$

$$= \begin{bmatrix} 1.9785 + 0.9z + 0.9z^{-1} & 0.009 + 0.4z + 0.4z^{-1} \\ 0.009 + 0.4z + 0.4z^{-1} & 1.9845 - 0.9z - 0.9z^{-1} \end{bmatrix}$$

There are many methods available to spectrally factorise the above. These include Shaked [27], Strintzis [86], Kučera [45] and Vostry [84]. For this example, the algorithm of Tuel [83] was used to obtain:

$$D(z^{-1})R_\varepsilon^{\frac{1}{2}} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix}$$

where

$$m_1 \triangleq 1.049 + 0.856z^{-1}$$

$$m_2 \triangleq 0.382z^{-1}$$

$$m_3 \triangleq 0.386z^{-1}$$

$$m_4 \triangleq 1.047 - 0.859z^{-1}$$

Now

$$D(z^{-1})R_\varepsilon^{\frac{1}{2}} = (I_2 + D_1 z^{-1})R_\varepsilon^{\frac{1}{2}}$$

thus

$$\lim_{z\to\infty} D(z^{-1})R_\varepsilon^{\frac{1}{2}} = R_\varepsilon^{\frac{1}{2}}$$

and thus

$$R_\varepsilon^{\frac{1}{2}} = \text{diag}\{1.049,\ 1.047\}$$

and the innovation signal variance becomes:

$$R_\varepsilon = R_\varepsilon^{\frac{1}{2}} R_\varepsilon^{T\frac{1}{2}}$$

$$= \text{diag}\{1.1,\ 1.09\}$$

Given R_ε, the matrix D_1 can be calculated as:

$$D_1 = \begin{bmatrix} 0.816 & 0.365 \\ 0.368 & -0.821 \end{bmatrix}$$

The matrix polynomial $P_\ell(z)$ is found using the algorithm in Appendix 6. For a smoothing lag of $\ell = 10$ steps and $n_a = 1$, this algorithm gives:

$$P_\ell(z) = \sum_{i=0}^{10} P_i z^i$$

where

$$P_0 = I_2$$

$$P_1 = A_1^T - D_1^T$$

and $P_2, P_3, \ldots, P_{10}$ are given as:

$$P_i = -P_{i-1} D_1^T \quad i = 2,3,\ldots, 10$$

The innovations sequence can be generated from (3.18)

$$\underline{\varepsilon}(k) = \underline{z}(k) + A_1 \underline{z}(k-1) - D_1 \underline{\varepsilon}(k-1)$$

and thence the optimal (10 step) fixed-lag smoother may be implemented using:

$$\hat{\underline{y}}_{op}(k - \ell | k) = \underline{z}(k - \ell) - R P_\ell(z) R_\varepsilon^{-1} \underline{\varepsilon}(k - \ell)$$

Simulation Results

The parameter estimates for the A_1 and D_1 matrices are shown in Figures 3.3 and 3.4. The true parameter values are shown in broken lines. Figure 3.5 shows the convergence of the measurement noise variance matrix R and Figure 3.6 shows the convergence of the innovations variance R_ε. Figures 3.7 and 3.8 show the signal, self-

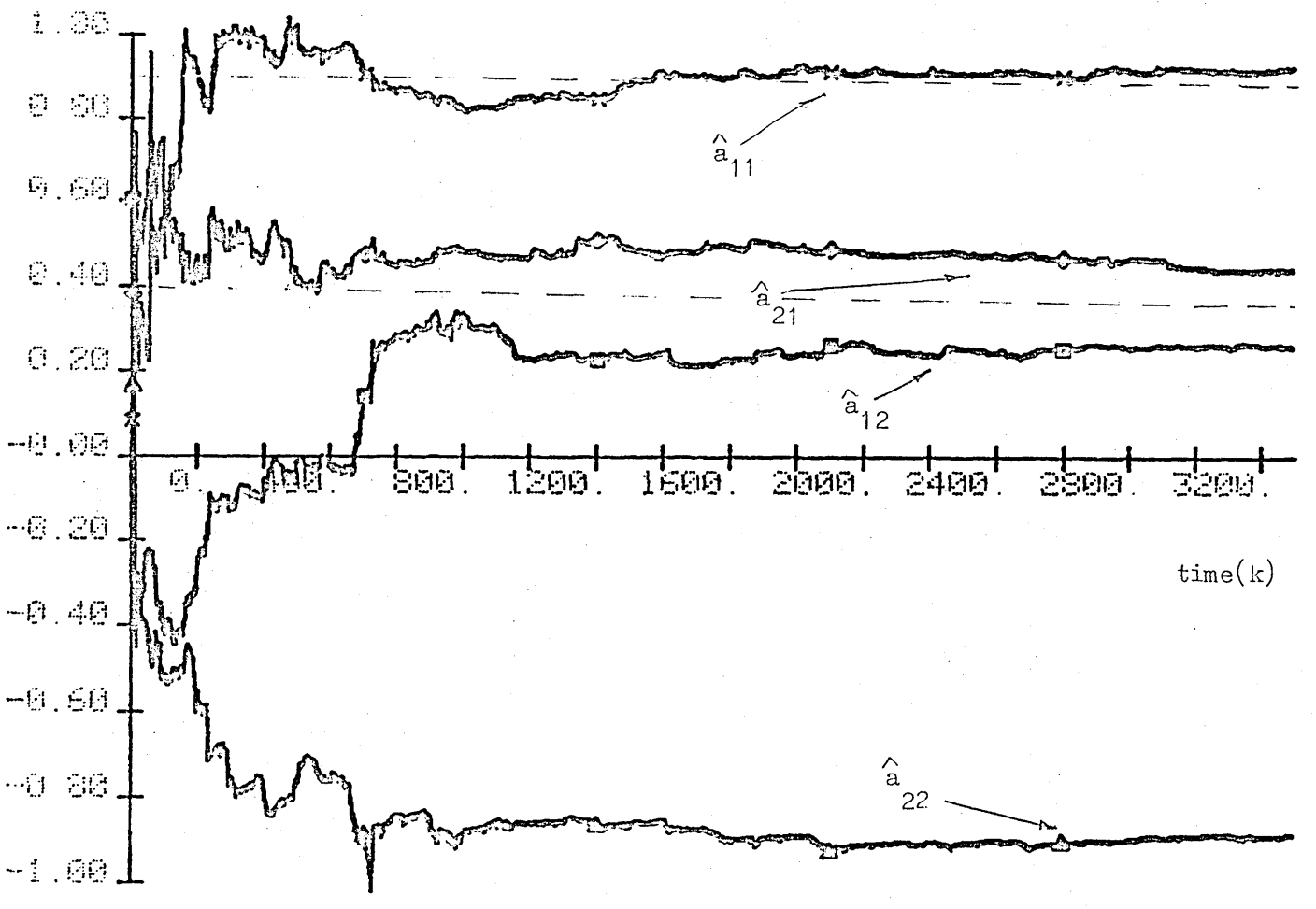


Figure 3.3 Parameter Estimates for A_{na}

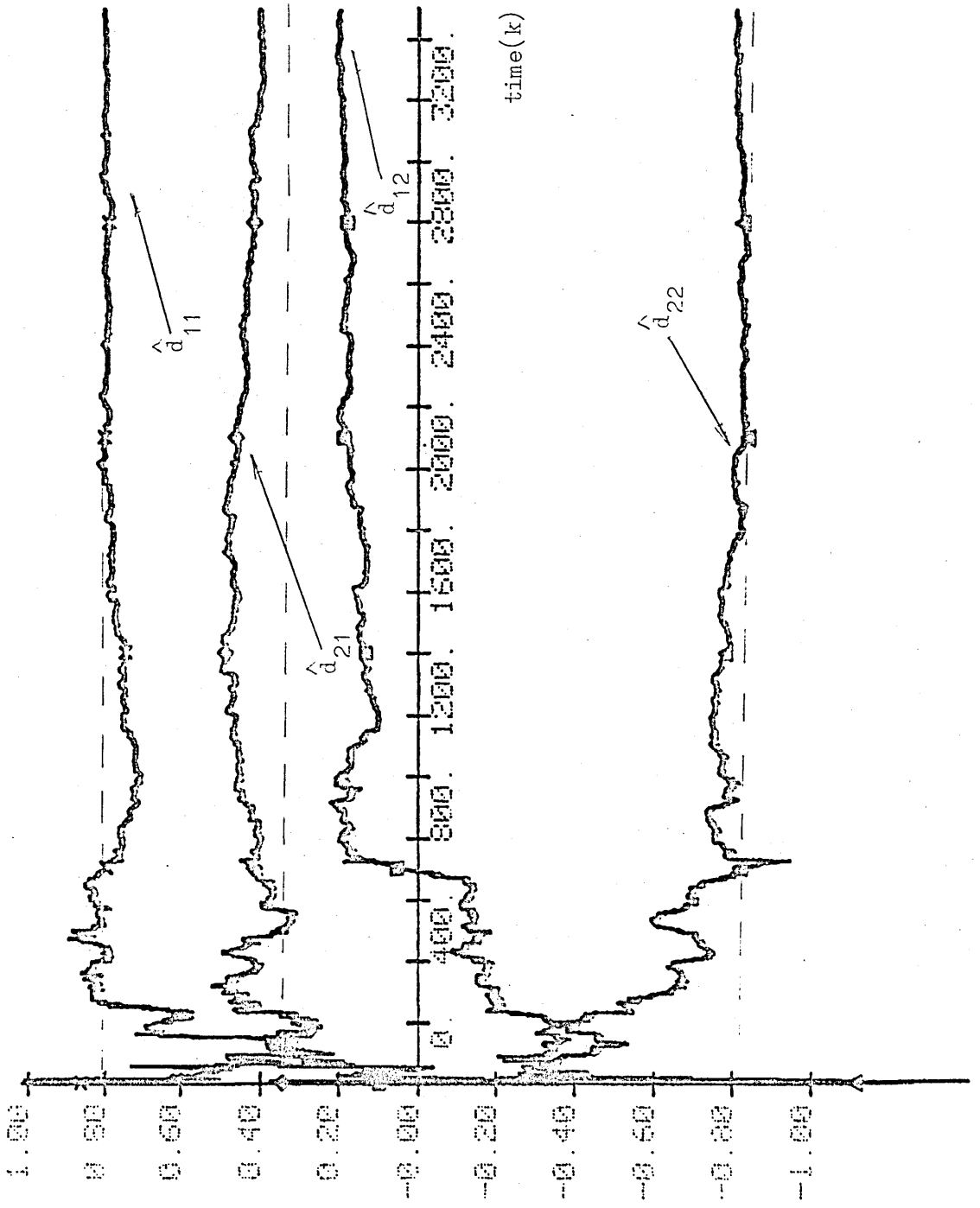


Figure 3.4 Parameter Estimation for D_{n_d}

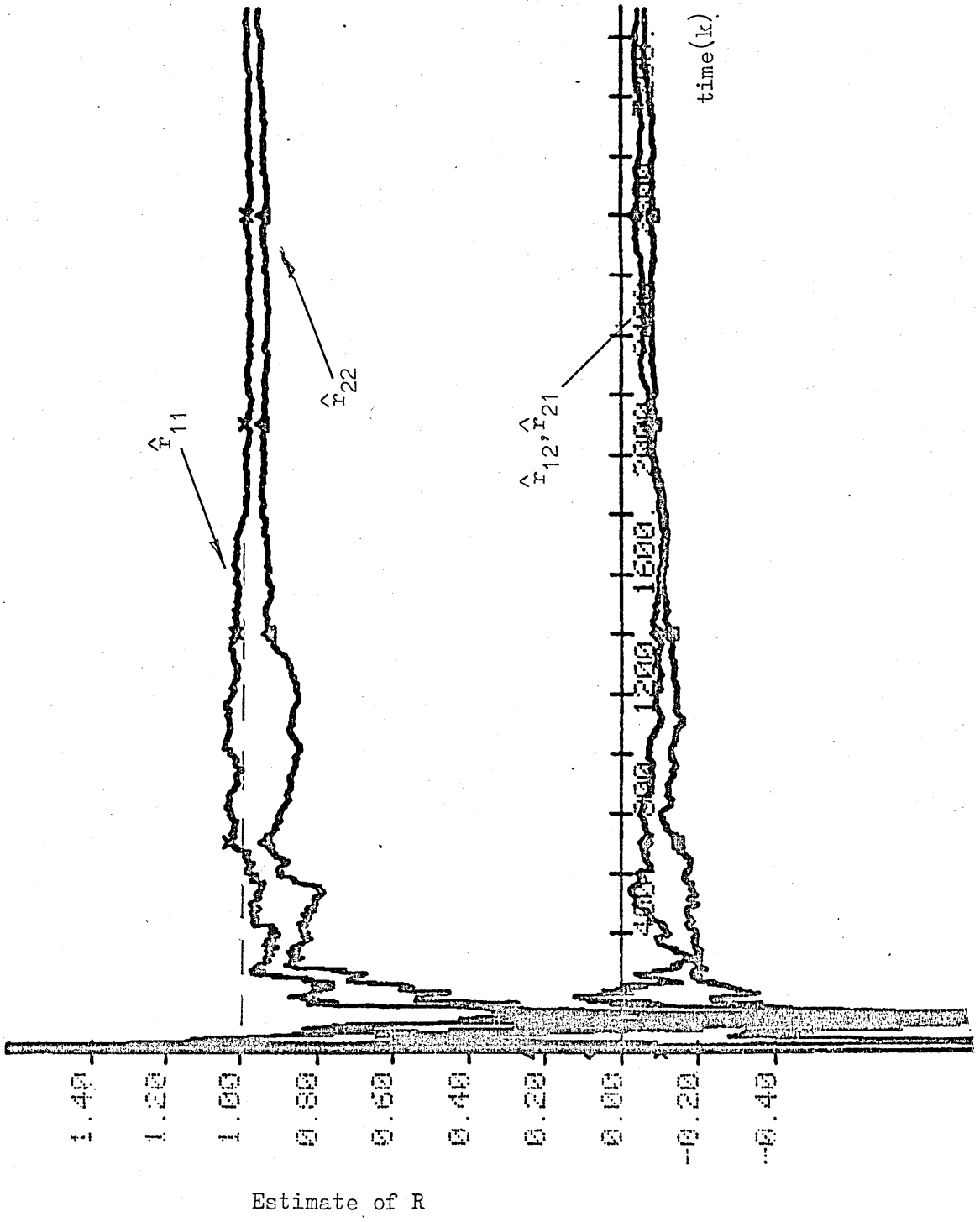


Figure 3.5 Convergence of R matrix parameters

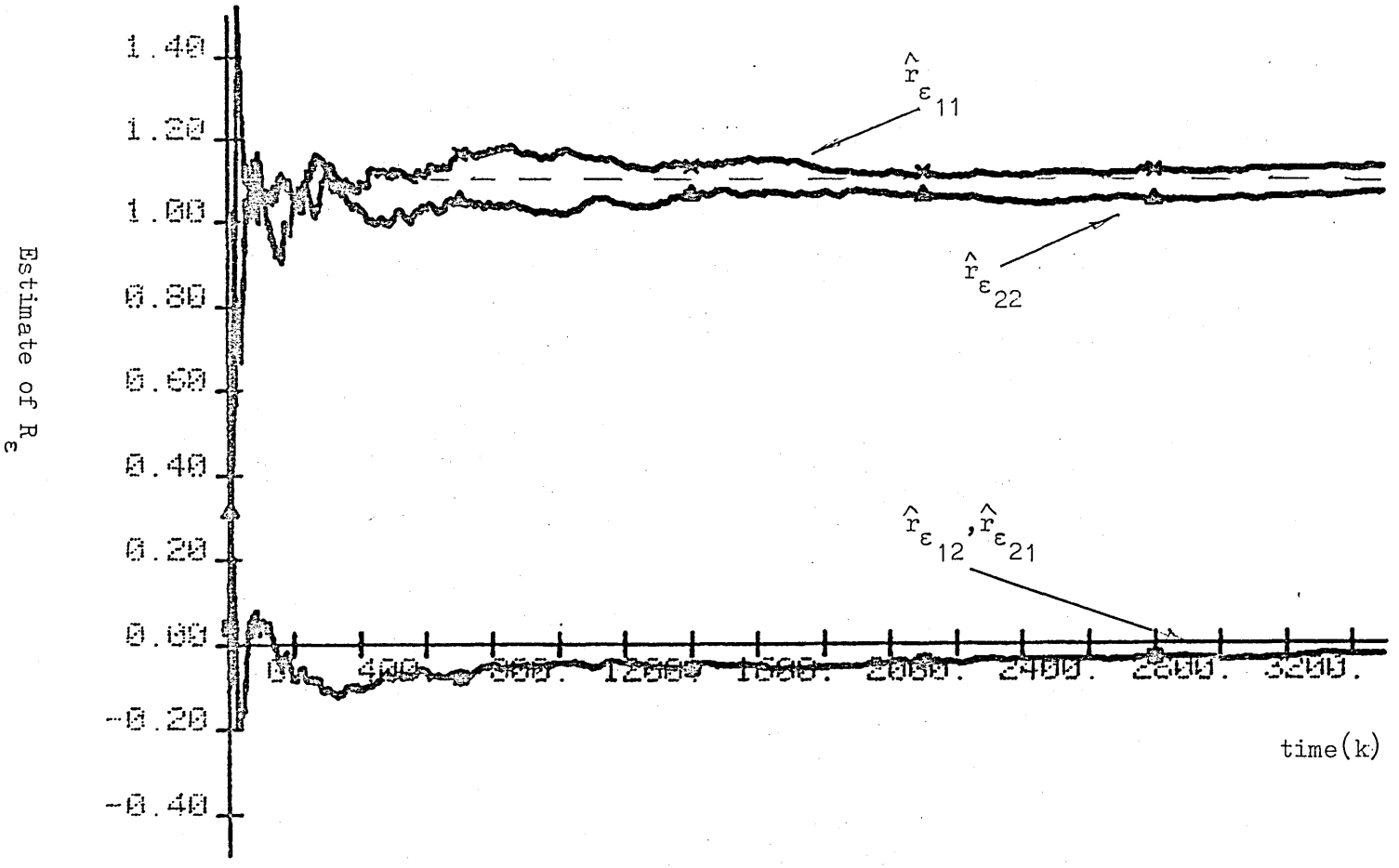


Figure 3.6 Convergence of R_ε matrix parameters

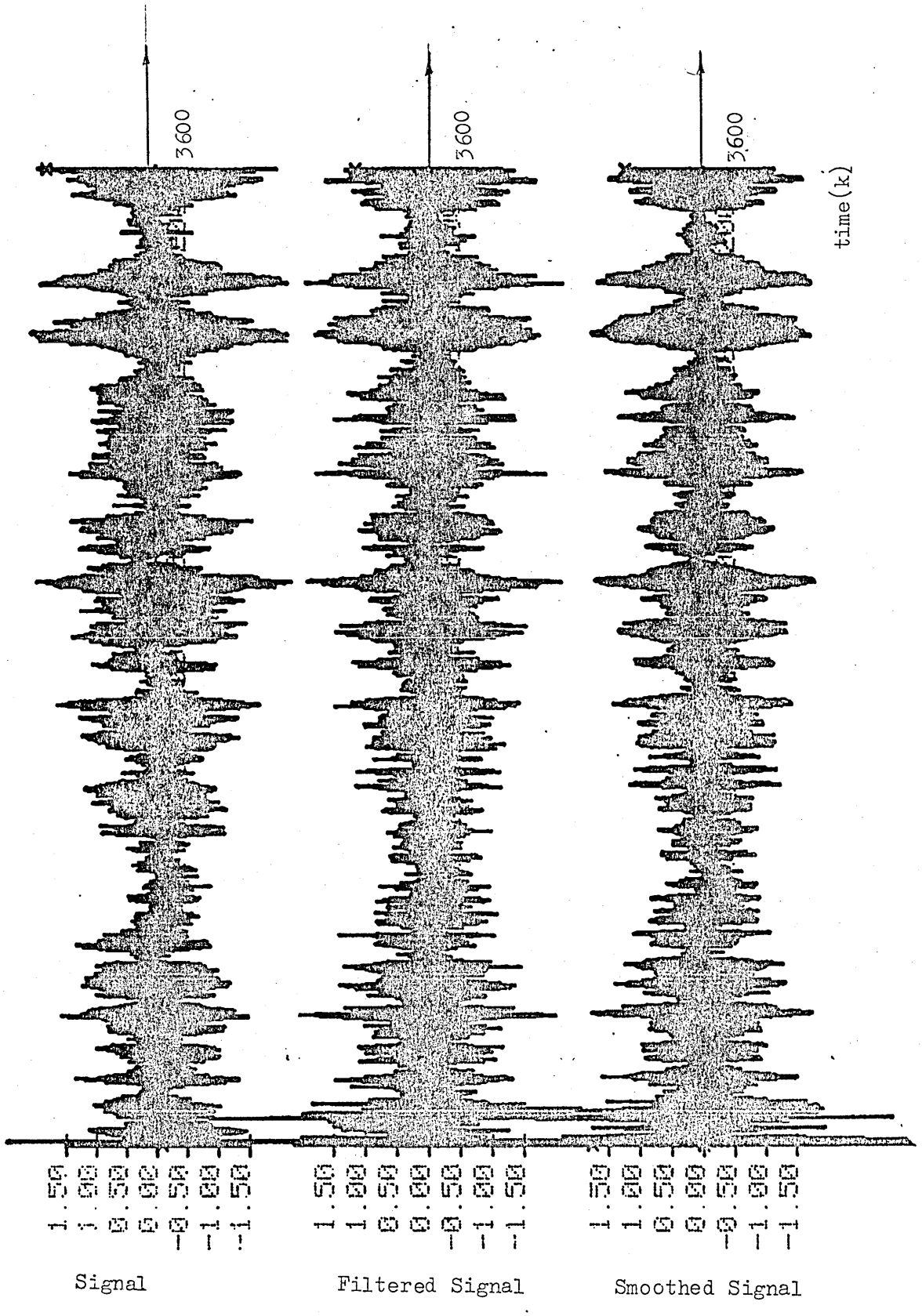


Figure 3.7 Signal, filtered & smoothed estimates for channel 1

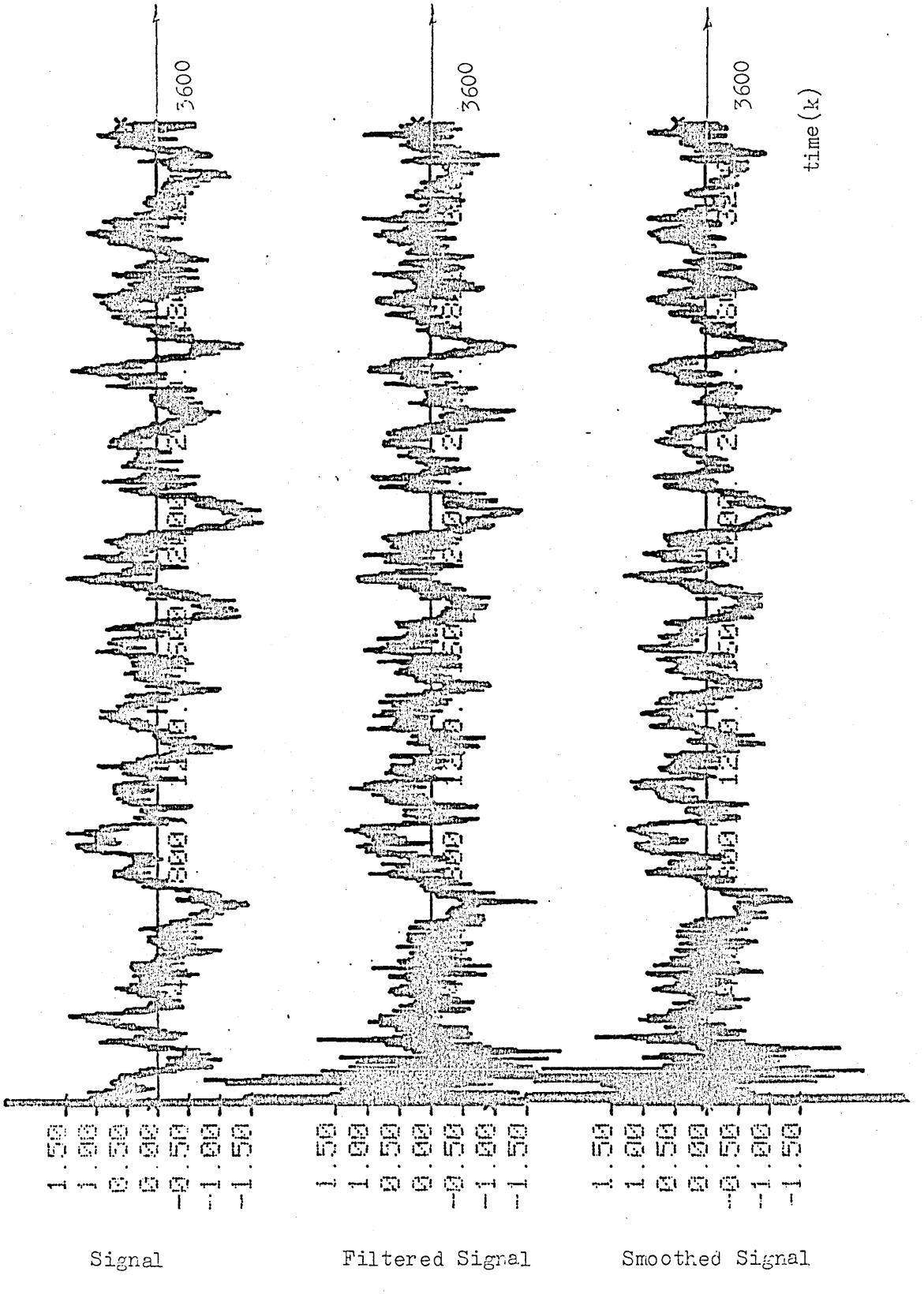


Figure 3.8 Signal, filtered & smoothed estimates for channel 2

tuning filtered and self-tuning smoothed estimates for channels 1 and 2 respectively. It can be seen (particularly with channel 2) how the filter and smoother 'tunes in' to the signal. The first few hundred estimates are erratic due to parameter uncertainty. To compare with the optimal cases, the cumulative loss per channel was computed from

$$L_i = \sum_{j=1000}^{2000} (y_i(j - \ell) - \hat{y}_i(j - \ell))^2$$

for $i = 1,2$

By starting at 1000, the initial errors associated with the parameter estimation is avoided. The loss per channel was computed for:

(a) the self-tuning Wiener filter ($\ell = 0$);

(b) the self-tuning fixed-lag smoother ($\ell = 10$ steps);

(c) the optimal Wiener filter ($\ell = 0$);

(d) the optimal fixed-lag smoother ($\ell = 10$ steps).

Figure 3.9 shows the results. For both channels, the self-tuning and optimal losses come very close to each other for the filter and smoother showing good performance.

Example 3.2

A simulation study using the following process is considered:

$$(I_2 + A_1 z^{-1} + A_2 z^{-2}) \underline{y}(K) = C_1 \underline{\omega}(K-1)$$

with

$$A_1 = \begin{bmatrix} -1 & -0.5 \\ -0.5 & -1.5 \end{bmatrix}$$

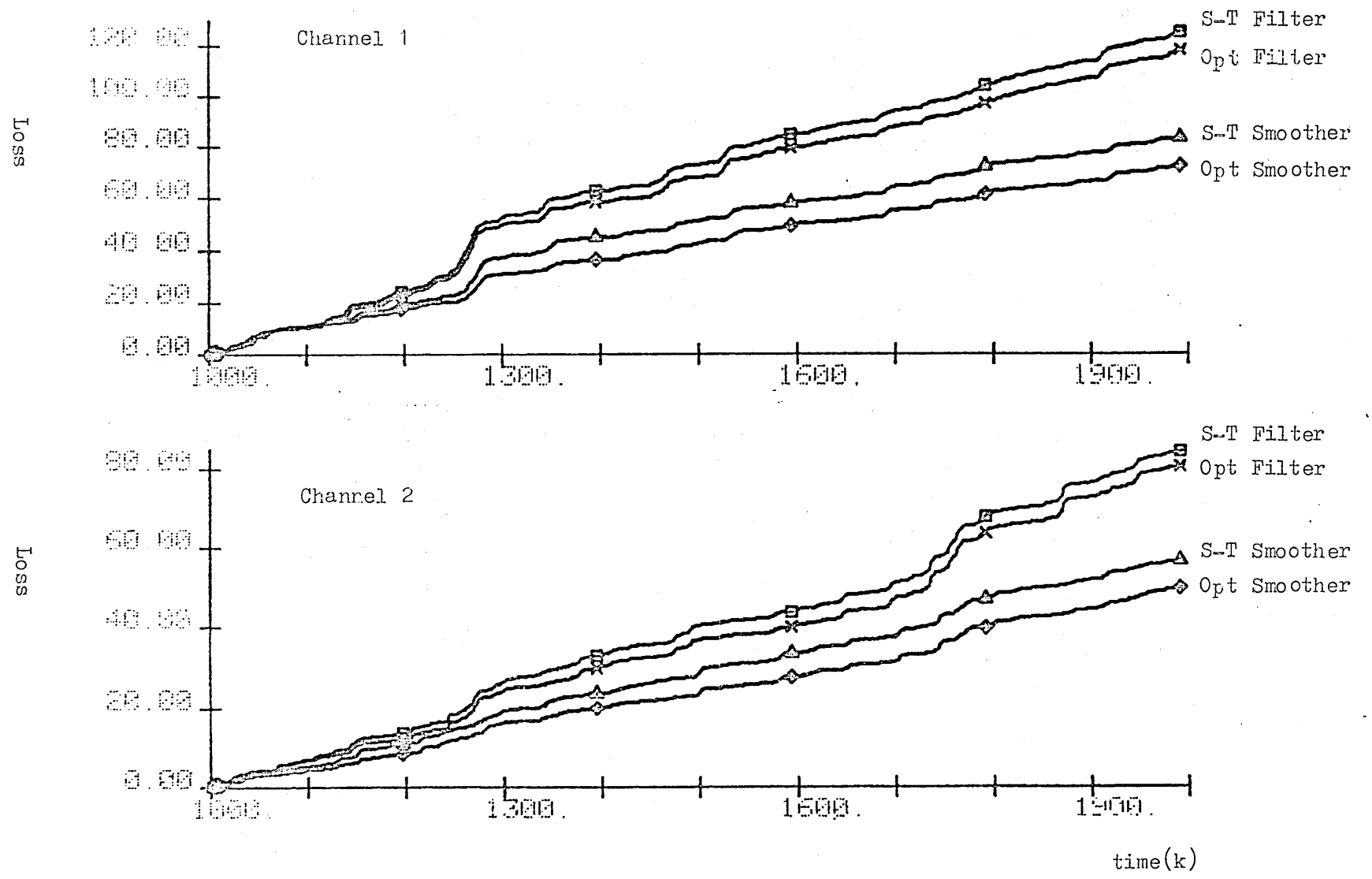


Figure 3.9 : cumulative loss per channel

$$A_2 = \begin{bmatrix} 0.8 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

$$C_1 = I_2$$

and the noise covariance matrices are given as:

$$Q = \text{diag}\{0.01, 0.01\}, \quad R = \text{diag}\{2, 1\}$$

A smoothing lag of $\ell = 10$ was used.

The signal and observation for each channel is shown in Figure 3.10, giving quite a low signal to noise ratio per channel.

The signal, self-tuning filtered and smoothed estimates for channels 1 and 2 are shown in Figures 3.11 and 3.12 respectively. These results, as with the previous example, are quite acceptable. Finally the R_ε and R matrix estimates are shown in Figures 3.13 and 3.14. Finally, the accumulative loss per channel is shown in Figure 3.15. This compares the self-tuning filter and smoother. As would be expected, the smoother gives a lower loss per channel than the filter.

