NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

CONTINUOUS-TIME CONSTRAINED LEAST-SQUARES ALGORITHMS FOR RECURSIVE PARAMETER ESTIMATION OF STOCHASTIC LINEAR SYSTEMS BY A STABILIZED OUTPUT ERROR METHOD

A.J. Udink ten Cate

September 1985
WP-85-54

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

ABSTRACT

Discrete-time least-squares algorithms for recursive parameter estimation have continuous-time counterparts, which minimize a quadratic functional. The continuous-time algorithms can also include (in)equality constraints. Asymptotic convergence is demonstrated by means of Lyapunov methods. The constrained algorithms are applied in a stabilized output error configuration for parameter estimation in stochastic linear systems.

CONTINUOUS-TIME CONSTRAINED LEAST-SQUARES ALGORITHMS FOR RECURSIVE PARAMETER ESTIMATION OF STOCHASTIC LINEAR SYSTEMS BY A STABILIZED OUTPUT ERROR METHOD

A.J. Udink ten Cate

1. INTRODUCTION

The subject of recursive parameter estimation in dynamic systems has received considerable attention in recent years (Eykhoff, 1974; Goodwin and Payne, 1977; Ljung and Söderström, 1983). In control, attention has been focused mainly on the estimation of parameters of discrete-time models from sampled data (Young, 1981) with a number of exceptions (Eykhoff, 1974; Bohn, 1982; Young, 1981). In many cases, however, the parameters of continuous-time models have to be estimated from experimental data. The advent of modern computer-operated dataloggers has made relatively high sampling rates feasible, stimulating interest in continuous-time algorithms which operate on quasi-continuous measurements and which can directly update a (physical) continuous model with some known and some unknown parameters. No constraints are generally imposed in recursive algorithms, although this is sometimes done in the framework of stochastic approximation (Kushner and Clarke, 1978). In many experimental situations, however, such constraints exist.

This report presents a class of continuous-time algorithms which minimize a quadratic functional of the difference between the observed and the predicted output. A modified form of the usual "equation error" adopted in, e.g., Young (1981), Lion (1967), Landau (1979) is taken as a measure of this difference and minimized. In contrast to other algorithms of this type (Young, 1981; Solo, 1980), the starting values of the estimates appear explicitly in the functional. The algorithms are characterized by exponential convergence of the parameter error (Anderson and Johnson, 1982) and can be regarded as continuous versions of the recursive least-squares method. These algorithms are then extended to handle equality constraints; inequalities can be handled using penalty functions. By the very nature of penalty functions, inequalities are treated mildly, which means that the estimates are not strictly confined to the feasible area. This feature makes the algorithms suitable for application in an output error parameter-estimation scheme for stochastic linear systems.

Output error estimation schemes are derived from model reference adaptive systems (Shackloth and Butchart, 1965; Parks, 1966; Landau, 1976, 1979). In output error methods, the error is filtered in order to ensure that the estimation procedure converges. The filter is based on *a priori* knowledge of the unknown system and is designed using Lyapunov or hyperstability theories. For discrete-time systems, Landau (1978) presented an approach which used an extended parameter vector in order to remove the need for *a priori* information. Results for discrete stochastic systems are given in Dugard and Landau (1980).

This paper proposes an output error scheme using the extended parameter vector approach for continuous-time algorithms. In the stochastic case it is not possible to ensure convergence for the extended parameters, which destabilizes the whole estimation scheme and leads to biased estimates. Using the constrained algorithms mentioned before, the extended parameters can be bounded within a set S_c, so that the scheme remains stable. The results of Ljung (1977) on the asymptotic convergence of stochastic systems (Ljung and Söderström, 1983) can then be used to formulate conditions on the set S_c for the convergence of the estimation procedure.

This paper is structured as follows. Section 2 presents the continuous-time least-squares algorithm and demonstrates its exponential convergence. Section 3 extends the algorithm to include constraints. The output error parameter-estimation algorithm is introduced in Section 4 for deterministic systems, and in Section 5 the analysis is carried out in a stochastic environment using a method proposed by Ljung (1977). Simulation examples are presented in Section 6 and the paper ends with some conclusions.

2. CONTINUOUS LEAST-SQUARES METHODS

A linear deterministic univariate system can be represented by the equation

$$y(t) = \frac{B(s)}{A(s)} u(t) \tag{2.1}$$

where y(t) is the measured output, u(t) is the input and s is the Laplace operator. The polynomials A(s) and B(s) are:

$$A(s) = a_0 + a_1 s + a_2 s^2 + \ldots + a_n s^n \quad , \qquad a_0 = 1$$

$$B(s) = b_0 + b_1 s + b_2 s^2 + \ldots + b_m s^m \quad , \qquad m \leqslant n \quad .$$

The coefficients a_i, b_j, are time-invariant or slowly time-varying. A more convenient notation is obtained by introducing a parameter vector

$$\underline{\theta}^T \triangleq (a_1, a_2, \ldots, a_n, b_0, b_1, \ldots, b_m) \quad , \quad \underline{\theta} \in \mathbb{R}^p \ , \ p = m+n+1$$

and a signal vector

$$\underline{\phi}^T(t) = (-y^{(1)}, -y^{(2)}, \ldots, -y^{(n)}, u^{(0)}, u^{(1)}, \ldots, u^{(m)}) \ ,$$

$$\underline{\phi} \in \mathbb{R}^p$$

where $y^{(k)} = s^k y = d^k y/dt^k$ and the superscript T denotes the transpose. Eqn. (2.1) can then be rewritten as

$$y(t) = \underline{\theta}^T \underline{\phi}(t) \quad . \tag{2.2}$$

The unknown parameter vector $\underline{\theta}$ is estimated using a model of the same dimensions:

$$\hat{y}(t) = \underline{\hat{\theta}}^T(t)\underline{\phi}(t) \tag{2.3}$$

where $\hat{y}(t)$ and $\underline{\hat{\theta}}(t)$ are estimates of $y(t)$ and $\underline{\theta}$, respectively. Introduce the parameter difference vector $\underline{\delta}(t) \hat{=} \underline{\theta} - \underline{\hat{\theta}}(t)$. The "modified equation error"

$$\varepsilon(t) \hat{=} y(t) - \hat{y}(t) \tag{2.4}$$

is minimized as a measure of $\|\underline{\delta}(t)\|$, and it is seen from eqns. (2.1) and (2.3) that

$$\varepsilon(t) = y(t) - \underline{\hat{\theta}}^T(t)\underline{\phi}(t) = \underline{\delta}^T(t)\underline{\phi}(t) \quad . \tag{2.5}$$

Remark 1. Formulation of the polynomial A(s) in eqn. (2.1) with $a_n = 1$ and $a_0 \neq 1$ leads to an equation error of a different type to that found in Young (1981), Lion (1967) or Mendel (1973, pp. 28-30). It is readily seen that both types are equivalent. As is demonstrated in Sections 4 and 5, the modified form offers considerable practical advantages in output error schemes.∇

The modified equation error given in eqn. (2.4) is minimized using eqn. (2.5) on the basis of the quadratic functional

$$J(\underline{\delta};t) = \|\underline{\hat{\theta}}(t)-\underline{\hat{\theta}}_s\|^2/g\cdot\exp(-\eta t)+\int_0^t c(\sigma)\,[y(\sigma)-\underline{\hat{\theta}}^T(t)\underline{\phi}(\sigma)]^2\exp(-\eta(t-\sigma))\,d\sigma \quad . \tag{2.6}$$

Here $\underline{\hat{\theta}}_s = \underline{\hat{\theta}}(0)$ is the starting value of $\underline{\hat{\theta}}(t)$, g and c are scalars and η is the decay parameter, with $g > 0$, $\eta \geqslant 0$ and $c(t) > 0$. A related form was presented by Minamide et al. (1983). The functional achieves a minimum when

$$\frac{\partial J(\underline{\delta};t)}{\partial \underline{\delta}} = -\frac{\partial J}{\partial \underline{\hat{\theta}}} = 0 \quad . \tag{2.7}$$

In Appendix A it is shown how this may be accomplished using the matrix-differential forms

$$\frac{d\underline{\hat{\theta}}(t)}{dt} = c(t)P(t)\underline{\phi}(t)\,[y(t) - \underline{\hat{\theta}}^T(t)\underline{\phi}(t)] \tag{2.8a}$$

$$\frac{dP(t)}{dt} = \eta(t)P(t) - c(t)P(t)\underline{\phi}(t)\underline{\phi}^T(t)P(t), \quad P \in \mathbb{R}^{p \times p} \tag{2.8b}$$

under the conditions $c(t) > 0$, $P(0) = gI$ (where I is the unit matrix). It will also be shown that $\eta \geqslant 0$ (eqn. 2.13). For reference purposes, recall that eqn. (2.8b) follows from (see Appendix A):

$$\frac{dP^{-1}(t)}{dt} = -\eta P^{-1}(t) + c(t)\underline{\phi}(t)\underline{\phi}^T(t) \quad . \tag{2.9}$$

A constant or time-decreasing function is selected as the scalar c(t) in this report; other possibilities can be found in Solo (1980).

The convergence of the parameter difference $\|\underline{\delta}(t)\|$ towards the origin after an initial disturbance can be investigated using Lyapunov's second method. For a process with bounded signals, a positive definite Lyapunov function

$$V(t) = \underline{\delta}^T(t)P^{-1}(t)\underline{\delta}(t) \tag{2.10}$$

is selected, where V(t) is a scalar. $P^{-1}(t)$ is a bounded symmetric matrix such that $P^{-1}(t) > 0$ and $\|P^{-1}(t)\| < L$, where L is a large positive number. Since $d\underline{\delta}/dt = -d\hat{\underline{\theta}}/dt$ from eqns. (2.8a) and (2.5), the time derivative of V(t) is obtained as

$$\frac{dV(t)}{dt} = -\underline{\delta}^T(t)\left(2c(t)\underline{\phi}(t)\underline{\phi}^T(t) - \frac{dP^{-1}(t)}{dt}\right)\underline{\delta}(t) \quad . \tag{2.11}$$

Convergence is ensured if eqn. (2.11) is negative semi-definite. There are several forms of P(t) and c(t) that lead to this result (Udink ten Cate, 1983). One possibility which leads to a continuous least-squares algorithm is

$$\frac{dP^{-1}(t)}{dt} = -\eta P^{-1}(t) + \gamma(t)\underline{\phi}(t)\underline{\phi}^T(t) \tag{2.12}$$

where $P^{-1}(0) > 0$, $P^{-T}(0) = P^{-1}(0)$ and $\eta \geqslant 0$, $\gamma(t) \geqslant 0$. It is demonstrated in Appendix B that under these conditions $P^{-1}(t) > 0$. Using eqns. (2.5), (2.10) and (2.12), eqn. (2.11) takes the form

$$\frac{dV(t)}{dt} = -\eta V(t) - [2c(t) - \gamma(t)]\varepsilon^2(t) \quad . \tag{2.13}$$

If $\gamma(t) < 2c(t)$, $c(t) > 0$, the time derivative of $V(t)$ is negative definite, provided $\underline{\delta}(t)$ and $\underline{\phi}(t)$ are non-orthogonal and nonzero. This occurs when the input signal is nonzero and contains a sufficient number of distinct frequencies (Lion, 1967; Anderson, 1977; Yuan and Wonham, 1977). In this case overall asymptotic stability is ensured, which means that after an initial disturbance $\|\underline{\delta}(t)\|$ will converge towards zero as $t \to \infty$. For $\eta > 0$ this convergence is exponential (Anderson, 1977).

