# MULTIVARIABLE ADAPTIVE PARAMETER AND STATE ESTIMATORS WITH CONVERGENCE ANALYSIS

J. B. MOORE and G. LEDWICH

### Abstract

The convergence properties of a very general class of adaptive recursive algorithms for the identification of discrete-time linear signal models are studied for the stochastic case using martingale convergence theorems. The class of algorithms specializes to a number of known output error algorithms (also called model reference adaptive schemes) and equation error schemes including extended (and standard) least squares schemes. They also specialize to novel adaptive Kalman filters, adaptive predictors and adaptive regulator algorithms. An algorithm is derived for identification of uniquely parameterized multivariable linear systems.

A passivity condition (positive real condition in the time invariant model case) emerges as the crucial condition ensuring convergence in the noise-free cases. The passivity condition and persistently exciting conditions on the noise and state estimates are then shown to guarantee almost sure convergence results for the more general adaptive schemes.

Of significance is that, apart from the stability assumptions inherent in the passivity condition, there are no stability assumptions required as in an alternative theory using convergence of ordinary differential equations.

## 1. Introduction

Consider a signal model with states $x_k$, a noise disturbance $\{v_k\}$ and unknown parameters $\theta$ driven by a known input sequence $\{v_k\}$. The adaptive estimation task is to determine from the (possibly vector) measurement sequence $\{z_k\}$, state estimates $\hat{x}_k$ and parameter estimates $\hat{\theta}_k$.

The task of simultaneously estimating parameters $\theta$ and states $x_k$ to minimize an index such as the conditional error variance is usually too formidable to attempt

176

even for simple model structures, and so there is strong motivation to derive sub-optimal schemes which work reasonably well. The following class of signal models lends itself to a very reasonable sub-optimal scheme.

Consider the class of signal models such that the unknown parameters $\theta$ can be readily estimated perhaps in some optimal fashion, on the assumption that the states are observable (known) and, likewise, state estimates (possibly optimal in some sense) can be achieved given knowledge of the model parameters. Let us denote such parameter and state estimates as $\hat{\theta}_{k/X}$ where $X$ denotes $\{x_0, x_1, \dots, x_k\}$, and $\hat{x}_{k/\theta}$ respectively.

Now a frequently used and very reasonable sub-optimal estimation scheme for simultaneous estimation of $\theta$ and $x_k$ is to implement the two estimation algorithms just referred to but coupled as now described. In estimating $\theta$, $x_k$ is replaced by an estimate of $x_k$ from the state estimator, and in estimating $x_k$, $\theta$ is replaced by an estimate of $\theta$ from the parameter estimation. A suitable notation for these estimates is $\hat{\theta}_{k/\hat{X}}$ and $\hat{x}_{k/\hat{\theta}}$, or more simply $\hat{\theta}_k$ and $\hat{x}_k$.

Many schemes for adaptive estimation in the engineering literature including "equation-error" or "series-parallel" schemes and "output-error" or "parallel" schemes have the general structure of the above sub-optimal arrangement. Examples of such schemes are given in [1–10], with the extended least squares algorithm [1, 7, 8] and its sub-optimal stochastic approximation derivatives [2] being perhaps the archetype of the "equation-error" methods, and the model reference adaptive schemes of [5] and the related instrumental variable algorithms [9, 10] being examples of the output error approach.

Analysis results for the various algorithms of [1–10] have been limited in [1–10] to the noise-free case, and when noise is present, to answering questions of whether or not bias exists in the estimates. It is clearly desirable to determine conditions in the stochastic case for the almost sure convergence of $\hat{x}_k$ to $\hat{x}_{k/\theta}$ as $k \to \infty$, and also of $\hat{\theta}_k$ to $\hat{\theta}$ as $k \to \infty$ for the case of uniquely parameterized models. This is particularly so in view of the fact that, for certain signal models, the adaptive algorithms diverge. Perhaps the most important role for a convergence analysis is to give some deeper explanation as to why the adaptive schemes designed using "engineering intuition" in fact work so well in practice.

In this paper, we first introduce a broad class of signal models including uniquely parameterized models for which suboptimal estimators as described above can be implemented. The class of estimators can be specialized to either "equation error" algorithms such as the extended least squares algorithm or to "parallel" algorithms much like the model-reference adaptive schemes of [5]. Also they can be viewed or re-organized as novel adaptive filters (for example, Kalman filters), predictors, and regulators with attractive convergence properties. Next, convergence analysis results are given via martingale convergence theorems for the stochastic case. A key convergence condition is that a certain system readily derived from the

signal model be strictly passive (or strictly positive real in the time-invariant model case). For consistent parameter estimation in uniquely parameterized models, very reasonable persistently exciting conditions are examined. These conditions are usually satisfied for sufficiently rich excitation signals and adequately modelled stochastic processes, and correspond to the persistently exciting conditions required for the almost sure convergence of least squares estimation algorithms.

A number of aspects of this paper bear some relationship to earlier work in the literature. For example, the continuous time signal models of [6] are specifically designed to have outputs which are bilinear in the model states and unknown parameters as in this paper. However, the signal models, estimators and convergence theory in [6] are deterministic and make no connection to the more realistic task of identification in a stochastic environment. In fact the algorithms do not converge to yield the true parameter estimates in the stochastic case. The model-reference adaptive algorithms of [5] are less general than those presented here but they do have good convergence properties in the stochastic case. The strong convergence results of [5] are limited to the deterministic situation. However, as one would expect, restrictions on the models required to achieve convergence in the deterministic case (passivity restrictions) are also required in the stochastic case here.

One pleasing result of this paper, and possibly unexpected, is that for convergence in the stochastic case no additional restrictions of the deterministic part of the signal model need be imposed. The only additional conditions required in the stochastic case are those of the type familiar in stochastic least squares theory [15].