Example 3.3

The purpose of this example is to study the effect of slowly time varying process parameters. Example 3.1 was used with $R = \text{diag}\{0.1, 0.1\}$ and the parameters a_{12} and a_{21} varying according to

$$a_{12} = a_{21} = 0.4\sin(\pi k/7000)$$

this gives for the number of points used in the simulation a quarter of a sine wave with peak value 0.4. Thus when $k = 3500$, $a_{12} = a_{21} = 0.4$ which is the same as used for

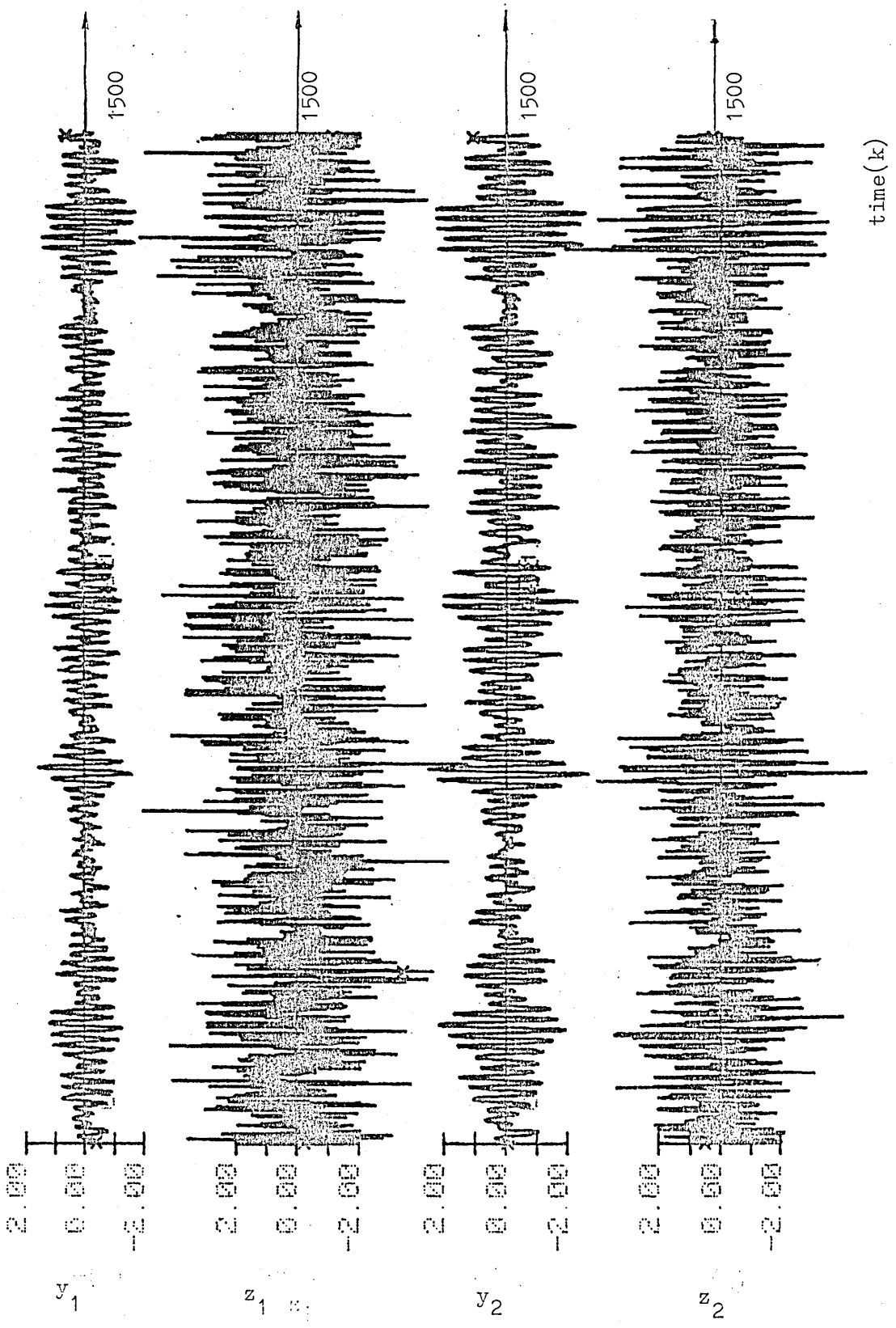


Figure 3.10 Signal & Observation for both Channels

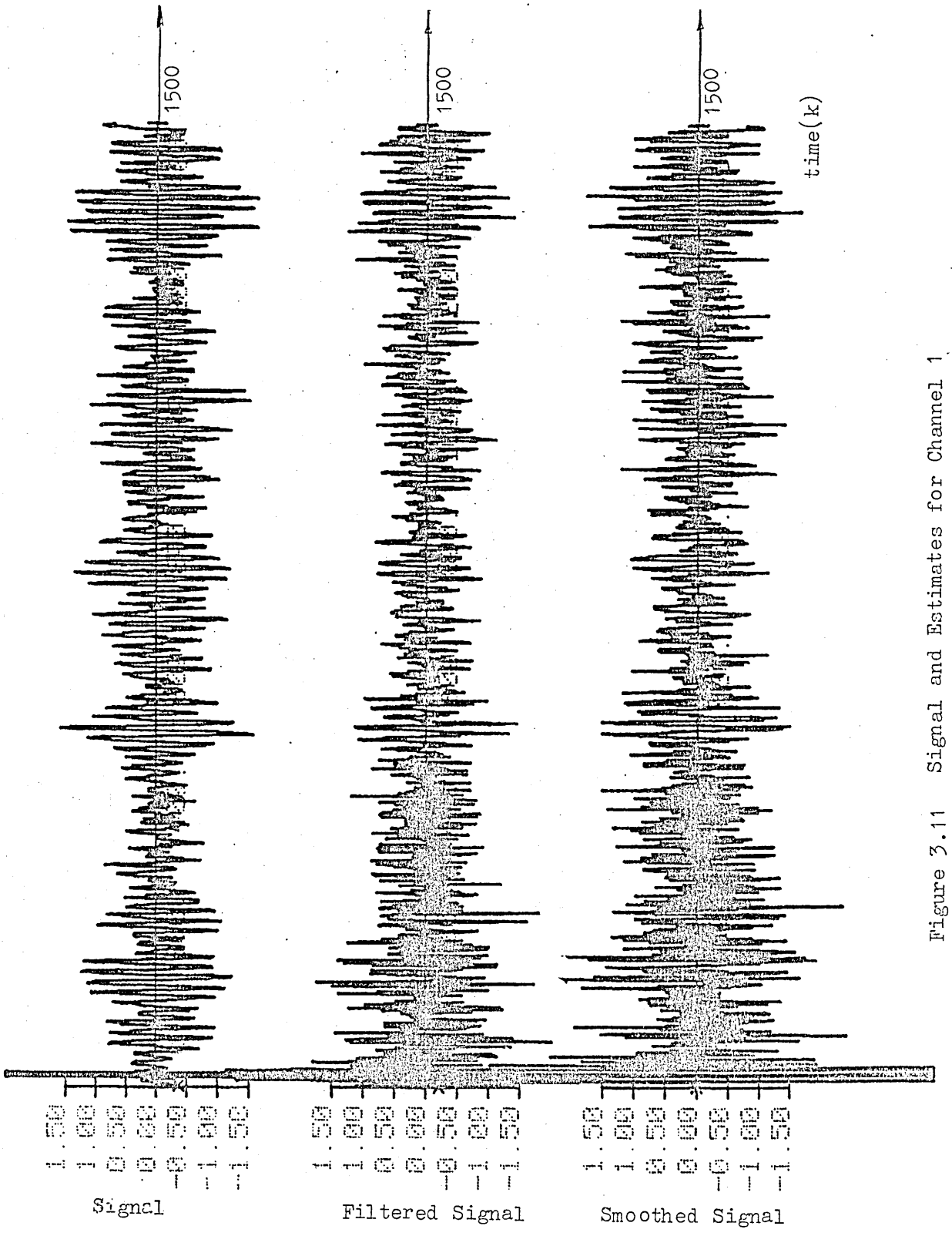


Figure 3.11 Signal and Estimates for Channel 1

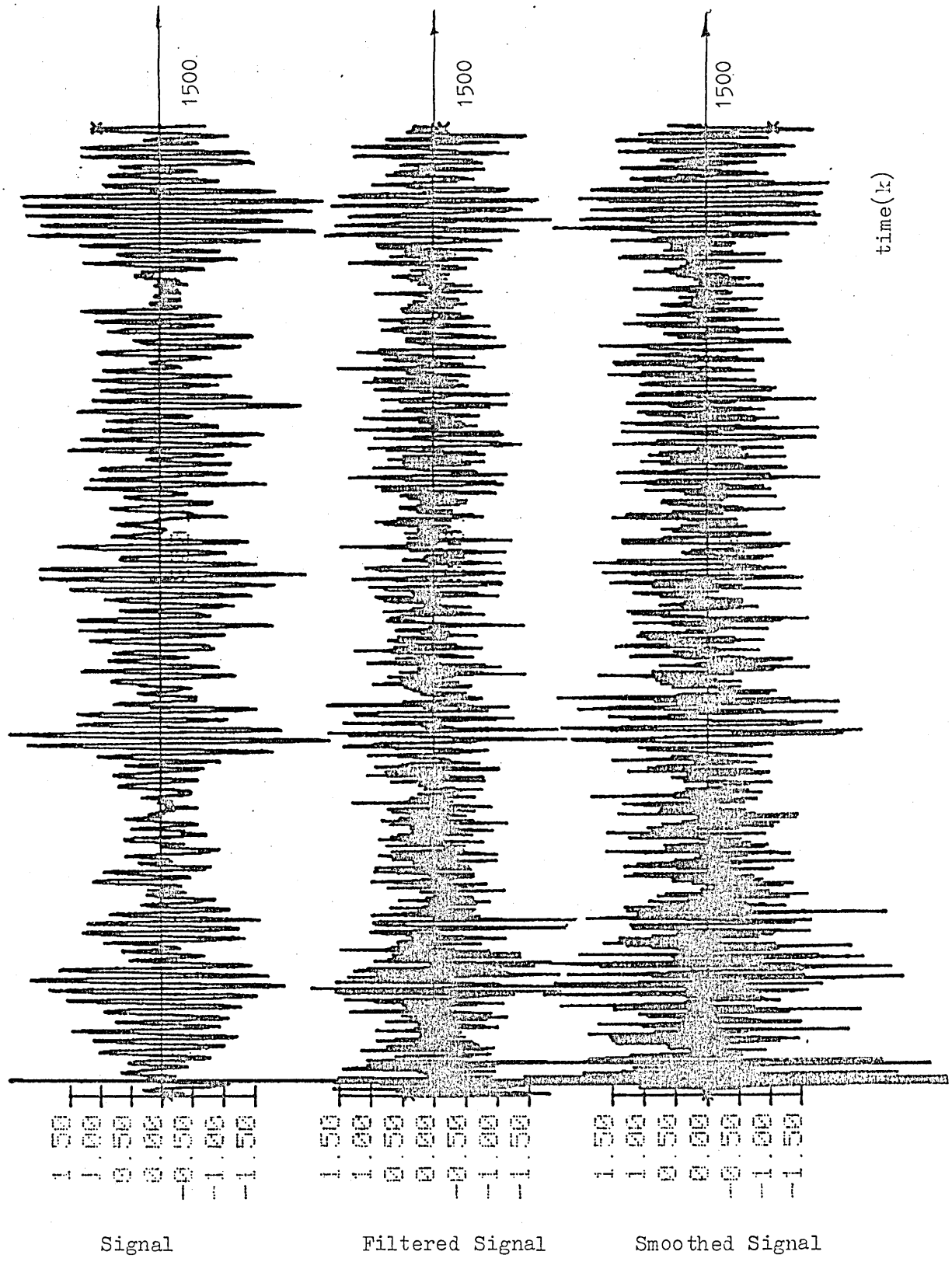


Figure 3.12 Signal and Estimates for Channel 2

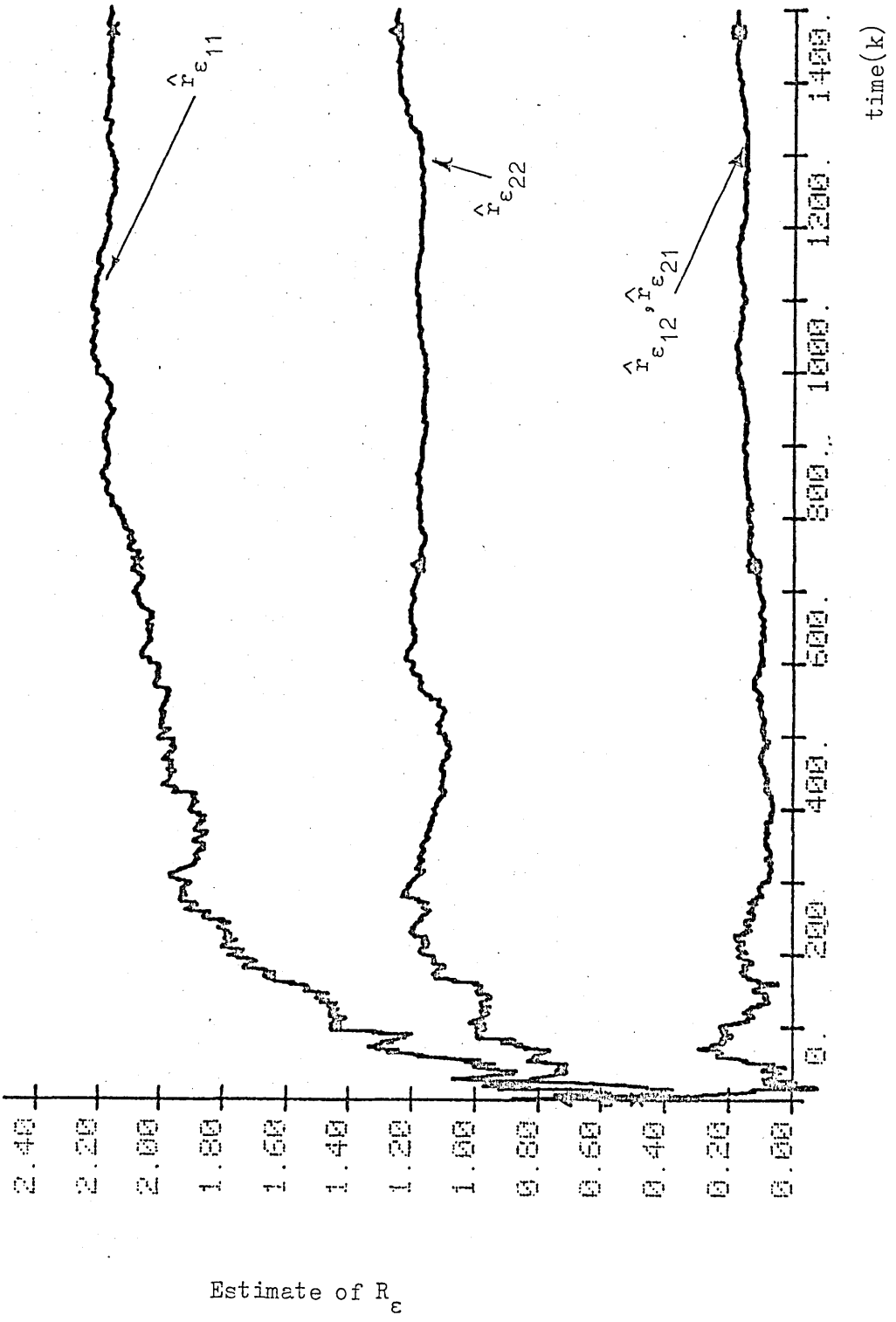


Figure 3.13 Convergence of R_ε matrix parameters

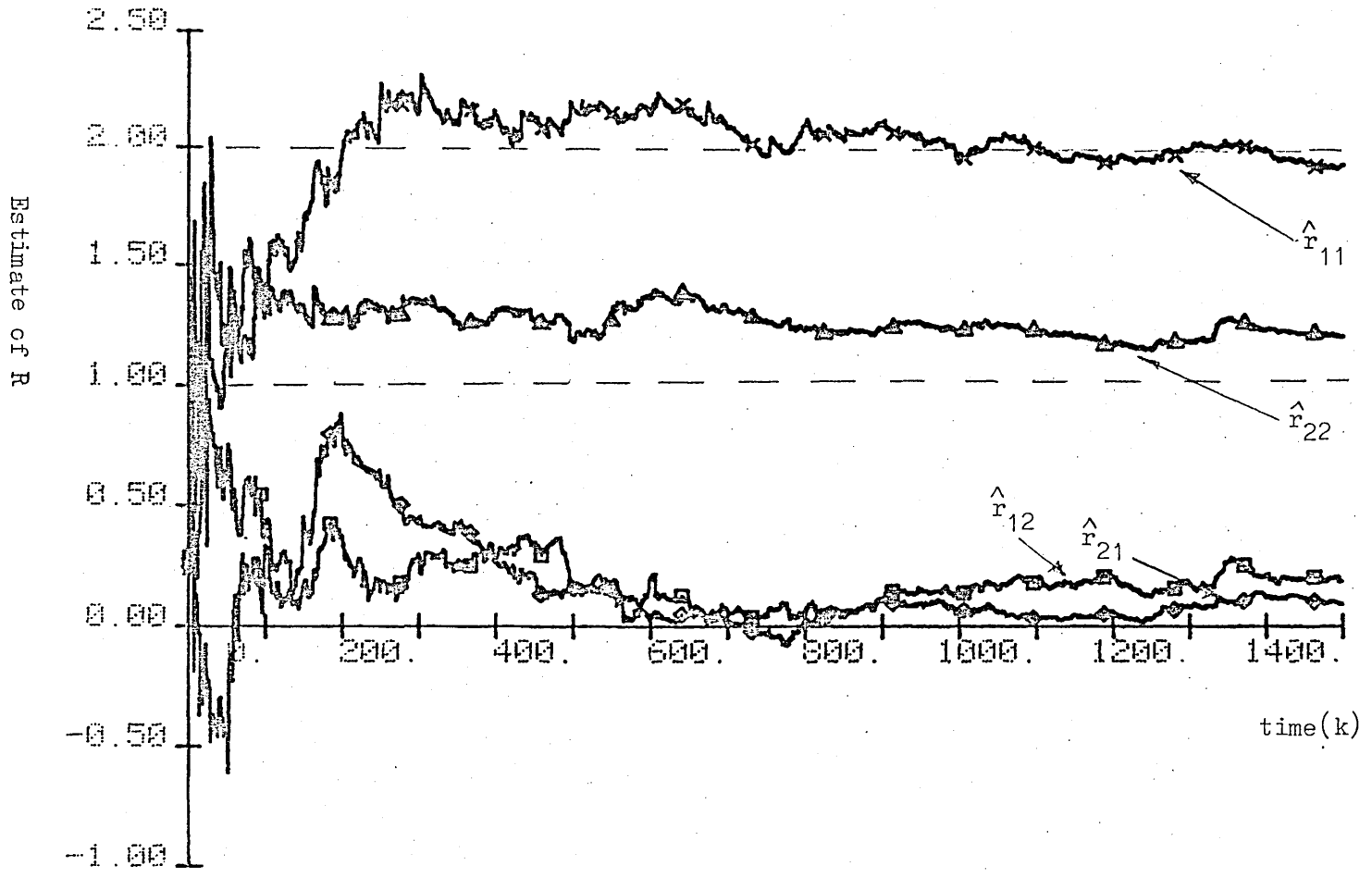


Figure 3.14 Convergence of R matrix parameters

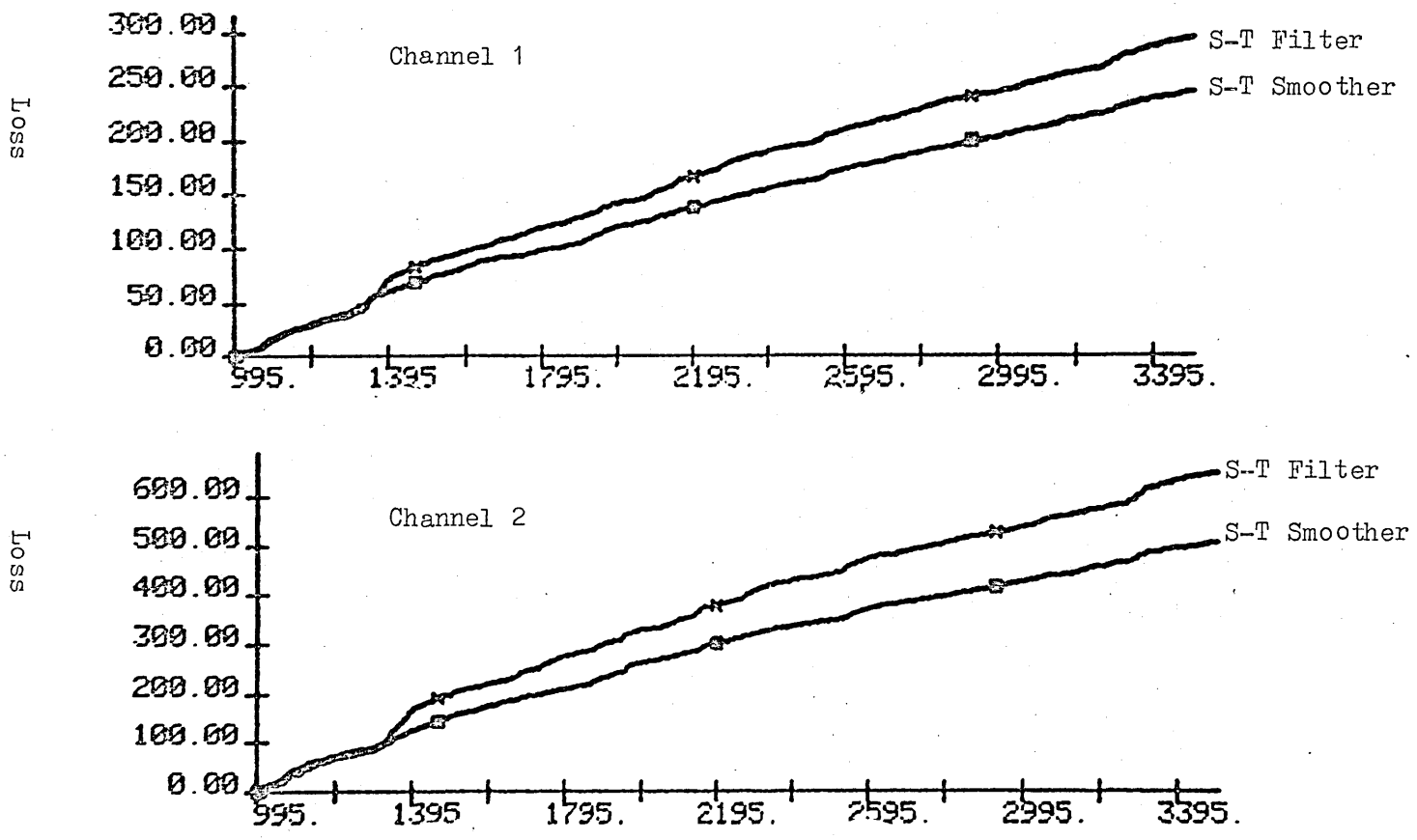


Figure 3.15 cumulative loss per channel

example 3.1.

In order to get a parameter estimation routine to track varying parameters, it is best to use a 'forgetting factor' of less than unity. In this way, past data can be weighted exponentially more than present data. For this example, extended recursive least squares [56] was used with a 'forgetting factor' of 0.997. Figure 3.17 shows the parameter estimates for A_1. The two time-varying parameters are both shown to be tracking the true sine curve $0.4\sin(\pi k/7000)$. Figure 3.18 shows the D_1 parameter estimates, two parameters of which are also varying with time. Figures 3.19 and 3.20 show convergence of R and R_ε. The self-tuning estimates are shown for channels 1 and 2 in figures 3.21 and 3.22. Note that the signal amplitudes are slowly increasing with time and how the self-tuning filter and smoother still manages to give good estimates for each channel.

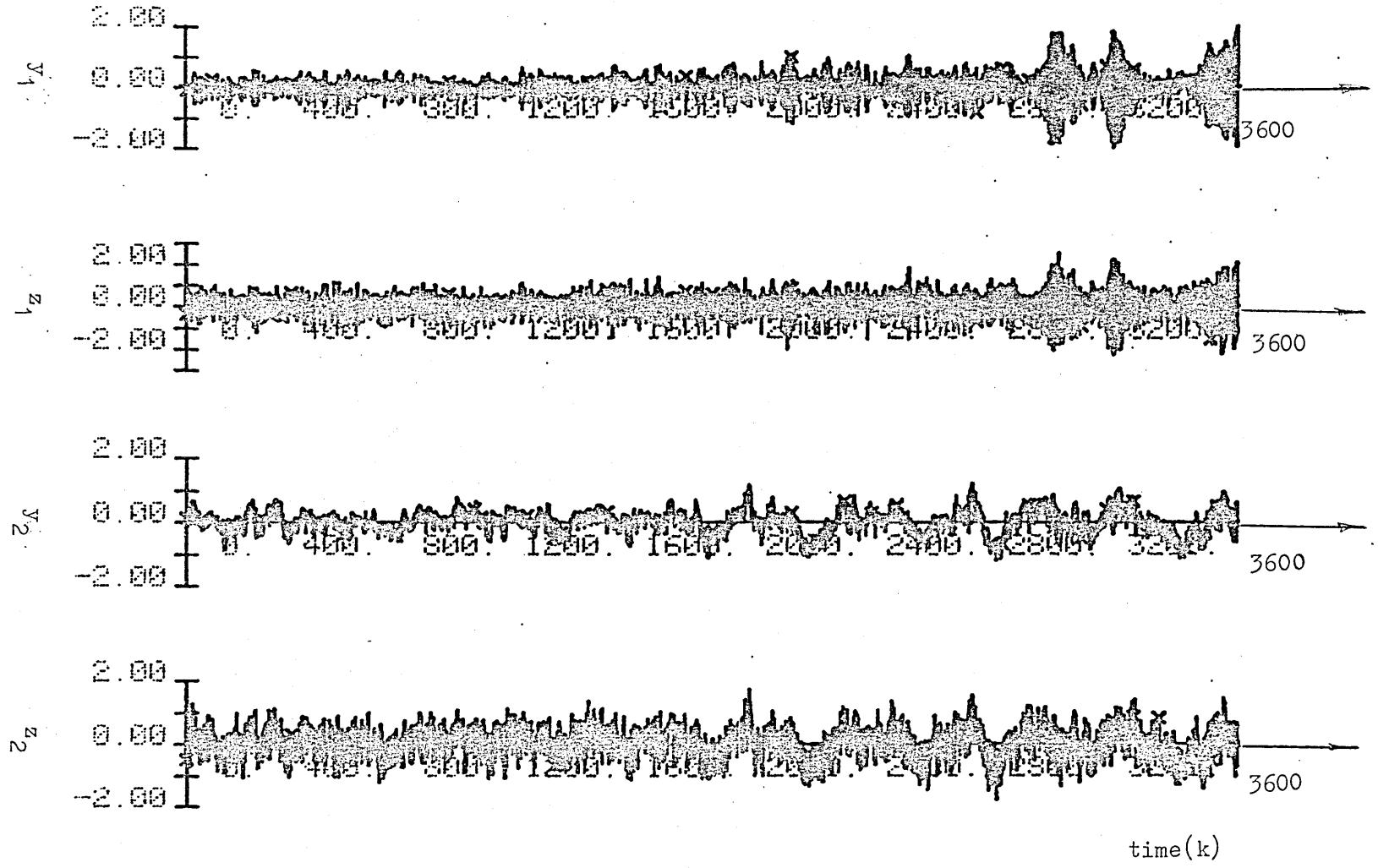
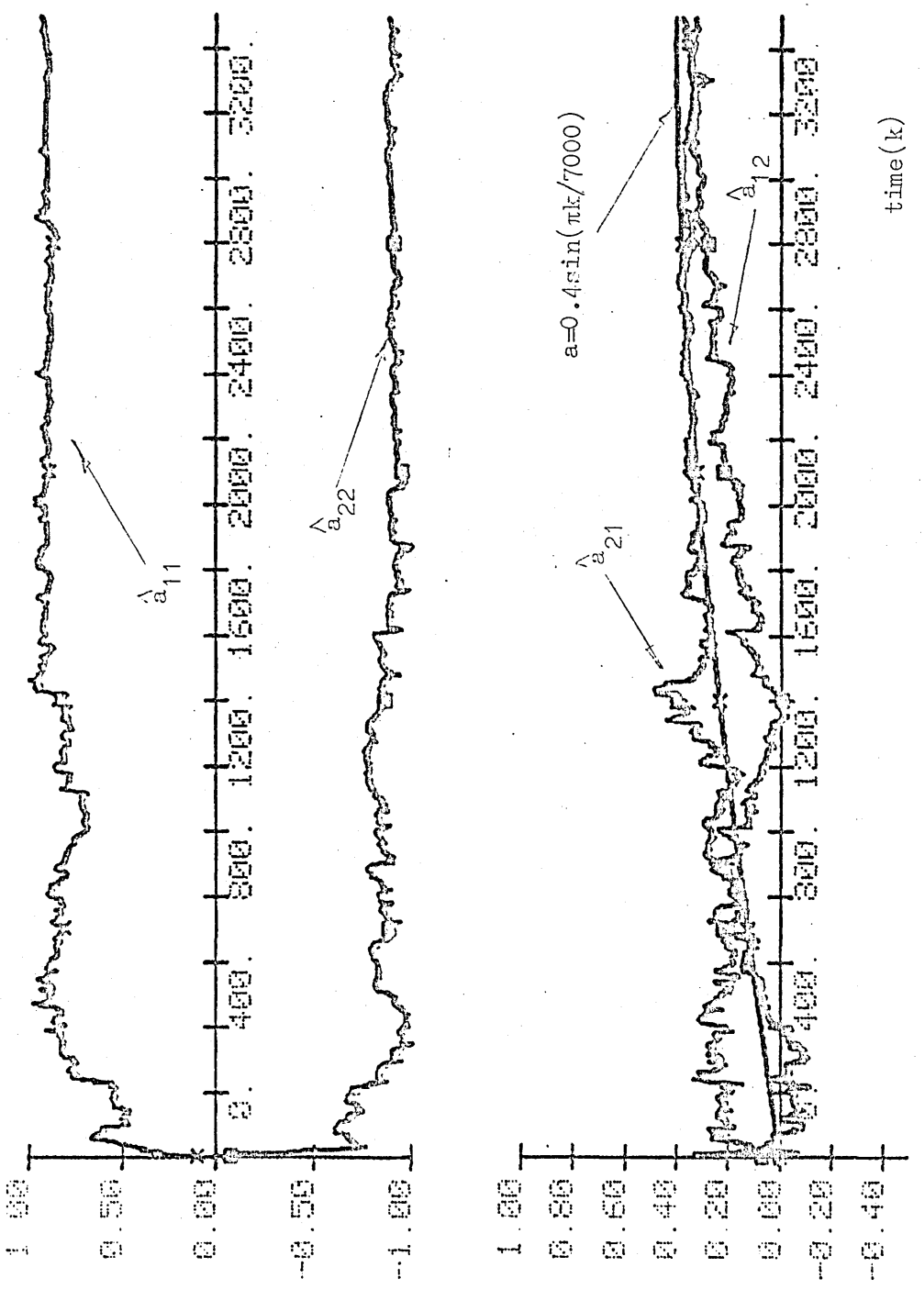