Restating the results, we arrive at the algorithms

$$\frac{d\underline{\hat{\theta}}(t)}{dt} = c(t)P(t)\underline{\phi}(t)[y(t) - \underline{\hat{\theta}}^T(t)\underline{\phi}(t)] \quad , \quad c(t) > 0 \tag{2.14a}$$

$$\frac{dP(t)}{dt} = \eta P(t) - \gamma(t)P(t)\underline{\phi}(t)\underline{\phi}^T(t)P(t) \quad , \quad P^{-1}(0) = P^{-T}(0) > 0$$

$$\eta \geq 0, \; 0 \leq \gamma(t) < 2c(t) \quad . \tag{2.14b}$$

Taking $\gamma(t) = c(t)$ and $P(0) = gI > 0$, we minimize the quadratic functional (2.6), demonstrating that the continuous least-squares algorithms have global exponential stability properties with respect to the parameter difference.

<u>Remark 2</u>. The above result can also be interpreted as a special case of minimization of the instantaneous equation error. Define the instantaneous error criterion as $J'(\underline{\delta};t) = \frac{1}{2}\varepsilon^2(t)$. The parameter vector $\underline{\hat{\theta}}(t)$ is adjusted according to the gradient

$$\operatorname{grad}_{\underline{\delta}} J'(\underline{\delta};t) = -\frac{\partial J'}{\partial \underline{\hat{\theta}}} = -\varepsilon(t)\frac{\partial \varepsilon(t)}{\partial \underline{\hat{\theta}}(t)} = -\varepsilon(t)\underline{\phi}(t)$$

and from eqn. (2.5)

$$\frac{d\underline{\hat{\theta}}(t)}{dt} = -\Lambda \operatorname{grad} J'(\underline{\delta};t) = \Lambda\underline{\phi}(t)[y(t) - \underline{\hat{\theta}}^T(t)\underline{\phi}(t)] \quad .$$

In most common gradient methods (Mendel, 1973; Lion, 1967) the so-called "gain" matrix Λ is defined as $\Lambda = \operatorname{diag}[\lambda_i] > 0$. However, a gain matrix such as $P(t) > 0$ in eqn. (2.14b) may be substituted for Λ (Udink ten Cate, 1983). Convergence may be demonstrated using Lyapunov stability methods. ∇

It should be noted that the construction of $\underline{\phi}(t)$ requires the generation of $m+n$ derivative signals for the process described by eqn. (2.1). This can be accomplished by means of "state variable filters" (Kohr, 1963; Young, 1981).

3. CONSTRAINED METHODS

The recursive estimation problem can also be formulated with equality constraints on the parameters. These constraints are derived from time-varying information on linear combinations of the unknown process parameters. Using suitable penalty functions, inequality constraints can also be treated.

The process described by eqn. (2.2) can be transformed into an augmented process incorporating equality constraints as follows:

$$\begin{bmatrix} y(t) \\ \hdashline \underline{\chi}(t) \end{bmatrix} = \begin{bmatrix} \underline{\phi}^T(t) \\ \hdashline F(t) \end{bmatrix} \underline{\theta} \quad , \quad \underline{\chi} \in \mathbb{R}^q \quad , \quad F \in \mathbb{R}^{q \times p} \quad . \tag{3.1}$$

Each row of the time-varying or constant matrix $F(t)$ contains a linear relation of the process parameters that equal a corresponding known element of the vector $\underline{\chi}(t)$. For convenience a matrix

$$M(t) \mathrel{\hat{=}} \begin{bmatrix} \underline{\phi}^T(t) \\ \hdashline F(t) \end{bmatrix} \quad , \qquad M \in \mathbb{R}^{(q+1) \times p}$$

and a vector

$$\underline{z}(t) \mathrel{\hat{=}} \frac{y(t)}{\underline{\chi}(t)} \quad , \qquad \underline{z} \in \mathbb{R}^{q+1}$$

are introduced such that

$$\underline{z}(t) = M(t)\underline{\theta} \quad . \tag{3.2}$$

A model of the same dimensions

$$\hat{\underline{z}}(t) = \begin{bmatrix} \hat{\underline{Y}}(t) \\ \hat{\underline{X}}(t) \end{bmatrix} = M(t)\hat{\underline{\theta}}(t) \qquad (3.3)$$

is also introduced, leading to the "augmented modified equation error vector"

$$\underline{\varepsilon}(t) \;\hat{=}\; \underline{z}(t) - \hat{\underline{z}}(t) = M(t)\underline{\delta}(t) \quad , \quad \underline{\varepsilon} \in \mathbb{R}^{q+1} \qquad (3.4)$$

which is similar in form to eqn. (2.5). The following quadratic functional is minimized in the estimation procedure:

$$J(\underline{\delta};t) = \|\hat{\underline{\theta}}(t)-\hat{\underline{\theta}}_s\|^2/g\cdot\exp(-\eta t) + \int_0^t (\underline{z}(\sigma) - M(\sigma)\hat{\underline{\theta}}(t))^T W(\sigma)(z(\sigma) - M(\sigma)\hat{\underline{\theta}}(t))\cdot$$

$$\cdot\exp(-\eta(t-\sigma))d\sigma \quad . \qquad (3.5)$$

Here $W(\cdot) = \text{diag}\,[w_i] > 0$ is a weighting matrix, the other variables having the same meanings as in eqn. (2.6). Following the arguments presented in Appendix A it can readily be demonstrated that minimization is achieved using the matrix differential equations:

$$\frac{d\hat{\underline{\theta}}(t)}{dt} = Q(t)M^T(t)W(t)\underline{\varepsilon}(t) \quad , \quad Q \in \mathbb{R}^{P\times P} \, , \quad W \in \mathbb{R}^{(q+1)\times(q+1)} \qquad (3.6a)$$

$$\frac{dQ^{-1}(t)}{dt} = -\eta Q^{-1}(t) + M^T(t)\Gamma(t)M(t) \quad , \quad \Gamma \in \mathbb{R}^{(q+1)\times(q+1)} \, . \qquad (3.6b)$$

To minimize $J(\underline{\delta};t)$ the diagonal matrix $\Gamma(t)$ must satisfy

$$\Gamma(t) = [\text{diag}\;\gamma_i] = I \; ; \quad \eta \geqslant 0 \quad .$$

Remark 3. The type of constraint represented by eqns. (3.2) and (3.3) may be a time-varying combination representing *a priori* knowledge of the process parameters. A penalty function falls within this category, leading to inequality constraints. Another type of equality constraint is of the form $G(\cdot)\underline{\phi}(t)$, where $G(\cdot)$ is a filter operating on the individual signals constituting $\underline{\phi}(t)$. For example, $G(\cdot)$ could be a low pass filter or a pure integral action. This leads to an interesting class of recursive algorithms. ∇

The convergence of the estimation procedure can again be investigated by stability methods. A Lyapunov function

$$V(t) = \underline{\delta}^T(t)Q^{-1}(t)\underline{\delta}(t) \tag{3.7}$$

is selected, where $Q^{-1}(t) = Q^{-T}(t) > 0$ for $\Gamma(t) \geqslant 0$ and $Q^{-1}(t)$ is bounded (see eqn. 2.10). This follows from eqn. (3.6b). Calculation of the time derivative using eqns. (3.4) and (3.6) yields

$$\frac{dV(t)}{dt} = -\eta V(t) - \underline{\varepsilon}^T(t)(2W(t) - \Gamma(t))\underline{\varepsilon}(t) \quad . \tag{3.8}$$

For $\eta \geqslant 0$, $\gamma_i \geqslant 0$, $w_i \geqslant 0$ and $\gamma_i(t) < 2w_i(t)$ this yields a negative definite form for $dV(t)/dt$, assuming that $\underline{\varepsilon}(t)$ and $M(t)$ are non-orthogonal. This can be ensured for the pair $\underline{\varepsilon}(t)$, $\underline{\phi}(t)$ in the same way as for $\underline{\delta}(t)$ and $\underline{\phi}(t)$ in eqn. (2.13). The equality constraints cannot ensure non-orthogonality. This leads to global asymptotic stability for $\|\underline{\delta}(t)\|$ after an initial disturbance. For $\eta > 0$, $\|\underline{\delta}(t)\|$ converges exponentially.

The rules proposed in eqns. (3.6) lead to the algorithms

$$\frac{d\hat{\underline{\theta}}(t)}{dt} = Q(t)M(t)W(t)(\underline{z}(t) - M(t)\hat{\underline{\theta}}(t)) \tag{3.9a}$$

$$\frac{dQ(t)}{dt} = \eta Q(t) - Q(t)M^T(t)\Gamma(t)M(t)Q(t) \tag{3.9b}$$

where $\eta \geqslant 0$ and $\Gamma(t) = \text{diag}\,[\gamma_i] \geqslant 0$, $W(t) = \text{diag}\,[w_i] \geqslant 0$, $\gamma_i(t) < 2w_i(t)$. Because the weighting function $W(t)$ is time variable, the weights of the individual equality constraints can be used in a penalty function procedure (which itself should be a continuous function).