The Lyapunov function approach of [6] and the hyperstability approach of [5] for the noise-free cases are built on here using martingale convergence theorems for the stochastic case. This approach was first explored by the authors in a conference paper [11] where somewhat weaker and less general results are reported than in the present paper. An alternative approach to using martingale convergence theorems is via ordinary differential equations [12], which is explored in [13, 14].†
The advantage of the approach using martingales is that, given the very accessible martingale convergence theorems, the derivations differ very little from the straightforward derivations of the deterministic theory. In contrast, the ordinary differential equation approach is based on extensive and highly technical derivations. Of perhaps more current interest however is the fact that the theory of this paper does not require the asymptotic stability assumptions in addition to the passivity conditions of [12–14].

---

† Since the first draft of the present paper, *viz.* [11] was presented as a conference paper, [13] has been independently written, based on [12]. Both [14] and the present paper have benefited from this earlier work. The authors wish to acknowledge benefit received from discussions with L. Ljung in revising the present paper.

In Section 2, for a broad class of signal models, associated adaptive estimators are introduced which specialize to known and novel adaptive algorithms. In Section 3, almost sure convergence analysis conditions are given for the stochastic case and in Section 4 specializations and extensions of the results to cover extended least squares, adaptive Kalman filtering, linear system identification and adaptive predictors are briefly discussed. In Section 5, discussions of the results and more specific comparisons with earlier work are presented.

## 2. Signal models and estimation algorithms

Here with the aim of carrying out a performance analysis we restrict attention to adaptive estimation schemes where we can write down state estimation error $(\tilde{x}_k = x_k - \hat{x}_k)$ equations coupled to parameter error $(\tilde{\theta}_k = \theta - \hat{\theta}_k)$ equations. In particular we consider the *model state equations* as

$$x_{k+1} = Fx_k + G_1 y_k + G_2 \nu_k + f_k(\nu_k, z_k),$$

$$z_k = y_k + \nu_k, \quad y_k = \theta'(x_k + \nu_k),$$

where $\theta$ is used here to denote a matrix with unknown elements. The matrices $F$, $G_1$ and $G_2$ are possibly time varying but for convenience the subscript $k$ is deleted. The vector inputs $\nu_k$ and outputs $z_k$ are observable (known) while the state $x_k$ and the noise disturbance $\nu_k$ are not. (The model is selected so that it can specialize to an innovations model and thus the notation $\nu_k$ which is frequently used to denote an innovations sequence.) Manipulations simplify the model equations as

$$x_{k+1} = Fx_k + G\nu_k + \bar{f}_k(\nu_k, z_k),$$

$$z_k = y_k + \nu_k, \quad y_k = \theta'(x_k + \nu_k),$$

where

$$G = G_1 - G_2 \quad \text{and} \quad \bar{f}_k(\nu_k, z_k) = f(\nu_k, z_k) + G_1 z_k.$$

We introduce the assumption that with $\mathscr{F}_k$ denoting the $\sigma$-algebra generated by $\nu_1, \nu_2, \ldots, \nu_k$, the noise term $\nu_k$ satisfies $E[\nu_k | \mathscr{F}_{k-1}] = 0$ with $E[\nu'_k \nu_k | \mathscr{F}_{k-1}] \leqslant 1$. This assumption is certainly satisfied when $\nu_k$ is zero mean, independently distributed, and bounded uniformly above in its covariance.

For the analysis to follow, it is only required that $y_k$ be a bilinear function of the elements of $\theta$ and $[x'_k \nu'_k]$, but for simplicity of presentation, only the special case $y_k = \theta'(x_k + \nu_k)$ is considered.

The signal model described above is so chosen that its inverse is readily constructed. The *state estimation equations* assuming $\theta$ is known are then simply the

state equations of the inverse of the signal model as follows

$$\hat{x}_{k+1/\theta} = F\hat{x}_{k/\theta} + G\hat{v}_{k/\theta} + \tilde{f}_k(v_k, z_k),$$

$$\hat{v}_{k/\theta} = z_k - \hat{y}_{k/\theta}, \quad \hat{y}_{k/\theta} = \hat{z}_{k/\theta} = \theta'(\hat{x}_{k/\theta} + v_k).$$

We now consider the adaptive estimator in which $\theta$ is replaced by some estimate $\hat{\theta}_k$ as

$$\hat{x}_{k+1} = F\hat{x}_k + G\hat{v}_k + \tilde{f}_k(v_k, z_k),$$

$$\hat{v}_k = z_k - \hat{y}_k, \quad \hat{y}_k = \hat{\theta}'_k(\hat{x}_k + v_k).$$

The *state estimation error equations* are now readily derived from a subtraction of the above sets of equations. Using the notation $\tilde{x}_k = x_k - \hat{x}_k$, $\tilde{\theta}_k = \theta - \hat{\theta}_k$, $\psi_k = \hat{x}_k + v_k$, then

$$\tilde{x}_{k+1} = (F - G\theta')\tilde{x}_k - G(-q_k), \quad q_k = -\theta'_k \psi_k. \tag{2.1}$$

Notice that in deriving (2.1), the possibly nonlinear and unbounded function $\tilde{f}(.,.)$ cancels out, and (2.1) is a linear state error equation driven by $q_k = -\tilde{\theta}'_k \psi_k$. The *parameter estimator equation* is taken to be

$$\hat{\theta}_k = \hat{\theta}_{k-1} + P_k \psi_k (z_k - \hat{\theta}'_{k-1} \psi_k)',$$

$$P_k^{-1} = \alpha_k^{-1} P_{k-1}^{-1} + \psi_k \psi'_k, \quad P_k^{-1} \geqslant 0, \quad \alpha_k \geqslant 1. \tag{2.2}$$

The *parameter estimation error equations* are (2.2) together with

$$\tilde{\theta}_k = (I - P_k \psi_k \psi'_k) \tilde{\theta}_{k-1} - P_k \psi_k (\tilde{x}'_k \theta + v'_k). \tag{2.3}$$

REMARKS. 1. In the case $x_k$ is known (that is, $\hat{x}_k = x_k$, $\tilde{x}_k = 0$), the equations are the standard least squares parameter estimation equations. An alternative expression for $P_k$ is given from the matrix inversion lemma as

$$P_k = \alpha_k \{ P_{k-1} - P_{k-1} \psi_k [\alpha_k^{-1} + \psi'_k P_{k-1} \psi_k]^{-1} \psi'_k P_{k-1} \}.$$

2. In practice square root versions of the algorithms (2.2) are used to avoid numerical difficulties ($P_k$ becoming 0) as $P_k$ becomes singular or approaches zero. Also as $P_k$ becomes closer to a singular matrix it is frequently intentionally made more positive than it otherwise would be by the addition of $\bar{\varepsilon}I$ for some $\bar{\varepsilon} > 0$ to allow the algorithm to track slowly varying $\theta$ and to avoid the possibility of round-off error causing $P_k$ to be non-positive definite at some $k$. The theory of this paper will assume that $\bar{\varepsilon} = 0$.