Figure 3.16 Signal & Observation for both channels



Parameter Estimation for A_1

Figure 3.17 Estimate of A_{n_a}

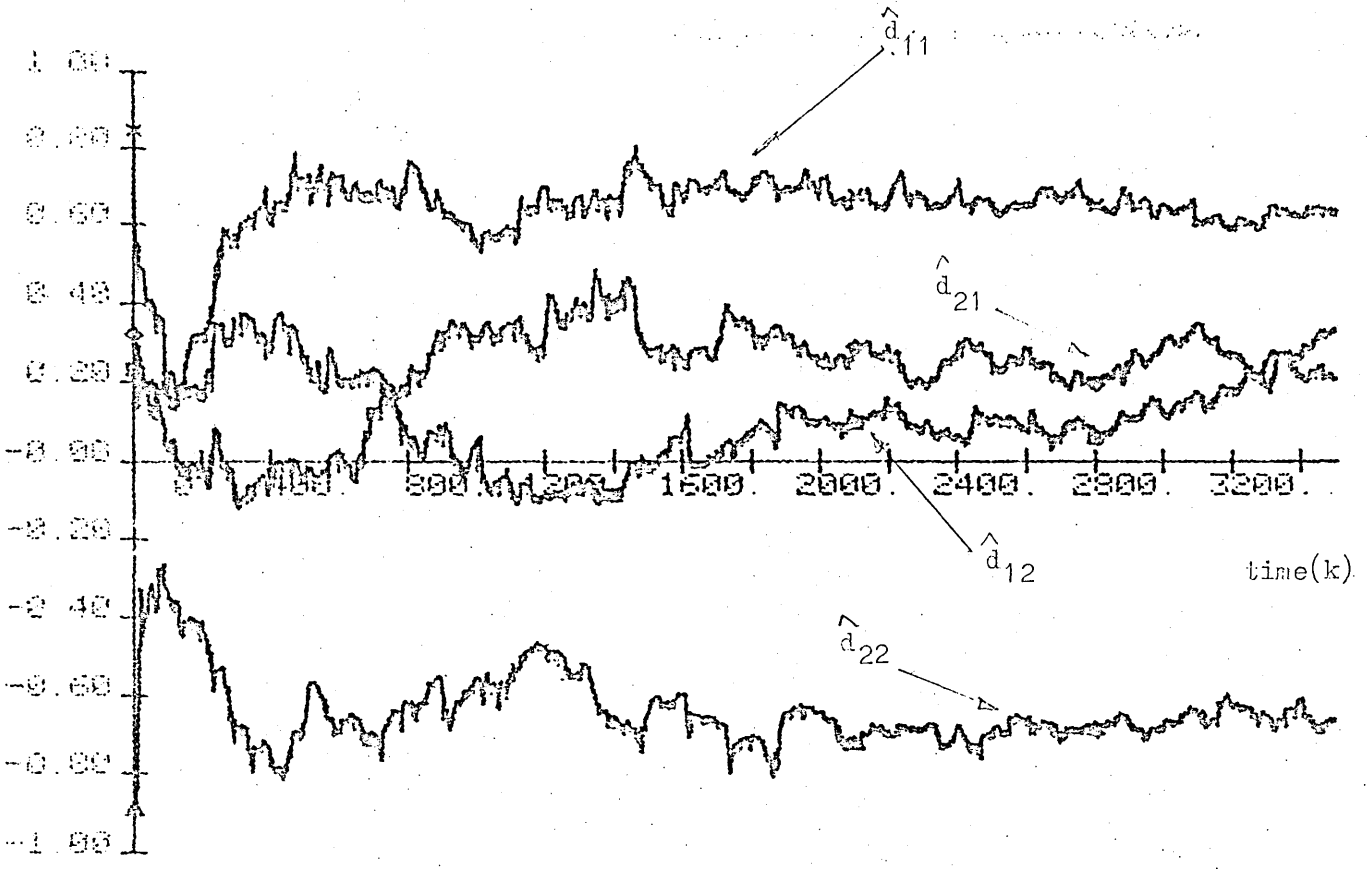


Figure 3.18 Estimate of D_{n_d}

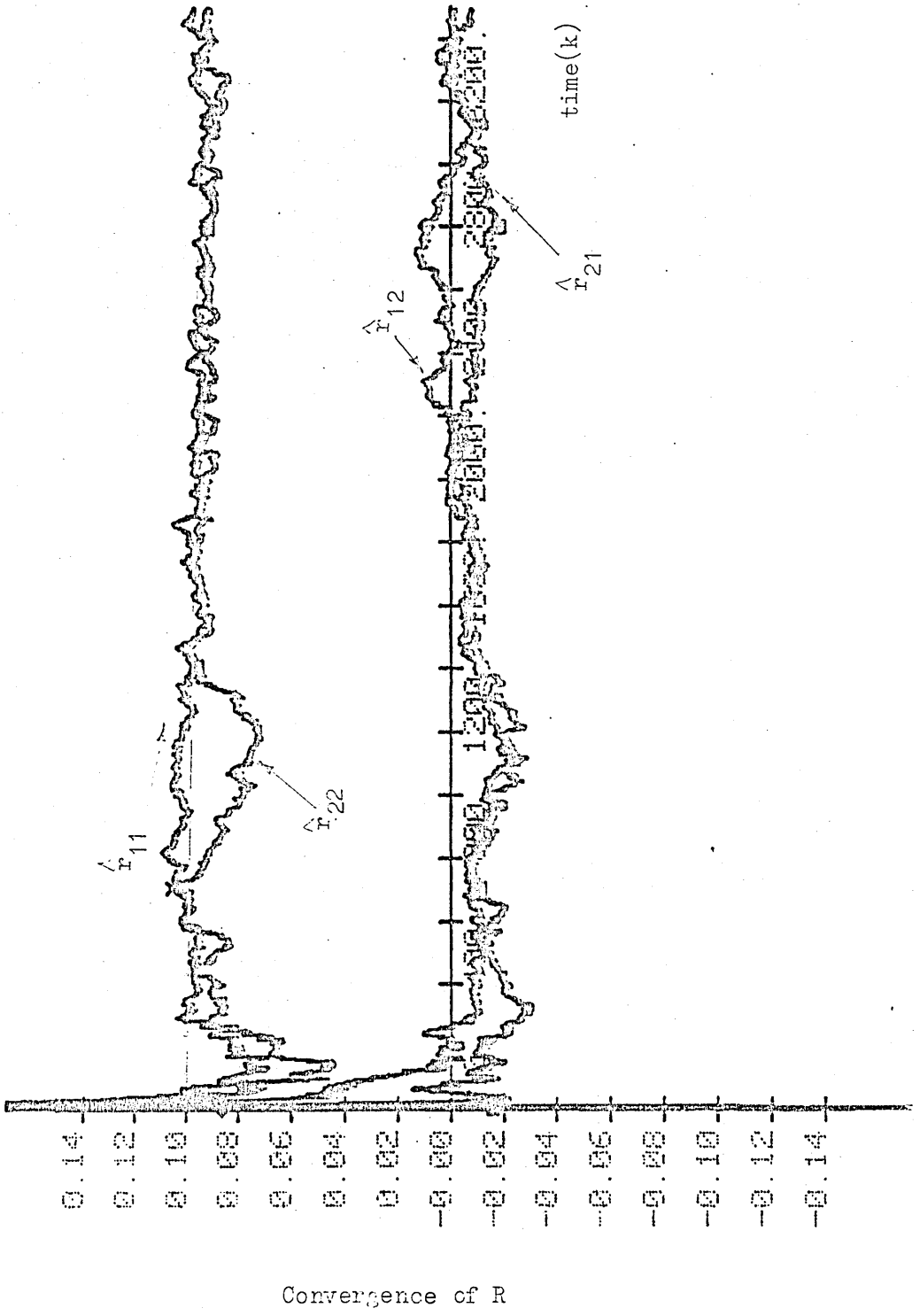


Figure 3.19 Convergence of R matrix parameters

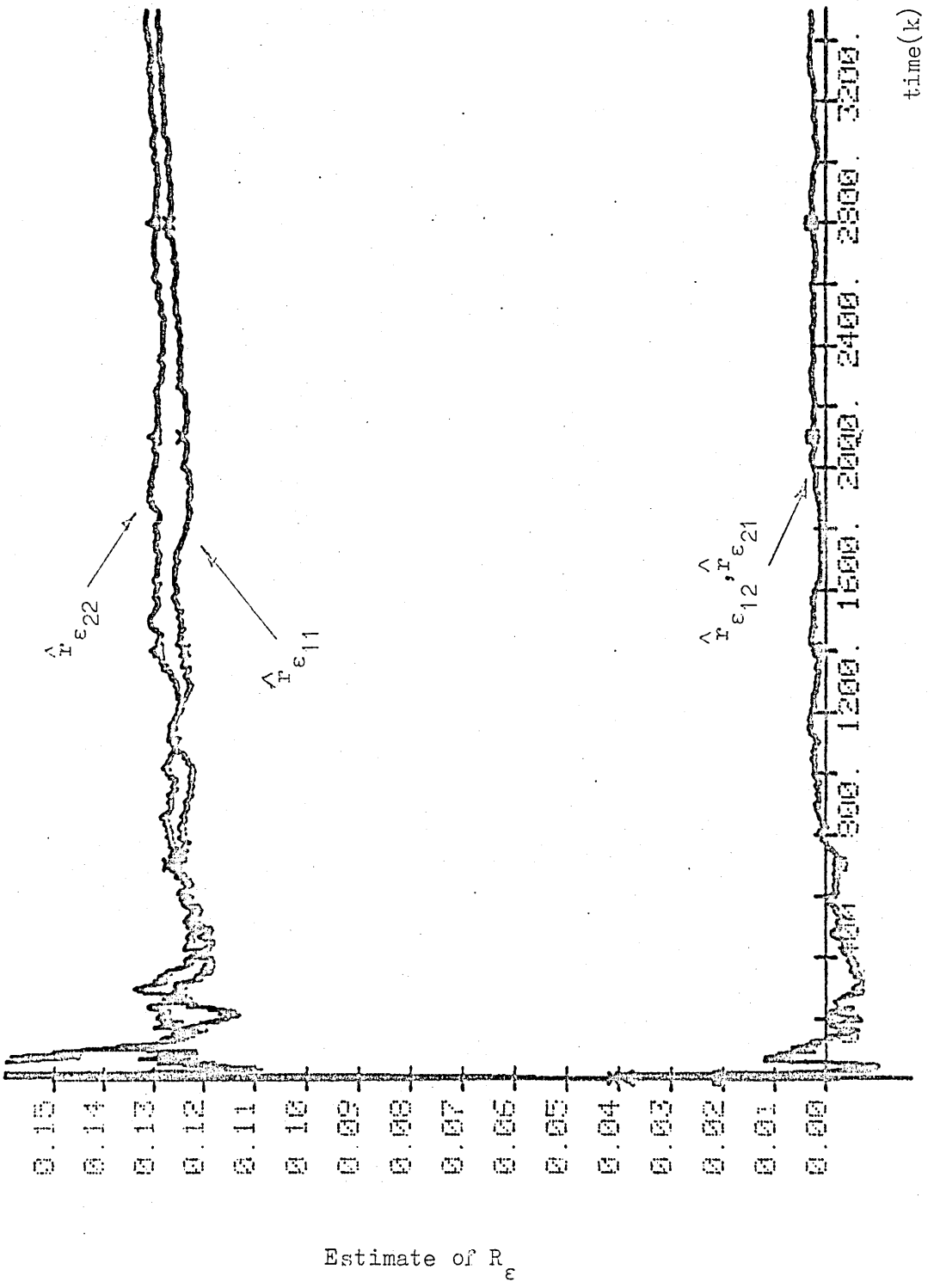


Figure 3.20 Convergence of R_ε matrix parameters

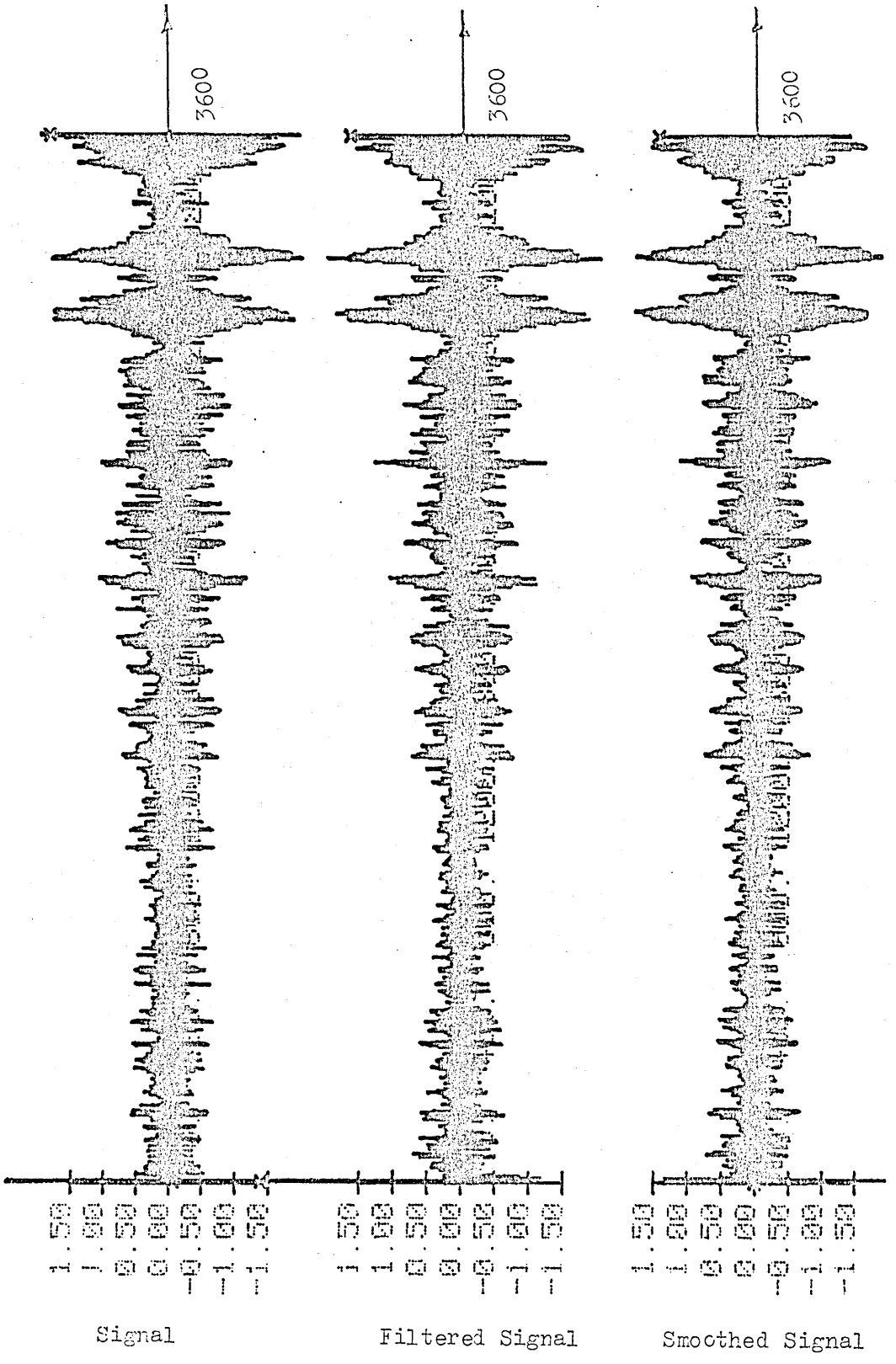


Figure 3.21 Signal, Filtered & Smoothed Estimates for Channel 1

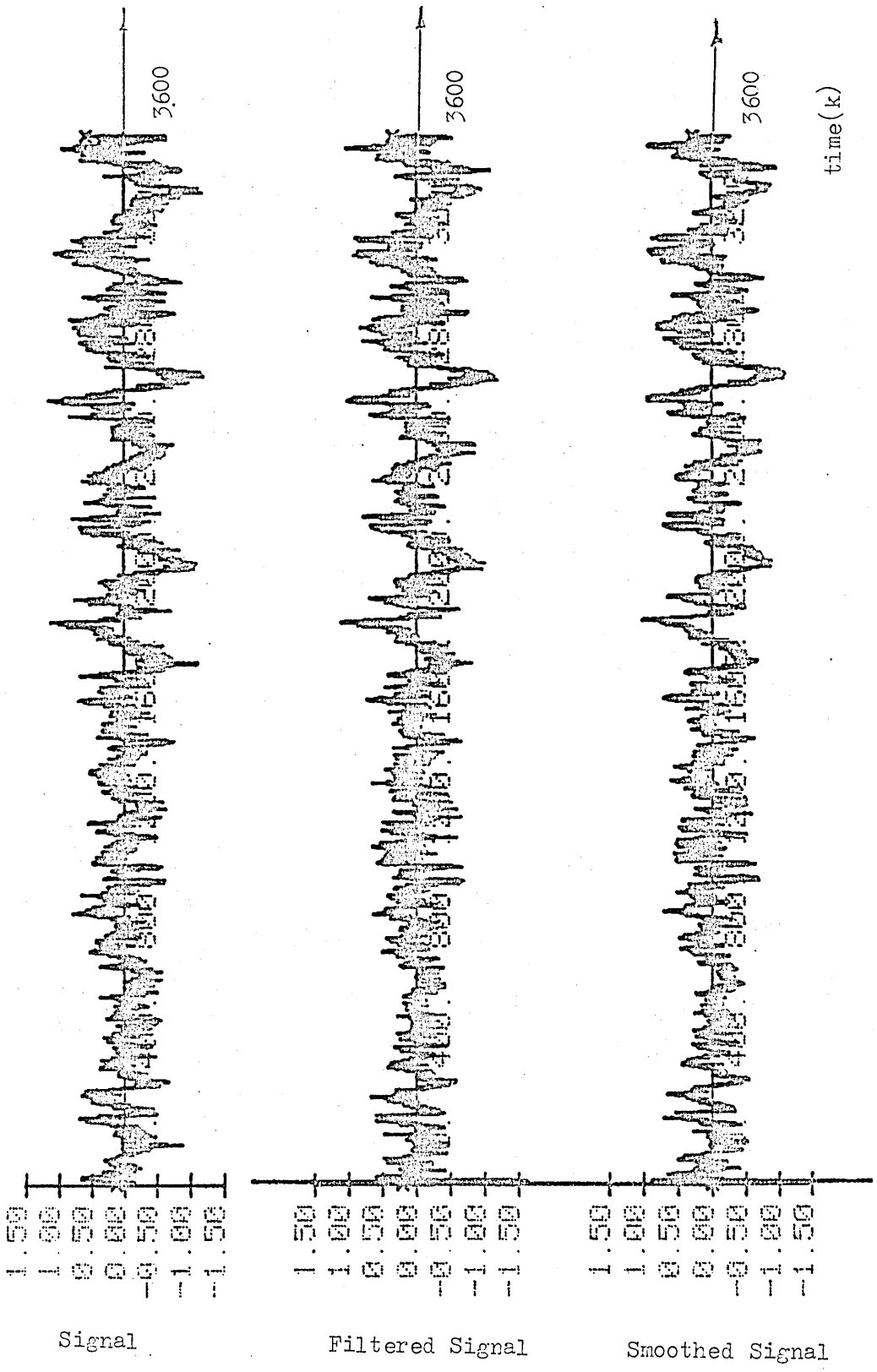


Figure 3.22 Signal, Filtered & Smoothed Estimates for Channel 2

3.6 Estimation of R when A_{n_a} is singular.

When A_{n_a} is singular algorithm 3.3 fails. This is due to the computation of the measurement noise covariance R in step 3 or the computation of S in step 3 for the filter. This section overcomes the problem for the special case when the measurement noise covariance matrix is diagonal. A simulation is shown to illustrate this.

If R is diagonal, let R be represented as $R = \text{diag}\{\alpha_1, \alpha_2 \ldots \alpha_r\}$ and let the vector $\underline{\theta}_m = [\alpha_1, \alpha_2 \ldots \alpha_r]^T$ be the vector of unknown parameters of R.

The problem may be defined as :

Given A_{n_a}, D_{n_d} and R_ε, find R as the solution of

$$A_{n_a} R = D_{n_d} R_\varepsilon \tag{3.65}$$

when $\det(A_{n_a}) = 0$.

Solution

Let the matrix A_{n_a} and matrix product $D_{n_d} R_\varepsilon$ have elements a_{ij} and β_{ij}, i, j ; 1, 2 ... r respectively.

Define the $r^2 \times r$ dimensional matrix A_x as

$$A_x = [N_1 \mid N_2 \mid \ldots \mid N_r]^T \tag{3.66}$$

where

$$N_i = \text{diag}\{a_{i1}, a_{i2} \ldots a_{ir}\} \tag{3.67}$$

$$i = 1, 2 \ldots r$$

and the r^2 dimensional vector $\underline{J}_x$ as the rows of $D_{n_d} R_\varepsilon$

$$\underline{J}_x = [\beta_{11}, \beta_{12} \ldots \beta_{1r}, \beta_{21}, \beta_{22} \ldots \beta_{2r}, \ldots \beta_{r1}, \beta_{r2} \ldots \beta_{rr}]^T \tag{3.68}$$

and rewrite (3.65) in terms of A_x, $\underline{\theta}_m$ and J_x thus :

$$A_x \underline{\theta}_m = \underline{J}_x \tag{3.69}$$

Equation (3.66) has a solution using least squares (Barnett [101]).

$$\underline{\theta}_m = (A_x^T A_x)^{-1} A_x^T \underline{J}_x \qquad (3.70)$$

Equation (3.70) holds under the assumption that A_x has full rank r.

Algorithm 3.3 must be altered in step 2 so that $R_\varepsilon(k-\ell)$ is estimated for any $\ell \geq 0$.

Step 3 now becomes :

Step 3 (singular A_{n_a})

For $\ell \geq 0$ compute $\hat{\underline{\theta}}_M(k-\ell) = (\hat{A}_x^T \hat{A}_x)^{-1} A_x^T \hat{\underline{J}}_x(k-\ell)$

giving $\hat{R}(k-\ell)$

if $\ell = 0$ compute $\hat{\underline{S}}(k) = \hat{R}(k)\, \hat{R}_\varepsilon^{-1}(k)$

Example 3.4 A_{n_a} singular

Consider the signal generating process given by :

$$(I_2 + A_1 z^{-1})\underline{y}(k) = \underline{\omega}(k-1)$$

where

$$A_1 = \begin{bmatrix} 0.49 & 0.49 \\ 0.49 & 0.49 \end{bmatrix}$$

and with Q = diag {0.01, 0.01} and R = diag {1.0, 1.0}

Evidently for this example Algorithm 3.3 fails when estimating R or S. The modified Algorithm in this section will be shown to be able to cope with that problem.

For this example $D_1 R_\varepsilon$ has the form

$$D_1 R_\varepsilon = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix}$$

hence from (3.68)

$$\underline{J}_x = [\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}]^T$$

A_x is found from (3.66) and (3.67) as

$$A_x = \begin{bmatrix} 0.49 & 0 \\ 0 & 0.49 \\ 0.49 & 0 \\ 0 & 0.49 \end{bmatrix}$$

which is full rank.

$\underline{\theta}_m = (\alpha_1, \alpha_2)^T$ where the true values of α_1 and α_2 are 1.0 as given by R. Using the modified algorithm to compute $\hat{\underline{\theta}}_m$ and thus $\hat{R}$, the estimate of α_1 and α_2 is shown in figure 3.23. Figure 3.24 shows the cumulative loss per channel for the self-tuning smoother ($\ell = 10$) and the filter. Figures 3.25 and 3.26 show the self-tuning smoothing and filtering responses for each channel. From the cumulative loss graphs, the smoother clearly performs better than the filter giving a lower mean square error.

3.6.1 Further difficulties arising

In the former analysis for A_{n_a} singular it was assumed that A_x was full rank r. Although this is a fair assumption to make there are cases where A_x is not full rank. For example if $A_1 = \begin{bmatrix} 0 & 0.49 \\ 0 & 0.49 \end{bmatrix}$ in example 3.4, then

A_x becomes

$$A_x = \begin{bmatrix} 0 & 0 \\ 0 & 0.49 \\ 0 & 0 \\ 0 & 0.49 \end{bmatrix}$$

which has rank one and makes $A_x^T A_x$ singular in (3.70). The following is a method of solving for this case although it is not intended to be in any way practical since it will require on-line testing of rank.

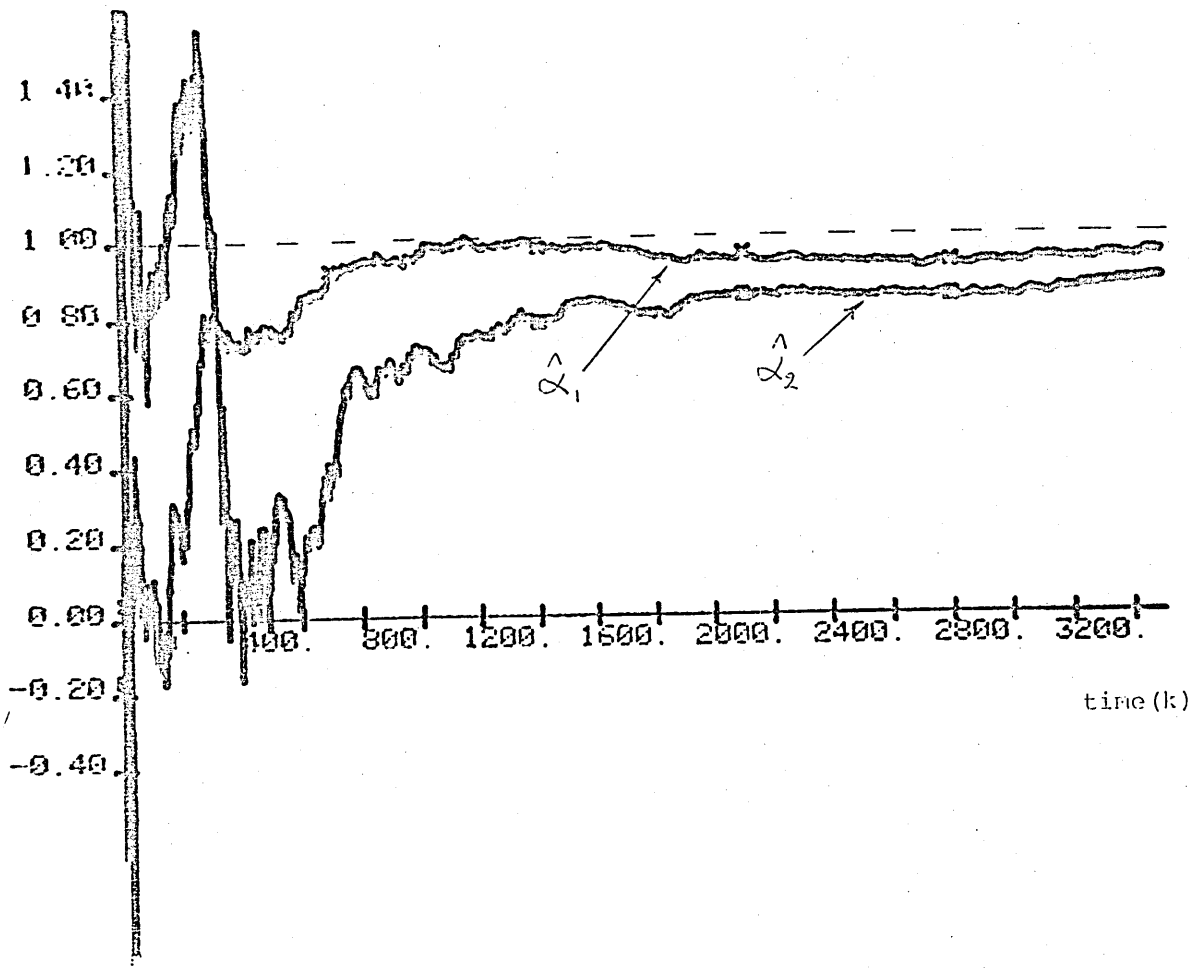


Figure 3.23 Estimate of R matrix

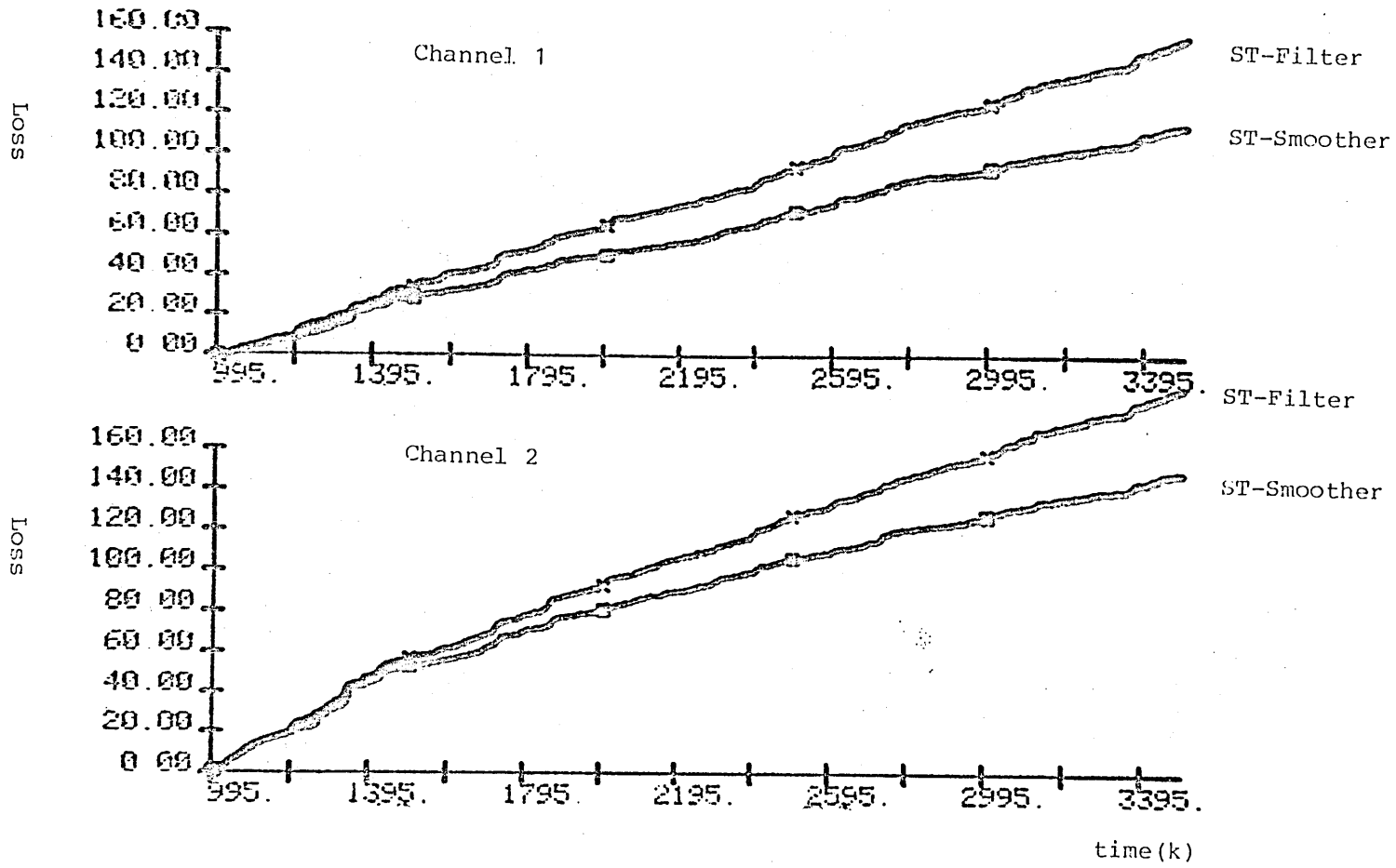


Figure 3.24 cumulative loss per Channel

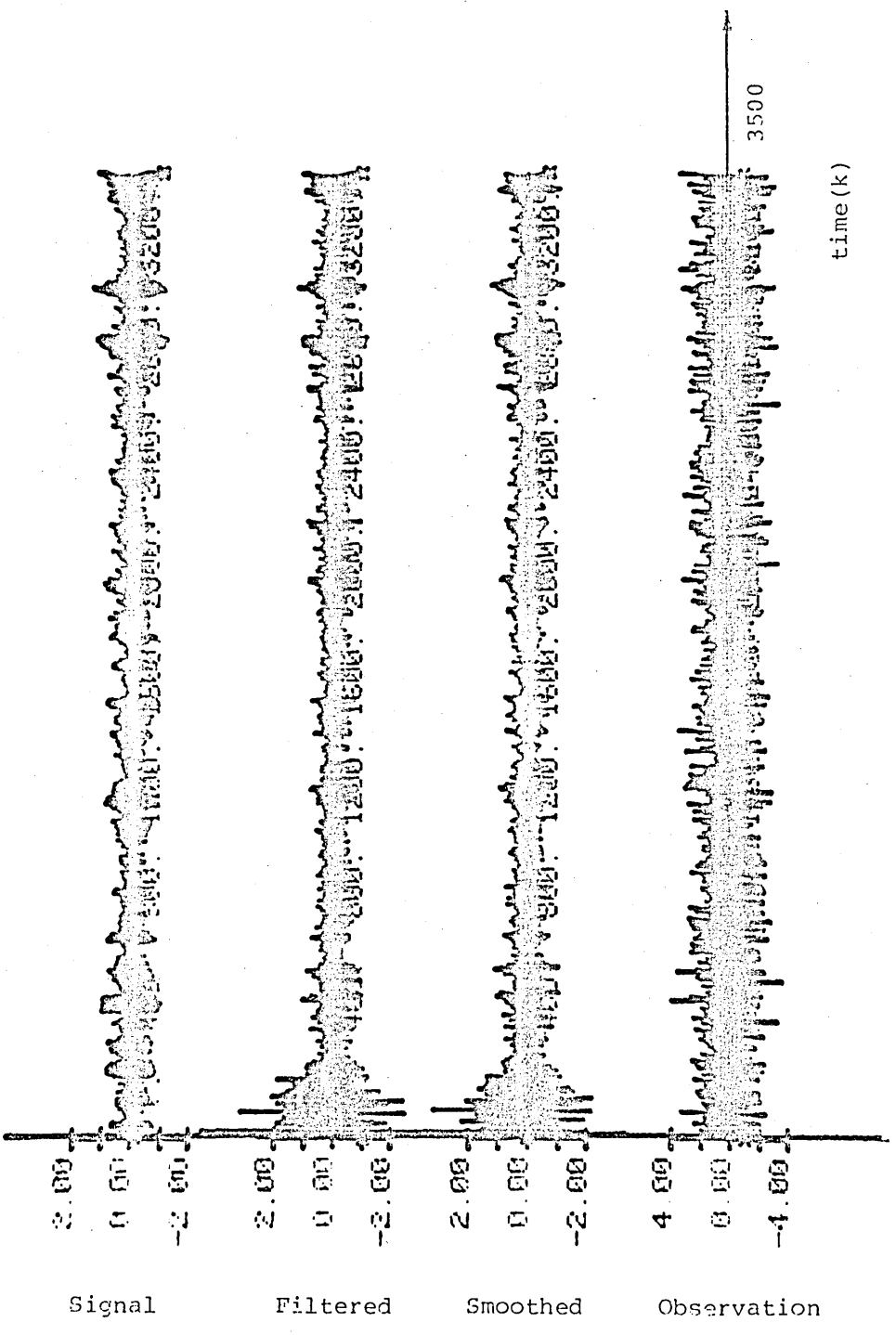


Figure 3.25 Signal, Estimates and Observation for Channel 1

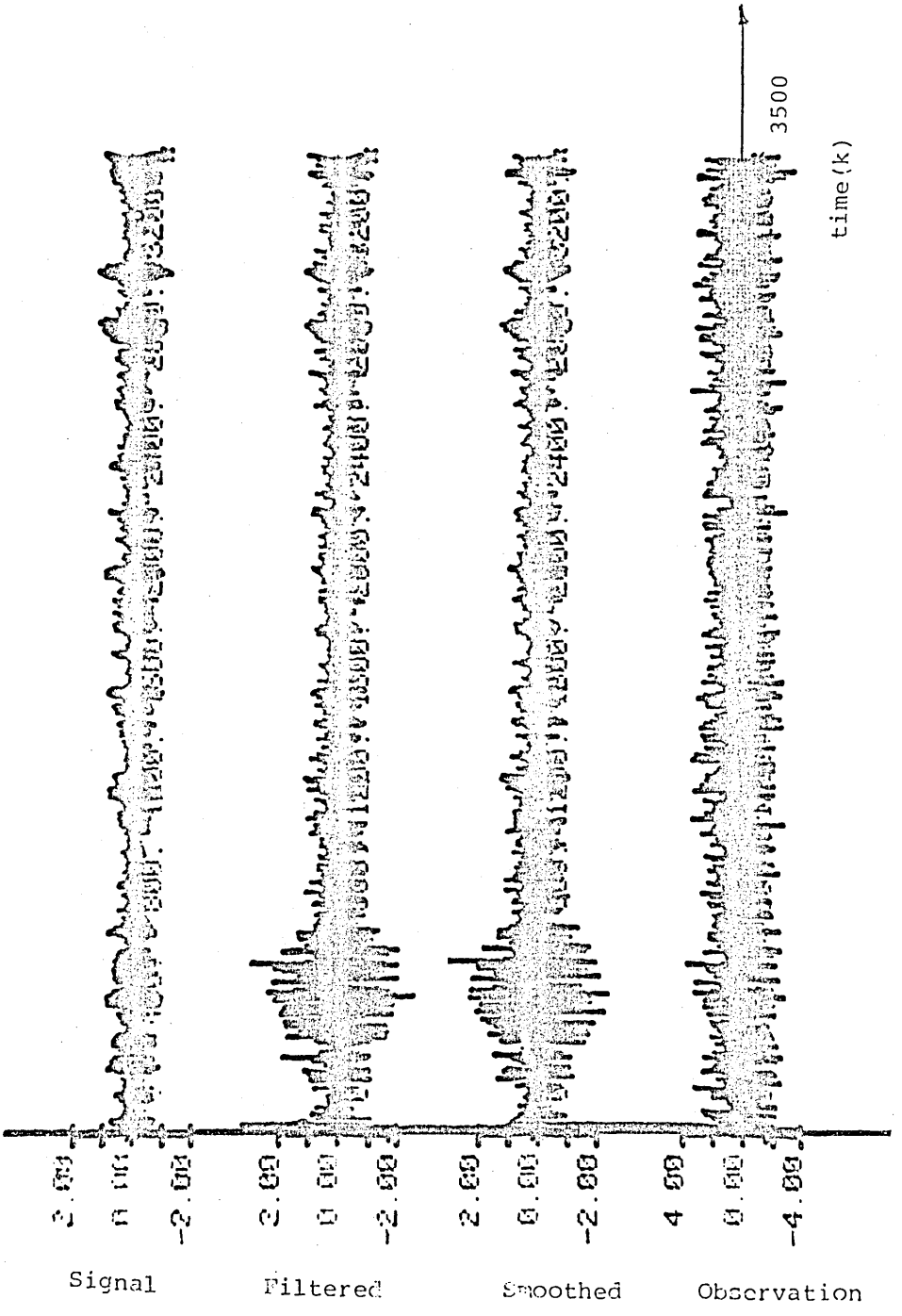


Figure 3.26 Signal,Estimates and Observation for Channel 2

Assume A_x has rank $p > 0$, then it is always possible to find matrices C_x and D_x such that

$$A_x = C_x D_x \tag{3.71}$$

where C_x is a $r^2 \times p$ matrix and D_x is a $p \times r$ matrix. Equation (3.69) then has solution (Barnett [101]).

$$\underline{\theta}_m = A_x^{\dagger} \underline{J}_x \tag{3.72}$$

where $A_x^{\dagger}$ is the generalised inverse of A_x given by

$$A_x^{\dagger} = D_x^T (D_x D_x^T)^{-1} (C_x^T C_x)^{-1} C_x^T \tag{3.73}$$

3.7 State Estimation

This section will briefly describe the possibilities of obtaining self-tuning filtered state estimates rather than self-tuning filtered signal estimates. The advantages in using state estimation lies mainly in optimal control or pole placement. For the single input/output case Tsay and Shieh [53] employ a type of self-tuning steady-state Kalman filter for pole-placement self-tuning control. A similar technique is used by Lam [55] for LQG optimal control. The approach used here is essentially that of Jakeman and Young [102].

From the state space description of the process (3.1), (3.2), (3.3) write the equivalent state-space innovations form as [103]:

$$\hat{\underline{y}}(k|k-1) = C\hat{\underline{x}}(k|k-1) \tag{3.74}$$

$$\hat{\underline{x}}(k+1|k) = A\hat{\underline{x}}(k|k-1) + K_F \underline{\varepsilon}(k) \tag{3.75}$$

without loss of generality the following canonical form for $\{A, K_F, C\}$ is found from equation (3.18)

$$A = \begin{bmatrix} -A_1 & I & 0 & 0 \\ -A_2 & 0 & I & 0 \\ \vdots & & & \\ & & & I \\ -A_{n_a} & 0 & 0 & 0 \end{bmatrix}$$

$$K_F = \begin{bmatrix} K_1 & K_2 & \cdots & K_{n_a} \end{bmatrix}^T$$

$$C = [I \; 0 \; \cdots \; 0]$$

where $K_i = D_i - A_i$, $i = 1, 2 \ldots, n_a$.