Remark 4. Eqn. (3.6b) suggests that, in a discrete-time version of the algorithm, the matrix inversion leading to eqn. (3.9b) in the continuous-time case will not lead to a form without explicit matrix inversion, since the matrix inversion lemma (Ljung and Söderström, 1983, p.19; Udink ten Cate and Verbruggen, 1978) cannot be applied. This can be considered as one of the main reasons for formulating algorithms (3.9) in continuous time. ∇

4. A STABLE OUTPUT ERROR METHOD

The estimation procedures presented in the previous sections assumed noise-free measurement of the system signals. When noise is present in autoregressive system identification the parameter estimates will generally be biased. An intuitively attractive approach is to feed a model with the same (noise-free) input signals as the system and minimize the output error (and its derivatives). This may be accomplished in a model reference adaptive control context. Here the output error is filtered using a filter designed according to stability theory, thus ensuring global convergence if the signals are deterministic. However, *a priori* knowledge of the system parameters is required in order to design the filters, a requirement that is not easy to satisfy in parameter estimation. Landau (1978, 1979) has proposed an output error procedure for discrete systems which requires no *a priori* knowledge. Global convergence is obtained using an extended unknown parameter vector.

In this section, this approach is used to develop a continuous-time algorithm based on an output error formulation analogous to the modified equation error (see Remark 1) for deterministic systems. Stochastic systems are treated in the next section.

The parameters of the system described by eqn. (2.1) are estimated by a model of the same dimensions

$$y_m(t) = \frac{\hat{B}(s)}{\hat{A}(s)} u(t) \tag{4.1}$$

which uses the same input signal as the system. The polynomials $\hat{A}(\cdot)$ and $\hat{B}(\cdot)$ contain the estimates of the parameters. This model can be rewritten as

$$y_m(t) = \underline{\hat{\theta}}^T(t)\underline{\phi}_m(t)$$

$$\underline{\phi}_m^T(t) = (-y_m^{(1)}, -y_m^{(2)}, \ldots, -y_m^{(n)}, u^{(0)}, u^{(1)}, \ldots, u^{(m)}),$$

$$\underline{\phi}_m \in \mathbb{R}^p \quad . \tag{4.2}$$

The output error is defined as

$$e(t) \hat{=} y(t) - y_m(t) \quad . \tag{4.3}$$

An error vector is introduced as follows:

$$\underline{e}_1^T(t) = (e^{(1)}, e^{(2)}, \ldots, e^{(n)}) = (y^{(1)} - y_m^{(1)}, \ldots, y^{(n)} - y_m^{(n)}), \; \underline{e}_1 \in \mathbb{R}^n \, . \tag{4.4}$$

For notational convenience, the parameter vector $\underline{\theta}$ will be divided into two sub-vectors containing the a_i and b_j parameters, respectively, i.e., $\underline{\theta}^T = (\underline{\theta}_a^T \vdots \underline{\theta}_b^T)$; $\underline{\theta}_a \in \mathbb{R}^n$, $\underline{\theta}_b \in \mathbb{R}^{m+1}$. The output error (eqn. 4.3) can be reformulated using eqns. (2.2) and (4.2) as follows:

$$e(t) = \underline{\theta}^T \underline{\phi}(t) - \underline{\hat{\theta}}^T(t)\underline{\phi}_m(t) = (\underline{\theta} - \underline{\hat{\theta}}(t))^T \underline{\phi}_m(t) - \underline{\theta}_a^T \underline{e}_1(t) = \underline{\delta}^T(t)\underline{\phi}_m(t) - \underline{\theta}_a^T \underline{e}_1(t) \, . \tag{4.5}$$

To ensure the stability of the estimation scheme, a parameter vector $\underline{\hat{\theta}}_c^T(t) = (\hat{c}_1(t), \hat{c}_2(t), \ldots, \hat{c}_n(t))$ is introduced, where $\underline{\hat{\theta}}_c \in \mathbb{R}^n$. The filtered output error is expressed in the following way:

$$\begin{aligned} \tilde{\varepsilon}(t) &\hat{=} e(t) + \underline{\hat{\theta}}_c^T(t)\underline{e}_1(t) = \hat{C}(s)e(t) \\ \hat{C}(s) &= 1 + \hat{c}_1(t)s + \hat{c}_2(t)s^2 + \ldots + \hat{c}_n(t)s^n \quad . \end{aligned} \tag{4.6}$$

This filtered output error will now be minimized. Using eqn. (4.5), the following relation holds for eqn. (4.6):

$$\tilde{\varepsilon}(t) = \underline{\delta}^T(t)\underline{\phi}_m(t) + (\underline{\hat{\theta}}_c(t) - \underline{\theta}_a)^T \underline{e}_1(t) = \underline{\tilde{\delta}}^T(t)\underline{\tilde{\phi}}_m(t) \tag{4.7}$$

where $\underline{\tilde{\delta}}^T(t) \hat{=} (\underline{\theta}^T - \underline{\hat{\theta}}^T(t) \vdots \underline{\hat{\theta}}_c^T(t) - \underline{\theta}_a) = (\underline{\delta}^T(t) \vdots \underline{\delta}_c^T(t))$, $\underline{\tilde{\phi}}_m^T(t) = (\underline{\phi}_m^T(t) \vdots \underline{e}_1^T(t))$; $\underline{\tilde{\delta}}, \underline{\tilde{\phi}}_m \in \mathbb{R}^{p+n}$. It can be seen that for $\underline{\hat{\theta}}_c(t) \equiv \underline{\hat{\theta}}_a(t)$ the filtered output error reduces to the modified equation error (eqn. 2.5). The parameter vector $\underline{\hat{\theta}}_c(t)$ is included to ensure stability. The error of eqn. (4.7) can also be written as

$$\tilde{\varepsilon}(t) = y(t) - \underline{\hat{\tilde{\theta}}}^T(t)\underline{\phi}_m(t) \tag{4.8}$$

where $\underline{\hat{\tilde{\theta}}}^T(t) = [\underline{\hat{\theta}}^T(t); \underline{\hat{\theta}}_c(t)]$, $\underline{\tilde{\theta}} \in \mathbb{R}^{p+n}$.

Taking a straightforward approach, consider the algorithm

$$\frac{d}{dt}\begin{bmatrix} \hat{\underline{\theta}}(t) \\ \hat{\underline{\theta}}_c(t) \end{bmatrix} = c(t)\tilde{P}(t)\tilde{\underline{\phi}}_m(t)\tilde{\varepsilon}(t) \tag{4.9a}$$

$$\frac{d\tilde{P}^{-1}(t)}{dt} = -\eta\tilde{P}^{-1}(t)+\gamma(t)\tilde{\underline{\phi}}_m(t)\underline{\phi}_m^T(t)\;,\quad \tilde{P}\in \mathbb{R}^{(p+n)\times(p+n)} \tag{4.9b}$$

where $\eta \geqslant 0, c(t) > 0, \gamma(t) \geqslant 0$. With $\gamma(t) = 1$ this algorithm is readily seen to minimize the functional

$$J(\underline{\delta};t) = \|\hat{\underline{\theta}}(t) - \hat{\underline{\theta}}_s\|^2/g\cdot\exp(-\eta t) + \int_0^t [y(\sigma) - \hat{\underline{\theta}}^T(t)\tilde{\underline{\phi}}_m(\sigma)]^2\exp(-\eta(t-\sigma))d\sigma \tag{4.10}$$

under conditions similar to those given for eqn. (2.6). However, eqn. (4.9a) suggests that when measurement noise is present in the observation of $y(t)$, correlation products will appear via $\tilde{\varepsilon}(t)$ and $\underline{e}_1(t)$, leading to biased parameter estimates. Therefore, eqn. (4.9b) will be reformulated in block-diagonal form, leading to

$$\frac{d}{dt}\begin{bmatrix} \hat{\underline{\theta}}(t) \\ \hat{\underline{\theta}}_c(t) \end{bmatrix} = \{c_p(t)\tilde{P}(t) + c_\lambda(t)\tilde{\Lambda}\}\tilde{\underline{\phi}}_m(t)\tilde{\varepsilon}(t)\;,\quad c_p(t), c_\lambda(t) \geqslant 0 \tag{4.11a}$$

$$\tilde{P}(t) = \left[\begin{array}{c|c} P_1(t) & 0 \\ \hline 0 & P_2(t) \end{array}\right],\quad P_1 \in \mathbb{R}^{p\times p}\;,\quad P_2 \in \mathbb{R}^{n\times n} \tag{4.11b}$$

where the matrices P_1^{-1} and P_2^{-1} are given by

$$\frac{dP_1^{-1}}{dt} = -\eta_1 P_1^{-1}(t) + \gamma_1(t)\underline{\phi}_m(t)\underline{\phi}_m^T(t)\;,\quad P_1^{-1}(0) > 0 \tag{4.11c}$$

$$\frac{dP_2^{-1}}{dt} = -\eta_2 P_2^{-1}(t) + \gamma_2(t)\underline{e}_1(t)\underline{e}_1^T(t)\;,\quad P_2^{-1}(0) > 0\;. \tag{4.11d}$$

Here $\tilde{\Lambda} = \mathrm{diag}\,[\lambda_i] \geqslant 0; \eta_1, \eta_2 \geqslant 0; \gamma_1(t), \gamma_2(t) \geqslant 0$. From eqn. (2.12) it can be demonstrated that $P_1^{-1} = P_1^{-T} > 0, P_2^{-1} = P_2^{-T} > 0$. It follows that $P_1, P_2 > 0$ and thus $\tilde{P} > 0$ and $\tilde{P}^{-1} > 0$.

The convergence of the parameter difference $\|\tilde{\underline{\delta}}(t)\|$ towards the origin can be demonstrated using a Lyapunov function candidate

$$V(t) = \frac{1}{2}\tilde{\underline{\delta}}^T(t)\tilde{\Lambda}^{-1}\tilde{\underline{\delta}}(t)\quad . \tag{4.12}$$

Calculation of the time derivative yields

$$\frac{dV(t)}{dt} = -c_p(t)\underline{\tilde{\delta}}^T(t)\tilde{\Lambda}^{-1}\tilde{P}(t)\underline{\tilde{\phi}}_m(t)\underline{\tilde{\phi}}_m^T\underline{\tilde{\delta}}(t) - c_\lambda(t)\tilde{\varepsilon}^2(t). \tag{4.13}$$

The first term on the right-hand side of eqn. (4.13) is negative semi-definite, since $\tilde{\Lambda}^{-1} > 0$, $\tilde{P}(t) > 0$ and $\underline{\tilde{\phi}}_m^T(t) \geqslant 0$. The second term is negative definite provided that $\underline{\tilde{\delta}}(t)$ and $\underline{\tilde{\phi}}_m(t)$ are non-orthogonal. This holds for the pair $(\underline{\delta}(t), \underline{\phi}_m(t))$ when the input signal contains a sufficient number of distinct frequencies; see also eqn. (2.13). However, it may not hold for the pair $(\underline{\delta}_c(t), \underline{e}_1(t))$, implying that $\|\underline{\delta}_c(t)\|$ may not converge towards the origin. Note that no exponential stability properties can be established from eqns. (4.11). In a practical situation, $c_\lambda(t) \equiv 0$ in eqns. (4.11). Because $\|\underline{e}_1(t)\| \to 0$ in eqn. (4.11d) when a good match between model and system is obtained, $\|P_2^{-1}(t)\| \to 0$ for $\eta_2 = 0$. This results in $\|P_2(t)\| \to \infty$ so that it is necessary to set $\gamma_2(t) = 0$.