3. If in the above parameter estimator equation, the matrix $P_k$ is replaced by the scalar $\{ \text{tr} [P_k^{-1}] \}^{-1}$ then the recurrence relations for $\text{tr} [P_k^{-1}]$ are scalar and relatively simple to implement. They are a stochastic approximation algorithm. Other

readily implementable schemes set $P_k = P$ a positive definite matrix or perhaps $P_k = P/(\alpha^{-1} + \psi_k' P \psi_k)$. Again, other schemes are more sophisticated adding memory terms consisting of $\psi_{k-i} \tilde{v}_{k-i}'$ for $i = 1, 2, \ldots, M$ in calculating $\hat{\theta}_k$ [5]. Another possibility explored in [11] is to set $\hat{y}_k = \hat{\theta}_{k-1}'(\hat{x}_k + v_k)$ rather than $\hat{y}_k = \hat{\theta}_k'(\hat{x}_k + v_k)$. Reference to these variations on the parameter estimation equations will be made in the body of the paper.

## 3. Convergence analysis

The estimation error equations (2.1), (2.3) upon manipulations† can be re-organized as in Fig. 3.1, namely as a feedback system with input $(-q_k)$ and output $p_k$ and a feedback system with input $(p_k + v_k)$ and output $q_k$. There is added and
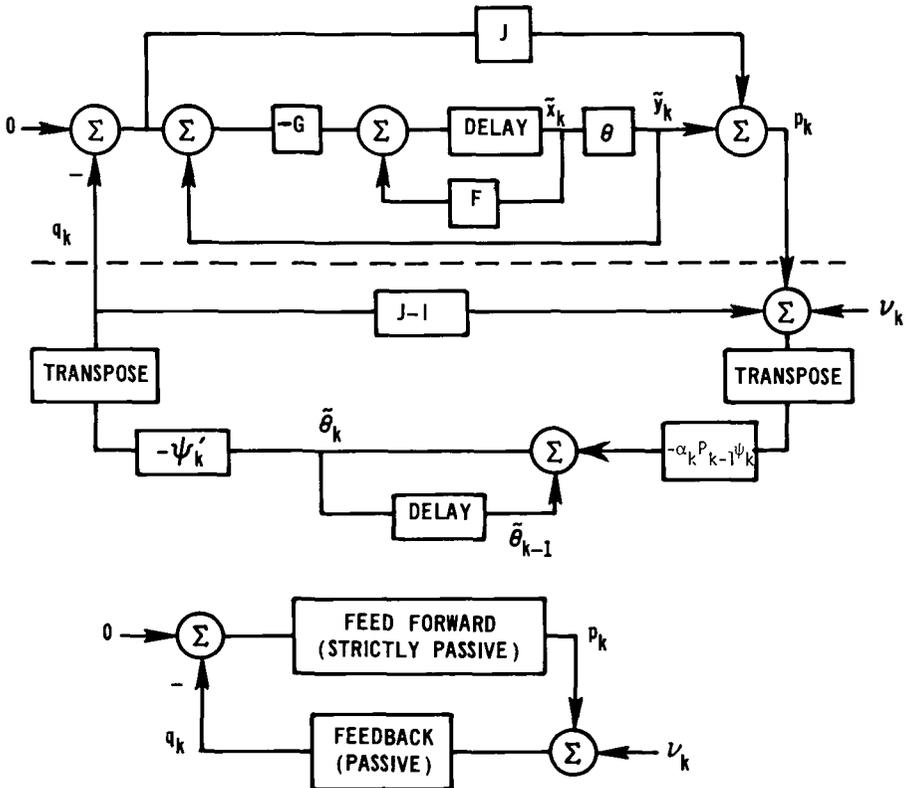


Fig. 3.1. Estimation error equations as two passive systems back to back.

---

† First premultiply both sides of (2.3) by $\alpha_k P_k^{-1} P_k^{-1}$, noting that $(I - P_k \psi_k \psi_k') = P_k P_k^{-1} \alpha_k^{-1}$ from (2.2), then apply the definition $p_k = \tilde{y}_k + J_k(-q_k)$. The remaining manipulations with the definition $q_k = -\tilde{\Theta}_k' \psi_k$ are immediate.

subtracted a feedthrough term $J_k q_k$ which of course does not affect the equations for $\tilde{\theta}_k$ and $\tilde{x}_k$ but is designed to ensure that the feedback system is passive. (Definitions follow.) An assumption that the feedforward system, given in terms of $[F, G, \theta, J]$, is strictly passive then allows us to study the convergence of $p_k, q_k$ and $\tilde{x}_k, \tilde{\theta}_k$ to zero as $k \to \infty$.

Let us now examine the passivity properties of the linear feedback system (3.2) using the fact that a linear system with state equations $x_{k+1} = A_k x_k + B_k u_k$, $y_k = C_k x_k + D_k u_k$ inputs $u_k$ and outputs $y_k$ is known [5] to be passive if for some sequence of positive definite matrices $Q_k$ and all $x_k$, $u_k$ and $k$

$$x_{k+1}' Q_{k+1} x_{k+1} - x_k' Q_k x_k - y_k' u_k - u_k' y_k \leqslant 0.$$

Conditions for the passivity of (3.2) are given in the following lemma.

LEMMA 3.1. *Consider the feedback system of Fig.* 3.1 *where* $P_k$ *is calculated via the least squares recursion* (2.2), *then this system is passive with* $J_k \leqslant \frac{1}{2} I$.