If A_i, D_i, $i = 1, 2 \ldots, n_a$ are estimated using Algorithm 3.2 together with $\underline{\varepsilon}(k)$ then the states may be reconstructed via equation (3.75). The signal estimate may be found using (3.74) but this is not the same estimate as would be obtained by using the self-tuning filter equation (3.52). This is because (3.74) computes the *a priori* estimate and (3.52) computes an *a posteriori* estimate. Equation (3.74) would be obtained using a one step ahead predictor (Tanttu [31]) and is equivalent to the work of Shaked (Appendix 8).

To obtain the *a posteriori* estimate of the state one must first compute [103]:

$$\hat{\underline{x}}(k|k) = \hat{\underline{x}}(k|k-1) + K\varepsilon(k) \tag{3.76}$$

where K is as equation (3.15). K is related to K_F by the relationship

$$AK = K_F \tag{3.77}$$

giving a filtered signal estimate from the state space as

$$\hat{\underline{y}}\,(k|k) = C\hat{\underline{x}}(k|k) \tag{3.78}$$

Equation (3.78) now gives the same estimate as the self-tuning filter equation (3.52). Equation (3.52) may be written as

$$\hat{\underline{y}}\,(k|k) = \underline{z}(k) - R\,R_{\varepsilon}^{-1}\underline{\varepsilon}(k) \tag{3.79}$$

The relationship between (3.79)(the self-tuning filter), (3.78) and (3.76) (the self-tuning state space filter), may be shown more clearly in the following.

Employing the state-space equations first :

From equation (3.76) and using (3.78)

$$\hat{\underline{y}}\,(k|k) = \hat{\underline{y}}\,(k|k-1) + CK\underline{\varepsilon}(k) \tag{3.80}$$

substituting (3.15) for K gives

$$\hat{\underline{y}}\,(k|k) = \hat{\underline{y}}\,(k|k-1) + CPC^{T}R_{\varepsilon}^{-1}\underline{\varepsilon}(k) \tag{3.81}$$

Now $R_{\varepsilon} = R + CPC^{T} \Rightarrow CPC^{T} = R_{\varepsilon} - R$

hence

$$\hat{\underline{y}}\,(k|k) = \hat{\underline{y}}\,(k|k-1) + \underline{\varepsilon}(k) - R\,R_{\varepsilon}^{-1}\underline{\varepsilon}(k) \tag{3.82}$$

From (3.29) with $\ell=0$

$$\underline{z}(k) = \hat{\underline{y}}\,(k|k-1) + \underline{\varepsilon}(k) \tag{3.83}$$

substituting (3.83) into (3.82) gives

$$\hat{\underline{y}}\,(k|k) = \underline{z}(k) - R\,R_{\varepsilon}^{-1}\underline{\varepsilon}(k)$$

which is (3.79), the equation for the self-tuning filter. Chen [76] has shown that *a posteriori* estimates give a lower mean squared error than *a priori* estimates and therefore the use of *a posteriori* estimates is justifiable.

3.8 Summary

A self-tuning fixed-lag smoother or filter for linear multivariable processes has been developed. A polynomial matrix solution to the fixed-lag smoothing problem was first presented. This solution involved the spectral factorisation of a polynomial matrix and assumed the process and noise statistics were known. The self-tuning smoother and filter assumes no such knowledge and the unknown parameters were estimated from an innovations signal model. These parameters are utilised directly in the smoother or filter and in so doing the process of spectral factorisation is avoided. In the special case of a single input/output process, the proposed smoothing algorithm has been shown to be equivalent to the work of Hagander and Wittenmark.

The convergence of the proposed algorithm has been proven using the asymptotic stochastic convergence theory of Ljung. The analysis could have been extended to employ the recent work of Solo [87] and this is an interesting topic for future research.

Examples have shown the successful operation of the self-tuning filters and smoothers. One simple example illustrates how the algorithm copes with slowly time varying process parameters.

The optimal multivariable predictor was derived by similar arguments to the above. However, the results were shown to be equivalent to those presented by Wittenmark [30] and Holst [82] in the scalar case and

Tanttu [31] in the multivariable case. Thus, this result provided some unification of different approaches but it was unnecessary to develop a self-tuning prediction algorithm since it had already been discussed by these authors.

CHAPTER 4

LQG Self-Tuning Regulators (STR)

4.1 Introduction

Linear Quadratic Gaussian regulators and controllers are often thought as being more robust than those which employ k-step ahead cost functions. The former can deal easily with non-minimum phase and unstable systems. The latter has problems when dealing with non-minimum phase systems (Astrom [34]) and even with the addition of control weighting (Clarke and Hasting-James [39]) has difficulty in dealing with unstable systems. The same conclusions may be drawn for self-tuning regulators and controllers. The widely used generalised minimum variance STR (Clarke and Gawthrop [48]) has to have an additional cost-function polynomial adjusted to cope with unstable systems and the non-minimum phase problems.

The state-space approach to LQG regulators employs a Kalman filter and a feedback gain matrix (Kwakernaak and Sivan [99].) This idea has been used by Lam [55] to give the first LQG STR. The STR is able to cope with non-minimum phase and unstable systems with sudden changes of time delay. The STR employed in this chapter is LQG but is considered from an input-output rather than a state-space point of view. The input-output LQG controller uses polynomial equations and has previously been developed by Peterka [40] and Kucera [45]. Two types of STR are developed here. The explicit STR requires the solution of one diophantine equation and one polynomial spectral factorisation. The implicit STR requires only one spectral

factorisation and no diophantine equations. By explicit, it is meant that the controller parameters are calculated from the diophantine equation whereas the implicit method estimates them directly.

4.2 System Description

Consider the single-input/output process and disturbance given by:

$$y(t) = \frac{B(z^{-1})}{A(z^{-1})} u(t) + \frac{C(z^{-1})}{A(z^{-1})} \omega(t) \qquad (4.1)$$

with a measured output given as:

$$z(t) = y(t) + v(t) \qquad (4.2)$$

The disturbance and measurement noise sequence $\omega(t)$ and $v(t)$ are stationary, have zero mean with variances Q and R respectively and are independent.

The polynomials $A(z^{-1}), B(z^{-1})$ and $C(z^{-1})$ are given as:

$$A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_{n_a} z^{-n_a} \qquad (4.3)$$

$$C(z^{-1}) = 1 + c_1 z^{-1} + c_2 z^{-2} + \ldots + c_{n_c} z^{-n_c} \qquad (4.4)$$

$$B(z^{-1}) = z^{-k}(b_0 + b_1 z^{-1} + \ldots + b_{n_{b-k}} z^{-(n_b-k)}) \qquad (4.5)$$

and n_a, n_b and n_c are assumed to be known and it is assumed that the A and B polynomials are relatively prime.

The process can be unstable, non-minimum phase and includes a time delay k_1 of magnitude one or greater. If the process is non-minimum phase the polynomial B is factored into $B = B_1 B_2$ where $B_1(z^{-1})$ has all its zeros inside or on the unit circle. B_2 includes both the time-delay and non-minimum phase terms and has the form

$$B_2(z^{-1}) = z^{-k}(b_{20} + b_{21} z^{-1} + b_{22} z^{-2} + \ldots + b_{2_{(n_{b2}-k)}} z^{-(n_{b_2}-k)}) \qquad (4.6)$$

The order n_{b_2} of the polynomial B_2 includes the delay k_1. If the delay k_1 is known then b_{20} is defined to be unity and $k \triangleq k_1$. If the delay k_1 is unknown ($k_1 < n_b$) then k must be set to unity in B and B_2 which is the minimum for physical realizability. Then if $k_1 > 1$ the coefficients

$b_0, b_1, \ldots, b_{k_1-2}$ and $b_{20}, b_{21}, \ldots, b_{2(k-2)}$ are zero and the polynomial B_2 can be normalized so that $b_{2(k_1-1)} = 1$.

Two generalised spectral factors introduced by Shaked [27, 43] are required for the analysis. They are:

$$\Delta(z)\Delta(z^{-1}) = \frac{C(z^{-1})C(z)}{A(z^{-1})A(z)} Q + R \tag{4.7}$$

$$= \frac{D(z^{-1})D(z)}{A(z^{-1})A(z)} \tag{4.8}$$

$$\Delta_1(z)\Delta_1(z^{-1}) = \frac{B(z^{-1})B(z)}{A(z^{-1})A(z)} Q_1 + R_1 \tag{4.9}$$

$$= \frac{D_1(z^{-1})D_1(z)}{A_1(z^{-1})A_1(z)} \tag{4.10}$$

Hence

$$\Delta(z^{-1}) = \frac{D(z^{-1})}{A(z^{-1})} \tag{4.11}$$

and

$$\Delta_1(z^{-1}) = \frac{D_1(z^{-1})}{A(z^{-1})} \tag{4.12}$$

The polynomials D and D_1 are Hurwitz [45] and are given as:

$$D(z^{-1}) = 1 + d_1z^{-1} + d_2z^{-2} + \ldots + d_{n_d}z^{-n_d} \tag{4.13}$$

$$D_1(z^{-1}) = 1 + d_{11}z^{-1} + d_{12}z^{-2} + \ldots + d_{1n_{d_1}}z^{-n_{d_1}} \tag{4.14}$$

4.3 LQG Regulator

This section will consider the case when the plant is assumed to be known. The optimal LQG regulator will then be given in polynomial equation form.

The LQG regulator is chosen to minimize the cost function:

$$J = E\{Q_1 e^2(t) + R_1 u^2(t)\} \tag{4.15}$$

where E is the unconditional expectation operator. The optimal LQG regulator is defined by the following:

Theorem 4.1 Optimal LQG Regulator

For the process (4.1), (4.2), the LQG optimal regulator has transfer function

$$C_0(z^{-1}) = \frac{G_0(z^{-1})}{H_0(z^{-1})} \tag{4.16}$$

where the polynomials G_0 and H_0 are given by the unique solution of the diophantine equation:

$$A(z^{-1})H_0(z^{-1}) + B(z^{-1})G_0(z^{-1}) = D_1(z^{-1})D(z^{-1}) \tag{4.17}$$

with

$$n_{g_0} = n_a - 1, \quad n_{h_0} = n_b - 1$$

and

$$H_0(z^{-1}) = 1 + h_1 z^{-1} + \ldots + h_{n_{h_0}} z^{-n_{h_0}} \tag{4.18}$$

$$G_0(z^{-1}) = g_0 + g_1 z^{-1} + \ldots + g_{n_{g_0}} z^{-n_{g_0}} \tag{4.19}$$

Proof : The proof is presented in Kučera [45] and Grimble [44]. A block diagram illustrating the LQG regulator is given in Figure 4.1.

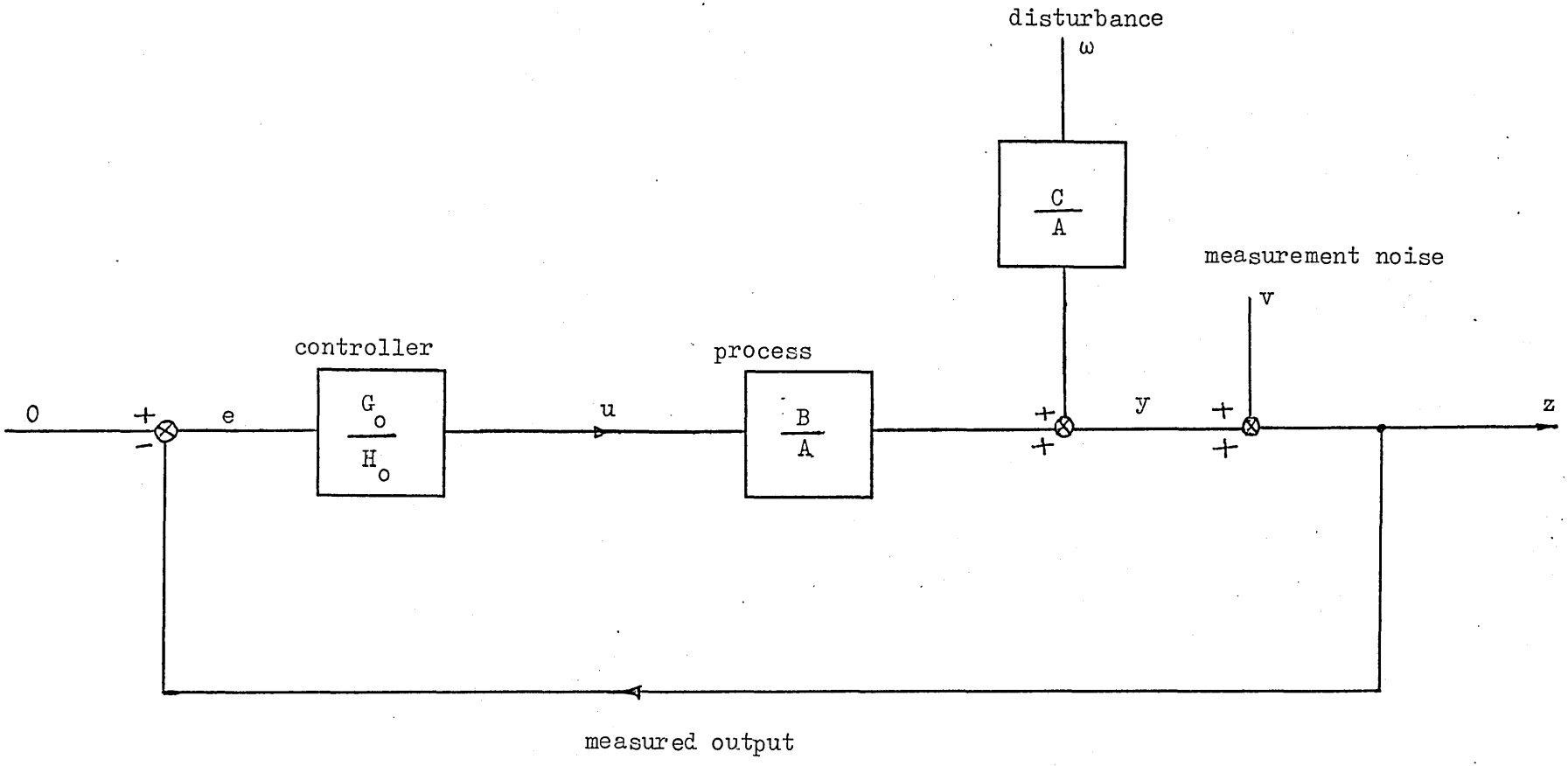


Figure 4.1 Structure of LQG Regulator

4.4 Self-Tuning LQG Regulators

To develop the self-tuning LQG regulator the innovations form of the process (4.1), (4.2) is written as:

$$e(t) = \frac{D(z^{-1})}{A(z^{-1})}\varepsilon(t) - \frac{B(z^{-1})}{A(z^{-1})}u(t) \qquad (4.20)$$

with the error $e(t) = -z(t)$.

4.4.1 Explicit LQG STR

An explicit (controller parameters calculated) LQG STR algorithm now follows based on theorem 4.1.

Algorithm 4.1 Explicit STR [64,65]

DATA : Choose the cost function weighting Q_1 and R_1.

Step 1 : Using extended recursive least squares estimate the parameters in the polynomials A, B and D from

$$A(z^{-1})e(t) = D(z^{-1})\varepsilon(t) - B(z^{-1})u(t)$$

Step 2 : Calculate the Hurwitz spectral factor D_1 using

$$D_1(z^{-1})D_1(z) = B(z^{-1})B(z)Q_1 + A(z^{-1})A(z)R_1$$

Step 3 : Calculate G_0 and H_0 from

$$A(z^{-1})H_0(z^{-1}) + B(z^{-1})G_0(z^{-1}) = D_1(z^{-1})D(z^{-1})$$

Step 4 : Compute the control signal

$$u(t) = \frac{G_0}{H_0}(z^{-1})e(t)$$

Return to Step 1.

Grimble [64] has termed the above algorithm Fixed Criterion since Q_1 and R_1 are assumed to be fixed. The term Fixed Spectrum has also arisen assuming that Step 2 is omitted and D_1 is chosen apriori, hence the STR then

becomes similar to the pole placement algorithms of Wellstead et al [50,51].

Assuming that the process parameters converge to their true values (this assumption may be made in all but a few counter-examples if using extended least squares, Ljung et al [100]) then the closed loop form of the regulator may be found by substitution of the control

$$u(t) = \frac{G_0(z^{-1})}{H_0(z^{-1})} e(t) \tag{4.21}$$

into the innovations form of the process (4.20), giving

$$e(t) = \frac{D(z^{-1})}{A(z^{-1})} \varepsilon(t) - \frac{B(z^{-1})}{A(z^{-1})} \frac{G_0(z^{-1})}{H_0(z^{-1})} e(t) \tag{4.22}$$

or

$$e(t)\{A(z^{-1})H_0(z^{-1}) + B(z^{-1})G_0(z^{-1})\} = D(z^{-1})H_0(z^{-1})\varepsilon(t) \tag{4.23}$$

by using the diophantine equation (4.17):

$$e(t) = \frac{H_0(z^{-1})}{D_1(z^{-1})} \varepsilon(t) \tag{4.24}$$

Equation (4.24) shows how the closed-loop poles are given by the polynomial $D_1(z^{-1})$. The STR is thus guaranteed to be closed-loop stable. It also follows that the steady-state variance of the regulator error signal is given by the complex integral

$$\sigma_e^2 = \sigma_\varepsilon^2 \frac{1}{2\pi i} \oint \frac{H_0(z)H_0(z^{-1})}{D_1(z)D_1(z^{-1})} \frac{dz}{z} \tag{4.25}$$

where the path of integration is around the unit circle. The above integral is easily computed by following the algorithm in Astrom [34].

The operation of the explicit LQG STR may be illustrated using the following examples.

Example 4.1

Consider the output regulator problem given in Grimble et al [65] and Tsay and Shieh [53]

$$y(t) = \frac{z^{-2}(1 - 2z^{-2})}{1 - z^{-1} + 2z^{-2}}\, u(t)$$

$$+ \left(\frac{1 - 0.5z^{-1}}{1 - z^{-1} + 2z^{-2}}\right) \omega(t)$$

For this example, there is no measurement noise and $\omega(t)$ is a zero-mean white noise process with variance $Q = 0.25$. The process is both unstable and non-minimum phase. Therefore a generalised minimum variance self-tuning regulator (Clarke and Gawthrop [48]) will work only with great difficulty on this process [65]. The reasoning behind this is as follows. Since the process is non-minimum phase the generalised minimum variance (GMV) self-tuning regulators control weighting may be made sufficiently large to stabilize the process. However, this action is similar to using open-loop control which is unstable for this process. Therefore using GMV self-tuning control on this process is difficult, requiring critical weighting on control for stability. Although a region of stability may be found using GMV self-tuning, it is unlikely that the corresponding values of control weighting would be known apriori.

The LQG STR of this chapter suffers from none of these difficulties. Selecting the performance criterion weightings as $Q_1 = 1$, $R_1 = 0.0625$ gives a D_1 polynomial as

$$D_1(z^{-1}) = 1 - 0.49132z^{-1} + 0.0289z^{-2}$$

which has poles at $z_1 = 0.423$, $z_2 = 0.0689$ resulting in a

stable closed loop process. Assuming the time delay ($k = 2$) is known, the G_0 and H_0 polynomials may be found from (4.17) to be

$$G_0(z^{-1}) = g_0 + g_1 z^{-1}$$

and

$$H_0(z^{-1}) = 1 + h_1 z^{-1} + h_2 z^{-2}$$

with

$$g_0 = 0.0159,\ g_1 = -1.7326 = h_2,\ h_1 = 0.00868$$

The fixed parameter or optimal regulator may be seen in Figure 4.2. This shows the control signal and output. A comparison may be made with the explicit LQG STR of Figure 4.3 which is almost identical with the exception of the erratic values for the first few steps. The process parameter estimates are shown in Figure 4.4, the true values shown with broken lines.

Finally, the calculated controller parameters are shown in Figure 4.5 and are close to the true values. The conclusion may be made that the LQG STR performs as well as the optimal LQG regulator for this example, after parameter convergence has taken place.

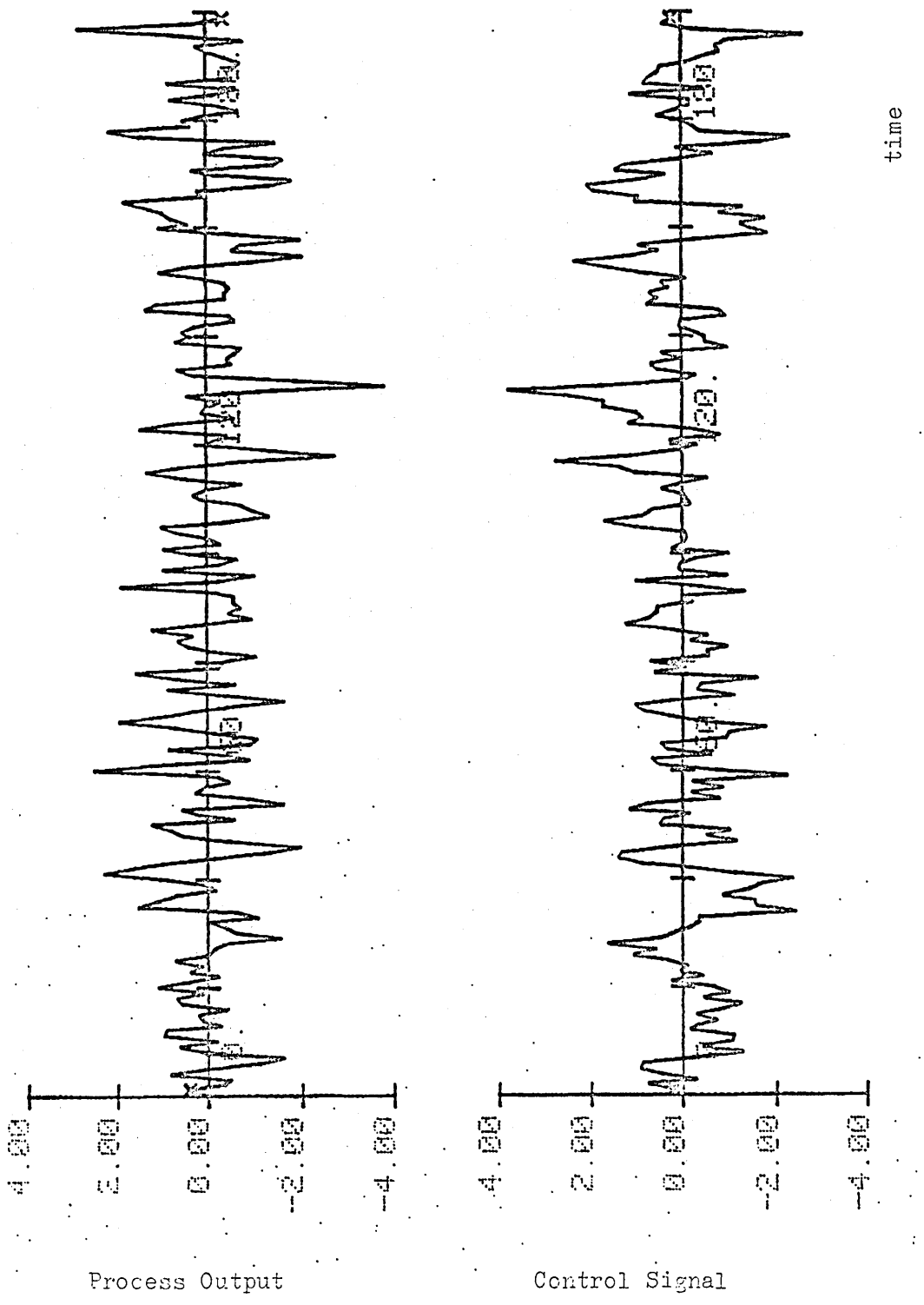


Figure 4.2 Fixed parameter LQG Regulator for example 4.1

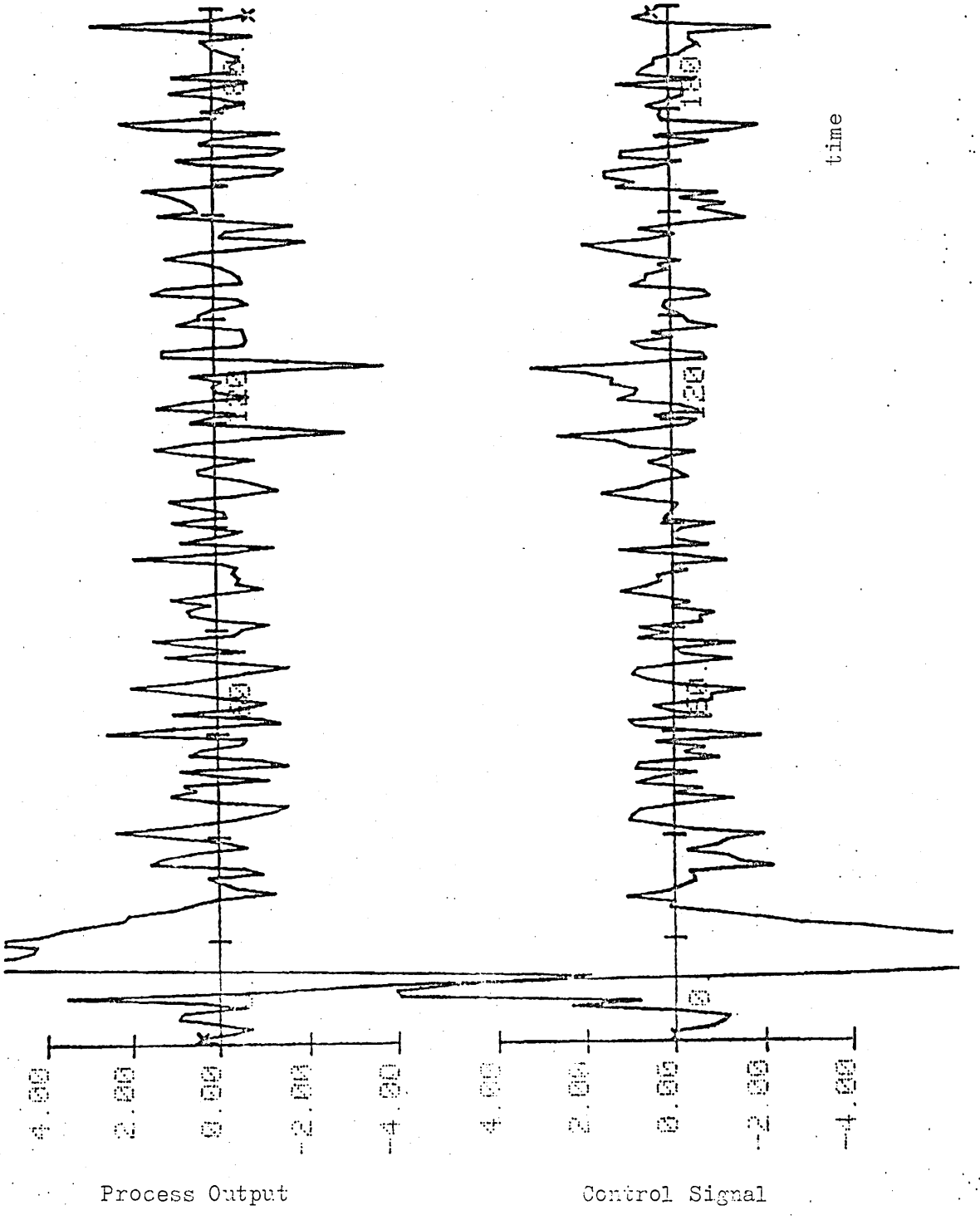


Figure 4.3 Explicit LQG STR for example 4.1

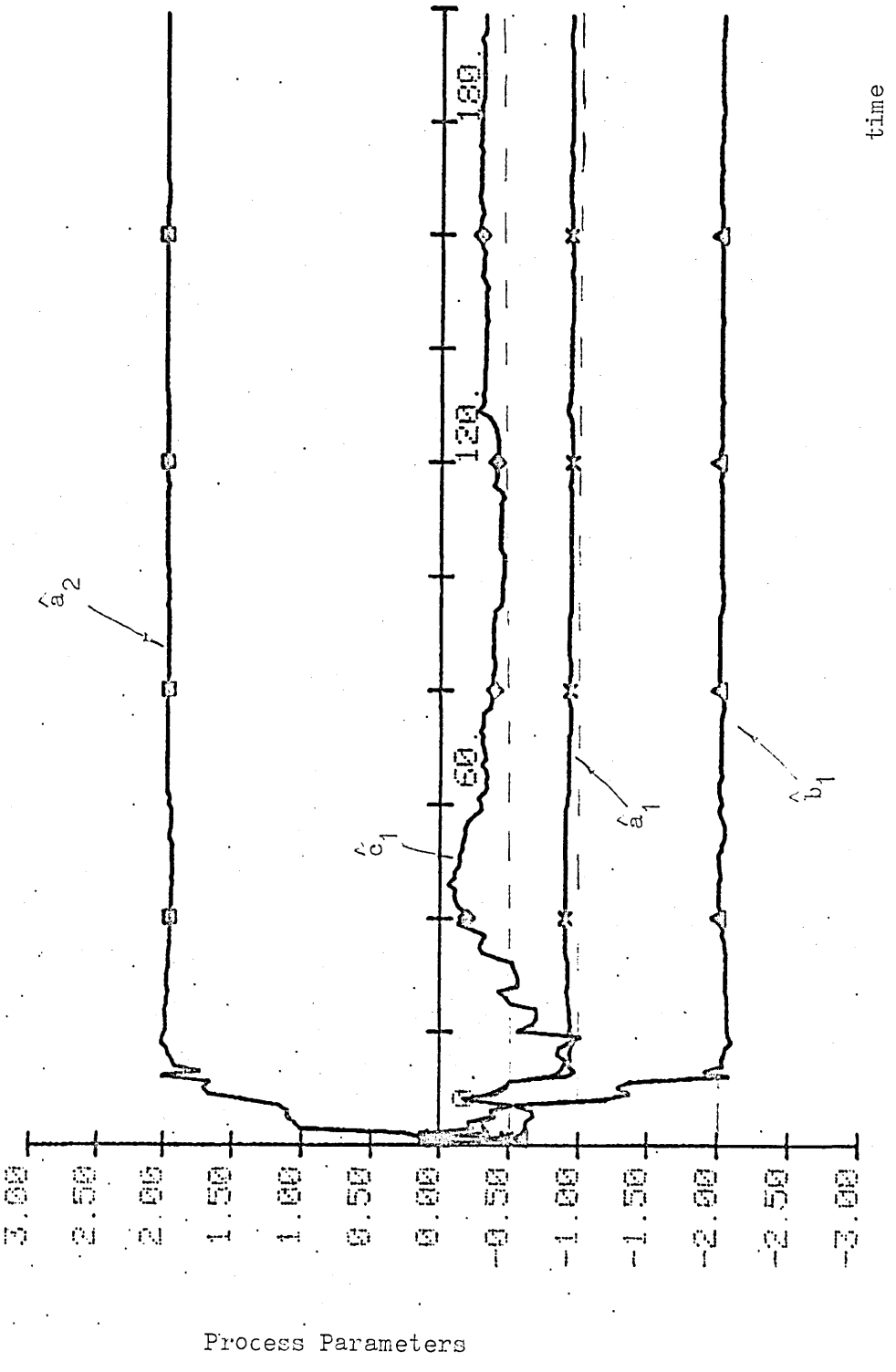


Figure 4.4 Estimates of process parameters for example 4.1

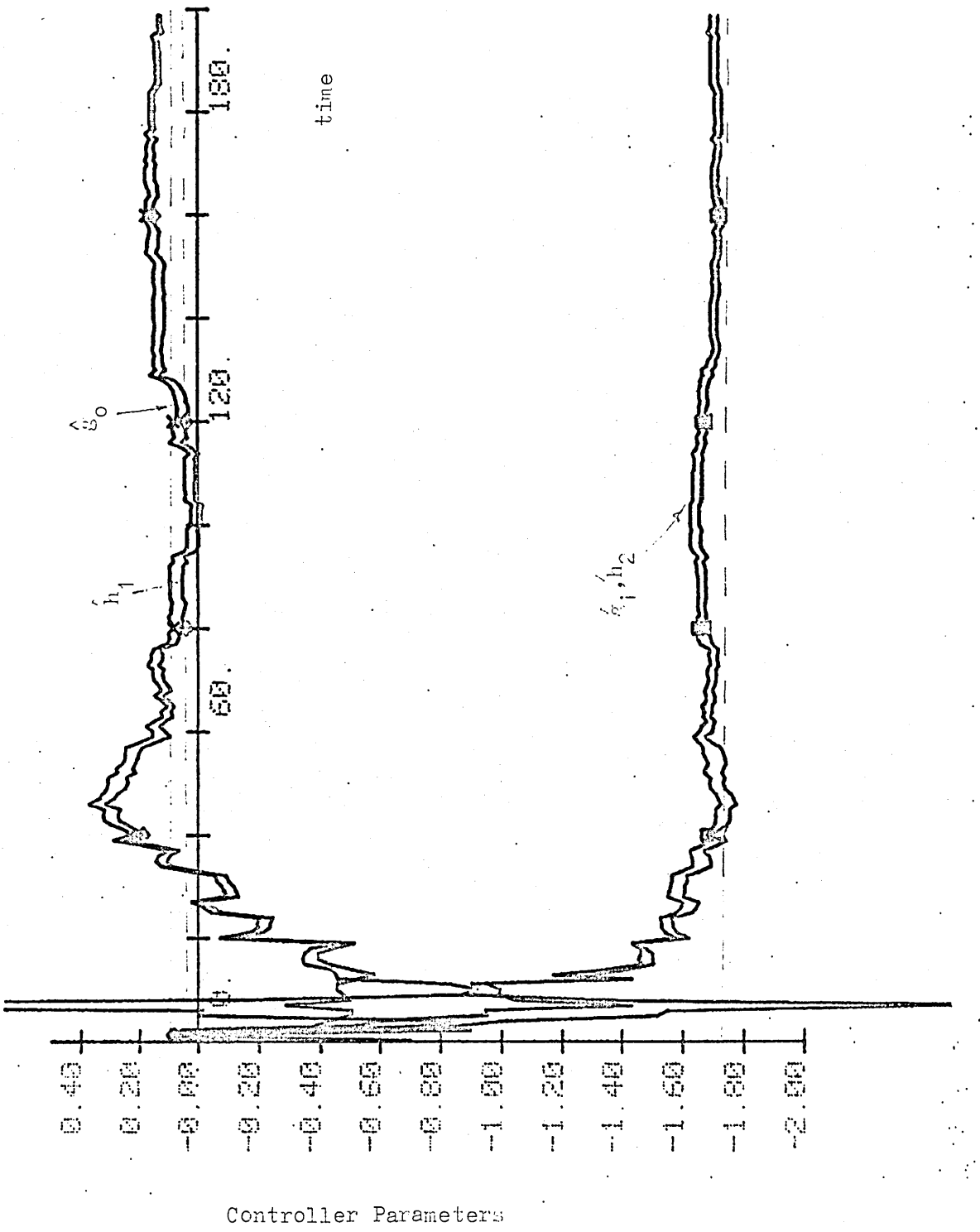


Figure 4.5 Convergence of controller parameters for example 4.1

Example 4.2

This example examines the problem of an unknown and variable time delay. It is shown how the explicit LQG STR can cope with this problem.