As already mentioned, $\underline{\hat{\theta}}_c(t)$ may not converge to its true value. In a practical situation it may drift, obscuring the stability properties of the estimation scheme. Therefore, the constrained estimation algorithm introduced in the previous section will be employed to keep $\underline{\hat{\theta}}_c(t)$ in a prespecified area S_c.

The system is written in augmented form (see eqn. (3.1)) as

$$\begin{bmatrix} y(t) \\ \hline \underline{\chi}(t) \end{bmatrix} = \left[\begin{array}{c|c} \underline{\phi}_m^T(t) & -\underline{e}_1^T(t) \\ \hline F(t) & F_c(t) \end{array}\right] \begin{bmatrix} \underline{\theta} \\ \hline \underline{\theta}_a \end{bmatrix} \qquad F_c \in \mathbb{R}^{q\times n} \quad . \tag{4.14}$$

An augmented model of the same dimensions leads to

$$\begin{bmatrix} y_m(t) - \underline{\hat{\theta}}_c^T(t)\underline{e}_1(t) \\ \hline \underline{\hat{\chi}}(t) \end{bmatrix} = \left[\begin{array}{c|c} \underline{\phi}_m^T(t) & -\underline{e}_1^T(t) \\ \hline F(t) & F_c(t) \end{array}\right] \begin{bmatrix} \underline{\hat{\theta}}(t) \\ \hline \underline{\hat{\theta}}_c(t) \end{bmatrix} \quad . \tag{4.15}$$

Using eqns. (4.14) and (4.15), the augmented equation error vector is defined as

$$\underline{\tilde{\varepsilon}}(t) \triangleq \begin{bmatrix} e(t) + \underline{\hat{\theta}}_c^T(t)\underline{e}_1(t) \\ \hline \underline{\chi}(t) - \underline{\hat{\chi}}(t) \end{bmatrix} \quad . \tag{4.16}$$

It may be seen that

$$\underline{\tilde{\varepsilon}}(t) = \tilde{M}(t)\underline{\tilde{\delta}}(t) \quad , \qquad \tilde{M} \in \mathbb{R}^{(q+1)\times(p+n)} \tag{4.17a}$$

$$\tilde{M}(t) = \left[\begin{array}{c|c} \underline{\phi}_m^T(t) & -\underline{e}_1^T(t) \\ \hline F(t) & F_c(t) \end{array}\right] = \left[\begin{array}{c|c} M_1 & M_2 \end{array}\right] \quad , \quad M_1 \in \mathbb{R}^{(q+1)\times p}, \ M_2 \in \mathbb{R}^{(q+1)\times n} \quad . \tag{4.17b}$$

Following the same procedure as in eqns. (3.6)-(3.8) leads to the recursive scheme

$$\frac{d\underline{\hat{\theta}}(t)}{dt} = \{c_p(t)Q_1(t) + c_\lambda(t)\Lambda_1\}M_1^T(t)W(t)\underline{\tilde{\varepsilon}}(t), \quad W \in \mathbb{R}^{(q+1)\times(q+1)} \tag{4.18a}$$

$$\frac{d\underline{\hat{\theta}}_c(t)}{dt} = \{c_p(t)Q_2(t) + c_\lambda(t)\Lambda_2\}M_2^T(t)W(t)\underline{\tilde{\varepsilon}}(t) \tag{4.18b}$$

$$\frac{dQ_1(t)}{dt} = \eta_1 Q_1(t) - Q_1(t)M_1^T(t)\Gamma_1(t)M_1(t)Q_1(t) \quad , \quad Q_1 \in \mathbb{R}^{p\times p} \tag{4.18c}$$

$$\frac{dQ_2(t)}{dt} = \eta_2 Q_2(t) - Q_2(t)M_2^T(t)\Gamma_2(t)M_2(t)Q_2(t) \quad , \quad Q_2 \in \mathbb{R}^{n\times n} \quad . \tag{4.18d}$$

In these eqns. $\eta_1, \eta_2 \geqslant 0; \Gamma_1(t) = \text{diag}\ [\gamma_{1,i}] \geqslant 0, \Gamma_2(t) = \text{diag}\ [\gamma_{2,i}] \geqslant 0,$ $\Gamma_1, \Gamma_2 \in \mathbb{R}^{(q+1)\times(q+1)}$; $W = \text{diag}\ [w_i] \geqslant 0$; $\Lambda_1 = \text{diag}\ [\lambda_{1,i}] \geqslant 0,$ $\Lambda_1 \in \mathbb{R}^{p\times p}$; $\Lambda_2 = \text{diag}\ [\lambda_{2,i}] \geqslant 0, \Lambda_2 \in \mathbb{R}^{n\times n}$; $c_p(t), c_\lambda(t) > 0$. When $W(t)$ is used as a penalty function, inequalities can also be treated with eqns. (4.18). In Appendix C it is demonstrated that the parameter difference converges towards the origin after an initial disturbance. This does not hold for $\|\underline{\delta}_c(t)\|$ in general but if inequality constraints are used, i.e., $F_c(t)$ is nonempty, the estimates $\underline{\hat{\theta}}_c(t)$ can be confined to an area S_c.

Because all the signals have to be bounded in order to ensure convergence, the model which generates $\underline{\phi}_m(t)$ must be stable. This means that the estimates $\underline{\hat{\theta}}(t)$ should be kept within a stable region S_θ, which can be achieved using the matrix F(t). Since the algorithms are implemented in a quasi-continuous way, the region S_θ must also be such as to prevent numerical instabilities.

5. STOCHASTIC LINEAR SYSTEMS

In the previous section it was decided not to adopt the straightforward approach as suggested in eqns. (4.9) because of the anticipated erroneous behavior of the algorithm when measurement noise is present in the observations of the system output signals y(t). Nevertheless, an analysis was carried out for deterministic systems. In the following, the behavior of the algorithm (4.18) is analyzed for stochastic linear systems. The procedure proposed by Ljung (1977) (see also Ljung and Söderström, 1983) is adopted, in which a deterministic differential equation is associated with the stochastic version of algorithm (4.18). Stability of this differential equation implies convergence of the algorithm.

We shall restrict ourselves to the unconstrained algorithms. The estimates $\underline{\hat{\theta}}$ and $\underline{\hat{\theta}}_c$ are assumed to be within the areas S_θ and S_c defined by F(t) and $F_c(t)$, respectively. Following Ljung and Söderström (1983, Ch. 4), let the model be described by the unconstrained relation

$$y_m(t|\underline{\theta}) = \frac{B(s)}{A(s)} u(t) \tag{5.1}$$

where $\underline{\theta}$ is a parameter vector belonging to the model set D_M which describes the observed data. The definitions of $\underline{\theta}$, A(s) and B(s) are analogous to those used in eqn. (2.1). From eqn. (4.2), eqn. (5.1) can be reformulated as

$$y_m(t|\underline{\theta}) = \underline{\phi}_m^T(t,\underline{\theta})\underline{\theta} \quad . \tag{5.2}$$

The data is described by

$$y(t) = \frac{B_0(s)}{A_0(s)} u(t) + v(t) \tag{5.3}$$

where the polynomials $A_0(s)$ and $B_0(s)$ contain the "true" parameters represented by the vector $\underline{\theta}_0$. The signal v(t) is a zero-mean disturbance.

When $y_0(t) = \frac{B_0(s)}{A_0(s)} u(t) = y(t) - v(t)$, eqn. (5.3) can be written as

$$y(t) = y_0(t) + v(t) = \underline{\phi}_0^T(t)\underline{\theta}_0 + v(t) \tag{5.4}$$

where $y_0(t)$ and $\underline{\phi}_0(t)$ are "undisturbed" signals. The output error is

$$e(t) \hat{=} y(t) - y_m(t|\underline{\theta}) = e_0(t) + v(t) \tag{5.5}$$

where $e_0(t)$ is the "undisturbed" signal. The error vector is defined as

$$\underline{e}_1(t) = \underline{e}_{1,0}(t) + \underline{v}_1(t) \tag{5.6}$$

where $\underline{v}_1^T(t) = (v^{(1)}, v^{(2)}, \ldots, v^{(n)})$. The augmented equation error is

$$\tilde{\varepsilon}(t) = \tilde{\varepsilon}_0(t) + C(s)v(t) = C(s)e_0(t) + C(s)v(t) = C(s)e(t) \ . \tag{5.7}$$

From eqn. (5.5) it follows that

$$e(t) = \underline{\phi}_0^T(t)\underline{\theta}_0 - \underline{\phi}_m^T(t)\underline{\theta} + v(t) = \underline{\phi}_m^T(t)(\underline{\theta}_0 - \underline{\theta}) + [A_0(s)-1](-e(t)+v(t))+v(t)$$

$$A_0(s)e(t) = \underline{\phi}_m^T(t)(\underline{\theta}_0 - \underline{\theta}) + A_0(s)v(t) \ . \tag{5.8}$$

Thus, using eqn. (5.7) we have

$$\frac{A_0(s)}{C(s)} \tilde{\varepsilon}(t) = \underline{\phi}_m^T(t)(\underline{\theta}_0 - \underline{\theta}) + A_0(s)v(t)$$

$$\tilde{\varepsilon}(t) = \left(\frac{C(s)}{A_0(s)}\underline{\phi}_m(t)\right)^T(\underline{\theta}_0 - \underline{\theta}) + C(s)v(t) \ . \tag{5.9}$$