PROOF. For the time-varying linear system (3.2) the left-hand side of (3.3) mildly generalized to handle matrix states yields

$$\delta_k = \operatorname{tr}\{\tilde{\theta}_k'(\alpha_{k+1} P_k)^{-1} \tilde{\theta}_k - \tilde{\theta}_{k-1}'(\alpha_k P_{k-1})^{-1} \tilde{\theta}_{k-1}\} - 2(p_k + \nu_k)' q_k$$

$$= -\operatorname{tr}\{\tilde{\theta}_k'[(\alpha_k P_{k-1})^{-1} - (\alpha_{k+1} P_k)^{-1} + 2\psi_k \psi_k'] \tilde{\theta}_k - 2\tilde{\theta}_k' \psi_k \psi_k' \tilde{\theta}_k J_k\}$$

$$- \alpha_k(\psi_k' P_{k-1} \psi_k) [(p_k + \nu_k)' \cdot q_k] \begin{bmatrix} I & (J_k - I) \\ (J_k - I) & (J_k - I)^2 \end{bmatrix} \begin{bmatrix} p_k + \nu_k \\ q_k \end{bmatrix}.$$

Since the second term is non-positive, we have from manipulation of the first term using (3.3) and the properties of the trace operation that

$$\delta_k \leqslant - \operatorname{tr}\{\tilde{\theta}_k'[1 - \alpha_{k+1}^{-1}] P_k^{-1}] \tilde{\theta}_k + \psi_k' \tilde{\theta}_k(I - 2J) \tilde{\theta}' \psi_k\}. \tag{3.1}$$

When $J_k \leqslant \frac{1}{2}$ and $\alpha_{k+1} \geqslant 1$ then $\delta_k \leqslant 0$, and the feedback system is passive as claimed in the lemma. This completes the proof of the lemma.

The feedforward system (3.1) has input $(-q_k)$, output $p_k$ and states $x_k$. With this system strictly passive, the desired asymptotic stability of the noise-free estimation error equations are immediate via the stability theorems of [17, see also 5]. Strict passivity of this feedforward system implies that there exists a $\bar{Q}_k = \bar{Q}_k' > 0$ such that for all $k$ and some $r \neq 0$

$$\Delta_k = \tilde{x}_{k+1}' \bar{Q}_{k+1} \tilde{x}_{k+1} - \tilde{x}_k' \bar{Q}_k \tilde{x}_k + p_k' q_k + q_k' p_k \leqslant -r^2(\tilde{x}_k' \tilde{x}_k + q_k' q_k). \tag{3.2}$$

For the stochastic case when $\nu_k \neq 0$, before giving the convergence results let us list the assumptions which will be referred to but defer comments on these until later.

**List of assumptions**

(i) The feedforward system of Fig. 3.1 in terms of the signal model parameters $\{F, G, \theta\}$ is strictly passive with $J_k = \frac{1}{2}I$.

(ii) The feedback system of Fig. 3.1 is passive with $J = \frac{1}{2}I$ (guaranteed for the parameter update algorithm of Lemma 3.1).

(iii) The noise conditions $E[\nu_k | \mathscr{F}_{k-1}] = 0$, $E[\nu_k' \nu_k | \mathscr{F}_{k-1}] \leqslant 1$ are satisfied.

(iv) For some $\beta \geqslant 0$, $0 < \gamma < 1$,

$$\lim_{k \to \infty} \sum_{i=0}^{k} i^{-(\gamma+\beta)} \psi_i' P_i \psi_i \leqslant M < \infty \quad \text{w.p. 1.}$$

(v) For some $\beta \geqslant 0$, $0 < \gamma < 1$,

$$\lim_{k \to \infty} k^{(\gamma+\beta)} \alpha_{k+1} P_k = 0 \quad \text{w.p. 1.}$$

(vi) $\psi_k' \psi_k < M_{k-1}$ for all $k$ and some $M > 0$. Also

$$\lim_{k \to \infty} \prod_{i=0}^{n} (F + G\theta')_i = 0, \quad \limsup_{k \to \infty} \sum_{j=0}^{k-1} \left\| \left[ \prod_{i=j+1}^{k-1} (F + G\theta')_i \right] G_j \right\| < \infty$$

(satisfied in the time-invariant model case when $|\lambda_{\max}(F + G\theta')| < 1$ which is guaranteed by (i)).

THEOREM 3.1. *Consider the adaptive estimation algorithms of Section 2 with estimation error equations depicted in the feedback arrangement of Fig. 3.1. Then with* (i)–(iv) *satisfied (except possibly for subsequences* $\{k_1, k_2, \ldots\}$ *with* $(k_{i+1} - k_i) \to \infty$ *as* $k_i \to \infty$)*, as* $k \to \infty$

$$(k-1)^{-\beta/2} \tilde{x}_k \to 0, \quad (k-1)^{-\beta/2} q_k \to 0, \quad (k-1)^{-\beta/2} (\hat{y}_k - y_k) \to 0 \quad (3.3)$$

*in mean square. With conditions* (i)–(v) *holding but relaxing the strict passivity in* (i) *to simply passivity, then as* $k \to \infty$, *almost surely*

$$\hat{\theta}_k \to \theta \quad (3.4)$$

*and with* (vi) *also holding* (3.3) *holds almost surely.*

PROOF. (Here given only for the case $\beta = 0$.) Consider $V_{k|\theta}$ defined from

$$V_{k|\theta} = \tilde{x}_k' \bar{Q}_k \tilde{x}_k + \text{tr}\{(\hat{\theta}_{k-1} - \theta)(\alpha_{k+1} P_k)^{-1}(\hat{\theta}_{k-1} - \theta)\}, \quad (3.5)$$

where $\hat{\theta}_k$, $\alpha_k$, $P_k$ are defined as earlier and are calculated without knowledge of $\theta$. With this definition, addition of (3.1) and (3.2) yields for $[\tilde{x}_k' - q_k'] \neq 0$.

$$\Delta_{k|\theta} + \delta_{k|\theta} = V_{k+1|\theta} - V_{k|\theta} - \nu_k' q_{k|\theta} - q_{k|\theta}' \nu_k.$$