The process is given as

$$y(t) = \frac{z^{-k}(1 - 2z^{-1})}{1 - z^{-1} + 2z^{-2}} u(t)$$

$$+ \left(\frac{1 - 0.5z^{-1}}{1 - z^{-1} + 2z^{-2}}\right)\omega(t)$$

With $D_1(z^{-1}) = 1 - z^{-1} + 0.5z^{-2}$ and

where $\omega(t)$ is a white noise sequence of variance 0.25 and, as before, R = 0. The time delay k is assumed to be unknown and will be varied from a minimum of k = 1 up to k = 3. Table 4.1 shows the various values of controller parameters and process parameters for these values. The following model for the process was used for simulation

$$y(t) = \frac{z^{-1}(b_0 + b_1z^{-1} + b_2z^{-2} + b_3z^{-3})}{1 - z^{-1} + 2z^{-2}} u(t)$$

$$+ \left(\frac{1 - 0.5z^{-1}}{1 - z^{-1} + 2z^{-2}}\right)\omega(t)$$

k	b_0	b_1	b_2	b_3	g_0	g_1	h_1	h_2	h_3	t
1	1	-2	0	0	+.5625	-.9375	-1.06	0	0	0-299
2	0	1	-2	0	-.375	-1.125	-0.5	-1.125	0	300-599
3	0	0	1	-2	-1.5	0.75	-0.5	-1.5	0.75	600-900

Table 4.1

The assumption is made of a one step delay for physical realizability of the controller. The actual time delay is varied according to Table 4.1 above and shows the

corresponding true values of the process model and controller parameters. The process parameter estimates are given in Figures 4.6 and 4.7. The spikes occurring in Figure 4.6 correspond to the sudden change in time delay. Figure 4.7 shows the changing $B(z^{-1})$ parameters and are easily verified to be consistent with the true values in Table 4.1 In this example, the parameter convergence is very rapid since the process is open loop unstable giving an added inherent perturbation to the control loop. For open loop stable processes, convergence will be much slower and may require a 'forgetting factor' of less than unity if the time delay changes, to aid adaptation. The computed LQG controller parameters are shown in Figure 4.8. These too, are close to the true values in Table 4.1. The process output and control signal for the explicit LQG STR are shown in Figure 4.9. Parameter uncertainty during the sudden time delay changes, results in an erratic output and control signal for these times. This is unavoidable and can only be remedied by using a very low initial parameter estimate covariance in the recursive least squares routine which leads to very slow parameter convergence and thus slow adaptation to variation in time delay.

4.4.2 Implicit LQG STR

A control algorithm combined with process identification may be classified as self-tuning control. The explicit LQG STR is one such example. Because of the computational burdens of such a philosophy (matrix inversions etc.), an implicit LQG STR will be derived in which the controller

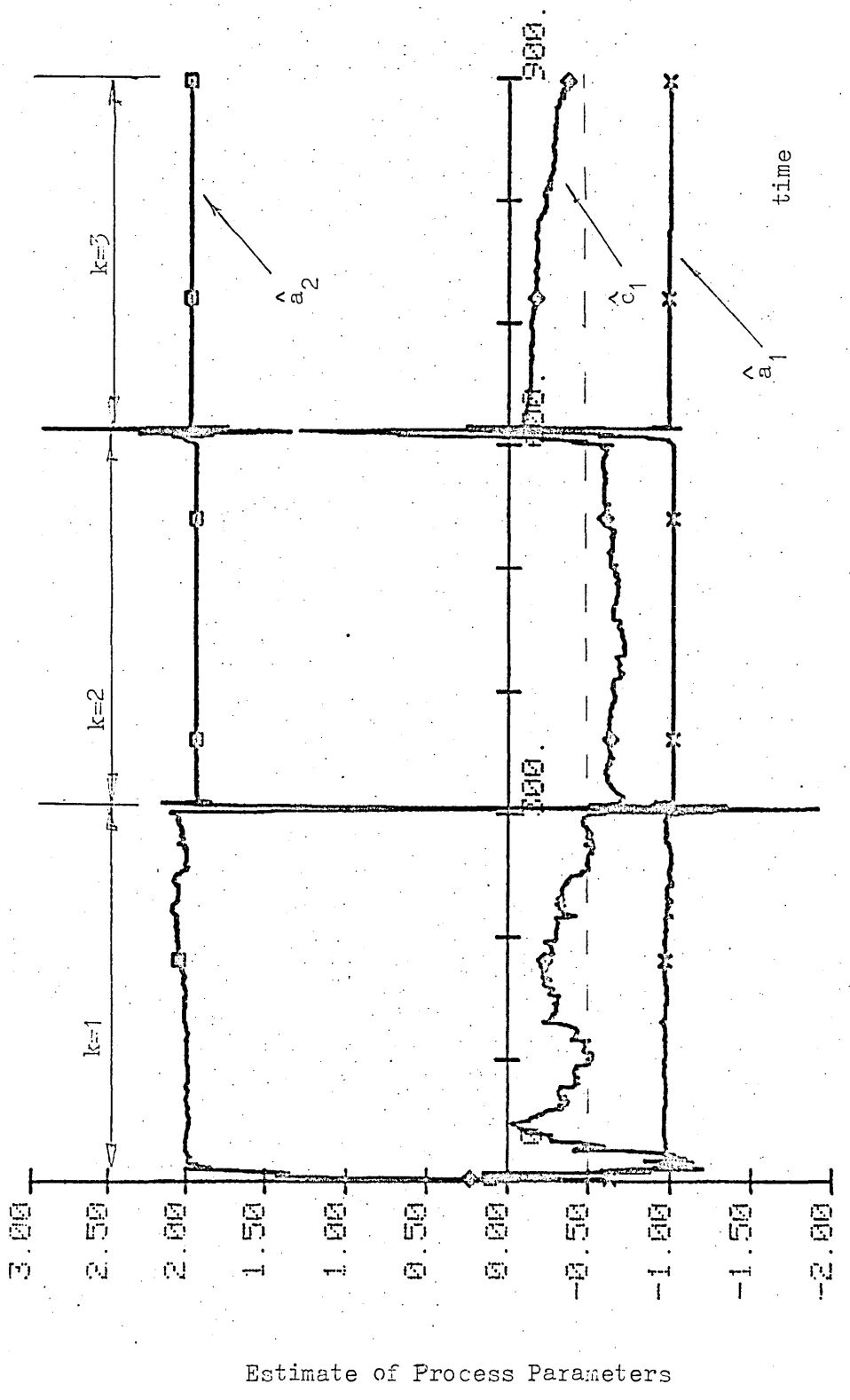


Figure 4.6 Process parameter estimates for example 4.2

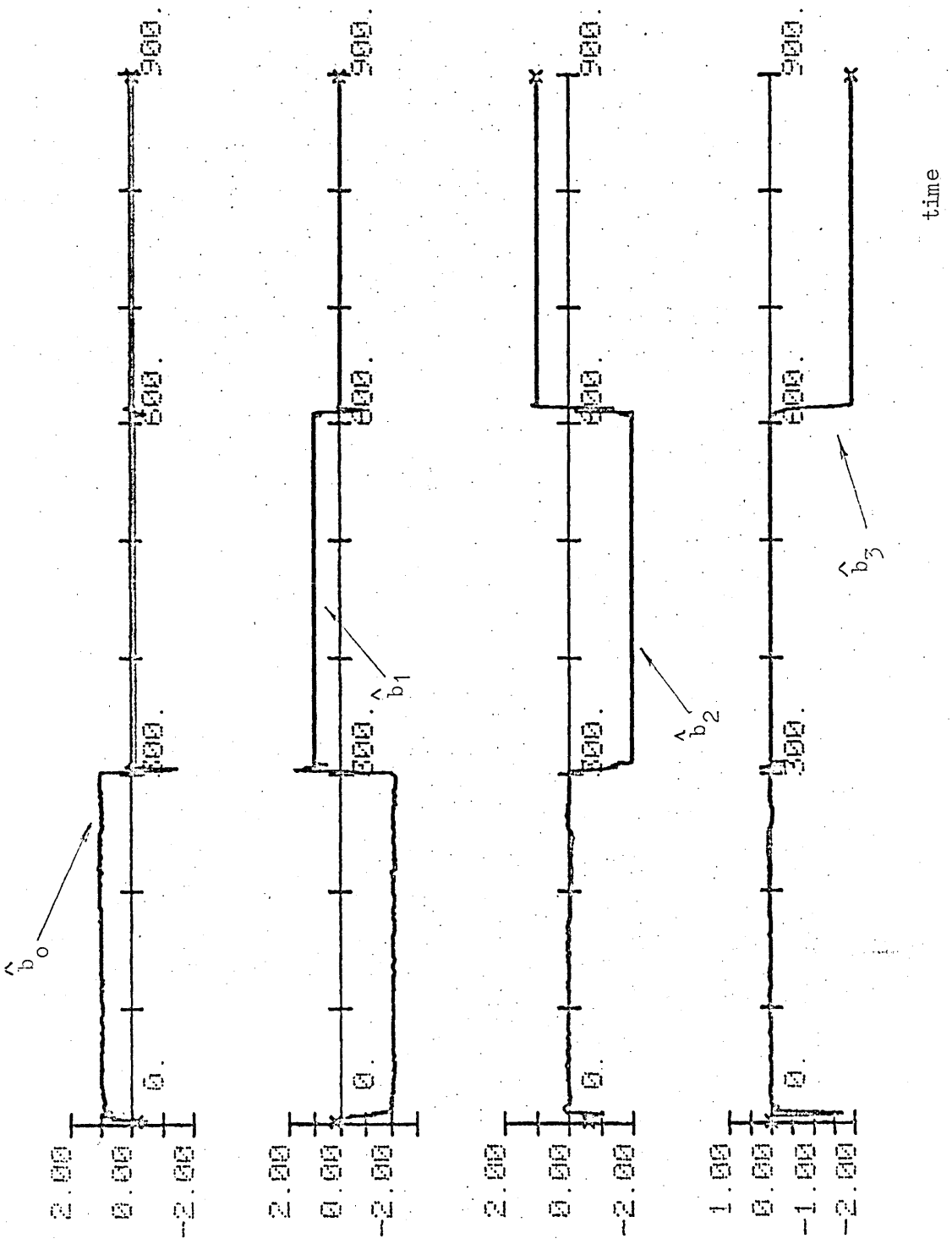


Figure 4.7 Estimate of B polynomial for example 4.2

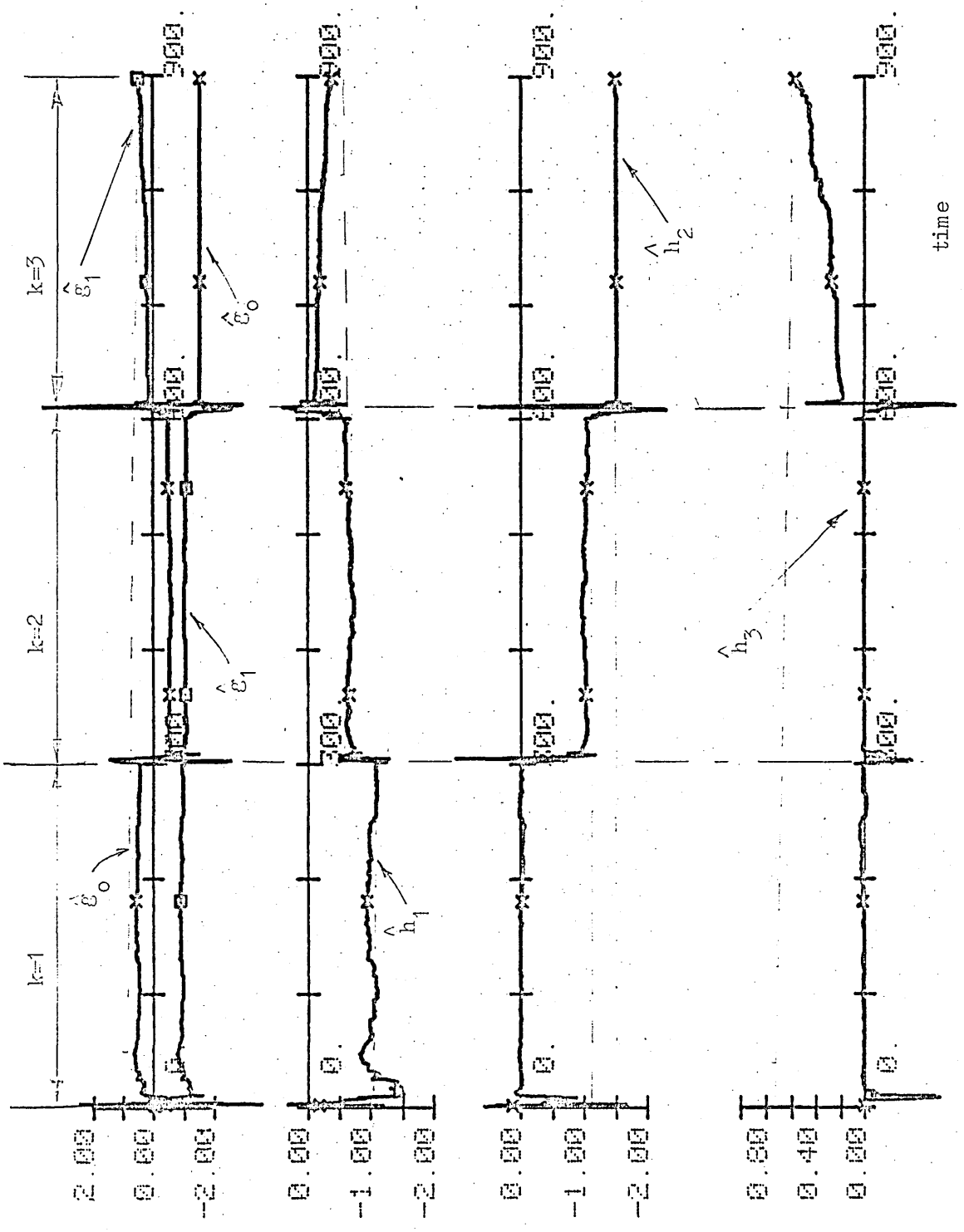


Figure 4.8 Controller parameters for example 4.2

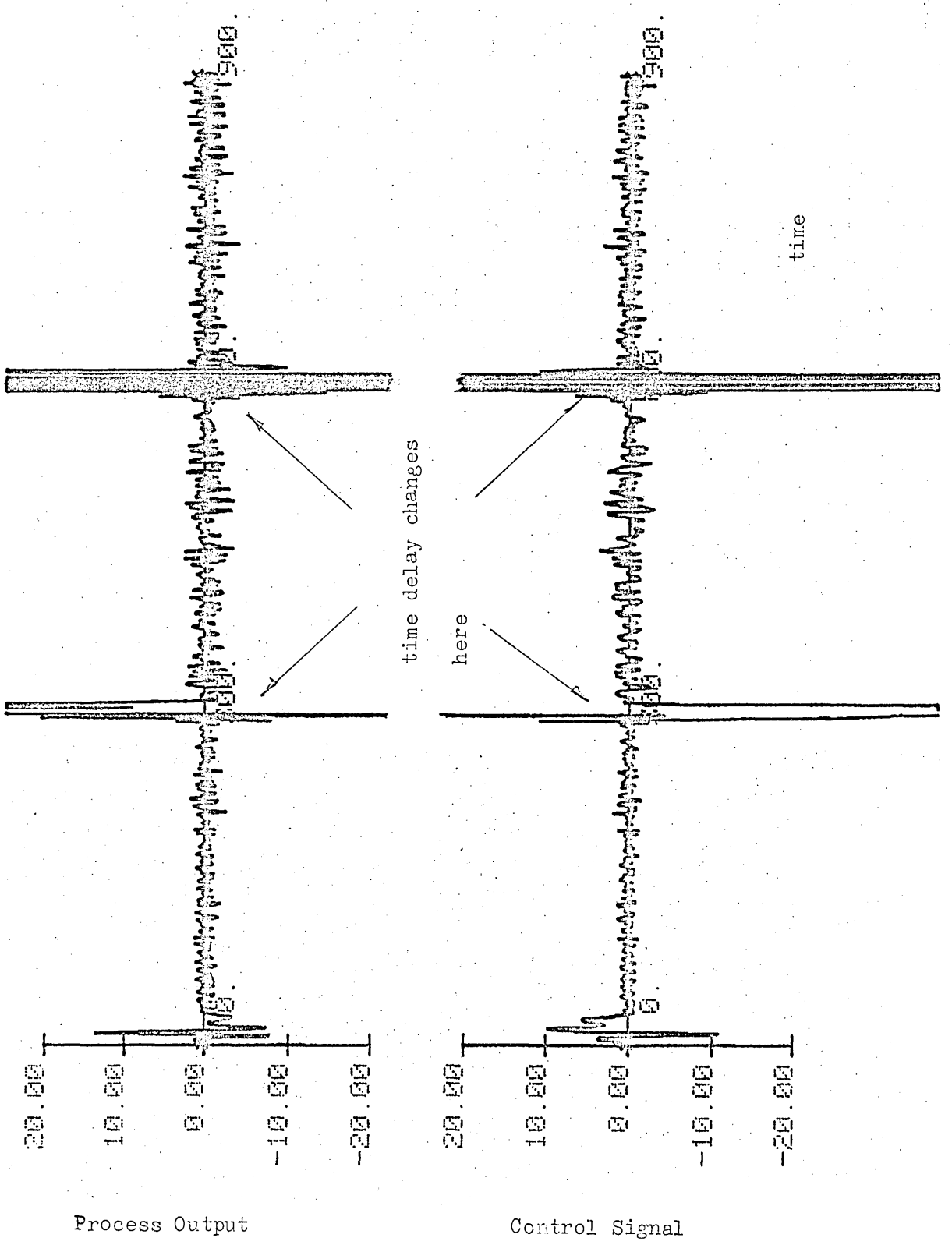


Figure 4.9 Process output & Control signal for STR with variable time delay

parameters may be estimated directly using recursive least squares, avoiding the solution of simultaneous equations. In what follows, the arguments of the polynomial equations will be dropped for convenience.

Theorem 4.2 Implicit Model

The implicit model from which G_0 and H_0 may be estimated is given as

$$\phi(t) = H_0\varepsilon(t) + \frac{B}{D}(G_0e(t) - H_0u(t))$$

where

$$\phi(t) = D_1e(t)$$

Proof:

From the innovations form of the process write

$$e(t) = \frac{D}{A}\varepsilon(t) - \frac{B}{A}u(t) \tag{4.26}$$

multiplying (4.26) by the A polynomial and substituting for D from the diophantine equation (4.17) gives

$$Ae(t) = \frac{(AH_0 + BG_0)}{D_1}\varepsilon(t) - Bu \tag{4.27}$$

or

$$D_1e(t) = \left(H_0 + \frac{BG_0}{A}\right)\varepsilon(t) - \frac{D_1Bu(t)}{A} \tag{4.28}$$

now let $\phi(t) \triangleq D_1e(t)$ so that

$$\phi(t) = H_0\varepsilon(t) + \frac{BG_0}{A}\varepsilon(t) - \frac{D_1Bu(t)}{A} \tag{4.29}$$

and substituting for $\varepsilon(t)$ from (4.26)

$$\phi(t) = H_0\varepsilon(t) + \frac{BG_0}{A}\left(\frac{Ae(t) + Bu(t)}{D}\right) - \frac{D_1Bu(t)}{A} \tag{4.30}$$

$$= H_0\varepsilon(t) + \frac{B}{D}\left(G_0e(t) + \frac{G_0Bu(t)}{A} - \frac{D_1Du(t)}{A}\right) \tag{4.31}$$

From the diophantine equation, (4.31) becomes

$$\phi(t) = H_0\varepsilon(t) + \frac{B}{D}\,(G_0e(t) - H_0u(t)) \qquad (4.32)$$

which completes the proof.

By writing the residual as $\tilde{\phi}(t) = H_0\varepsilon(t)$ and the predicted output $\hat{\phi}(t|t-k) = \frac{B}{D}\,(G_0e(t) - H_0u(t))$

Hence

$$\phi(t) = \tilde{\phi}(t) + \hat{\phi}(t|t-k) \qquad (4.33)$$

Consider $\hat{\phi}(t|t-k)$. Assume the B polynomial is known or can be estimated separately. Now define $e_b(t) = -Bz(t)$ and $u_b(t) = Bu(t)$. The 1/D term is of the form (1 + delay terms) of which only unity is of significance since the optimal control has set the predictor to zero at all previous times. This argument was originally employed by Clarke and Gawthrop [48]. Assume also for the moment that the residual $\tilde{\phi}(t)$ and the predictor $\hat{\phi}(t|t-k)$ are uncorrelated after convergence so that any estimates will be unbiased.

Estimates of G_o and H_o by regression can now be provided using the following regression equation:

$$-D_1z(t) = G_oe_b(t) + H_ou_b(t) + H_o\varepsilon(t) \qquad (4.34)$$

where the regressand $-D_1z(t)$ is available, the residual is given by $H_o\varepsilon(t)$ and the regressors of $e_b(t)$ and $u_b(t)$ are known. The following LQG STR may now be given in a two stage identification algorithm.

Algorithm 4.2 Implicit LQG Self-Tuning Regulator 1

Date: Choose the cost function weighting Q_1 and R_1.

Step 1: Estimate the process polynomials A, B and D using extended least squares and the innovations model:

$$Ae(t) = D\varepsilon(t) - Bu(t)$$

Step 2 : Calculate the Hurwitz spectral factor D_1 using:

$$D_1(z^{-1})D_1(z) = B(z^{-1})B(z)Q_1 + A(z^{-1})A(z)R_1$$

Step 3 : Estimate G_0 and H_0 using recursive least squares and the prediction model:

$$\hat{\phi}(t|t-k) = G_0 e_b(t) - H_0 u_b(t) = \underline{X}^T(t-k)\underline{\theta}$$

Step 4 : Calculate the control signal as $u(t) = (G_0/H_0)e(t)$ and return to Step 1.

The so-called fixed spectrum STR [64] neglects Step 2 and places the closed-loop poles in desired regions within the unit circle. In Step 3 above the vectors $\underline{X}$ and $\underline{\theta}$ are defined as:

$$\underline{X}^T(t-k) = (e_b(t), e_b(t-1) \ldots e_b(t-n_{g_0}), - u_b(t-1), - u_b(t-2) \ldots - u_b(t-n_{h_0})$$

$$\underline{\theta}^T = (g_0, g_1 \ldots g_{n_{g_0}}, h_1, h_2 \ldots h_{n_{h_0}})$$

The drawback with Algorithm 4.2 as will be illustrated by an example, is one of parameter bias of G_0 and H_0. The assumption is made that after convergence $\hat{\phi}(t|t-k) = 0$ and hence $\tilde{\phi}(t)$ and $\hat{\phi}(t|t-k)$ are uncorrelated. That implies that $E\{\varepsilon \cdot \underline{X}^T\underline{\theta}\} = 0$. However, there are regressors in $\tilde{\phi}$ which depend upon data within $\hat{\phi}$. That is, $E\{\varepsilon \cdot \underline{X}^T\} \neq 0$ which violates the condition for unbiased estimates [49,56] and implies that $E\{\hat{\underline{\theta}}\} \neq 0$. Moreover, $\tilde{\phi}(t)$ and $\hat{\phi}(t|t-k)$ are correlated for all terms in $\tilde{\phi}(t)$ occurring at times less than, or equal to, t-k. This problem is easily overcome by dividing the residual into two terms $H_{01}\varepsilon(t)$ and $H_{02}\varepsilon(t)$.

$$\tilde{\phi}(t) = H_0\varepsilon(t) = H_{01}\varepsilon(t) + H_{02}\varepsilon(t)$$

where

$$H_{01}(z^{-1}) = 1 + h_1 z^{-1} + \ldots + h_{k-1} z^{-(k-1)} \qquad (4.35)$$

$$H_{02}(z^{-1}) = h_k z^{-k} + h_{k+1} z^{-(k+1)} + \ldots + h_{n_{b-1}} z^{-(n_b-1)} \qquad (4.36)$$

giving

$$H_{01}\varepsilon(t) = \varepsilon(t) + h_1\varepsilon(t-1) + \ldots h_{k-1}\varepsilon(t-k+1) \qquad (4.37)$$

$$H_{02}\varepsilon(t) = h_k\varepsilon(t-k) + h_{k+1}\varepsilon(t-k-1) + \ldots + h_{n_{b-1}}\varepsilon(t-n_b+1) \qquad (4.38)$$

The term $H_{02}\varepsilon(t)$ now contains all residuals occuring from time t-k, t-k-1 etc. These terms are correlated with the predictor $\hat{\phi}(t|t-k)$ and hence must be included within it. The modified predictor becomes:

$$\hat{\phi}_1(t|t-k) = \hat{\phi}(t|t-k) + H_{02}\varepsilon(t) \qquad (4.39)$$

and therefore the implicit model may be rewritten as:

$$\phi(t) = H_{01}\varepsilon(t) + \hat{\phi}_1(t|t-k) \qquad (4.40)$$

The regressions within $H_{01}\varepsilon(t)$ now are in the future ($t \geq t-k+1$) relative to the terms within $\hat{\phi}_1(t|t-k)$ which occur in the past ($t \leq t-k$). The left-hand side of the expression is therefore uncorrelated with the right-hand side. This condition will give unbiased estimates. The new implicit Algorithm follows.

Algorithm 4.3 Implicit LQG STR 2

As Algorithm 4.2 except Step 3 is altered to accommodate the modified predictor.

Step 3 : Estimate G_0 and H_0 using least squares and the

prediction model:

$$\hat{\phi}_1(t|t-k) = G_0 e_b(t) - H_{01} u_b(t) - H_{02}(u_b(t) - \hat{\varepsilon}(t))$$

where $\hat{\varepsilon}(t)$ is the t step estimate of the innovation, computed from Step 1 using extended recursive least squares.

From the above:

$$\hat{\phi}_1(t|t-k) = \underline{X}_1^T(t-k)\underline{\theta}$$

where

$$\underline{X}_1^T(t-k) = [e_b(t),\ e_b(t-1) \ldots e_b(t-ng_0),$$
$$- u_b(t-1) \ldots - (u_b(t-k) - \varepsilon(t-k)),$$
$$- (u_b(t-k-1) - \varepsilon(t-k-1)) \ldots .$$
$$\ldots - (u_b(t-nh_0) - \varepsilon(t-nh_0))]$$

and $\underline{\theta}$ remains as before.

From the previous argument

$$E\{\varepsilon \underline{X}_1^T\} = 0 \qquad (4.41)$$

which implies unbiased parameter estimates or [56]

$$E\{\hat{\underline{\theta}}\} = \underline{\theta} \qquad (4.42)$$

Example 4.3

The following example illustrates the material discussed in 4.4.2. The operation of an implicit LQG STR is simulated using both Algorithm 4.2 and 4.3. The example is chosen in such a way as to highlight the drawbacks of Algorithm 4.2 by illustrating the bias problem when estimating the G_0 and H_0 polynomials.

The process is given as

$$y(t) = z^{-1} \frac{(1 + b_1 z^{-1} + b_2 z^{-2})}{(1 + a_1 z^{-1})} u(t)$$

$$+ \left[\frac{1 + c_1 z^{-1}}{1 + a_1 z^{-1}}\right] \omega(t)$$

$\omega(t)$ is a white noise sequence with standard deviation 10.0 and R = 0.

The true parameter values are given as $a_1 = -0.9$, $c_1 = -0.5$, $b_1 = -1$, $b_2 = 1$. For simplicity, the poles of the closed loop system may be pre-specified as

$$D_1(z^{-1}) = 1 - 0.6z^{-1}$$

This type of controller has been termed an implicit fixed spectrum LQG STR [64]. From the diophantine equation (4.17) the controller parameters are computed as $g_0 = 0.1186$, $h_1 = -0.3186$ and $h_2 = 0.13186$. Initially, the implicit STR employing Algorithm 4.2 is simulated. Figure 4.10 gives the controller parameter estimates showing them to be either diverging or biased. This is due to correlation between the $\tilde{\phi}$ and $\hat{\phi}$ terms when using this algorithm. In particular, the h_1 parameter is the most in error. The process parameter estimates of Figure 4.11 are close to the true values. Algorithm 4.3 is now

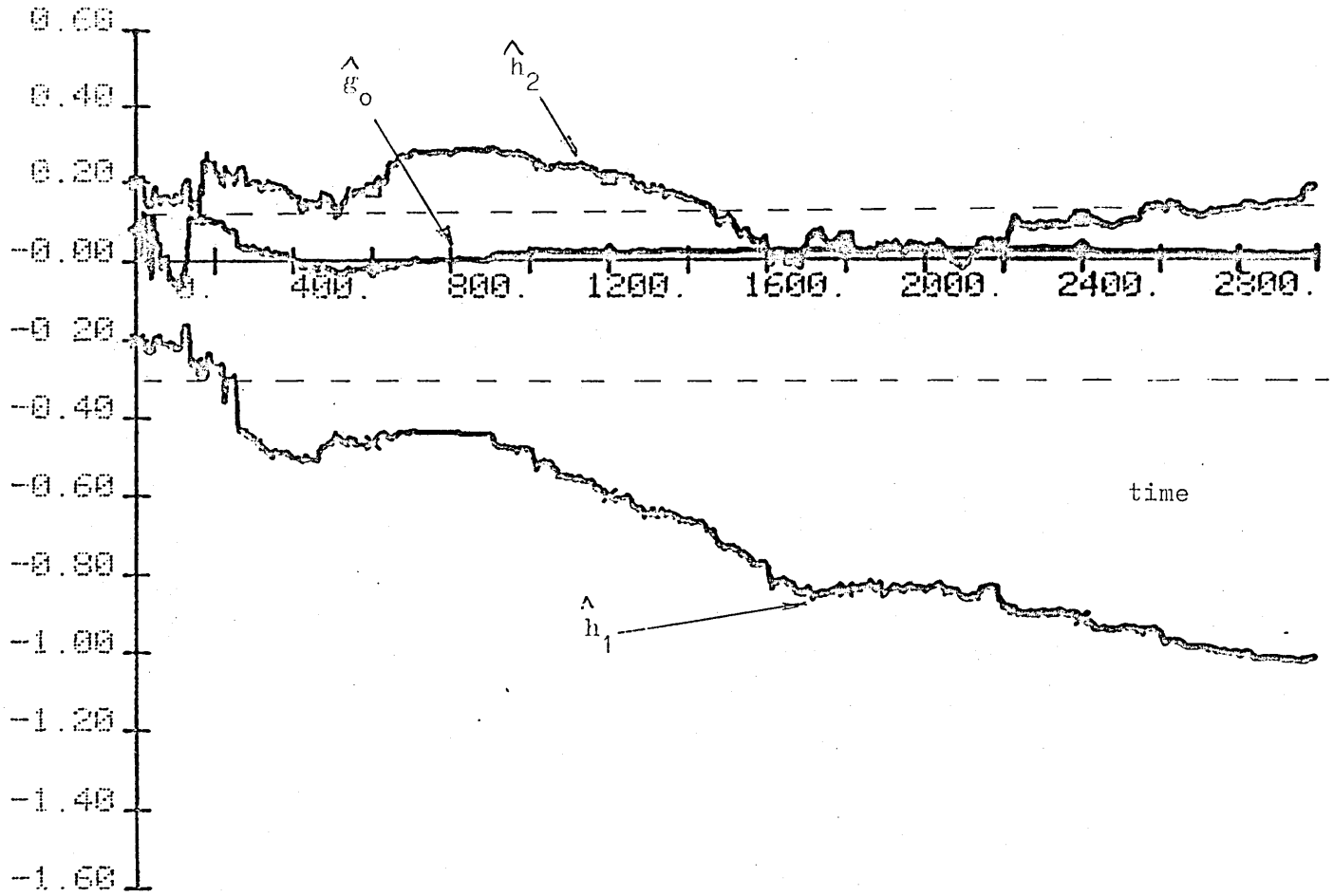


Figure 4.10 Non Convergence of controller parameters using Algorithm 4.2

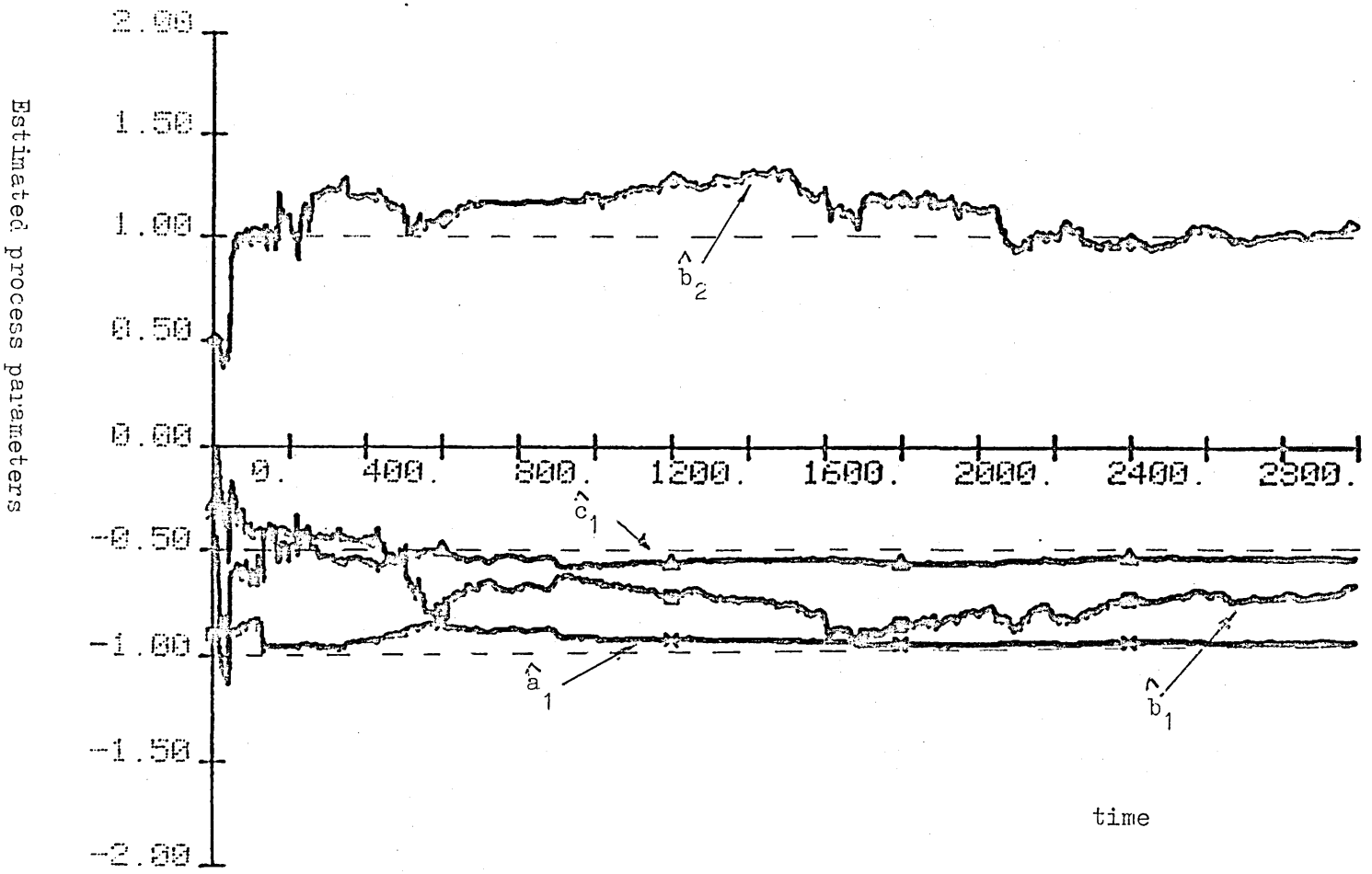


Figure 4.11 Convergence of process parameters using Algorithm 4.2

employed and will overcome the problem of bias in H_0 and G_0.