After these preliminaries we shall now investigate the convergence of the algorithm for an associated differential equation

$$\frac{d\tilde{\underline{\theta}}(t)}{dt} = \frac{d}{dt}\left[\begin{array}{c}\underline{\theta}(t)\\ \hline \underline{\theta}_c(t)\end{array}\right] = \left[\begin{array}{c|c}R_1(t) & 0\\ \hline 0 & R_2(t)\end{array}\right]\left[\begin{array}{c}f_1(\underline{\theta})\\ \hline f_2(\underline{\theta}_c)\end{array}\right] = \tilde{R}(t)f(\tilde{\underline{\theta}}),\ \tilde{R}(t) > 0 \tag{5.10}$$

where $f_1(\underline{\theta}) = \bar{E}\underline{\phi}_m(t)\tilde{\varepsilon}(t)$, $f_2(\underline{\theta}_c) = \bar{E}\underline{e}_1(t)\tilde{\varepsilon}(t)$, $f(\tilde{\underline{\theta}}) = \bar{E}\hat{\underline{\phi}}_m(t)\tilde{\varepsilon}(t)$, $\bar{E}f(\cdot) = \lim_{T\to\infty}\frac{1}{T}\int_0^T Ef(\cdot)$. We now evaluate $f_1(\cdot)$ and $f_2(\cdot)$, taking $v_F(t) \hat{=} C(s)v(t)$ and making use of eqns. (5.5) and (5.6):

$$f_1(\underline{\theta}) = \bar{E}\underline{\phi}_m(t)\tilde{\varepsilon}(t) = \bar{E}\underline{\phi}_m(t)\tilde{\varepsilon}_0(t) + \bar{E}\underline{\phi}_m(t)v_F(t) = \bar{E}\underline{\phi}_m(t)\tilde{\varepsilon}_0(t) \tag{5.11a}$$

$$f_2(\underline{\theta}_c) = \bar{E}\underline{e}_1(t)\tilde{\varepsilon}(t) = \bar{E}(\underline{e}_{1,0}(t)+\underline{v}_1(t)(\tilde{\varepsilon}_0(t)+v(t))) =$$

$$= \bar{E}\underline{e}_{1,0}(t)\tilde{\varepsilon}(t) + \bar{E}\underline{v}_1(t)v_F(t) \quad . \tag{5.11b}$$

Here we used the assumption that the pairs $(\underline{\phi}_m,v_F);(\underline{e}_{1,0},v_F)$ and $(\underline{v}_1,\tilde{\varepsilon}_0)$ are independent. The possible convergence of $\|\tilde{\underline{\theta}}_0-\tilde{\underline{\theta}}\|$, $\tilde{\underline{\theta}}_0^T = (\underline{\theta}_0^T \vdots \underline{\theta}_{a,0}^T)$, towards the origin is now investigated using the Lyapunov function

$$V(t) = \frac{1}{2}(\tilde{\underline{\theta}}_0-\tilde{\underline{\theta}})^T(\tilde{\underline{\theta}}_0-\tilde{\underline{\theta}}) \quad . \tag{5.12}$$

Using the fact that $\tilde{\varepsilon}_0(t) = \underline{\phi}_{m,0}^T(t)(\tilde{\underline{\theta}}_0-\tilde{\underline{\theta}})$, $\underline{\theta}_{m,0}^T(t) = [\underline{\phi}_m^T(t)\vdots\underline{e}_{1,0}^T(t)]$ and eqns. (5.10) and (5.11), the time derivative is calculated as

$$\frac{dV(t)}{dt} = -(\tilde{\underline{\theta}}_0-\tilde{\underline{\theta}})^T\tilde{R}(t)\bar{E}\tilde{\underline{\phi}}_{m,0}(t)\tilde{\underline{\phi}}_{m,0}^T(t)(\tilde{\underline{\theta}}_0-\tilde{\underline{\theta}})-(\underline{\theta}_{a,0}-\underline{\theta}_c)^T R_2(t)\bar{E}\underline{v}_1(t)v_F(t) \quad . \tag{5.13}$$

This expression is not always negative definite. For $E\underline{v}_1(t)v_F(t) \neq 0$, mismatch of $\underline{\theta}_c$ and its true value $\underline{\theta}_{a,0}$, together with small terms containing $\underline{e}_{1,0}(t)$ in $\bar{E}\tilde{\underline{\phi}}_{m,0}(t)\underline{\phi}_{m,0}^T(t)$ (which occurs when the model and the system are close) can cause the norm $\|\underline{\theta}_{a,0}-\underline{\theta}_c\|$ to become relatively large. However, with constrained estimation $\underline{\theta}_c$ is confined to S_c. We shall now investigate which set S_c leads to convergence of $\underline{\theta}$ to its true value $\underline{\theta}_0$. Consider the following Lyapunov function, which is a function of $\underline{\theta}$ only:

$$V_1(t) = \frac{1}{2}(\underline{\theta}_0-\underline{\theta})^T(\underline{\theta}_0-\underline{\theta}) \quad . \tag{5.14}$$

From eqns. (5.7),(5.9),(5.10) and (5.11a), the time derivative is

$$\frac{dV_1(t)}{dt} = -(\underline{\theta}_0-\underline{\theta})^T R_1(t)\overline{E}\underline{\phi}_m(t)\underline{\phi}_F^T(t)(\underline{\theta}_0-\underline{\theta}) \quad , \quad R_1(t) > 0 \tag{5.15}$$

where

$$\underline{\phi}_F(t) = \frac{C(s)}{A_0(s)}\,\underline{\phi}_m(t) = H(s)\underline{\phi}_m(t) \left(H(s) \hat{=} \frac{C(s)}{A_0(s)} \right) \quad .$$

This term is negative (semi)definite when $\overline{E}\underline{\phi}_m(t)\underline{\phi}_F^T(t) \geq 0$, which occurs when $H(s)$ is strictly positive real.

Proof (see Ljung and Söderström, 1983, p.212). Define any column vector $\underline{\ell} \neq 0$ and define $x(t) = \underline{\ell}^T\underline{\phi}_m(t), x_F(t) = \underline{\ell}^T\underline{\phi}_F(t) = H(s)x(t)$. We have $\overline{E}\underline{\phi}_m(t)\underline{\phi}_F^T(t) \geq 0$ when (by definition)

$$\underline{\ell}^T\overline{E}\underline{\phi}_m(t)\underline{\phi}_F^T(t)\underline{\ell} = \overline{E}\underline{\ell}^T\underline{\phi}_m(t)\underline{\phi}_F^T(t)\underline{\ell} = \overline{E}x(t)H(s)x(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\Phi_{zz}(\omega)H(j\omega)d\omega =$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_{zz}(\omega)\mathrm{Re}(H(j\omega))d\omega \geq 0 \quad . \tag{5.16}$$

Here $\Phi_{zz}(\omega)$ is the spectral density of $x(t)$. The inequality in (5.16) will hold when $\mathrm{Re}\,H(j\omega) \geq 0$, which means that $H(j\omega)$ is strictly positive real. Equality occurs for $\overline{E}x^2(t) = \overline{E}(\underline{\ell}^T\underline{\phi}_m(t))^2 = 0$. Q.E.D.

This means that $\|\underline{\theta}_0-\underline{\theta}\|$ will converge to 0 when $C(s)/A_0(s)$ is strictly positive real. With this result the convergence of algorithm (4.18) is established both for a deterministic system (after an initial disturbance) and for a stochastic system with a zero-mean disturbance. In the case of an initial disturbance global convergence can be ensured, while in the stochastic case it is necessary for the estimates $\hat{\underline{\theta}}_c$ to converge such that $C(s)/A_0(s)$ is strictly positive real. This may be achieved by putting constraints on $\hat{\underline{\theta}}_c$. This combination of global deterministic convergence and constrained stochastic convergence produces an algorithm which performs satisfactorily.

6. SIMULATION

Simulations were run in order to test the proposed algorithm (4.18). It has already been stated that the proposed algorithm is expected to have applications in small personal computers. For this reason, the signal vectors were generated by straightforward difference methods rather than sophisticated techniques. Integration was performed by Euler's method. The straightforward numerical methods can be implemented using the DYNAMO simulation language.

The parameters of a first-order system were estimated using the transfer function

$$H(s) = \frac{y_0(s)}{u(s)} = \frac{K}{\tau s+1} \tag{6.1}$$

where $K = 1$ and $\tau = 10$ sec. The system parameters were assumed to be unconstrained so eqns. (4.11a-c) could be used to estimate $\underline{\hat{\theta}}^T(t) = [\hat{\tau},\hat{K}]$. The parameter $\hat{\theta}_c = [\hat{c}]$ was estimated using eqns. (4.18b,d) although the bounds on $\hat{c} \in [-20,20]$ were never reached, which makes eqns. (4.18b,d) equivalent to (4.11a,b,d). We took $c_\lambda \equiv 0$, $\eta_1 = \eta_2 = \eta$ and $\gamma_1 = \gamma_2 = \gamma$. The output signal $y_0(t)$ was disturbed by noise $v(t)$ from a gaussian random generator with zero mean. The noise/signal ratio $v^2(t)/y_0^2(t)$ was 0.1 (see also eqns. 5.3 - 5.4).

The results of the parameter estimation are presented in Tables 1-4. Simulation was carried out up to time t = 1000 sec.

TABLE 1. Results obtained with $\eta = 0.01$, $\gamma = 1$, $P_1(0) = \text{diag}\ [0.1, 0.001]$, $P_2(0) = 0.1$.

Estimate	Average	Rel. variance	Rel. error	Parameter distance criterion	Final value of P matrices
$\hat{K}$	1.079	0.0057	0.0795	$E_{250} = 402.8$	
$\hat{\tau}$	11.49	0.023	0.149	$E_{500} = 593.6$	$P_1(1000) = \begin{bmatrix} 5.03 & 0.31 \\ 0.31 & 0.029 \end{bmatrix}$
$\hat{c}$	-0.038	0.336	-	$E_{750} = 614.6$	
				$E_{1000} = 625.0$	$P_2(1000) = 0.30$

TABLE 2. Results obtained with $\eta = 0.02$, other values as Table 1.

Estimate	Average	Rel. variance	Rel. error	Parameter distance criterion	Final value of P matrices
$\hat{K}$	0.987	0.0096	0.013	$E_{250} = 343.2$	$P_1(1000) = \begin{bmatrix} 8.86 & 0.48 \\ 0.48 & 0.048 \end{bmatrix}$
$\hat{\tau}$	9.81	0.022	0.019	$E_{500} = 369.8$	
$\hat{c}$	-0.029	1.099	-	$E_{750} = 371.18$	
				$E_{1000} = 371.80$	$P_2(1000) = 0.50$

TABLE 3. Results obtained with $\eta = 0.02$, $\gamma = 1$, $P_1(0) = \text{diag}\ [0.3, 0.003]$, $P_2(0) = 0.3$.