Taking conditional expectations with respect to the $\sigma$-algebra $\mathscr{F}_{k-1}$, noting that $\tilde{x}_k$, $\psi_k$, $P_k$, $\hat{\theta}_{k-1}$ and thus $V_{k|\theta}$ belong to $\mathscr{F}_{k-1}$ and that $E[\nu_k|\mathscr{F}_{k-1}] = 0$, then manipulations show that

$$E[V_{k+1|\theta}|\mathscr{F}_{k-1}] = V_{k|\theta} + E[\Delta_{k|\theta} + \delta_{k|\theta}|\mathscr{F}_{k-1}] + 2(\psi'_k P_k \psi_k)\, E[\nu'_k \nu_k|\mathscr{F}_{k-1}]. \quad (3.6)$$

(The intermediate result $E[q'_{k|\theta}\nu_k|\mathscr{F}_{k-1}] = (\psi'_k P_k \psi_k)\, E[\nu'_k \nu_k|\mathscr{F}_{k-1}]$ is derived by substituting $\tilde{\theta}_{k|\theta}$ from (2.3) into $q'_{k|\theta}\nu_k = \psi'_k \tilde{\theta}_{k|\theta}\nu_k$ and taking conditional expectations.)

Now with the passivity conditions $\Delta_{k|\theta} \leqslant 0$, $\delta_{k|\theta} \leqslant 0$ and the bound

$$E[\nu'_k \nu_k|\mathscr{F}_{k-1}] \leqslant 1,$$

(3.6) yields the inequality

$$E[V_{k+1|\theta}|\mathscr{F}_{k-1}] \leqslant V_{k|\theta} + 2\psi'_k P_k \psi_k. \quad (3.7)$$

To study the convergence properties of $V_{k|\theta}$, let us under (iv) define for some scalar $0 < \gamma < 1$

$$S_{k|\theta} = k^{-\gamma} V_{k|\theta} + 2M - 2\sum_{i=0}^{k-1} i^{-\gamma} \psi'_i P_i \psi_i \quad (3.8)$$

and observe using (3.7) that for $k \geqslant 0$

$$E[S_{k+1|\theta}|\mathscr{F}_{k-1}] = (k+1)^{-\gamma} E[V_{k+1|\theta}|\mathscr{F}_{k-1}] + 2M - 2\sum_{i=0}^{k} i^{-\gamma} E[\psi'_i P_i \psi_i|\mathscr{F}_{k-1}]$$

$$\leqslant k^{-\gamma}(V_{k|\theta} + 2\psi'_k P_k \psi_k) + 2M - 2\sum_{i=0}^{k} i^{-\gamma} \psi'_i P_i \psi_i$$

$$= S_{k|\theta}.$$

Moreover, taking expectations yields that $E[S_{k+1|\theta}] \leqslant E[S_{k|\theta}]$ and thus

$$E[S_{k|\theta}] \leqslant E[S_{1|\theta}] = S_{1|\theta} = \operatorname{tr}\{(\hat{\theta}_1 - \theta)' P_1^{-1}(\hat{\theta}_1 - \theta)\} + \tilde{x}'_1 Q \tilde{x}_1.$$

With finite initial conditions, then $S_{1|\theta} < \infty$ and $E[S_{k|\theta}] < \infty$. Thus since

$$E[S_{k+1|\theta}|\mathscr{F}_k] \leqslant S_{k|\theta}, \quad S_{k|\theta} \geqslant 0 \quad \text{and} \quad E[S_{k|\theta}] < \infty$$

for all $k$, $S_{k|\theta}$ is a positive super martingale (w.p.1.) on $\mathscr{F}_k$ and converges almost surely [16]. We conclude that

$$\limsup_{k \to \infty} k^{-\gamma} V_{k|\theta} < \infty, \quad \text{w.p.1.}$$

from which

$$\limsup_{k \to \infty} \operatorname{tr}\{(\hat{\theta}_k - \theta)'\, [k^{-\gamma}(\alpha_{k+1} P_k)^{-1}](\hat{\theta}_k - \theta)'\} < \infty, \quad \text{w.p.1.}$$

Condition (v) yields immediately (3.4) as required. The bound on $\psi_k$ of (vi) implies $q_k \to 0$ almost surely as $k \to \infty$ and applications of the Toeplitz Lemma to the non-recursive expression for $\tilde{x}_k$ readily obtained from Fig. 3.1 yields that $\tilde{x}_k \to 0$ almost surely as $k \to \infty$ under the remaining conditions of (vi).

Taking expectations of both sides of (3.6), we have

$$E[V_{k+1|\theta}] < E[V_{k|\theta}] - E[|\Delta_{k|\theta}|] + 2E[\psi_k' P_k \psi_k]$$

or with $\gamma > 0$

$$(k+1)^{-\gamma} E[V_{k+1|\theta}] \leqslant k^{-\gamma} E[V_{k|\theta}] - k^{-\gamma} E[|\Delta_{k|\theta}|] + 2E[k^{-\gamma} \psi_k' P_k \psi_k].$$

Recursive application of this inequality for $k = 1, 2, ..., n$ yields

$$(n+1)^{-\gamma} E[V_{n+1|\theta}] \leqslant E[V_{1|\theta}] = \sum_{k=1}^{n} k^{-\gamma} E[|\Delta_{k|\theta}|] + 2E\left[\sum_{k=1}^{n} k^{-\gamma} \psi_k' P_k \psi_k\right].$$

Taking limits as $n \to \infty$, we see that unless

$$\lim_{n \to \infty} \sum_{i=1}^{n} E[|\Delta_{k|\theta}|] k^{-\gamma} < \infty, \quad \text{w.p.l.}$$

then the upper bound for $(n+1)^{-\gamma} E[V_{n+1|\theta}]$ would be negative (at least with (iv) satisfied) violating the non-negativity constraint on $E[V_{k+1|\theta}]$.

We conclude that with (i)–(iv), $E[\Delta_{k|\theta} \to 0$ as $k \to \infty$ (except possibly for subsequences $\{k_1, k_2, ...\}$ where $k_{i+1} - k_i \to \infty$ as $k_i \to \infty$), and in turn from (3.2) that (3.3) is satisfied as desired. Note that $y_k - \hat{y}_k = \theta' \tilde{x}_k - q_k$.