Figure 4.12 shows the convergence of controller parameters using this scheme and shows the elimination of the bias and divergence problem. For this example the H_0 polynomial is divided into

$$H_{01} = 1$$

$$H_{02} = h_1 z^{-1} + h_2 z^{-2}$$

thus two terms in the regressors of the residual are correlated with the predictor when using Algorithm 4.2. When using Algorithm 4.3, these two terms are included within the predictor and the resulting $\hat{\phi}_1$ is uncorrelated with $\tilde{\phi}$.

The explicit Algorithm has also been simulated and the results are shown in Figures 4.15 to 4.17. The cumulative loss function given by

$$L = \sum_{i=1000}^{2000} (Qe^2(i) + Ru^2(i))$$

is compared for the various schemes including the fixed parameter or optimal regulator. Figure 4.19 shows the loss for the implicit Algorithm 4.2 to be the highest whilst the fixed parameter, Algorithm 4.3 and the explicit Algorithm 4.1 give identical results of much lower loss.

By fixing the first identification in Algorithms 4.2 and 4.3 (i.e. working with a known process), very similar results were obtained and are shown in Figures 4.20 and 4.21 respectively.

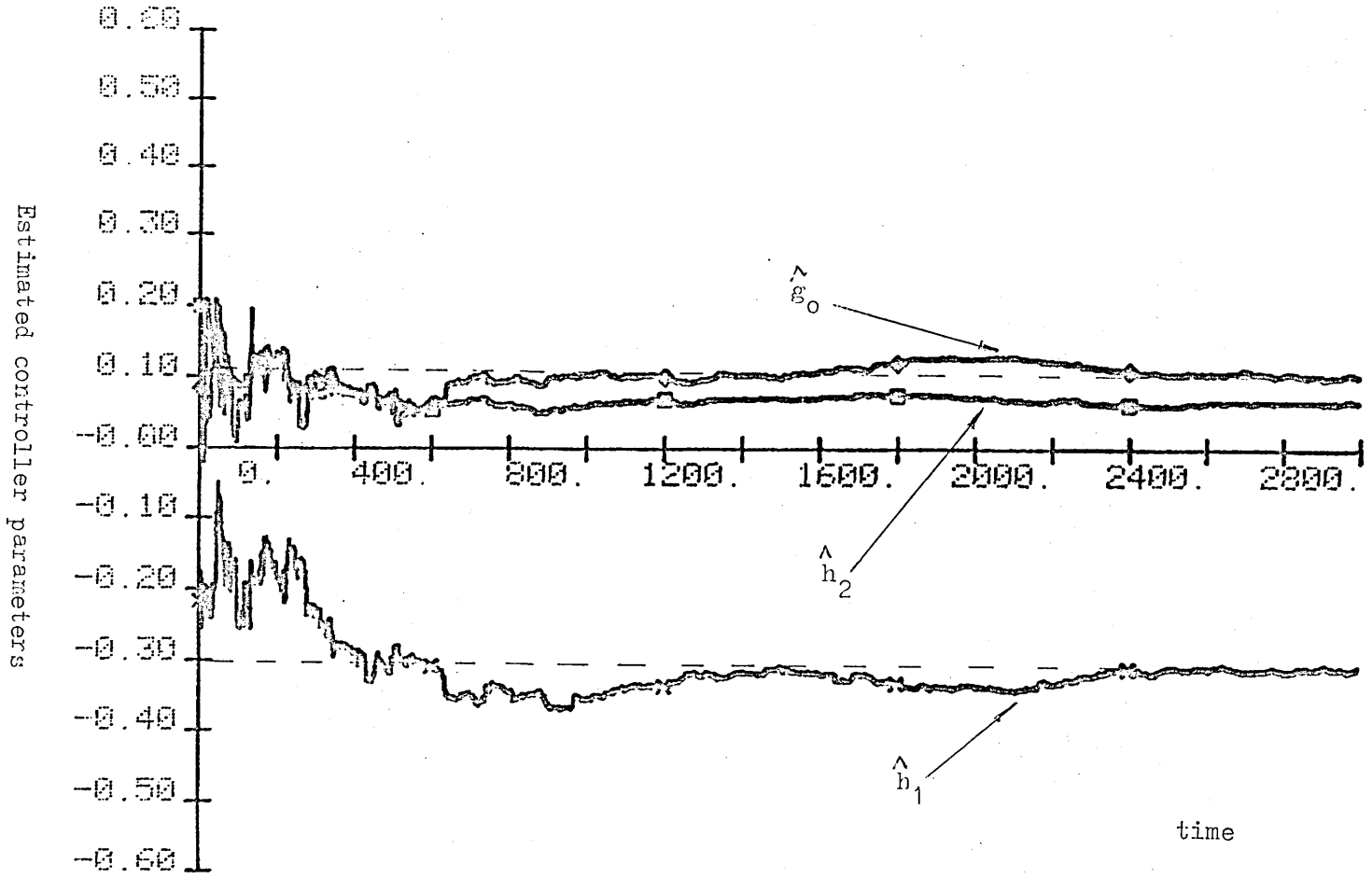


Figure 4.12 Convergence of controller parameters using Algorithm 4.3

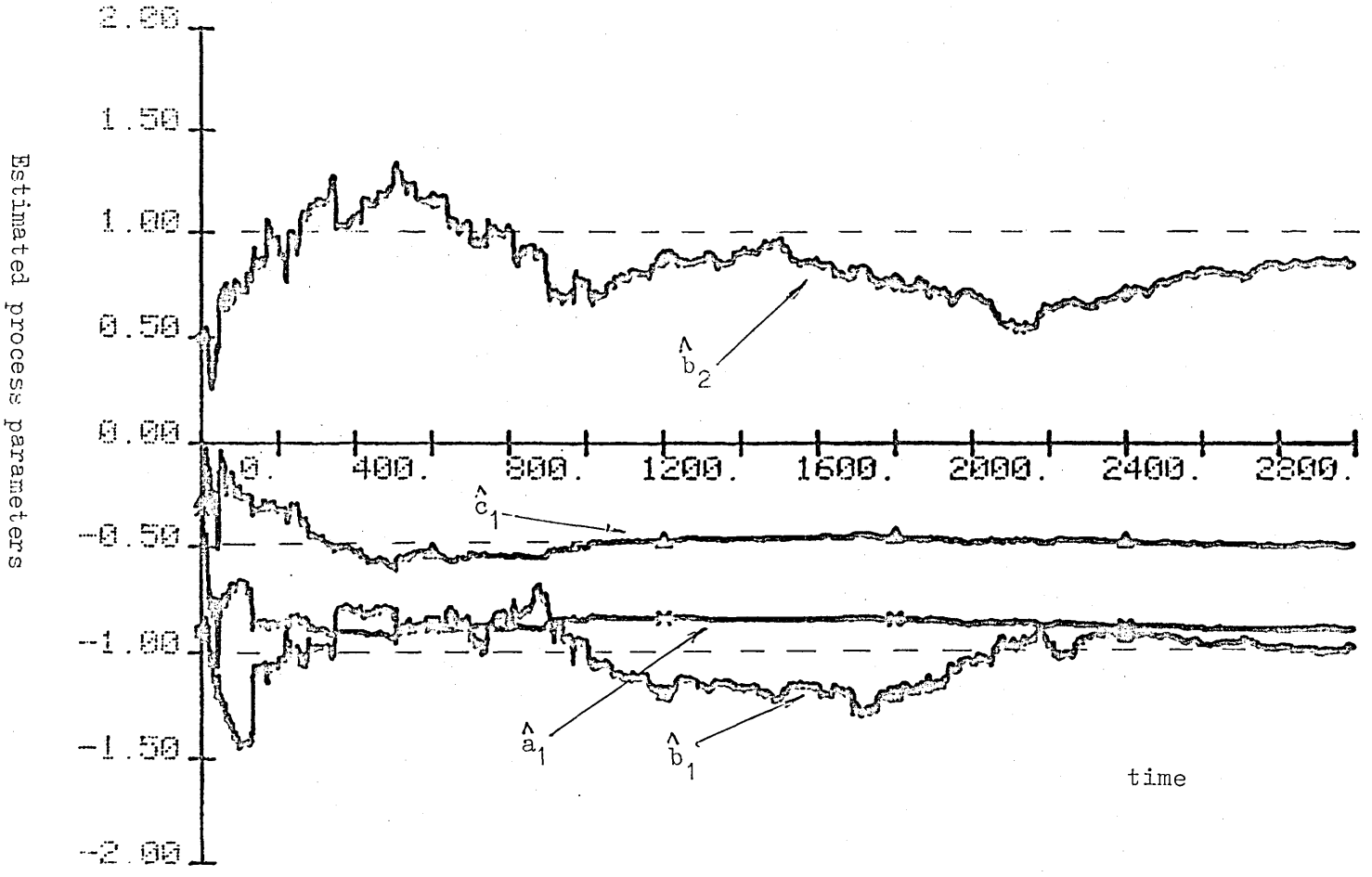


Figure 4.13 Convergence of process parameters using Algorithm 4.3

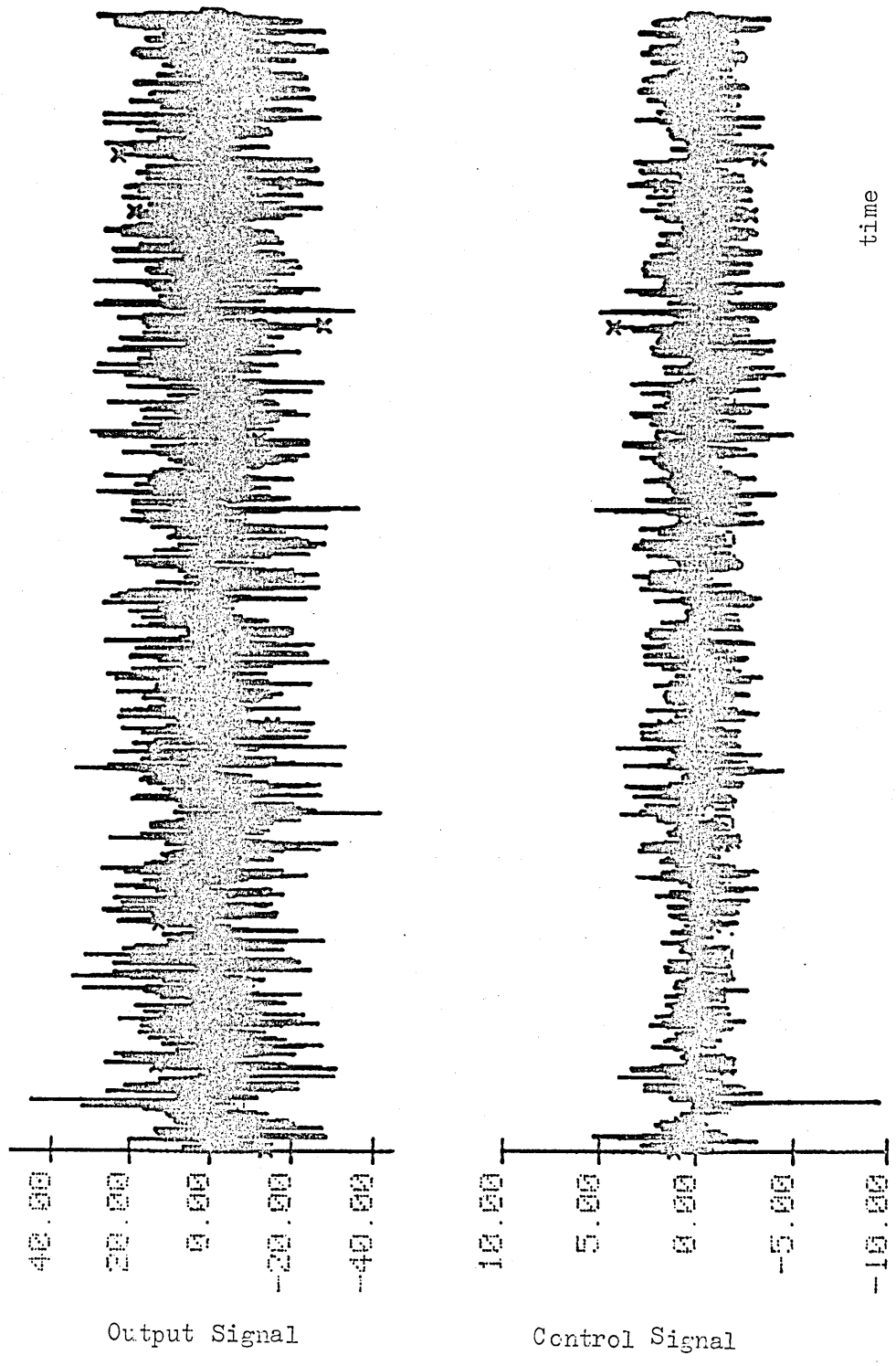


Figure 4.14 Output & Control Signals for implicit Algorithm 4.3

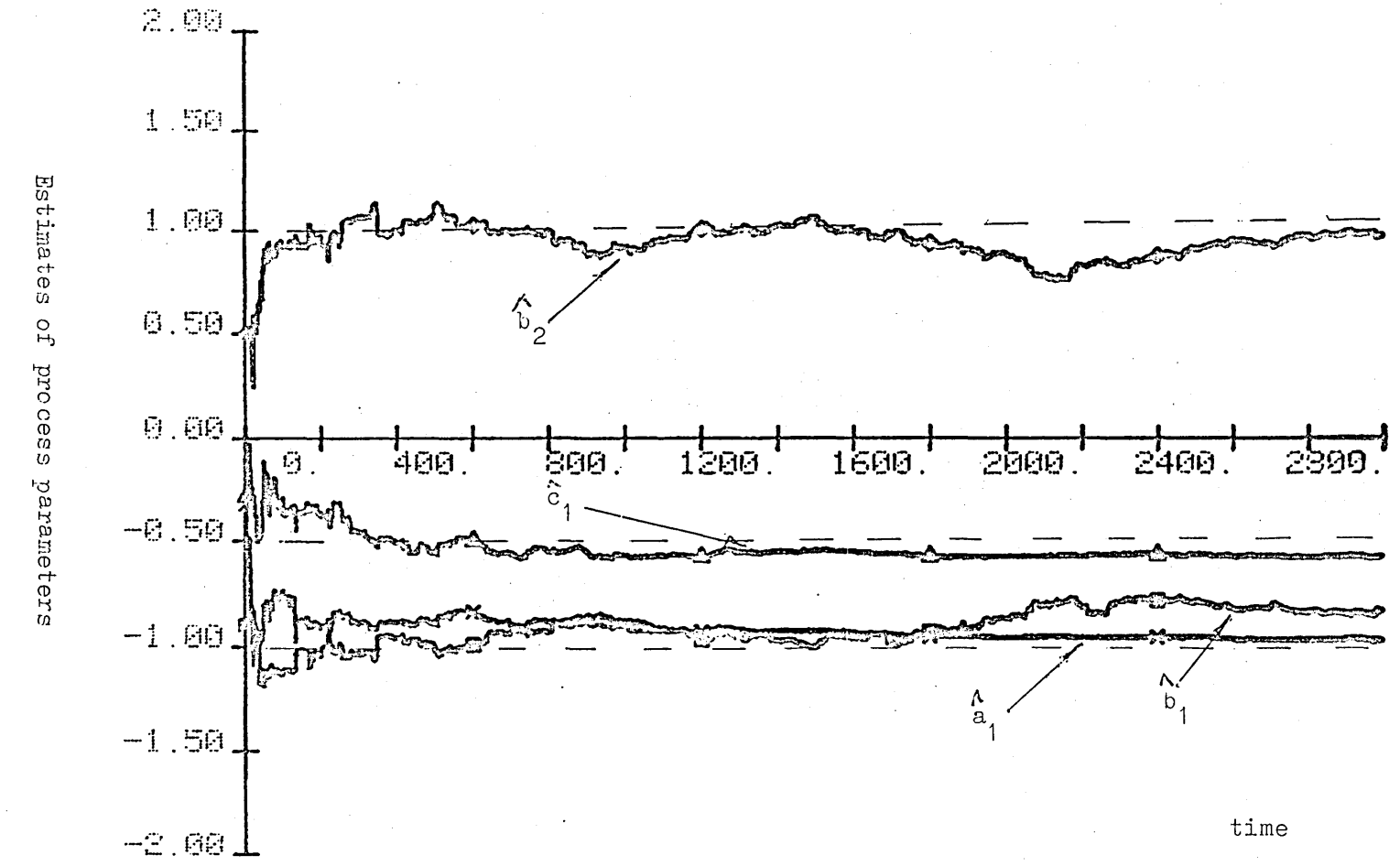


Figure 4.15 Convergence of process parameters for explicit Algorithm

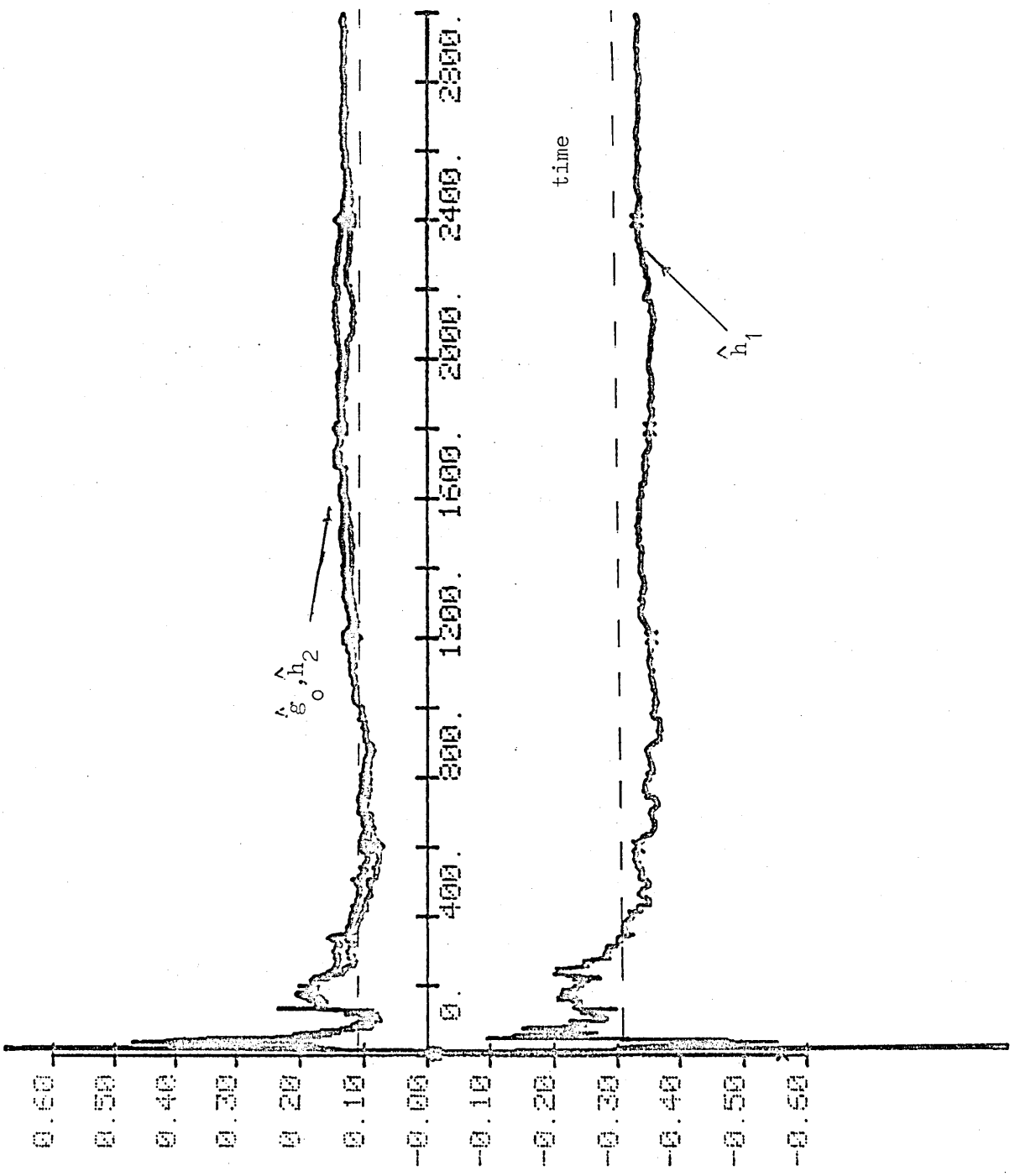


Figure 4.16 Process controller parameter convergence for explicit Algorithm

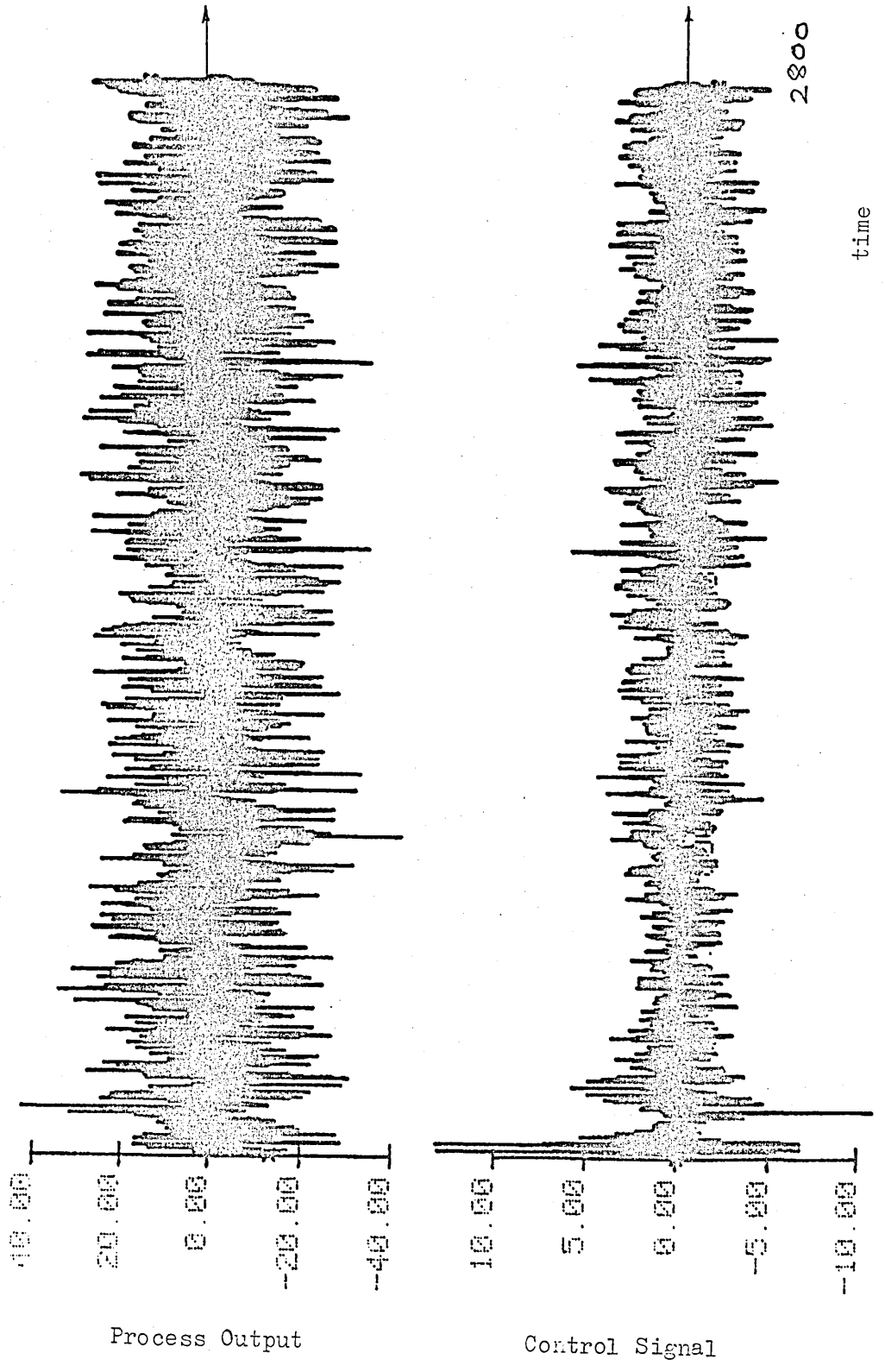


Figure 4.17 Process output & Control signal for explicit Algorithm

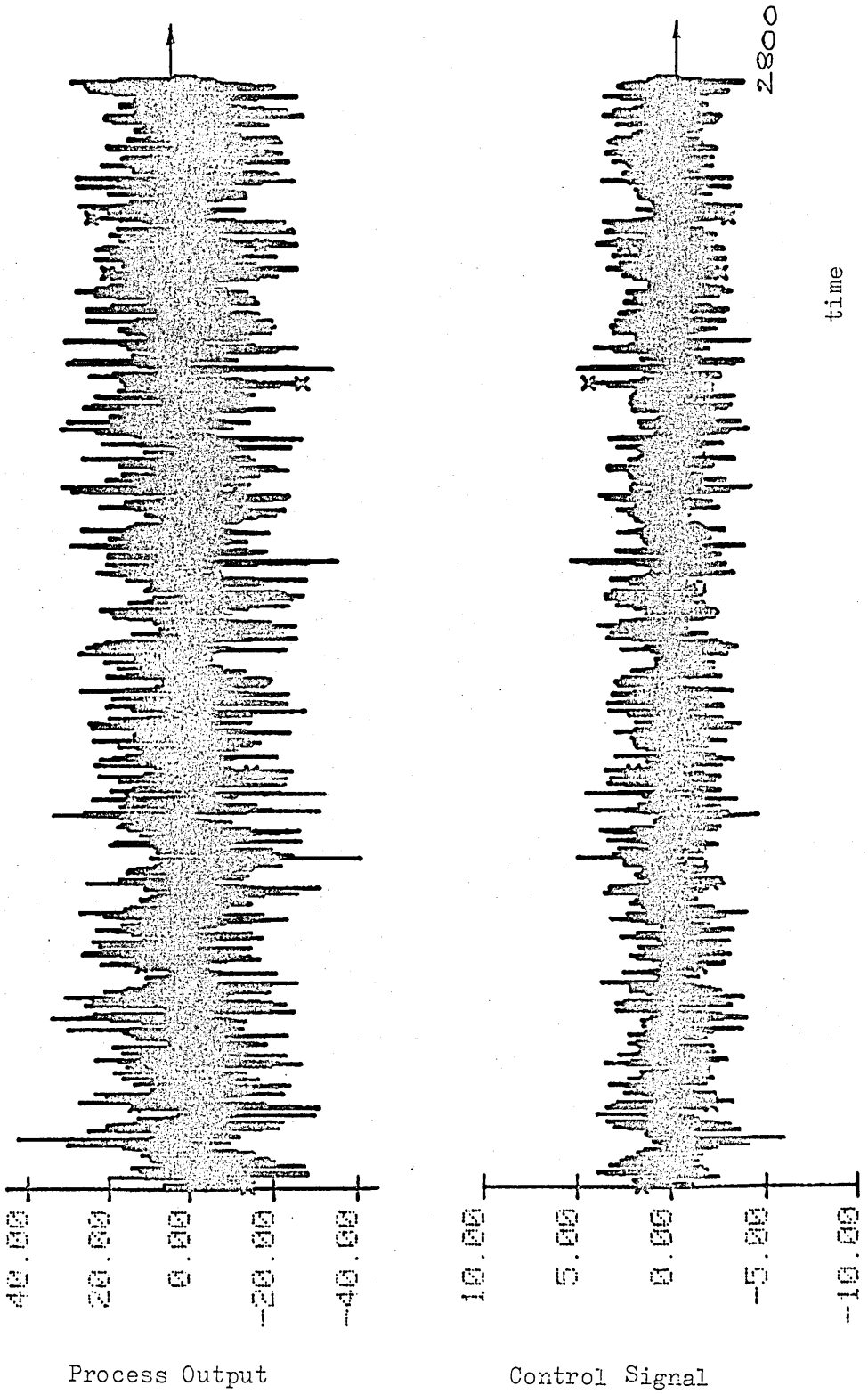


Figure 4.18 Process Output & Control Signal for fixed parameters

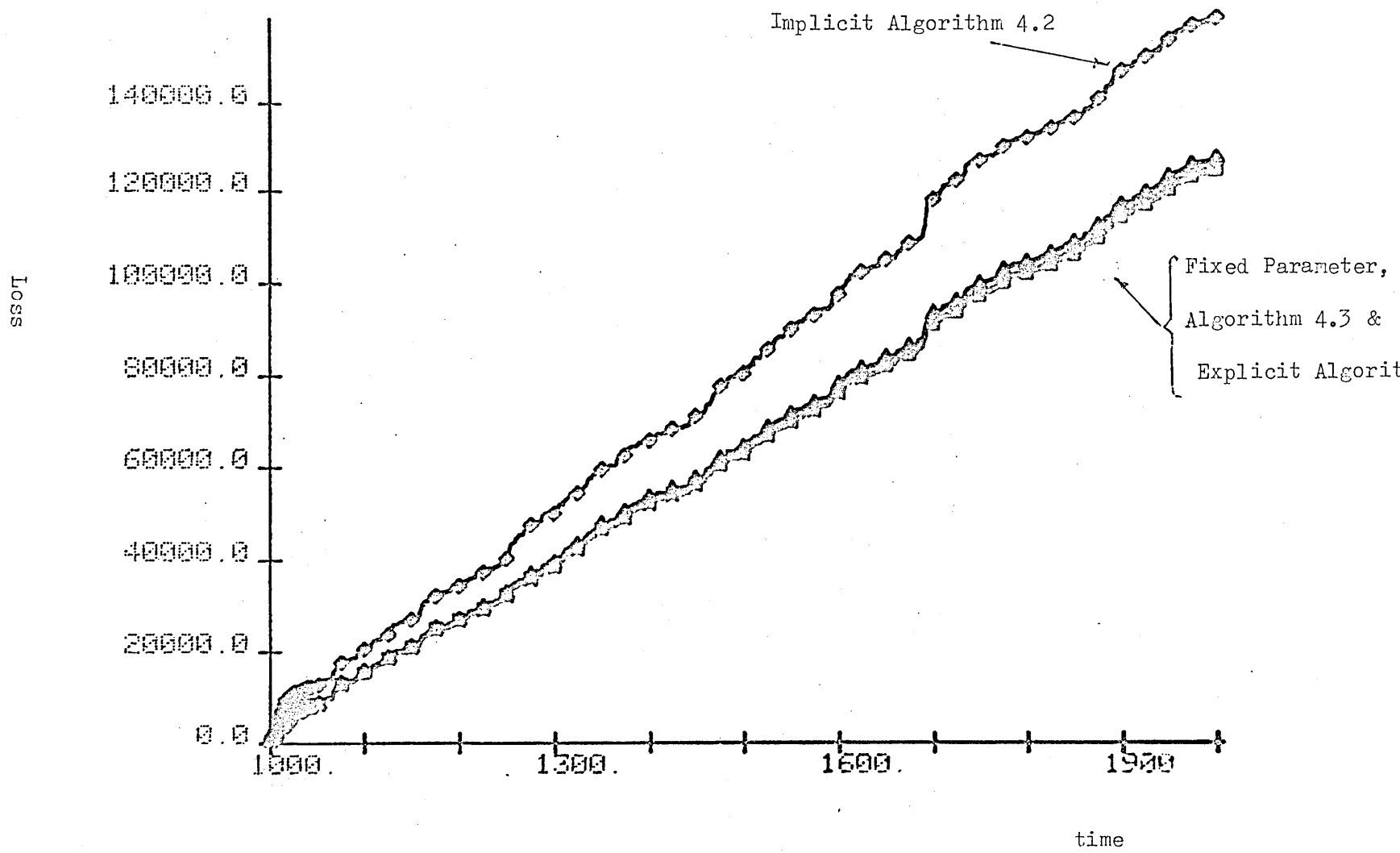


Figure 4.19 cumulative Loss

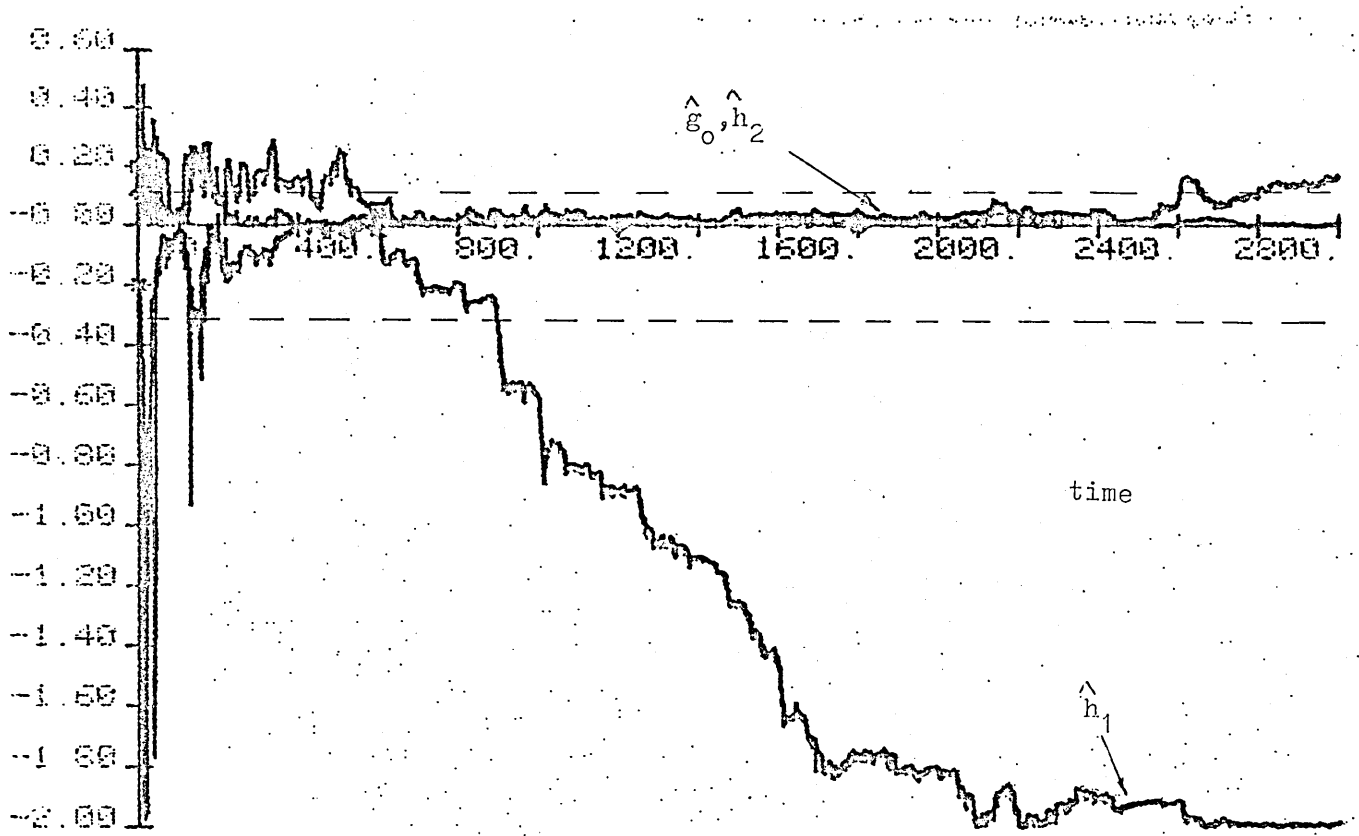


Figure 4.20 Using Algorithm 4.2 with known process parameters

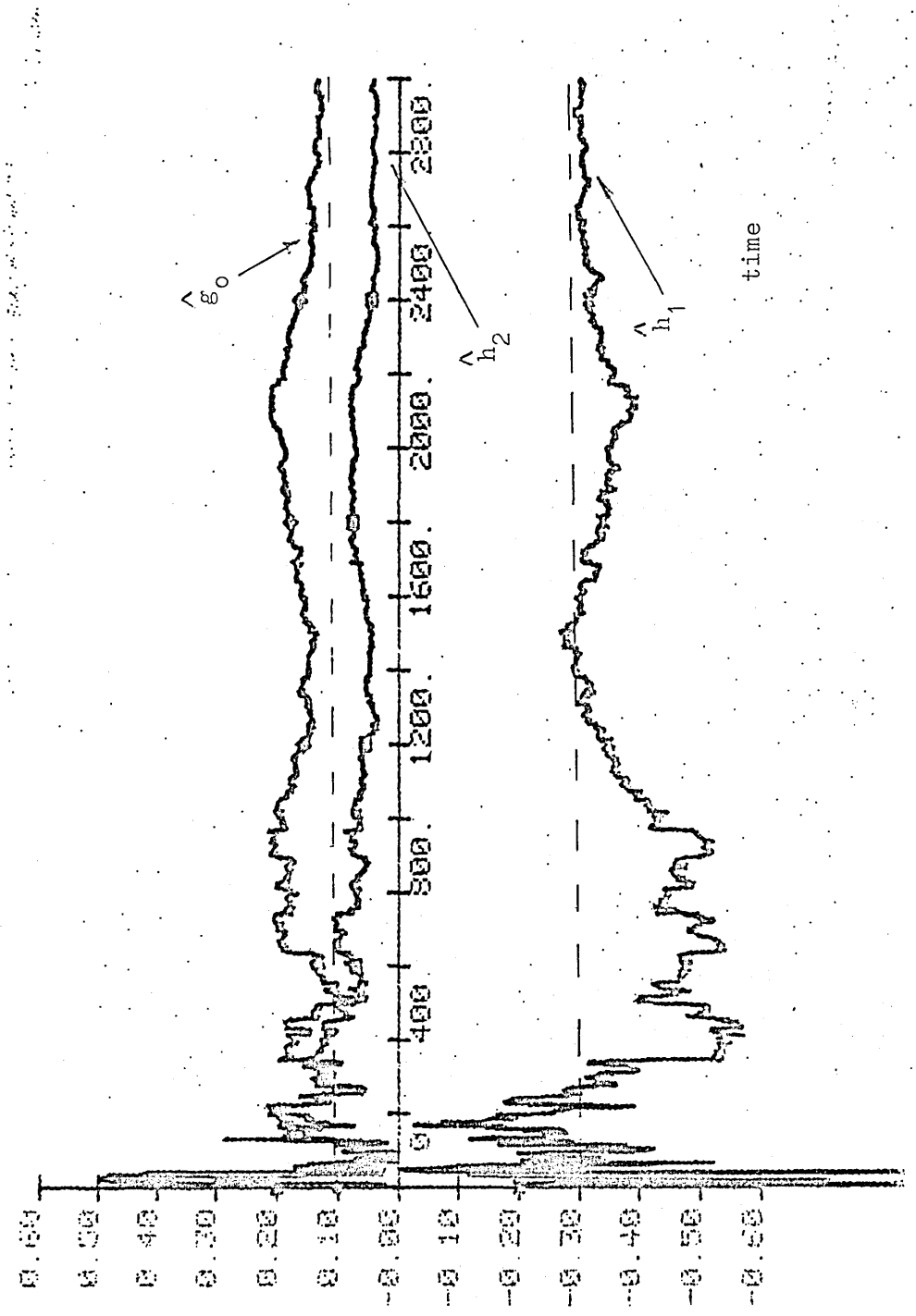


Figure 4.21 Using Algorith 4.3 with known process parameters

Example 4.4

The examples in this chapter have so far not included measurement noise. This example follows from example 4.3 but with the inclusion of measurement noise. Consider example 4.3 but with Q = 1 and R = 10, thus giving a significant amount of measurement noise. The process model is identical to the previous example except the measurement now becomes:

$$z(t) = y(t) + v(t)$$

The explicit LQG STR is only considered here since the implicit case follows as before. All the algorithms previously given can already cope with measurement noise in their present form. If the measurement noise is zero, then $D(z^{-1}) = C(z^{-1})$ and the innovations becomes the disturbance noise $\omega(t)$. For this example, the only change is that the D polynomial alters to cope with the measurement noise as does the covariance of the innovations sequence.

The modified D polynomial may be found from (4.7) and (4.8) as:

$$A(z^{-1})A(z)R + C(z^{-1})C(z)Q = D(z^{-1})D(z^{-1})R_{\varepsilon} \quad (4.43)$$

where R_{ε} is the innovations covariance. From Chapter 3, it follows that the D polynomial is acting as part of a self-tuning filter which is implicit in the controller design to eliminate the effect of measurement noise. For this example the D polynomial was found from (4.43) to be:

$$D(z^{-1}) = 1 - 0.8254\ z^{-1}$$

which leads to an LQG optimal controller with parameters given as $g_0 = 0.022$, $h_1 = -0.548$ and $h_2 = 0.0246$. The

process parameter estimates are shown in Figure 4.22 where the innovations representation of the process is estimated. This avoids any spectral factorisation in the form of equation (4.43).

The computed controller parameters, found by the solution of the diophantine equation, are shown in Figure 4.23. These parameters are very close to their true values. Finally, the measured process output, the process output and the control signal are displayed in Figure 4.24. In practice, the process output would not be accessible but it is shown here to illustrate the amount of measurement noise for this example.

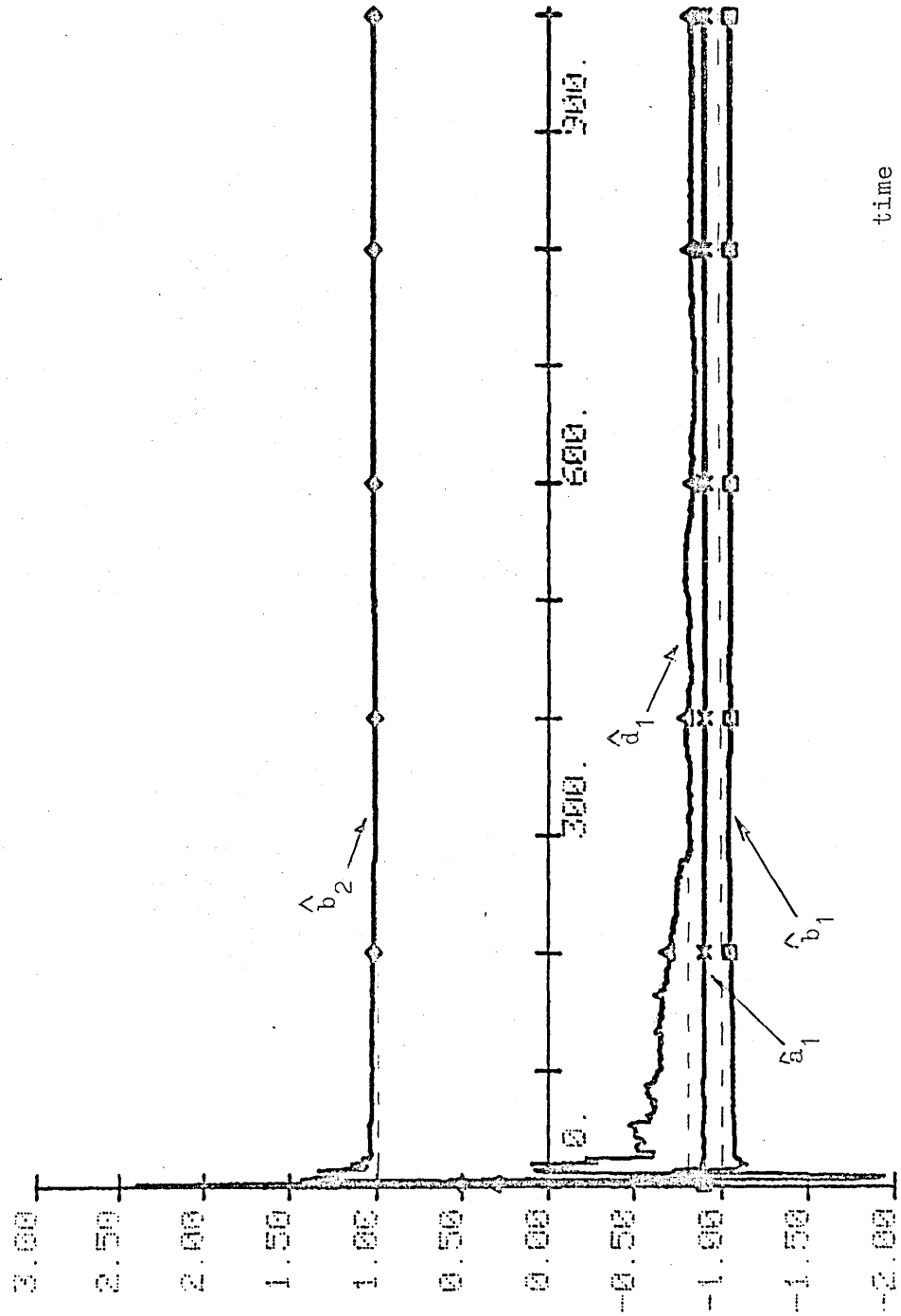


Figure 4.22 Process parameter estimates for example 4.4

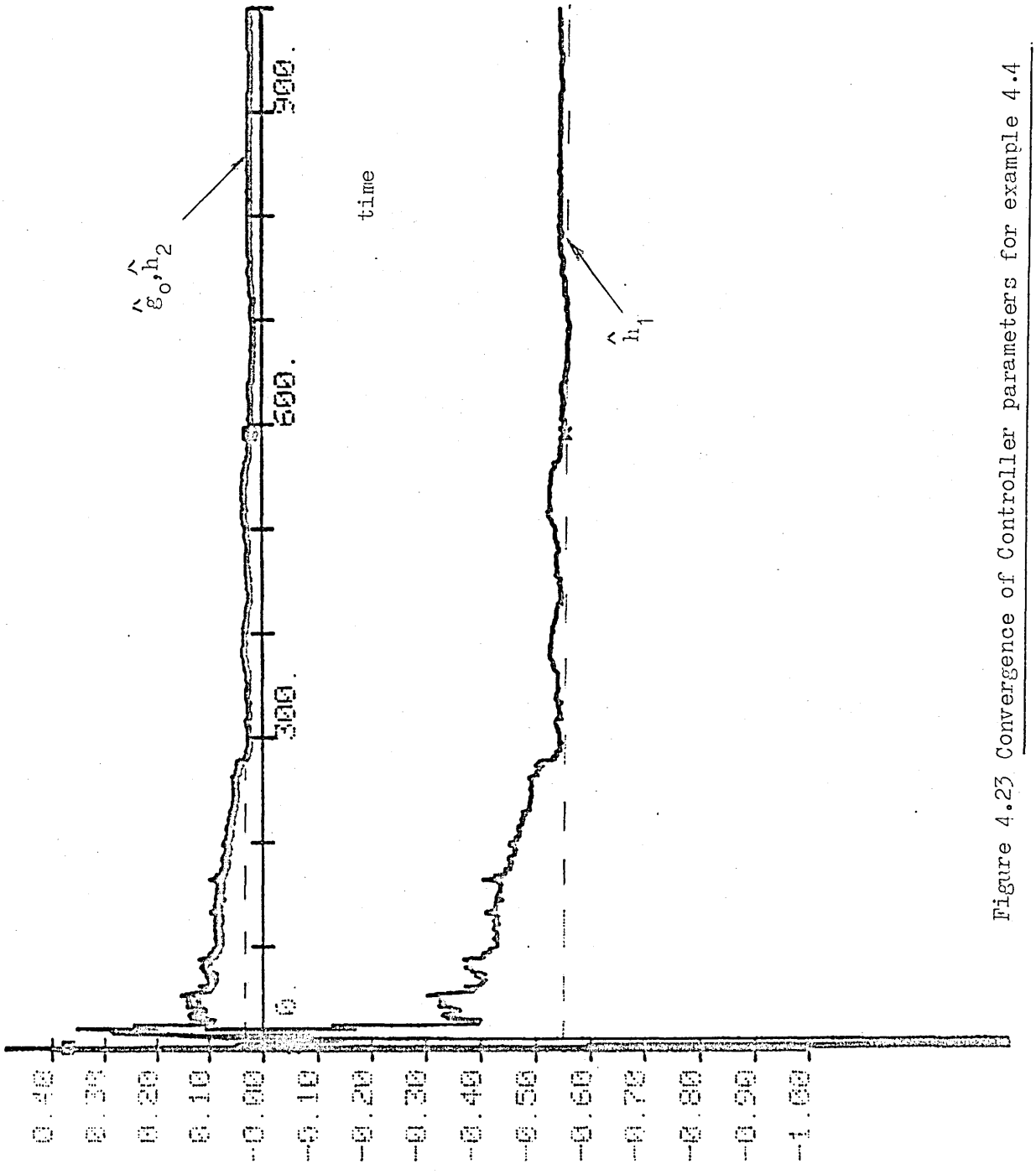


Figure 4.23 Convergence of Controller parameters for example 4.4

Convergence of Controller Parameters

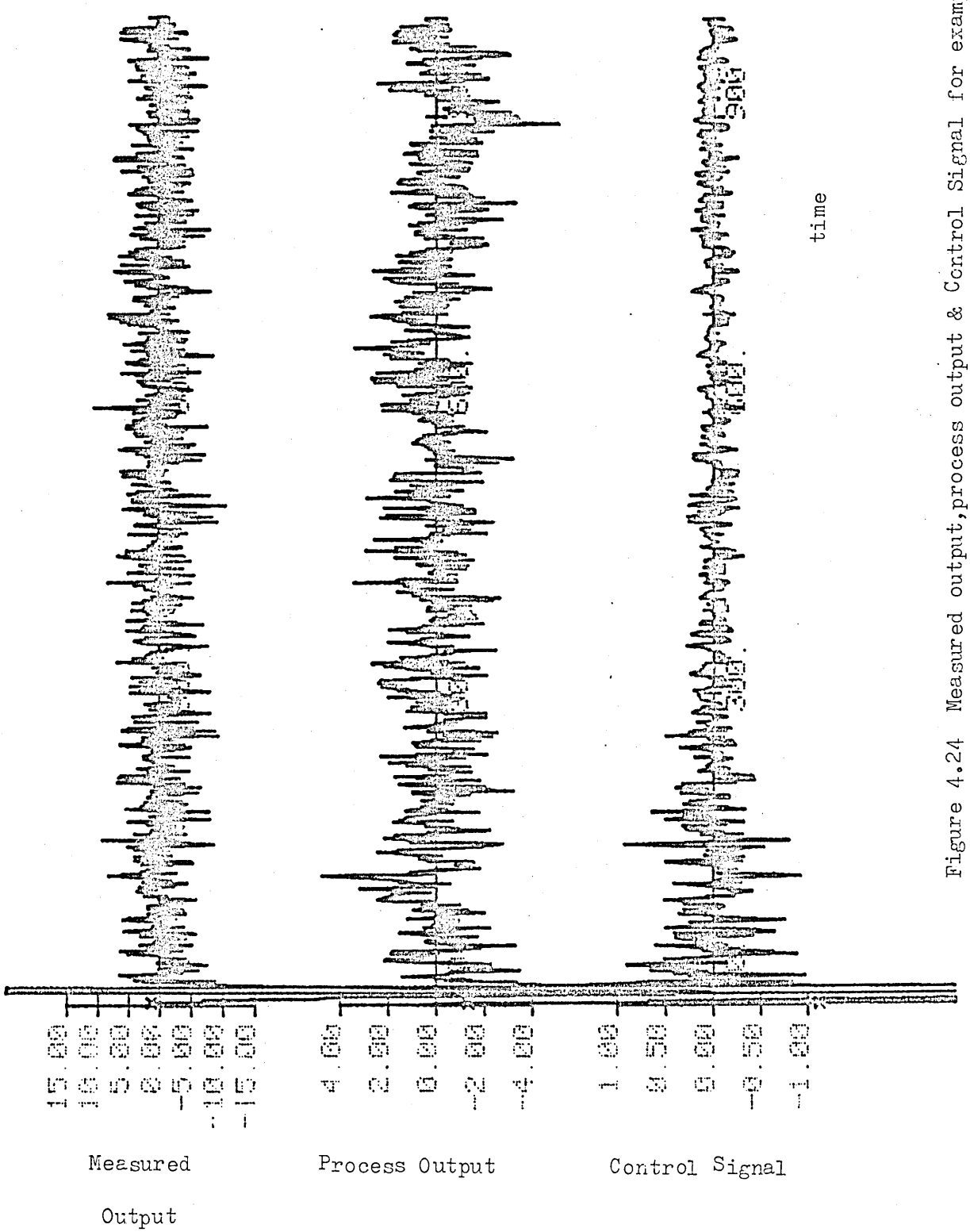


Figure 4.24 Measured output, process output & Control Signal for example 4.4

4.5 Summary

A new LQG STR has been demonstrated using both explicit and implicit algorithms. The explicit algorithm requires the solution of a diophantine equation to compute the controller parameters. Using least squares, the implicit algorithms can estimate the controller parameters directly. This avoids the solution of the diophantine equation. One of the implicit algorithms has been shown to create parameter bias and this is overcome in the second implicit algorithm. The implicit algorithms are new, requiring a two-stage identification procedure. The first identification is that of the process and using these estimates the controller parameters may be estimated from the second identification.

The LQG STR has been shown to cope easily with non-minimum phase and open-loop unstable processes. Previous STR have had difficulty dealing with this problem. In addition, measurement noise can be allowed for without any complications by using the innovations form of the process. This acts much like a self-tuning filter.

The time-delay in a process is usually assumed to be known or estimated beforehand. This assumption need not be made here. Moreover, an example illustrating this point has shown how an unknown and variable time-delay can be accommodated.

Chapter 5

Concluding Remarks and Further Work

In this thesis, new theoretical results have been presented in the field of linear least-squares estimation and stochastic optimal control. The least-squares estimation results overcome some existing problems for filtering and smoothing in discrete time processes. The finite-interval smoother is of mainly theoretical importance due to its complexity but it has interesting properties and advantages over existing Riccati equation based solutions.

The section on multivariable self-tuning estimation is intended for many real processes where the model and signal statistics are unknown. Successful operation is shown with various examples which may be encountered in industry. Fung and Grimble [104] have applied this work to the dynamic ship positioning problem where a multivariable self-tuning filter is used to estimate the high-frequency motion of a vessel at sea. The filter has been shown to reliably adapt to various sea conditions giving encouraging results.

LQG regulators and controllers can cope with unstable and non-minimum phase processes without any difficulty. This thesis has considered LQG self-tuning regulators using an input-output rather than state-space approach. It has advantages over the pole-placement methods of Wellstead

and co-workers since an implicit algorithm has been derived. A greater awareness is achieved for the control of a process if using LQG control since the process output and control signal can be separately weighted to the designer's wishes. Pole placement ideas, although robust do not give the designer any of this versatility.

Future research concerns the study of the asymptotic convergence of the LQG self-tuning regulator algorithms. In particular, the implicit algorithm. This employs a novel two-stage parameter identification. The first estimates the process parameters and the second the regulator parameters. In the analysis it would be necessary to fix the parameters of the first identification while studying the analysis of the second, and vice-versa, using probably the methods of Ljung. As an alternative to this two-stage identification technique, the bilinear estimation algorithm of Astrom [89] could be used.

REFERENCES

1. WIENER N. 'The extrapolation, interpolation and smoothing of a stationary time series, with engineering applications'. Wiley, 1949.

2. KOLMOGOROV, A N. 'Sur l'interpolation et extrapolation des suits stationnaires'. C R Acad Sci, Vol 208, 1939.

3. KALMAN, R E. 'A new approach to linear filtering and prediction problems'. Journal of Basic Engineering, 82D, 1960, pp 35-45.

4. KALMAN, R E and BUCY, R S. 'New results in linear filtering and prediction theory'. Journal of Basic Engineering, 83D, 1961, pp 95-107.

5. KALMAN, R E. 'On the general theory of control systems'. Moscow, 1960, First IFAC Cong, Proceedings 1960. Butterworth.

6. WONG, E and THOMAS, J B. 'On the multidimensional prediction and filtering problems and the factorization of spectral matrices'. Journal of the Franklin Institute, Vol 272, No 2, 1961, pp 87-99.

7. RAUCH, H E. 'Solutions to the smoothing problem'. IEEE Trans Aut Control, Vol AC-8, 1963, pp 371-372.

8. MEDITCH J. 'Orthogonal projection and discrete optimal linear smoothing'. Sc Am J Contr. Vol 5, No 1, 1967, pp 74-89.

9. MEDITCH, J. 'On optimal fixed point linear smoothing'. Int J Cont Vol 6, No 2, 1967, pp 189-199.

10. MOORE, J B. 'Discrete-time fixed-lag smoothing algorithms'. Automatica, Vol 9, 1973, pp 163-173.

11. KAILATH, T. 'Lectures on linear least-squares estimation'. Springer-Verlag, New York, 1976.

12. JAZWINSKI, A. 'Limited memory optimal filtering'. IEEE Trans Aut Cont Vol AC, Oct 1968, pp 556-563.

13. GRIMBLE, M J. 'A finite-time filter for discrete time systems'. Int J Cont, 1980 Vol 32, No 3, pp 413-432.

14. GRIMBLE, M J. 'Solution to the Kalman filtering problem for stationary noise and finite data records'. Int J Systems Science, Vol 10, No 2, 1979, pp 177-196.