Estimate	Average	Rel. variance	Rel. error	Parameter distance criterion	Final value of P matrices
$\hat{K}$	0.986	0.009	0.014	$E_{250} = 268.4$	$P_1(1000) = \begin{bmatrix} 8.62 & 0.48 \\ 0.48 & 0.048 \end{bmatrix}$
$\hat{\tau}$	9.80	0.049	0.020	$E_{500} = 277.2$	
$\hat{c}$	-0.029	1.099	-	$E_{750} = 278.76$	
				$E_{1000} = 279.37$	$P_2(1000) = 0.50$

TABLE 4. Results obtained with time-invariant P_1 and P_2 (compare with Tables 2 and 3).

Estimate	Average	Rel. variance	Rel. error	Parameter distance criterion	Final value of P matrices
$\hat{K}$	0.985	0.0082	0.015	$E_{250} = 32.56$	$P_1 = \begin{bmatrix} 8.7 & 0.48 \\ 0.48 & 0.048 \end{bmatrix}$
$\hat{\tau}$	9.76	0.015	0.024	$E_{500} = 32.72$	
$\hat{c}$	-0.030	1.138	-	$E_{750} = 33.58$	
				$E_{1000} = 33.99$	$P_2 = 0.50$

The averages are obtained from data points taken at 5 sec intervals between t = 905 and t = 1000 so that 20 data points are used. The relative variance is $\sigma\theta/\overline{\theta}$, where θ is the unknown parameter; the relative error is $|\theta_0 - \overline{\theta}|/\theta_0$, where θ_0 is the true value. A parameter distance criterion

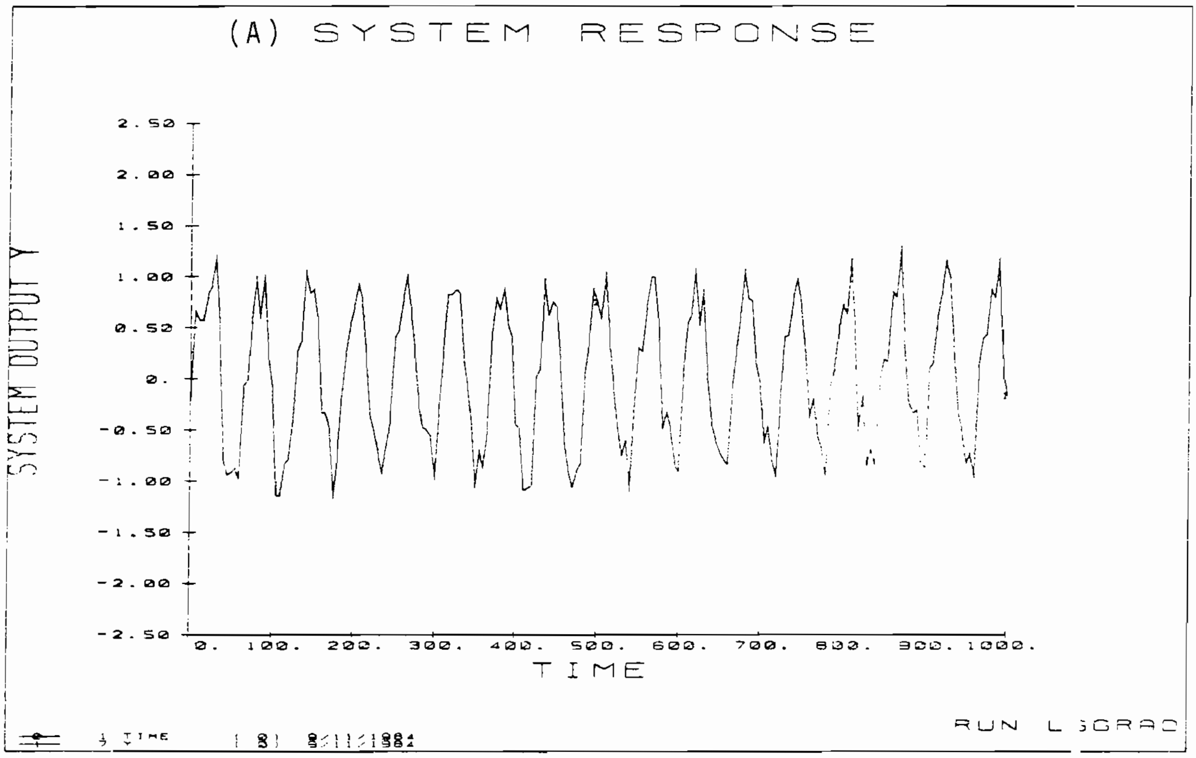
$$E(\underline{\delta};t) = \int_0^t \left[\left(\frac{K-\hat{K}(\sigma)}{K}\right)^2 + \frac{(\tau-\hat{\tau}(\sigma))^2}{\tau}\right] d\sigma \tag{6.2}$$

is also given for several values of t; this gives an indication of the rate of convergence. From the tables it can be seen that the estimates of K and τ are accurate with a low relative variance, which indicates their usefulness for online application. The estimate $\hat{c}$ has a larger relative variance. Note that the initial values of the P matrices are rather low; larger values lead to numerical difficulties. Interesting results are obtained on using fixed P matrices (see Table 4). If a proper constant matrix is known (e.g., by evaluating the P matrices in a separate run) rapid convergence is obtained yielding comparable results (Tables 2 and 3). A diagonal matrix $P_1 = \text{diag}\ (0.8,\ 0.008)$ with $P_2 = 0.8$ was found to exhibit slow convergence.

Figure 1 presents the responses listed in Table 3. The graphs show the disturbed system output y(t), the parameter estimates and the error criterion. From the response of $\hat{c}$ (denoted by - CHAT in Fig. 1d) it can be seen that for small output errors $\hat{c}$ tends to drift away.

7. CONCLUSIONS

This paper presents a continuous-time version of the well-known recursive discrete-time least-squares algorithm for the estimation of parameters of continuous-time systems. The algorithm is shown to minimize a quadratic functional representing a cost function in terms of the parameter error between the system and its model. The estimation scheme is demonstrated to have exponential convergence properties. The error considered is a modified version of the usual equation error.



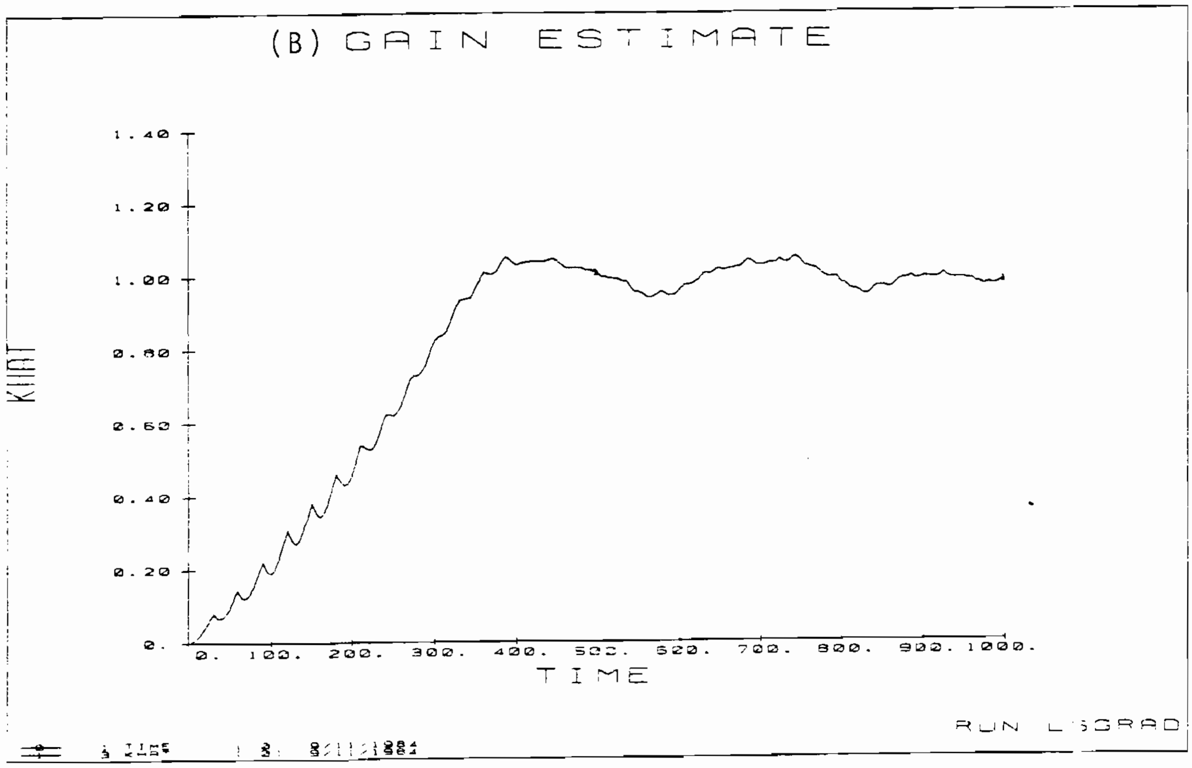
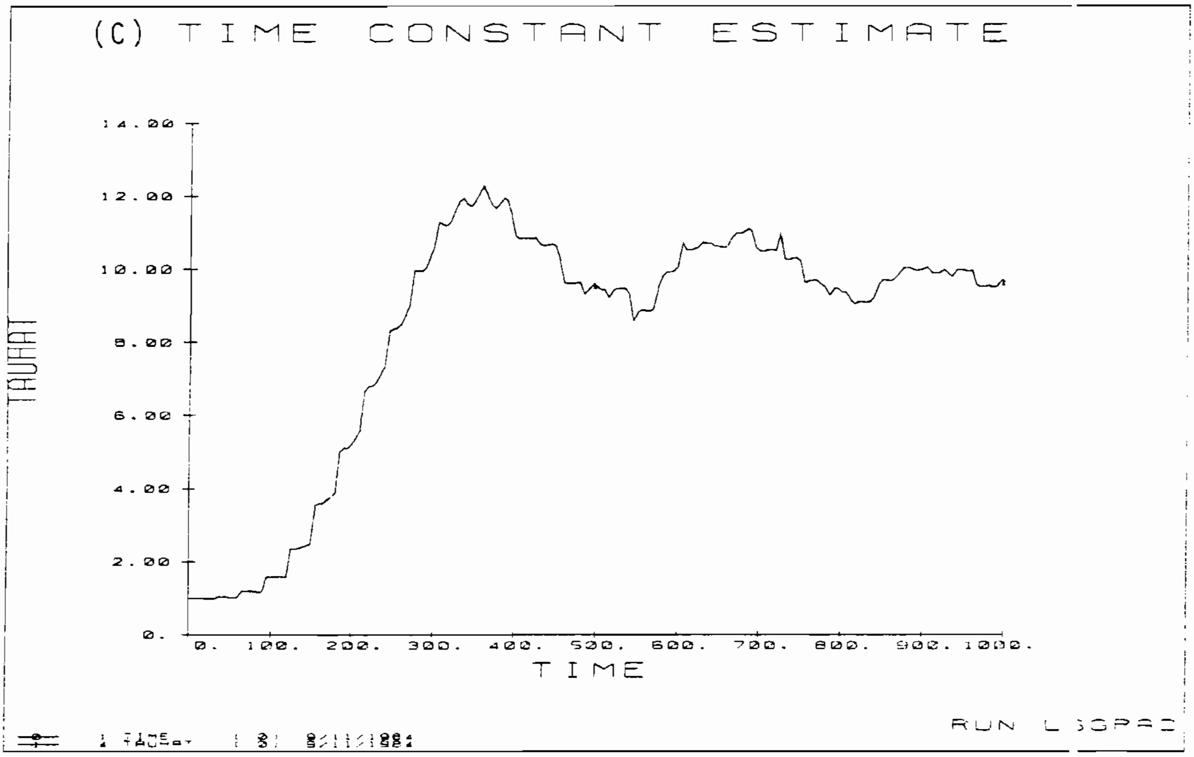