REMARKS. 1. For the case when $F$, $G$, $J$, $\theta$ are constant the transfer function of the feedforward system in Fig. 3.1 is $W(z) = J - \theta'[zI - (F - G\theta')]^{-1}$. Now $W(z)$ is strictly passive, equivalently strictly positive real if and only if (i) $W(z)$ is real for real $z$, (ii) $W(z)$ has no poles in $|z| \geqslant 1$, (iii) $W(e^{j\omega}) + W(e^{-j\omega}) > 0$ for all real $\omega$.

2. For the case when $\tilde{x}_k = 0$ for all $k$, as in standard least squares, then $\Delta_k = 2p_k' q_k = -2q_k J_k q_k \leqslant 0$ and the passivity of $W(z)$ is assured. The results reduced to those given in [15b].

3. For the case $\nu_k = 0$, $E[V_{k+1|\theta} | \mathscr{F}_{k-1}] \leqslant V_{k|\theta}$, and so $V_k$ converges. It immediately follows that $\Delta_k \to 0$. Thus (3.3) holds almost surely under (i) and (ii) and with the additional condition (v), then (3.4) holds almost surely.

4. The restriction $\beta = 0$ in the above proof can be relaxed. We simply apply the theorem for the case $\beta = 0$ to a modified signal model where the output is now $z_k^* = (k-1)^{-\beta/2} z_k$ and the results interpreted to yield the theorem for the case $\beta > 0$.

5. The above convergence results tell us that for some signal models (those with an associated passivity condition satisfied) convergence can be achieved. Studies elsewhere reported in [12, 13] give some evidence that the passivity

condition is virtually a necessary condition for convergence. It appears that if the passivity condition fails then the algorithm diverges or vacillates between appearing to converge over a significant time period and then appearing to diverge for a short time period.

6. The passivity condition is automatically satisfied when $G = 0$, or when $G = G_2 - G_1$ is sufficiently small. To gain some insight into when $G$ may be small, recall that when $G_1 = 0$, $G_2$ is the Kalman gain of the conditional estimator. The Kalman gain is known to be small when the output measurement noise in the usual state space signal model is large. We conclude that for signal models with $G_1 = 0$ and for sufficiently high measurement noise, then the passivity condition is satisfied. Also we could comment that in the high noise case, the persistently exciting conditions are more likely to be satisfied.

7. Notwithstanding the above remarks, it should be noted that in general the passivity condition does depend on $\theta$ which is of course unknown, but if $\theta$ belongs to a known compact set, then of course it can be checked (albeit tediously) over this set.

8. For the output-error algorithm on which $G_2 = 0$, it may be that the passivity condition is not satisfied for the parameter update algorithms studied in this paper, but it will be satisfied for ones involving the memory terms $\psi_{k-i} J_{k-i}$ for $i = 1, 2, ..., M$ as studied in [5].

9. The real power of the theory is that it does give us a tangible explanation as to why extended least squares and related algorithms work so well most of the time but for unusual models violating the passivity condition they fail. We see that when an algorithm fails, it is probably not simply a matter of poor initial conditions. The word "probably" is used here to cover cases when the signals may not be sufficiently rich or the numerical calculations ill-conditioned in some way not explored in the theory here. Future theoretical work on these algorithms could well yield robustness and finite-time results.

10. Condition (iv) is not as simple as one would like, being hard to verify and current research efforts are to replace the condition by one which simply requires that $P_\infty = 0$, at least to ensure that $\hat{\theta}_k \to \infty$ w.p.1. However, even in the standard least squares case when $x_k = 0$ for all $k$, it is not yet shown that such a condition is all that is needed to prove that $\theta_k \to \infty$.p.1.

Of course, condition (iv) is satisfied for arbitrary $\beta + \gamma > 0$ when $\psi'_k P_k \psi_k \to 0$ as $1/k$. Such a situation arises when $\psi_k$ is derived from an asymptotically stable signal model and is sufficiently rich in the sense that $\lambda_i[P_k] \to 0$ as $1/k$ for all $i$. Note, however, that there is an implied restriction that $k^{-(\gamma+\beta)} \psi' P_k \psi_k$ itself be bounded, for otherwise as when $\psi_k$ is gaussian there is a finite probability, perhaps negligibly small, that the bound (iv) is exceeded. In other words, there is this hidden restriction that the noise term $k^{-(\gamma+\beta)} P_k v_k^2$ be bounded. The condition (iv) is of course satisfied for arbitrary $\beta + \gamma > 1$ since from (2.2) we have that $\psi'_k P_k \psi_k \leq 1$. A more precise

explanation of the class of systems for which (iv) is satisfied is beyond the scope of this paper.

11. The theory of [13] provides for almost sure convergence in (3.3) rather than mean squares convergence here. Note, however, that in the unique parameterized model case when (3.4) holds, then under (vi) there is also almost sure convergence of in (3.3).

12. A theory along the lines given above for the case $\beta = 0$ can be worked out for stochastic approximation versions, at the expense of a few additional complications not discussed here.

13. Observe that when $\beta + \gamma > 1$, then (iv) is satisfied since $\psi_k' P_i \psi_i \leqslant 1$ for all $i$ and (v) is satisfied when $\psi_k$ is derived from a system with all modes unstable forcing $P_k$ to approach zero exponentially [15]. Thus we can violate the stablity restrictions of [13] requiring $\psi_k$ to be derived from an asymptotically stable system and still achieve convergence. Results can be obtained for the mixed stable/unstable mode case but these are not explored here because of space limitation. Actually the results of [13] can also be weakened to avoid this stability restriction as discussed in a later paper.

## 4. Useful specializations and generalizations

In this section, a number of useful signal models and adaptive estimation algorithms, which are specializations of the more general case discussed so far, are now described. Convergence conditions for these cases are particularly simple ones. We assume throughout this Section that the least squares parameter update scheme (2.2) is employed.