15. GRIMBLE, M J. 'A new finite time linear smoothing filter'. Int J Systems Science, 1980, Vol 11, No 10, pp 1189-1212.

16. MEHRA, R K. 'Approaches to adaptive filtering'. IEEE Trans Aut Cont Vol AC-17, 1972, pp 693-698.

17. MEHRA, R K. 'On the identification of variances and adaptive Kalman filtering'. IEEE Trans Aut Control Vol AC-15, No 2, April 1970, pp 175-184.

18. MEHRA, R K. 'On line identification of linear dynamic systems with applications to Kalman filtering'. IEEE Trans Aut Cont Vol AC-16, Feb 1971, pp 12-21.

19. ALSPACH, D L and ABIRI, A. 'Optimal non-linear estimation for a linear system with unknown plant and measurement noise covariances'. PROC 3rd Symposium Non-Linear Estimation Theory, 1972, San-Diego.

20. BÉLANGER, P R. 'Estimation of noise covariance matrix for a linear time-varying stochastic process'. Automatica, Vol 10, 1974, pp 267-275.

21. CAREW, B and BÉLANGER, P R. 'Identification of optimum filter steady state gain for systems with unknown noise covariances'. IEEE Trans Aut Cont Vol AC-18, Dec 1973, pp 582-587.

22. LEONDES, C T and PEARSON, J O. 'Kalman filtering of systems with parameter uncertainties - a survey'. Int J Cont Vol 17, No 4, 1973, pp 785-801.

23. ALSPACH, D L. 'Comments on "on the identification of variances and adaptive Kalman filtering" ' IEEE Trans Aut Cont, Vol AC-17, No 6, 1972, pp 843-845.

24. MacFARLANE, A G J. 'Return-difference matrix properties of optimal stationary Kalman-Bucy filter'. Proc IEE, 1970, 118, No 2, pp 373-376.

25. ARCASOY, C C. 'Return-difference matrix properties for optimal stationary discrete Kalman filter'. Proc IEE, 1971, Vol 118, pp 1831-1834.

26. BARRETT, J F. 'Construction of Wiener filters using the return difference matrix'. Int J Contr Vol 26, No 5, 1977, pp 797-803.

27. SHAKED, U. 'A transfer function approach to the linear discrete stationary filtering and the steady-state discrete optimal control problems'. Int J Cont, 1979, Vol 29, No 2, pp 279-291.

28. KUCERA, V. 'Transfer function solution of the Kalman-Bucy filtering problem'. Kybernetika, Vol 14, No 2, 1978, pp 110-122.

29. AKAIKE, H. 'A new look at the statistical model identification'. IEEE Trans, AC-19, 1974, pp 716-723.

30. WITTENMARK, B. 'A self-tuning predictor'. IEEE Trans Aut Cont Vol AC-19, 1974, pp 848-851.

31. TANTTU, J T. 'A self-tuning predictor for a class of multivariable stochastic processes'. Int J Control, Vol 32, No 2, 1980, pp 359-370.

32. LEDWICH G and MOORE, J B. 'Multivariable self-tuning filters'. 2nd Conf on Differential Games and Control Theory, 1976, July.

33. HAGANDER, P and WITTENMARK, B. 'A self-tuning filter for fixed lag smoothing'. IEEE Trans INF Theory Vol IT-23, No 3, May 1977, pp 377-384.

34. ASTROM, K J. 'Introduction to stochastic control theory'. Academic Press, 1970.

35. LJUNG, L. 'Analyses of Recursive stochastic algorithms'. IEEE Trans Aut Cont, Vol AC-22, No 4, Aug 1977, pp 551-575.

36. LJUNG, L. 'Asymptotic behaviour of the extended Kalman filter as a parametric estimator for linear systems'. IEEE Trans Aut Cont, Vol AC-24, No 1, Feb 1979, pp 36-50.

37. LJUNG, L. 'Convergence analysis of parametric identification methods'. IEEE Trans Aut Contr, Vol AC-23, No 5, Oct 1978, pp 770-783.

38. NEWTON, G C., GOULD, L A. and KAISER, J F. 'Analytical design of linear feedback systems'. Wiley, 1957.

39. CLARKE, D W. and HASTING-JAMES, R 'Design of digital controllers for randomly disturbed systems'. Proc IEE, Vol 118, 1971, pp 1503-1506.

40. PETERKA, V. 'On steady state minimum variance control strategy'. Kybernetika, 1972, Vol 8, No 3, pp 219-232.

41. ASTROM, K J. and WITTENMARK, B. 'Analysis of a self-tuning regulator for non-minimum phase systems'. IFAC Symposium on Stochastic Control, Budapest, 1974.

42. GRIMBLE, M J. 'A control weighted minimum variance controller for non-minimum phase systems', Int J Contr, 1981, Vol 33, No 4, pp 751-762.

43. SHAKED U. 'A general transfer-function approach to the discrete-time steady-state linear quadratic Gaussian control problem'. Int J Control, 1979, Vol 29, No 3, pp 361-386.

44. GRIMBLE, M J. 'Solution of the discrete-time stochastic optimal control problem in the Z-domain'. Int J Syst Sci 1979, Vol 10, No 12, pp 1369-1390

45. KUCERA, V. 'Discrete linear control, the polynomial equation approach'. Wiley, 1979.

46. ASTROM, K J and WITTENMARK, B. 'On self-tuning regulators'. Automatica, Vol 9, 1973, pp 185-199.

47. ASTROM, K J., BORISSON U, LJUNG L and WITTENMARK, B. 'Theory and application of self-tuning regulators'. Automatica, Vol 13, 1977, pp 457-476.

48. CLARKE, D W and GAWTHROP, P J. 'Self-tuning controller'. Proc IEE, Vol 122, No 9, Sept 1975, pp 929-934.

49. CLARKE, D W and GAWTHROP, P J. 'Self-tuning control'. PROC IEE, Vol 126, No 6, June 1979, pp 633-640.

50. WELLSTEAD, R E., PRAGER, D L and ZANKER, P. 'Pole assignment self-tuning regulator'. PROC IEE, 1979, 126 (8), pp 781-789.

51. WELLSTEAD, P E, EDMUNDS, J M., PRAGER, D L and ZANKER, P. 'Self-tuning pole/zero assignment regulators'. IJC 1979, 30, pp 1-26.

52. GRIMBLE, M J 'A control weighted self-tuning controller for non-minimum phase systems'. Int J Contr, 1982, Vol 36, No 4, pp 597-609

53. TSAY, Y T and SHIEH, L S. 'State-space approach for self-tuning feedback control with pole assignment'. PROC IEE, Vol 128, pt D, No 3, May 1981, pp 93-101.

54. WARWICK, K. 'Self-tuning regulators - A state space approach'. Int J Contr Vol 33, 1981, No 5, pp 839-859.

55. LAM, K.P. 'Implicit and explicit self-tuning controllers'. D Phil Thesis Univ of Oxford, 1980.

56. ASTROM, K J and EYKHOFF, P. 'System identification - a survey'. Automatica, Vol 7, pp 123-162, 1971

57. SYM, N. and WELLSTEAD, P E. 'Self-tuning prediction - multivariable applications in statistical forecasting'. PROC IEE, Int Conf Control and Applications, Univ of Warwick, 1981.

58. PRICE, C F. 'An analysis of the divergence problem in the Kalman filter'. IEEE Trans Aut Contr Vol AC-13, Dec 1968, pp 699-702.

59. WOODSIDE, C M. 'Estimation of the order of linear systems'. Automatica, Vol 7, 1971, pp 727-733.

60. SAGE, A P and MELSA, J L. 'Estimation theory with applications to communications and control', New York, McGraw-Hill.

61. ANDERSON, B D O and MOORE, J B. 'Optimal filtering', New Jersey, Prentice-Hall, 1979.

62. MOIR, T J and GRIMBLE, M J. 'Finite interval smoothing for discrete time systems'. Presented at the Int Conf Systems Science, Sept 15-18 1981, Wroclaw, Poland.

63. MOIR, T J and GRIMBLE M J. 'Optimal self-tuning filtering, prediction and smoothing for discrete multivariable processes'. SCP Research Rep No EEE/76/July 1981. Accepted for publication, IEEE Trans Aut Control.

64. GRIMBLE, M J. 'Observations weighted and LQG controllers for self-tuning systems'. Sheffield City Polytechnic, Research Report No EEE/72/January 1981.

65. GRIMBLE, M J., MOIR, T J and FUNG, P T K. 'Comparison of weighted minimum variance and LQG self-tuning controllers'. IEEE Conf, Hull, 19-21 July 1982.

66. ZADEH, L A. and RAGAZZINI, J R. 'An extension of Wiener's theory of prediction'. Journal of Applied Physics, Vol 21, July 1950, pp 645-655.

67. GRIMBLE, M J. 'New finite-time linear filters and smoothers'. Presented at the third IMA Conf on Control Theory, Univ of Sheffield, Sept 1980, Paper No 24.

68. O'REILLY, J. 'A finite time linear filter for discrete-time systems in singular case'. Submitted for publication in Int J Systems Science.

69. BLUM, M. 'Fixed memory least squares filters using recursion methods'. IRE Trans on Inf Theory, 1957, Vol IT-3, Sept, pp 178-182.

70. BUXBAUM, P J. 'Fixed-memory recursive filters', IEEE Trans Inf Theory, Vol IT-20, Jan 1974, pp 113-115.

71. BIERMAN, G J. 'Fixed memory least-squares filtering'. IEEE Trans Inf Theory, Vol IT-21, Nov 1975, pp 690-692.

72. LEE, R C K. 'The "Moving Window" approach to the problems of estimation and identification'. USA Air Force Report No SAMSO-TR-68-78, El Segundo, California.

73. LEVY B, KAILATH, T, LJUNG L and MORF M (1979). 'Fast time-invariant implementations for linear least-squares smoothing filters'. IEEE Trans. Aut Contr. AC-24, 5. Octoberpp770-774.

74. KAILATH, T, LEVY, B, LJUNG, L, MORF, M. 'Fast time-invariant implementations of Gaussian signal detectors'. IEE Transactions on Inf Theory, Vol IT-24, No 4, July 1978.

75. FRASER, D C. and POTTER, J E. 'The optimum linear smoother as a combination of two optimum linear filters, IEEE Trans Aut Contr, Vol AC-14, 1969, pp 387-390.

76. CHEN, C. 'On digital Wiener filters'. Proc IEEE, Dec 1976, pp 1736-1737.

77. PANUSKA V. 'A new form of the extended Kalman filter for parameter estimation in linear systems with correlated noise'. IEEE Trans Aut Control, Vol AC-25, No 2, 1980, pp 229-235.

78. FUNG, P T K and GRIMBLE, M J 'Self-tuning control of ship positioning systems', IEE Workshop on 'The Theory and Application of Self-tuning and Adaptive Control', 1981, University of Oxford.

79. GEVERS, M R. and KAILATH, T. 'An innovations approach to least-squares estimation - Part VI: Discrete-time innovations representations and recursive estimation'. IEEE Trans Aut Contr Vol AC-18, No 6, pp 588-560.

80. AASNAES, H B. and KAILATH T. 'An innovations approach to least-squares estimation - Part VII: Some applications of Vector ARMA models', IEEE Trans Aut Contr Vol AC-18, 1973, pp 601-609.

81. GUPTA, N and KAILATH, T. 'Corrections and extensions to innovations - Part VII: Some applications of vector ARMA models'. IEEE Trans Aut Contr, 23, 1978, pp 511-512.

82. HOLST, J. 'Adaptive prediction and recursive estimation'. Report LUTFD2/(TFRT-1013) 1-206, Lund Institute of Technology, Lund, Sweden.

83. TUEL, W G. 'Computer algorithm for spectral factorisation of rational matrices'. IBM Journal, March 1968, pp 163-170.