FIGURE 1: The responses corresponding to Table 3



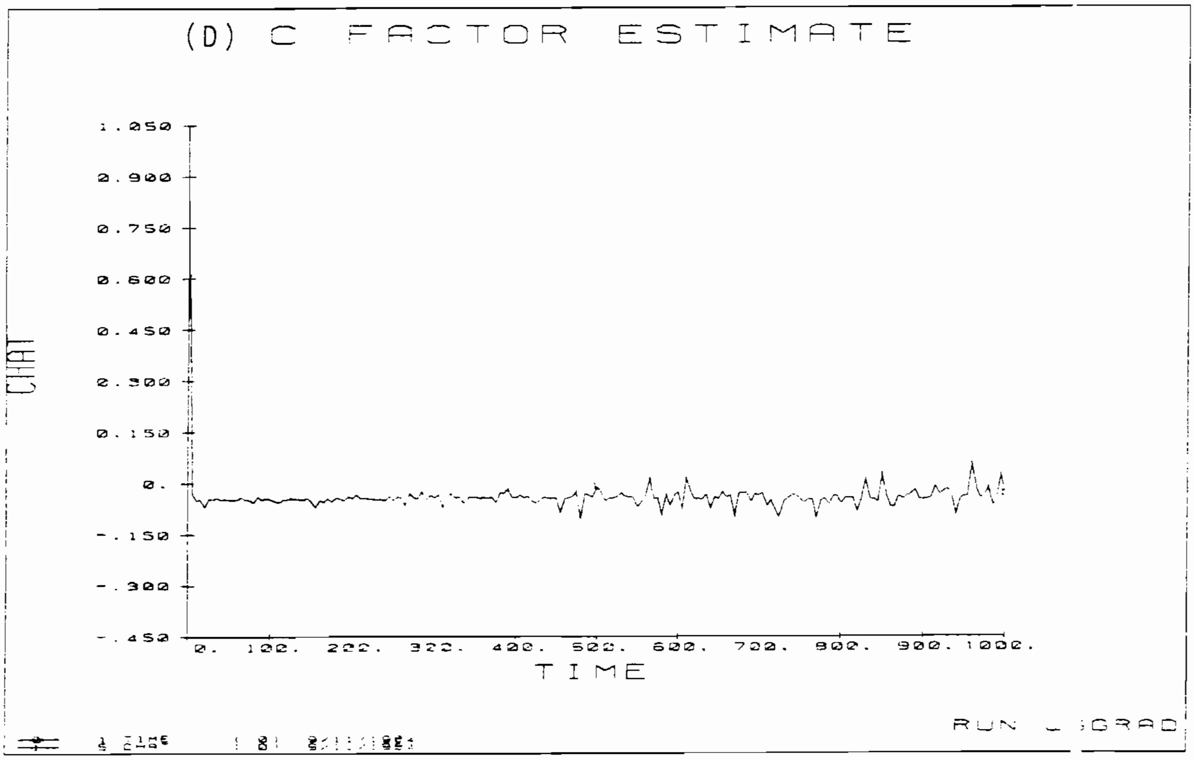


FIGURE 1 (continued)

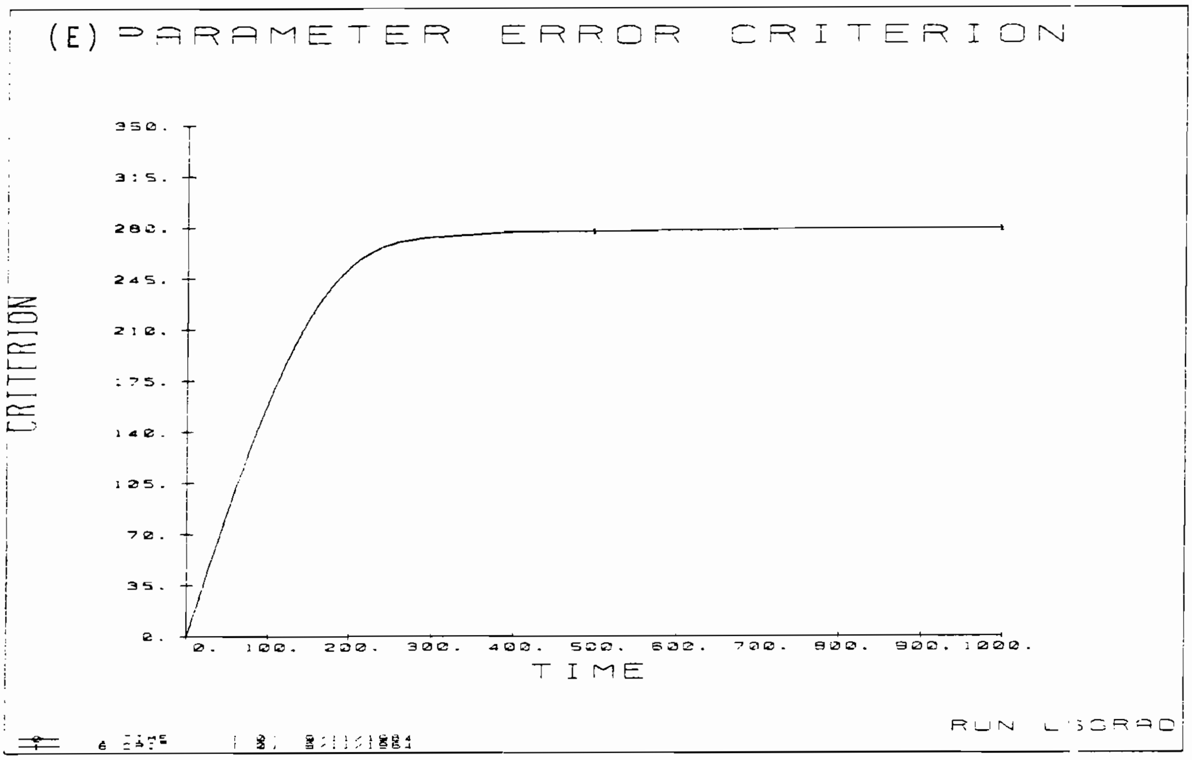


FIGURE 1 (continued)

The algorithm is then extended in order to handle equality constraints. This algorithm is also shown to minimize a quadratic functional and to possess exponential convergence properties. By the use of a penalty function inequality constraints can also be treated. The constrained version of the algorithm is used in an output error scheme for parameter estimation. On applying this scheme to deterministic systems, the parameter error displays global convergence when the output error is filtered by an adjustable filter. The parameters of this filter are estimated by an extended parameter vector. In stochastic systems, convergence is obtained when a transfer function associated with the adjustable filter parameters and the (unknown) system is strictly positive real (following the analysis proposed by Ljung). This can be achieved using an algorithm with inequality constraints, where the filter parameters are confined to a set S_c. Simulation shows this method to be feasible.

Because slowly time-varying parameters can be tracked, the output error method could be very useful in (model reference) adaptive control schemes.

REFERENCES

Anderson, B.D.O. (1977). Exponential stability of linear equations arising in adaptive identification. *IEEE Trans. Autom. Control* AC-22:83-88.

Anderson, B.D.O. and C.R. Johnson (1982). Exponential convergence of adaptive identification and control algorithms. *Automatica* 18(1):1-13.

Bohn, E.V. (1982). Estimation of continuous time linear system parameters from periodic data. *Automatica* 18(2):27-36.

Brockett, R.W. (1970). *Finite-Dimensional Linear Systems.* John Wiley, New York.

Dugard, L. and I.D. Landau (1980). Recursive output error identification algorithms: theory and evaluation. *Automatica* 16:443-462.

Eykhoff, P. (1974). *System Identification - Parameter and State Estimation.* John Wiley, London.

Goodwin, G.C. and R.L. Payne (1977). *Dynamic System Identification: Experiment Design and Data Analysis.* Academic Press, New York.

Kohr, R.H. (1963). A method for the determination of a differential equation model for simple nonlinear systems. *IEEE Trans. Electron. Computers* 12(4):395-400.

Kushner, H.J. and D.S. Clark (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems.* Series on Applied Mathematical Sciences, Vol 26. Springer-Verlag, New York.

Landau, I.D. (1976). Unbiased recursive identification using model reference adaptive techniques. *IEEE Trans. Autom. Control* AC-21:194-202.

Landau, I.D. (1978). Elimination of the real positivity condition in the design of parallel MRAS. *IEEE Trans. Autom. Control* AC-23:1015-1020.

Landau, I.D. (1979). *Adaptive Control: the Model Reference Approach.* Marcel Dekker, New York.

Lion, P.M. (1967). Rapid identification of linear and nonlinear systems. *Am. Inst. Aeronaut. Astronaut. J.* 5(10):1835-1842.