### An output error algorithm

Consider the signal model

$$y_k = -a_1 y_{k-1} - a_2 y_{k-2} \ldots - a_k y_{k-n} + v_k + b_1 v_{k-1} + \ldots b_m v_{k-m},$$

$$z_k = y_k + v_k \quad (v_k \text{ zero mean and white})$$

or

$$y_k = \theta' x_k, \quad z_k = y_k + v_k,$$

$$\theta' = [-a_1 - a_2 \ldots - a_k \vdots b_1 b_2 \ldots b_m], \quad x_k' = [y_{k-1} y_{k-2} \ldots y_{k-n} \vdots v_{k-1} v_{k-2} \ldots v_{k-n}].$$

An output error algorithm or parallel algorithm for this model akin to those of [3, 5] and others is simply

$$\hat{y}_k = \hat{\theta}'_k \hat{x}_k, \quad \hat{\theta}_k = \hat{\theta}_{k-1} + P_k \hat{x}_k (z_k - \hat{\theta}'_{k-1} \hat{x}_k), \quad P_k^{-1} = P_{k-1}^{-1} + \hat{x}_k \hat{x}'_k,$$

where, of course, $\hat{x}'_k = [\hat{y}_{k-1} \ldots \hat{y}_{k-n} v_{k-1} \ldots v_{k-m}]$. This is clearly a specialization of the scheme of Section 2, and the feedforward system of Fig. 3.1, which is required to be passive, simplifies as

$$W(z) = \tfrac{1}{2}I - \theta'[zI - (F - G\theta')]^{-1} G = A^{-1}(z) - \tfrac{1}{2}I,$$

where $A(z) = 1 + a_1 z^{-1} + \ldots a_n z^{-n}$, and

$$F = \begin{bmatrix} 0 & 0 & 0 & 0 \\ I_{n-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & I_{m-1} & 0 \end{bmatrix}, \quad G = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

## A novel adaptive Kalman filter

Consider the state space signal model

$$x_{k+1} = Ax_k + Bu_k, \quad z_k = Cx_k + w_k \quad \text{(scalar)},$$

$$E\left\{ \begin{bmatrix} u_k \\ w_k \end{bmatrix} [u_k \, w_k] \right\} = \begin{bmatrix} Q & S \\ S' & R \end{bmatrix}, \quad E\begin{bmatrix} u_k \\ w_k \end{bmatrix} = 0,$$

with states $x_k$, noise driving terms $u_k$, $w_k$, known parameters $[A, B, C]$ but unknown noise covariance matrices $[Q, R, S]$. Then the conditional minimum variance one-step-ahead predictor has the structure indicated in Fig. 4.1(a) where $\bar{W}(z) = C(zI - A)^{-1}$ and the steady state Kalman gain $K$ is unknown. An alternative structure is given in Figure 4.1(b) since $\bar{W}(z) K = K\bar{W}(z)$ (a scalar transfer function).

It is clear from Fig. 4.1 that only the second structure of Fig. 4.1(b) is suitable to be made adaptive along the lines taken in Section 2. Figure 4.2 depicts such an adaptive scheme. Notice that we have not in the first instance worked with the state space signal model above or the signal model of Section 2 for that matter.

For the scheme of Fig. 4.2, the convergence condition that the feedforward system of Fig. 3.1 be passive simplifies to requiring that

$$W(z) = \tfrac{1}{2}I - \theta'[zI - (F - G\theta')]^{-1} G$$

be passive where

$$F = A, \quad G = -C', \quad \theta' = K'.$$

For the case of vector measurements no longer is $W(z) K = K' W'(z)$. However, the appropriate re-organization is possible as illustrated in the following example. Let

$$W(z) = \begin{bmatrix} \bar{w}_{11}(z) & \bar{w}_{12}(z) \\ \bar{w}_{21}(z) & \bar{w}_{22}(z) \end{bmatrix}, \quad K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}$$
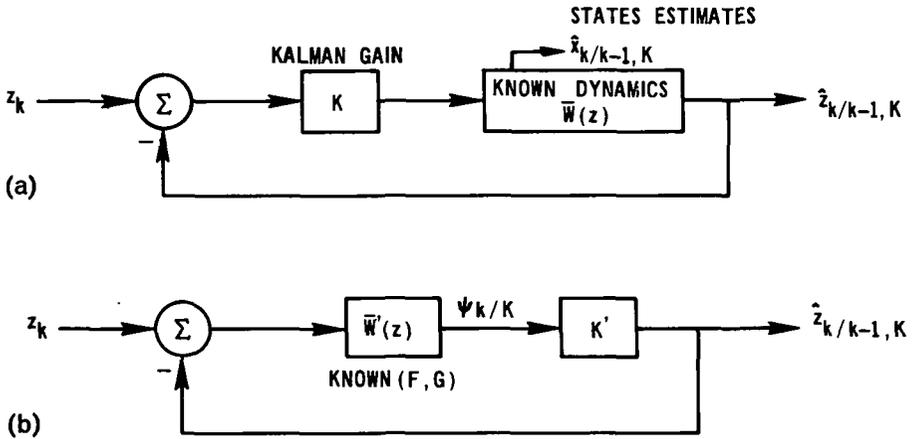


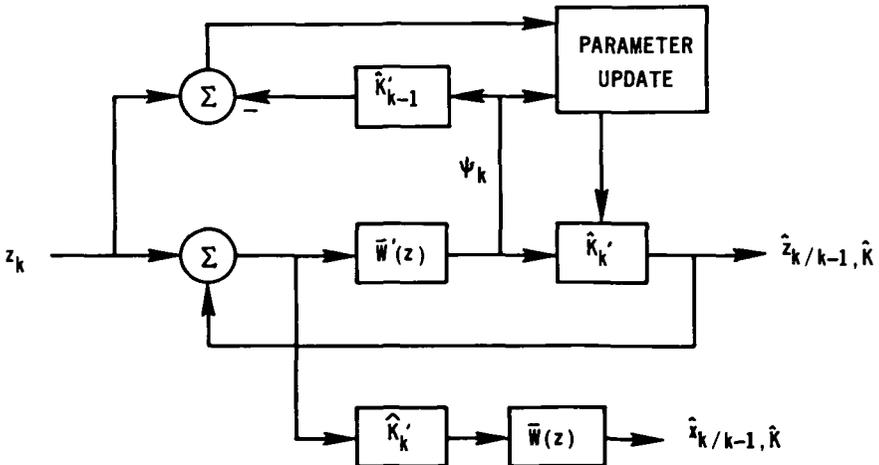Fig. 4.1. Alternative structures of the Kalman one-step-ahead predictor.