84. VOSTRY, Z. 'A numerical method of matrix spectral factorisation'. Kybernetika, Vol 8, No 5, pp 448-470.

85. GOODWIN, G C, RAMAGE, P J, and CAINES, P E. 'A globally convergent adaptive predictor'. Automatica, Vol 17, No 1, 1981, pp 135-140.

86. STRINTZIS, M G. 'A solution to the matrix factorisation problem'. IEEE Trans. Inf. Theory, Vol. IT-18, pp 225-232.

87. SOLO, V. 'Some aspects of recursive parameter estimation'. Int. J. Control, 1980, Vol. 32, No 3, pp 395-410.

88. SHAKED, U. 'Transfer function approach to the fixed-point continuous smoothing problem'. IEEE Trans. Aut. Control. Vol. AC-23, No 5, October 1978, pp 945-948.

89. ASTROM, K J. 'New implicit adaptive pole-zero-placement algorithms for non-minimum phase systems'. Lund Institute of Technology, Sweden, Report No. LUTFD2/(TFPT-7172/1-021) 1979.

90. ANDERSON, B D O and CHIRATTANANON. 'New linear smoothing formulas'. IEEE Trans. Aut. Cont. Vol. AC-17, Feb. 1972, pp 160-161.

91. KAILATH T. and FROST P. 'An innovations approach to least-squares estimation - Part II: Linear smoothing in additive white noise'. IEEE Trans. Aut. Control. Vol. AC-13, No 6, Dec. 1968, pp 655-660.

92. ANDERSON, B D O and MOORE J B. 'The Kalman-Bucy filter as a true time-varying Wiener filter'. IEEE Trans. on Systems, Man and Cybernetics, April 1971, Vol. SMC-1, No 2, pp 119-128.

93. ANDERSON, B D O and MOORE J B. 'Solution of a time-varying Wiener filtering problem'. Electronics letters, Vol. 3, No 12, Dec. 1967, pp 562-563.

94. ANDERSON, B D O, MOORE J B and LOO, S G. 'Spectral factorisation of time-varying covariance functions'. IEEE Trans. Inf. Theory, Vol. IT-15, Sept. 1969, pp 550-557.

95. GUIDORZI, R. 'Canonical structures in the identification of multi-variable systems'. Automatica, Vol. 11, pp 361-374, 1975

96. SINHA, N K and KWONG, Y H. 'Recursive estimation of the parameters of linear multivariable systems', Automatica, Vol. 15, 1979, pp 471-475.

97. TSE, E and ANTON, J J. 'On the identifiability of parameters'. IEEE Trans. Aut. Control. Vol AC-17, Oct. 1972, pp 637-646.

98. TSE, E and WEINERT, H. 'Correction and extension of "on the identifiability of parameters" '. IEEE Trans. Aut. Contr. (corresp.), Vol. AC-18, Dec. 1973, pp 687-688.

99. KWAKERNAAK, H and SIVAN R. 'Linear optimal control systems'. Wiley - Interscience, New York, 1972.

100. LJUNG, L., SODERSTROM, T. and GUSTAVSSON, I. 'Counterexamples to general convergence of commonly used recursive identification method'. IEEE Trans. Aut. Control, Vol. AC-20, No 5, 1975, pp 643-652.

101. BARNETT, S. 'Matrices in Control Theory'. Van Nostrand Reinhold, London, 1971.

102. JAKEMAN, A. and YOUNG, P. 'Joint parameter/state estimation'. Electronics letters, 13th September 1979, Vol. 15, No.19, pp 582-583.

103. KAILATH, T. 'Some alternatives in recursive estimation'. Int J Cont, 1980, Vol.32, No.2, pp 311-328.

104. FUNG, P.T.K. and GRIMBLE, M.J. 'Dynamic Ship Positioning Using a Self-Tuning Kalman Filter' Accepted for publication, IEEE Trans. Aut.Control.

APPENDIX 1 : PROOF OF THEOREM 2.1

Let the weighting sequence of the optimal smooting filter by denoted by $h(k;N,\ell)$ for all k, where N is the interval length and ℓ is the smoothing delay. The filter response for $\forall$ k then becomes

$$\hat{\underline{x}}(k_1/k) \triangleq \sum_{i=-\infty}^{k} \left[\underline{h}(k-i;N,\ell)\underline{z}(i) + \psi(k-i+1;\ell)\delta(i-1)\hat{\underline{x}}_0'\right] \quad \text{(A1.1)}$$

$$= \psi(k;\ell)\hat{\underline{x}}_0 + \sum_{i=-\infty}^{k} h(k-i;N,\ell)\underline{z}(i) \quad (k > 0) \quad \text{(A1.2)}$$

Note that since the plant has an impulse input at time zero, the smoother has an input of magnitude $\hat{\underline{x}}_0'$ in order to optimize the zero input response. In the finite time problem, $\hat{\underline{x}}_0'$ is the initial state estimate. Define the vector

$$\underline{d}_0(k;N,\ell) = \psi(k;\ell)\hat{\underline{x}}_0' \quad \forall\ k > 0 \quad \text{(A1.3)}$$

$$= 0 \quad \forall\ k \leq 0 \quad \text{(A1.4)}$$

Equation (A1.2) then becomes

$$\hat{\underline{x}}(k_1/k) = \underline{d}_0(k;N,\ell) + \sum_{i=-\infty}^{k} h(k-i;N,\ell)\underline{z}(i) \quad \text{(A1.5)}$$

Necessary and sufficient conditions for optimality

A necessary and sufficient condition that $\hat{\underline{x}}(Kl/N)$ be a minimum variance smoother for $\underline{x}_i(Kl)$ is that the estimator impulse response must satisfy the Wiener-Hopf equation.

$$\mathrm{COV}\left[\underline{x}_i(Kl),\underline{z}(\lambda)\right] = \sum_{i=-\infty}^{N} \hat{h}(N-i;N,\mathscr{L})\mathrm{COV}\left[\underline{z}(i),\underline{z}(\lambda)\right] \quad \text{(A1.6)}$$

and the smoother zero-input response $\hat{\underline{d}}_0(k;N,\ell)$ be such that:

$$E\{\hat{\underline{x}}(K1/N)\} = \bar{\underline{x}}_i(K1) \qquad (A1.7)$$

This is proved using the projection theorem in Appendix 3. The Wiener-Hopf equation can now be expanded and then transformed into the frequency domain, as follows:

Expansion of the Wiener-Hopf Equation

The LHS of equation (A1.6) can now be expanded by using the results in [13]

$$\begin{aligned}\mathrm{COV}\left[\underline{x}_i(K1), \underline{z}(\lambda)\right] = {} & \phi(K1-\lambda-1)DGU(K1-\lambda-1) \\ & + \Phi(K1)A\Sigma_0 A^T\Phi^T(\lambda)C^T \\ & - \psi(N;\ell)\left[\phi(-\lambda-1)DGU(-\lambda-1)\right. \\ & \left. + \phi_{x_0x}(-\lambda)C^T\right] \\ & + \sum_{i=-\infty}^{\lambda-1}(\Phi(K1-i)DQD^T\Phi^T(\lambda-i)C^T\end{aligned} \qquad (A1.8)$$

where

$$\phi_{x_0x}(-\lambda) = E\{\underline{x}(0)\underline{x}^T(\lambda)\} \qquad (A1.9)$$

The RHS of the Wiener-Hopf equation is given as [13]

$$\sum_{i=-\infty}^{N}\hat{h}(N-i;N,\ell)\mathrm{COV}\left[\underline{z}(i),\underline{z}(\lambda)\right]$$

$$\begin{aligned} = {} & \sum_{i=-\infty}^{N}\hat{h}(N-i;N,\ell)\sum_{j=-\infty}^{i-1}\sum_{r=-\infty}^{\lambda-1}w(i-j)Q\delta(j-r)w^T(\lambda-r) \\ & + \sum_{i=-\infty}^{N}\hat{h}(N-i;N,\ell)\left[R\delta(i-\lambda) + w(i-\lambda)G + G^Tw^T(\lambda-i)\right] \\ & + \sum_{i=-\infty}^{N}\hat{h}(N-i,N,\ell)C\Phi(i)A\Sigma_0A^T\Phi^T(\lambda)C^T \end{aligned} \qquad (A1.10)$$

The fully expanded Wiener-Hopf equation now comprises equations (A1.9) and (A1.10) and can be simplified by using the result

$$\left[\phi(K1) - \sum_{i=1}^{N} \hat{h}(N-i;N,\ell)C\phi(i)\right]\Sigma_0 A^T\Phi^T(\lambda)C^T$$

$$= \psi(N;\ell)\Sigma_0 A^T\Phi^T(\lambda)C^T U(\lambda) \qquad \text{(A1.11)}$$

for $\lambda > 0$. The complete Wiener-Hopf equation can be considerably simplified by substituting for $\lambda = K1 - k$ and by using the substitution of equation (A1.11). The condition for optimality may now be expressed as:

$$G_N(k) = 0 \quad \forall\ k \geq -\ell \qquad \text{(A1.12)}$$

where

$$G_N(k) \triangleq$$

$$\psi(N;\ell)\Big[\Sigma_0 A^T\Phi^T(K1-k)C^T U(K1-k) - \phi(k-K1-1)DGU(k-K1-1)$$

$$- \phi_{x_0 x}(k-K1)C^T\Big] + \phi(k-1)DGU(k-1)$$

$$- \sum_{i=-\infty}^{N} \hat{h}(N-i;N,\ell)\Bigg\{\sum_{j=-\infty}^{i-1}\sum_{r=-\infty}^{K1-k-1} w(i-j)Q\delta(j-r)w^T(K1-k-r)$$

$$+ R\delta(i+k-K1) + w(i+k-K1)G + G^T w^T(K1-k-i)\Bigg\}$$

$$+ \sum_{i=-\infty}^{K1-k-1} \Phi(K1-i)DQw^T(K1-k-i) \qquad \text{(A1.13)}$$

where the smoothing lag is defined to be

$$\ell = N - K1 \qquad \text{(A1.14)}$$

Transforming into the frequency domain

The transformation into the frequency domain is obtained by taking bilinear z transforms of equation (A1.10). Using the results of Appendix 2 the frequency domain expression becomes

$$G_N(z) = \psi(N;\ell)\Big[\Sigma_0 A^T\Phi^T(z^{-1})z^{-K1}C^T - \Phi(z)z^{-K1}DG - \Phi(z)DQD^T\Phi^T(z^{-1})C^Tz^{-K1}\Big] + \Phi(z)DG + \Phi(z)DQW^T(z^{-1}) - \hat{H}(z;N,\ell)\Big[W(z)QW^T(z^{-1}) + R + W(z)G + G^TW^T(z^{-1})\Big]z^{\ell} \qquad (A1.15)$$

The generalised spectral factor, $\Delta(z)$, of the discrete, observation spectral density is such that [27]

$$\Delta(z)\Delta^T(z^{-1}) = W(z)QW^T(z^{-1}) + R + W(z)G + G^TW^T(z^{-1}) \qquad (A1.16)$$

where $\Delta(z)$ is a rational function of z whose poles are those of W(z) and whose inverse is analytic outside the unit circle.

Multiplying both sides of equation (A1.15) by $\Delta^{-T}(z^{-1})z^{-\ell}$, the frequency domain expression becomes:

$$G_N(z)z^{-\ell}\Delta^{-T}(z^{-1}) = \psi(N;\ell)\Big[\Sigma_0 A^T\Phi^T(z^{-1})C^T - \Phi(z)DG - \Phi(z)DQD^T\Phi^T(z^{-1})C^T\Big]\Delta^{-T}(z^{-1})z^{-N} + \Phi(z)D(G + QW^T(z^{-1}))\Delta^{-T}(z^{-\ell})z^{-\ell} - \hat{H}(z;N,\ell)\Delta(z) \qquad (A1.17)$$

Referring to equation (A1.12), it can be seen that the cross spectral density term $G_N(z)\Delta^{-T}(z^{-1})z^{-\ell}$ is zero for all $k \geq 0$. The smoothing filter is required to be causal, so that all uncausal terms must be removed. This is achieved in the frequency domain by removing all terms in expression (A1.17) whose poles lie outside the unit disk (assuming that the signal model is stable). Equation (A1.17), solved for the causal optimal smoother, then becomes:

$$H(z;N,\ell) = \Big\{\Phi(z)D(G + QW^T(z^{-1}))\Delta^{-T}(z^{-1})z^{-\ell} - \psi(N;\ell)\big[\overline{\Phi}(z)D(G + QW^T(z^{-1}))\Delta^{-T}(z^{-1}) - \Sigma_0 A^T\Phi^T(z^{-1})C^T\Delta^{-T}(z^{-1})\big]z^{-N}\Big\}_+ \Delta^{-1}(z) \quad \text{(A1.18)}$$

And by using the simplifications of equations (2.28-2.31), equation (A1.18) can be expressed as:

$$\hat{H}(z;N,\ell) = n_{11}(z,\ell)\Delta^{-1}(z) + \psi(N;\ell)\big[\overline{n}_{12}(z,N) + n_2(z,N)\big]\Delta^{-1}(z) \quad \text{(A1.19)}$$

This is the optimal, finite time smoothing filter in the z-domain. The calculation of the $\psi(N;\ell)$ matrix may now be considered.

Calculation of $\psi(N;\ell)$ matrix

From the necessary and sufficient conditions for optimality (Appendix 3), the zero input response must satisfy

$$\hat{\underline{d}}_0(N;N,\ell) = \overline{\underline{x}}_1(K1) - \sum_{i=-\infty}^{N} \hat{h}(N-i;N,\ell)\overline{\underline{z}}(i) \quad \text{(A1.20)}$$

and since $\underline{\bar{z}}(i) = C\underline{\bar{x}}(i) + \underline{\bar{v}}(i)$ then

$$\underline{\hat{d}}_0(N;N,\ell) = \underline{\bar{x}}_i(K1) - \sum_{i=-\infty}^{N} \hat{h}(N-i;N,\ell)C\underline{\bar{x}}(i) \qquad (A1.22)$$

Since $\underline{\bar{x}}(i) = 0$ for $i \leq 0$ and by using equation (2.18):

$$\underline{\hat{d}}_0(N:N,\ell) = \left[\phi(K1) - \sum_{i=1}^{N} \hat{h}(N-i;N,\ell)C\phi(i)\right]\underline{m}_0 \qquad (A1.23)$$

and by comparison of equations (A1.23) and (A1.3), the $\psi(N;\ell)$ matrix becomes

$$\psi(N;\ell) = \phi(K1) - \sum_{i=1}^{N} \hat{h}(N-i;N,\ell)C\phi(i) \qquad (A1.24)$$

and

$$\underline{\hat{x}}_0' = \underline{m}_0$$

From equation (A1.19)

$$H(z;N,\ell)C\Phi(z)A = \left[n_{11}(z,\ell) + \psi(N;\ell)(n_{12}(z,N) + n_2(z,N)\right]M(z) \qquad (A1.25)$$

Using the following results

$$\{M^T(z^{-1})z^{-N}\}_+ = \mathcal{Z}_1\left[\bar{M}^T(N-k)U(N-k)\right]$$

$n_{11}(k,\ell) = \mathcal{Z}_1^{-1}(n_{11}(z,\ell))$ and $n_{12}(k,N) = \mathcal{Z}_1^{-1}(n_{12}(z,N))$. The inverse transform of equation (A1.25) into equation (A1.24) gives

$$\psi(N;\ell) = \phi(K1) - \sum_{k=1}^{N} n_{11}(k,\ell)M(N-k) - \psi(N;\ell)\sum_{k=1}^{N} n_{12}(k,N)M(N-k) - \psi(N;\ell)\Sigma_0\left(\sum_{k=1}^{N} M^T(N-k)M(N-k)\right) \qquad (A1.26)$$

Hence after re-arranging equation (A1.26)

$$\psi(N;\ell) = \left[\bar{\phi}(K1) - I_{11}(N,\ell)\right]\left[\underline{\bar{I}}_n + I_{12}(N) + \Sigma_0 S(N)\right]^{-1} \qquad (A1.27)$$

where $I_{11}(N,\ell)$, $I_{12}(N)$ and $S(N)$ are given in equations (2.35 - 2.37) respectively.

APPENDIX 2 : $\mathcal{Z}_2$ TRANSFORMATIONS OF SUMMATION TERMS

Consider the following

$$\mathcal{Z}_2\left[\sum_{i=-\infty}^{N}\sum_{j=-\infty}^{i=1}\sum_{r=-\infty}^{K1-k-1}\hat{h}(N-i;N,\ell)w(i-j)Q\delta(j-r)\right.$$

$$\left.x\ w^T(K1-k-r)\right] \qquad (A2.1)$$

By definition of the two sided z transform, equation (A2.1) becomes

$$\sum_{k=-\infty}^{\infty}\sum_{i=-\infty}^{N}\sum_{j=-\infty}^{i-1}\sum_{r=-\infty}^{K1-k-1}\hat{h}(N-i;N,\ell)w(i-j)Q\delta(j-r)$$

$$x\ w^T(K1-k-r)z^{-k} \qquad (A2.2)$$

After the removal of the delta function, by substituting:

$$p = K1-k-j$$

$$m = i-j$$

$$s = N-i$$

(A2.2) now becomes

$$\sum_{s=0}^{\infty}\sum_{m=1}^{\infty}\sum_{p=1}^{\infty}\hat{h}(s;N,\ell)w(m)Qw^T(p)z^{-K1}z^{N}z^{-s}z^{-m}z^{p} \qquad (A2.3)$$

Since $N-K1 = \ell$ and $w(0) = 0$

$$\sum_{s=0}^{\infty}\sum_{m=0}^{\infty}\sum_{p=0}^{\infty}\hat{h}(s;N,\ell)z^{-s}w(m)z^{-m}Qw^T(p)z^{p}z^{\ell}$$

$$= \hat{H}(z;N,\ell)W(z)QW^T(z^{-1})z^{\ell} \qquad (A2.4)$$

Employing the same technique

$$\sum_{k=-\infty}^{\infty}\sum_{i=-\infty}^{N}\hat{h}(N-i;N,\ell)\left[R\delta(i+k-K1) + w(i+k-K1)G\right.$$

$$\left. + G^Tw^T(K1-k-i)\right]z^{-k}$$

$$= \hat{H}(z;N,\ell)(R + W(z)G + G^TW^T(z^{-1}))z^{\ell} \qquad (A2.5)$$

and

$$\sum_{k=-\infty}^{\infty} \sum_{i=-\infty}^{K1-k-1} \Phi(K1-i)DQw^T(K1-k-i)z^{-k}$$

$$= \Phi(z)DQW^T(z^{-1}) \qquad \text{(A2.6)}$$

To find $\mathcal{Z}_2(\phi_{x_0x}(k-K1)$

From equation (A1.10)

$$\phi_{x_0x}(-\lambda) = E(\underline{x}(0)\underline{x}^T(\lambda))$$

and from equation (2.17)

$$E(\underline{x}(0)) = \sum_{i=-\infty}^{-1} \Phi(-i)DE(\underline{\omega}(i))$$

and from equation (2.16)

$$E(\underline{x}^T(\lambda)) = \sum_{j=-\infty}^{\lambda-1} E(\underline{\omega}^T(j))D^T\Phi^T(\lambda-j)$$

which gives

$$E(\underline{x}(0)\underline{x}^T(\lambda)) = \sum_{i=-\infty}^{-1} \sum_{j=-\infty}^{\lambda-1} \Phi(-i)DQ\delta(i-j)D^T\Phi^T(\lambda-j)$$

Thus

$$\mathcal{Z}_2(\phi_{x_0x}(k-K1))$$

$$= \sum_{k=-\infty}^{\infty} \sum_{i=-\infty}^{-1} \sum_{j=-\infty}^{K1-k-1} \Phi(-i)DQ\delta(i-j)D^T\Phi^T(\lambda-j)z^{-k}$$

$$= \sum_{r=1}^{\infty} \sum_{t=1}^{\infty} \Phi(r)DQD\Phi^T(t)z^{-r}z^{t}z^{-K1}$$

$$= \Phi(z)DQD^T\Phi^T(z^{-1})z^{-K1} \qquad \text{(A2.7)}$$

APPENDIX 3 : THE DISCRETE WIENER-HOPF EQUATION

Theorem

A necessary and sufficient condition that $\underline{\hat{x}}(K1/N)$ be a minimum variance smoother for $\underline{x}_i(K1)$ is that the matrix function $\hat{h}(N-i;N,\ell)$ should satisfy the Wiener-Hopf equation ($K1 < N$)

$$\mathrm{COV}\left[\underline{x}_i(K1),\ \underline{z}(\lambda)\right] = \sum_{i=-\infty}^{N} \hat{h}(N-i;N,\ell)\ \mathrm{COV}\left[\underline{z}(i),\ \underline{z}(\lambda)\right] \tag{A3.1}$$

for all $\lambda\varepsilon(-\infty,N)$, and the smoother zero input response $\underline{\hat{d}}_0(N;N,\ell)$ satisfy the equation:

$$\underline{d}_0(N;N,\ell) = \underline{\bar{x}}_i(K1) - \sum_{i=-\infty}^{N} \hat{h}(N-i;N,\ell)\underline{\bar{z}}(i) \tag{A3.2}$$

Proof

Let X denote the space generated by the random vectors $\{\underline{x}_i(K1)\}$ and let $U \subseteq X$ represent the subspace generated by the set of observations $\{\underline{z}(k)\}$ where the elements $u(K1)\varepsilon U$ are given by

$$u(K1) = \underline{d}_0(N;N,\ell) + \sum_{\lambda=-\infty}^{N} h(N-\lambda;N,\ell)z(\lambda) \tag{A3.3}$$

where $\underline{d}_0(N;N,\ell)$ is an unknown deterministic vector. From the orthogonal projection theorem and the decomposition theorem [12], the norm $||\underline{x}_i-u||$ is minimised by a vector $\underline{\hat{x}}\ \varepsilon U$ if $(\underline{x}_i-\hat{x})$ is orthogonal to U.

From the above, the inner product becomes

$$E\{(\underline{x}_i(K1) - \underline{\hat{x}}(K1/N))u^T(K1)\} = 0 \tag{A3.4}$$

Substituting for u(K1) from (A3.3) gives

$$(\underline{\bar{x}}_i(K1) - E\{\hat{\underline{x}}(K1/N)\})\underline{d}_0^T(N;N,\ell) + \sum_{\lambda=-\infty}^{N} E\{(\underline{x}_i(K1) - \hat{\underline{x}}(K1/N))\underline{z}^T(\lambda)\} \times h^T(N-\lambda;N,\ell) = 0 \quad \text{(A3.5)}$$

The sufficient condition for optimality becomes:

$$E\{\hat{\underline{x}}(K1/N)\} = \underline{\bar{x}}_i(K1) \quad \text{(A3.6)}$$

and

$$E\{(\underline{x}_i(K1) - \hat{\underline{x}}(K1/N))\underline{z}^T(\lambda)\} = 0 \text{ for } \lambda\varepsilon(-\infty,N) \quad \text{(A3.7)}$$

since $\underline{d}_0(N;N,\ell)$ and $h(N-\lambda;N,\ell)$ are arbitrary matrices. The necessity of the theorem can be verified by noting that if these matrices are chosen as

$$\underline{d}_0(N;N,\ell) = \underline{\bar{x}}_i(K1) - E\{\hat{\underline{x}}(K1/N)\} \quad \text{(A3.8)}$$

and

$$h(N-\lambda;N,\ell) = E\{(\underline{x}_i(K1) - \hat{\underline{x}}(K1/N))\underline{z}^T(\lambda)\} \quad \text{(A3.9)}$$

then the terms in (A3.5) are positive definite. It follows that these terms must be identically equal to zero, establishing the necessity of the theorem.

The first condition for optimality requires that the smoother be unbiased and the second gives:

$$E\{\underline{x}_i(K1)\underline{z}^T(\lambda)\} = \hat{d}_0(N;N,\ell)\underline{\bar{z}}^T(\lambda) + \sum_{i=-\infty}^{N} \hat{h}(N-i;N,\ell)E\{\underline{z}(i)\underline{z}^T(\lambda)\} \quad \text{(A3.10)}$$

Combining (A3.10) and (A3.8) gives

$$E\{\underline{x}_i(K1)\underline{z}^T(\lambda)\} = \sum_{i=-\infty}^{N} \hat{h}(N-i;N,\ell)E\{\underline{z}(i)\underline{z}^T(\lambda)\} \quad \text{(A3.11)}$$

This is the required form of the discrete Wiener-Hopf

APPENDIX 4

To prove

$$\Phi(z)DQD^{T}W^{T}(z^{-1})\Delta^{-T}(z^{-1}) = P\left[A^{T}\Phi^{T}(z^{-1}) + I\right] \times C^{T}\Delta^{-T}(z^{-1}) + \Phi(z)AKR_{\varepsilon}^{\frac{1}{2}} \quad (A4.1)$$

After manipulation of (2.44) (Arcasoy [25])

$$(zI-A)P(z^{-1}I-A^{T}) + AP(z^{-1}I-A^{T}) + (zI-A)PA^{T} + AKCPA^{T} = DQD^{T} \quad (A4.2)$$

or

$$\Phi(z)^{-1}P\Phi^{-T}(z^{-1}) + AP\Phi^{-T}(z^{-1}) + \Phi(z)^{-1}PA^{T} + AKCPA^{T} = DQD^{T} \quad (A4.3)$$

Thus pre-multiplying (A4.3) by $\Phi(z)$ and post multiplying by $\Phi^{T}(z^{-1})C^{T}$ gives

$$PC^{T} + \Phi(z)APC^{T} + PA^{T}\Phi^{T}(z^{-1})C^{T} + \Phi(z)AKCPA^{T}\Phi^{T}(z^{-1})C^{T} = \Phi(z)DQW^{T}(z^{-1}) \quad (A4.4)$$

$$= PC^{T} + PA^{T}\Phi^{T}(z^{-1})C^{T} + \Phi(z)\left[AKR_{\varepsilon}K^{T}A^{T}\Phi^{T}(z^{-1})C^{T} + AKR_{\varepsilon}\right] = \Phi(z)DQW^{T}(z^{-1}) \quad (A4.5)$$

with K defined as in (2.45) and R_{ε} as in (2.46).

(A4.5) now becomes

$$P\left[A^{T}\Phi^{T}(z^{-1}) + I\right]C^{T} + \Phi(z)AKR_{\varepsilon}\left[I + K^{T}A^{T}\Phi^{T}(z^{-1})C^{T}\right] = \Phi(z)DQW^{T}(z^{-1}) \quad (A4.6)$$

From (2.48)

$$F^{T}(z^{-1}) = K^{T}A^{T}\Phi^{T}(z^{-1})C^{T} + I \quad (A4.7)$$

and using (2.50)

$$R_{\varepsilon}^{-\frac{1}{2}}\Delta^{T}(z^{-1}) = F^{T}(z^{-1})$$

Hence (A4.6) can be re-written as

$$\Phi(z)DQW^T(z^{-1}) = P\left[\underline{A}^T\Phi^T(z^{-1}) + \underline{I}\right]C^T + \Phi(z)AKR_\varepsilon^{\frac{1}{2}}\Delta^T(z^{-1}) \qquad \text{(A4.8)}$$

Clearly

$$\Phi(z)DQW^T(z^{-1})\Delta^{-T}(z^{-1}) = P\left[\underline{A}^T\Phi^T(z^{-1}) + \underline{I}\right]C^T\Delta^{-T}(z^{-1}) + \Phi(z)AKR_\varepsilon^{\frac{1}{2}} \qquad \text{(A4.9)}$$

which proves equation (A4.1).

APPENDIX 5 : PROOF OF THEOREM 3.1

Using the result of Theorem 2.1 as $N \to \infty$ and $G = 0$ gives the infinite interval or Wiener smoother. From (2.39) this becomes

$$H(z;\ell) = \{W(z)QW^T(z^{-1})\Delta^{-T}(z^{-1})z^{-\ell}\}_+\Delta^{-1}(z) \qquad (A5.1)$$

Substituting for $\Delta(z)$ and $\Delta^T(z^{-1})$ from (2.49) and (2.50) (Arcasoy [27]) gives

$$H(z;\ell) = \{W(z)QW^T(z^{-1})F^T(z^{-1})^{-1}R_\varepsilon^{-\frac{1}{2}}z^{-\ell}\}_+ \times R_\varepsilon^{-\frac{1}{2}}F^{-1}(z) \qquad (A5.2)$$

$$= \{W(z)QW^T(z^{-1})F^T(z^{-1})^{-1}z^{-\ell}\}_+R_\varepsilon^{-1}F^{-1}(z) \qquad (A5.3)$$

APPENDIX 6 : POLYNOMIAL MATRIX EXPANSION ALGORITHM

This algorithm enables the computation of the $\ell + 1$ coefficient matrices of the polynomial matrix $P(z)$, where

$$P(z) \triangleq A(z)^T D^T(z)^{-1}$$

$$= I_r + P_1 z + P_2 z^2 + \ldots + P_\ell z^\ell + \ldots$$

and

$$A(z) = I_r + A_1 z + A_2 z^2 + \ldots A_{n_a} z^{n_a}$$

$$D(z) = I_r + D_1 z + D_2 z^2 + \ldots D_{n_d} z^{n_d}$$

The coefficient matrices A_i, D_i, $i = 1,2, \ldots$. are r-square and it is assumed that n_a and n_d.

The solution for the matrix coefficients is an extension of the scalar case given in [45]. The kth P matrix coefficient is given by the formulae:

$$P_k = A_k^T - \sum_{j=1}^{s} P_{k-j} D_j^T \qquad \text{for } k \leq n_a$$

$$= - \sum_{j=1}^{s} P_{k-j} D_j^T \qquad \text{for } k > n_a$$

where

$$s = \text{MIN}(n_a, k)$$

This leads to the following computational algorithm.

Algorithm

Data required: smoothing lag ℓ and degree of process n_a.

Step

1 $k = 0$

2 Set $R_\sigma = \underline{0}$

3 If $k = 0$ GO TO 8

4 $j = 0$

5 $j = j + 1$

6 $R_\sigma = R_\sigma + P_{k-j} D_j^T$

7 If $j < \text{MIN}(n_a, k)$ GO TO 5

8 If $k \leq n_a$ then $P_k = A_k^T - R$ ELSE COMPUTE $P_k = -R_\sigma$

9 If $k = \ell$ STOP

10 $k = k + 1$

11 GO TO 2

To initialise the algorithm $P_o = A_o = I_r$. Finally, the matrix $P_\ell(z)$ may be defined as:

$$P_\ell(z) = I_r + P_1 z + P_2 z^2 + \ldots + P_\ell z^\ell$$

APPENDIX 7 : SMOOTHED AND FILTERED ESTIMATES

The smoothed estimate $\underline{\hat{y}}(k - \ell|k)$ can be expressed in terms of the filtered estimate $\underline{\hat{y}}(k|k)$ as shown below. From (3.28):

$$\underline{\hat{y}}(k - \ell|k) = \underline{z}(k - \ell) - RP_{\ell}(z)R_{\varepsilon}^{-1}\underline{\varepsilon}(k - \ell)$$

and

$$\underline{z}(k) = \underline{\hat{y}}(k|k) + RR_{\varepsilon}^{-1}\underline{\varepsilon}(k)$$

thence

$$\underline{\hat{y}}(k - \ell|k) = \underline{\hat{y}}(k - \ell|k - \ell) + R(I_r - P_{\ell}(z))R_{\varepsilon}^{-1}\underline{\varepsilon}(k - \ell) \qquad (A7.1)$$

$$= \underline{\hat{y}}(k - \ell|k - \ell) - R\sum_{i=1}^{\ell} P_i R_{\varepsilon}^{-1}\underline{\varepsilon}(k - \ell + i) \qquad (A7.2)$$

or alternatively

$$\underline{\hat{y}}(k|k + \ell) = \underline{\hat{y}}(k|k) - \sum_{i=1}^{\ell} RP_i R_{\varepsilon}^{-1}\underline{\varepsilon}(k + i) \qquad (A7.3)$$

by substituting $k + \ell = b$ in (A7.3) above

$$\underline{\hat{y}}(k|b) = \underline{\hat{y}}(k|k) - \sum_{i=1}^{b-k} RP_i R_{\varepsilon}^{-1}\underline{\varepsilon}(k + i) \qquad (A7.4)$$

$$= \underline{\hat{y}}(k|k) - \sum_{i=1+k}^{b} RP_{i-k} R_{\varepsilon}^{-1}\underline{\varepsilon}(i) \qquad (A7.5)$$

and (A7.5) may be compared with other more well-known results, e.g. Anderson and Chirarattananon [90] and Kailath and Frost [91].

APPENDIX 8 : COMPARISON OF TRANSFER-FUNCTION ESTIMATORS

This appendix compares the derived transfer-function estimators of Chapter 3 with the related work of Shaked and Barrett.

Consider first the predictor. From (3.30) with $\ell_1 = 1$, for a one step ahead predictor:

$$\hat{H}(z;1) = \{F(z)z\}_+ F^{-1}(z) \tag{A8.1}$$

Expanding F(z) in a convergent matrix series in z^{-1} gives

$$\hat{H}(z;1) = \{(I_r + F_1 z^{-1} + F_2 z^{-2} + \ldots)z\}_+ F^{-1}(z) \tag{A8.2}$$

where F_i, i = 1,2 .... are Markov parameters. Now taking the causal part gives:

$$\hat{H}(z;1) = (F(z) - I_r)zF^{-1}(z) \tag{A8.3}$$

$$= z(I_r - F^{-1}(z)) \tag{A8.4}$$

From (3.34)

$$\underline{\hat{y}}(k+1|k) = \hat{H}(z;1)\underline{z}(k)$$

$$= \hat{H}'(z;1)\underline{z}(k+1) \tag{A8.5}$$

where

$$\hat{H}'(z;1) = I_r - F^{-1}(z) \tag{A8.6}$$

Equation (A8.6) for the one step ahead predictor is the same as which has been derived by Shaked [27] and Barrett [26] for continuous time. The aim of their work is to show how the steady-state return difference matrix F(z) of the Kalman filter is useful in determining the transfer function solution to linear estimation problems. Clearly, the predictor is defined entirely by only F(z).

The optimal filter transfer function matrix may also be written solely in terms of F(z). From (3.27) with $\ell = 0$)

$$H(z;0) = I_r - RR_\varepsilon^{-1}F^{-1}(z) \tag{A8.7}$$

from (3.14)

$$F(0) = I_r + C\Phi(0)AK \tag{A8.8}$$

$$= I_r - CK \tag{A8.9}$$

Using (3.15) for K

$$F(0) = I_r - CPC^TR_\varepsilon^{-1} \tag{A8.10}$$

since

$$R_\varepsilon = R + CPC^T$$

on substituting $CPC^T = R_\varepsilon - R$ into (A8.10) yields

$$F(0) = I_r - (R_\varepsilon - R)R_\varepsilon^{-1}$$

$$= RR_\varepsilon^{-1} \tag{A8.11}$$

Hence the optimal filter becomes

$$H(z;0) = I_r - F(0)F^{-1}(z) \tag{A8.12}$$

which is expressed exclusively in terms of the return difference matrix.