Ljung, L. (1977). Analysis of recursive stochastic algorithms. *IEEE Trans. Autom. Control* AC-22:551-575.

Ljung, L. and T. Söderström (1983). *Theory and Practice of Recursive Identification.* MIT Press, Cambridge, Mass.

Mendel, J.M. (1973). *Discrete Techniques of Parameter Estimation: The Equation Error Formulation.* Marcel Dekker, New York.

Minamide, N., P.N. Nikiforuk and M.M. Gupta (1983). Design of an adaptive observer and its application to an adaptive pole placement controller. *Int. J. Control* 37(2):349-366.

Parks, P.C. (1966). Liapunov redesign of model reference adaptive control systems. *IEEE Trans. Autom. Control* AC-11:362-367.

Shackloth, B. and R.L. Butchart (1965). Synthesis of model reference adaptive systems by Liapunov's second method. In P.H. Hammond (Ed.), *Theory of Self-Adaptive Control Systems.* Plenum Press, New York.

Solo, V. (1980). Some aspects of recursive parameter estimation. *Int. J. Control* 32(3):395-410.

Udink ten Cate, A.J. (1983). Modeling and (adaptive) control of greenhouse climates. Dissertation, Department of Physics and Meteorol., Agric. Univ., Wageningen, The Netherlands.

Udink ten Cate, A.J. and H.B. Verbruggen (1978). A least-squares-like gradient method for discrete process identification. *Int. J. Control* 28(6):933-952.

Young, P. (1981). Parameter estimation for continuous-time models - a survey. *Automatica* 17(1):23-39.

Yuan, J.S.-C. and W.M. Wonham (1977). Probing signals for model reference identification. *IEEE Trans. Autom. Control* AC-22:530.

APPENDIX A

This appendix demonstrates that the functional (2.6) is minimized by eqns. (2.8). Recall that the functional is given by

$$J(\underline{\hat{\vartheta}};t)=\|\underline{\hat{\vartheta}}(t)-\underline{\hat{\vartheta}}_S\|^2/g\cdot\exp(\eta t)+ \tag{A.1}$$

$$+\int_0^t c(\sigma)[y(\sigma)-\underline{\hat{\vartheta}}^{\mathrm{T}}(t)\underline{\varphi}(\sigma)]^2\exp(-\eta(t-\sigma))\mathrm{d}\sigma$$

with $g>0,\ \eta\geq 0,\ c(\cdot)\geq 0$. Eqn. (A.1) is minimized when

$$\frac{\partial J}{\partial\underline{\hat{\vartheta}}}=\frac{-\partial J}{\partial\underline{\hat{\vartheta}}}=-2(\underline{\hat{\vartheta}}(t)-\underline{\hat{\vartheta}}_S)/g\ \exp(-\eta t)+ \tag{A.2}$$

$$+2\int_0^t c(\sigma)[y(\sigma)-\underline{\hat{\vartheta}}^{\mathrm{T}}(t)\underline{\varphi}(\sigma)]\exp(-\eta(t-\sigma))\underline{\varphi}(\sigma)\mathrm{d}\sigma=0.$$

After rearrangement of terms, this yields

$$\underline{\hat{\vartheta}}(t)=[I/g\cdot\exp(-\eta t)+\int_0^t\underline{\varphi}(\sigma)\underline{\varphi}^{\mathrm{T}}(\sigma)c(\sigma)\exp(-\eta(t-\sigma))\mathrm{d}\sigma]^{-1}\times \tag{A.3}$$

$$\times[\underline{\hat{\vartheta}}_S/g\cdot\exp(-\eta t)+\int_0^t c(\sigma)y(\sigma)\underline{\varphi}(\sigma)\exp(-\eta(t-\sigma))\mathrm{d}\sigma=$$

$$=P(t)[\underline{\hat{\vartheta}}_S/g\exp(-\eta t)+\int_0^t c(\sigma)y(\sigma)\underline{\varphi}(\sigma)\exp(-\eta(t-\sigma))\mathrm{d}\sigma.$$

Here the matrix $P(t)$ is given by

$$P^{-1}(t)=I/g\exp(-\eta t)+\int_0^t\underline{\varphi}(\sigma)\underline{\varphi}^{\mathrm{T}}(\sigma)c(\sigma)\exp(-\eta(t-\sigma))\mathrm{d}\sigma. \tag{A.4}$$

This may be written in matrix differential form as

$$\frac{\mathrm{d}P^{-1}(t)}{\mathrm{d}t}=-\eta P^{-1}(t)+c(t)\underline{\varphi}(t)\underline{\varphi}^{\mathrm{T}}(t),\quad P_0^{-1}=I/g. \tag{A.5}$$

Since for any non-singular matrix $A(t)$ we have

$$\frac{\mathrm{d}A^{-1}(t)}{\mathrm{d}t}=-A^{-1}(t)\frac{\mathrm{d}A(t)}{\mathrm{d}t}A^{-1}(t),$$

it follows that

$$\frac{dP(t)}{dt} = \eta P(t) - c(t)P(t)\underline{\varphi}(t)\underline{\varphi}^{T}(t)P(t) \tag{A.6}$$

where $\eta \geq 0$. Using eqn. (A.6), we can deduce from eqn. (A.3) (after some calculation) that

$$\frac{d\hat{\underline{\vartheta}}(t)}{dt} = c(t)P(t)\underline{\varphi}(t)[y(t) - \underline{\varphi}^{T}(t)\hat{\underline{\varphi}}(t)]. \tag{A.7}$$

APPENDIX B

This appendix shows that the matrix $P^{-1}(t)$ in eqn. (2.12) is positive definite. Let

$$\frac{dP^{-1}(t)}{dt} = -\eta P^{-1}(t) + \gamma(t)\underline{\varphi}(t)\underline{\varphi}^{T}(t) \tag{B.1}$$

where $P^{-1}(0) = (P^{-1}(0))^{T} > 0$. It will be shown that $P^{-1}(t) = P^{-T}(t) > 0$.

Proof. We use a theorem by Brockett (1970, p. 59). The solution of a linear matrix equation of the form

$$\frac{dX(t)}{dt} = A_1(t)X(t) + X(t)A_2(t) + F(t) \tag{B.2}$$

where $A_1(t), A_2(t)$ and $F(t)$ are known is given by

$$X(t) = \Phi_1(t,t_0)X(t_0)\Phi_2^{T}(t,t_0) + \int_{t_0}^{t} \Phi_1(t,\sigma)F(\sigma)\Phi_2^{T}(t,\sigma)d\sigma \tag{B.3}$$

where $\Phi_1(t,t_0)$ is the transitional matrix of $d\underline{x}(t)/dt = A_1(t)\underline{x}(t)$ with solution $\underline{x}(t) = \Phi_1(t,t_0)\underline{x}(t_0)$ and $\Phi_2(t,t_0)$ is the analogous matrix for $d\underline{x}(t)/dt = A_2(t)\underline{x}(t)$. $X(t_0)$ is the initial value of $X(t)$.

Rewriting eqn. (B.2) with $X(t) = P^{-1}(t)$, $A_1(t) = A_2(t) = -\frac{1}{2}\eta I$ and $F(t) = \gamma(t)\underline{\varphi}(t)\underline{\varphi}^{T}(t)$ gives eqn. (B.1). Since $P^{-1}(t_0) > 0$ and $\Phi_1(t,t_0) = \Phi_2(t,t_0)$, the first term on the right-hand side of eqn. (B.3) is a decaying matrix which is positive (semi)definite. The second term is positive definite because $\int_{t_0}^{t} \underline{\varphi}(t)\underline{\varphi}^{T}(t) > 0$ is related to the process signal covariance matrix. Q.E.D.

APPENDIX C

This appendix demonstrates the convergence of the parameter difference using eqns. (4.18) and a suitable Lyapunov function. Consider the function

$$V(t)=\frac{1}{2}\underline{\delta}^{\mathrm{T}}(t)\Lambda_1^{-1}\underline{\delta}(t)+\frac{1}{2}\underline{\delta}_c^{\mathrm{T}}(t)\Lambda_2^{-1}\underline{\delta}_c(t)=\frac{1}{2}\underline{\tilde{\delta}}(t)\tilde{\Lambda}^{-1}\underline{\delta}(t) \tag{C.1}$$

where

$$\tilde{\Lambda}^{-1}=\left[\begin{array}{c|c}\Lambda_1^{-1} & 0\\ \hline 0 & \Lambda_2^{-1}\end{array}\right]$$

Using eqns. (4.18), the time derivative is

$$\begin{aligned}\frac{\mathrm{d}V(t)}{\mathrm{d}t} &= -c_p(t)\underline{\delta}^{\mathrm{T}}(t)\Lambda_1^{-1}Q_1(t)M_1^{\mathrm{T}}(t)W(t)\underline{\tilde{\varepsilon}}(t)\\ &\quad -c_p(t)\underline{\delta}_c^{\mathrm{T}}(t)\Lambda_2^{-1}Q_2(t)M_2(t)W(t)\underline{\tilde{\varepsilon}}(t)\\ &\quad -c_\lambda(t)\underline{\delta}^{\mathrm{T}}(t)M_1^{\mathrm{T}}(t)W(t)\underline{\tilde{\varepsilon}}(t)-c_\lambda(t)\underline{\delta}_c^{\mathrm{T}}(t)M_2^{\mathrm{T}}(t)W(t)\underline{\tilde{\varepsilon}}(t)=\\ &\quad -c_p(t)\underline{\tilde{\delta}}^{\mathrm{T}}(t)\tilde{\Lambda}^{-1}\tilde{Q}(t)\tilde{M}^{\mathrm{T}}(t)W(t)\tilde{M}(t)\tilde{\delta}(t)-c_\lambda(t)\underline{\tilde{\varepsilon}}^{\mathrm{T}}(t)W(t)\underline{\tilde{\varepsilon}}(t)\end{aligned} \tag{C.2}$$

where

$$\tilde{Q}=\left[\begin{array}{c|c}Q_1 & \\ \hline & Q_2\end{array}\right].$$

This form is negative definite with respect to $\|\tilde{\delta}\|$ provided that $\underline{\tilde{\delta}}(t)$ and $\tilde{M}(t)$ are non-orthogonal. This can be ensured by sufficiently rich input for $\underline{\delta}(t)$ and $M_1(t)$, but not for $\underline{\delta}_c(t)$ and $M_2(t)$.