Fig. 4.2. Adaptive one-step-ahead predictor.

for scalar $\bar{w}_{ij}(z)$ and $K_{ij}$. Now $\tilde{W}(z)\,K = \theta'_p\,W(z)$ for

$$
\tilde{W}(z) =
\begin{bmatrix}
\bar{w}_{11}(z) & 0 \\
0 & \bar{w}_{11}(z) \\
\bar{w}_{12}(z) & 0 \\
0 & \bar{w}_{12}(z) \\
\bar{w}_{21}(z) & 0 \\
0 & \bar{w}_{21}(z) \\
\bar{w}_{22}(z) & 0 \\
0 & \bar{w}_{22}(z)
\end{bmatrix},
\quad
\theta_p =
\begin{bmatrix}
K_{11} & 0 \\
K_{12} & 0 \\
K_{21} & 0 \\
K_{22} & 0 \\
0 & K_{11} \\
0 & K_{12} \\
0 & K_{21} \\
0 & K_{22}
\end{bmatrix}.
$$

Here $\theta_p$ is partially specified and a further re-arrangement allows use of the alternative output equation $\hat{y}_k = \hat{X}'_k\,\hat{\theta}_k$ for

$$
\hat{\theta}' = [\hat{K}_{11}\,\hat{K}_{12}\,\hat{K}_{21}\,\hat{K}_{22}], \quad \hat{X}_k = [\hat{X}_{12} \vdots \hat{X}_{22}],
$$

where

$$
\hat{X}_{1k} =
\begin{bmatrix}
\bar{w}_{11} & 0 \\
0 & \bar{w}_{11} \\
\bar{w}_{21} & 0 \\
0 & \bar{w}_{21}
\end{bmatrix} \hat{v}_k,
\quad
\hat{X}_{2k} =
\begin{bmatrix}
\bar{w}_{21} & 0 \\
0 & \bar{w}_{21} \\
\bar{w}_{22} & 0 \\
0 & \bar{w}_{22}
\end{bmatrix} \hat{v}_k.
$$

Here $\bar{w}_{ij}$ denotes the operation corresponding to $\bar{w}_{ij}(z)$. Clearly, identification of $\theta$ gives directly the Kalman gain. (More precisely, a least squares index for the vector measurement case weighted by the innovations covariance or a sampled asymptotic approximation to this gives the Kalman gain.) The complete details are readily worked out.

We believe the scheme described above is the first description of a Kalman filter with an adaptive Kalman gain where precise on-line convergence results are given.

### Extended least squares

Consider the signal model

$$
z_k = -\sum_{i=1}^{n} a_i z_{k-i} + \sum_{i=1}^{m} b_i v_{k-i} + \sum_{i=1}^{p} c_i v_{k-i} + v_k
$$

with $\nu_k$ zero mean and white. These equations can be re-expressed as $z_k = \theta' x_k + \nu_k$ where

$$\theta' = [-a_1 \dots a_n \vdots b_1 \dots b_m \vdots c_1 \dots c_m],$$

$$x'_k = [z_{k-1} \dots z_{k-n} \vdots v_{k-1} \dots v_{k-m} \vdots \nu_{k-1} \dots \nu_{n-p}].$$

The extended least squares estimator is

$$\hat{z}_k = \hat{\theta}'_k \hat{x}_k, \quad \hat{\theta}_k = \hat{\theta}_{k-1} + P_k x_k(z_k - \hat{\theta}'_{k-1} \hat{x}_k), \quad P_{k-1}^{-1} = P_{k-1}^{-1} + \hat{x}_k \hat{x}'_k$$

where

$$\hat{x}'_k = [z_{k-1} \dots z_{kn-} \vdots v_{k-1} \dots v_{k-m} \vdots \hat{v}_{k-1} \dots \hat{v}_{k-p}] \quad \text{and} \quad \hat{v}_i = z_i - \hat{z}_i.$$

Again this is a specialization of the scheme of Section 2 and the feedforward system of Fig. 3.1 required to be passive by the convergence theory simplifies as

$$W(z) = \tfrac{1}{2}I - \theta'[zI - (F - G\theta')]^{-1} G = C^{-1}(z) - \tfrac{1}{2}I,$$

where $C(z) = 1 + c_1 z^{-1} + \dots c_p z^{-p}$,

$$F = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ I_{n-1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_{m-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & {}_{p-1} & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

## Novel self-tuning predictor

It is almost trivial to convert a one-step-ahead Kalman filter into an $N$-step-ahead predictor since with $x_{k+1} = Fx_k + Gw_k$, $\hat{x}_{k+n/k} = F^N \hat{x}_{k/k}$. However, if only an estimate of $F$, viz. $\hat{F}$, is available from an adaptive Kalman filter, then there may be considerable errors in calculating an approximate prediction using $\hat{F}_k^N$ and there is considerable computational effort involved. An alternative approach to adaptive prediction is given in [19]. This we build upon here to obtain a novel predictor so that the convergence analysis for self-tuning filters can be applied directly. In this way good convergence properties are assured.

Consider the predictor of Fig. (4.3) re-organized as the feedback structure of Fig. 4.3(b). The feedforward sub-system can be viewed as an arrangement as in Fig. 4.4 with a known linear dynamical system $W$ and an unknown parameter

matrix $\theta'$. Self-tuning versions of these can be constructed but there is at present no convergence analysis. Consider the non-minimal re-organization of Fig. 4.5(a) with a known block $\bar{W}$ consisting of an $N$-delay and the block $\bar{W}$ and the unknown parameters $\theta'$. This re-organization can now be made adaptive as indicated in Fig. 4.5(b) and it can be seen that such an adaptive prediction is a specialization of the adaptive schemes of Section 2 with augmentation to achieve the desired sub-optimal $N$-step-ahead prediction estimate $z^s_{k+N/k}$. The augmentations do not
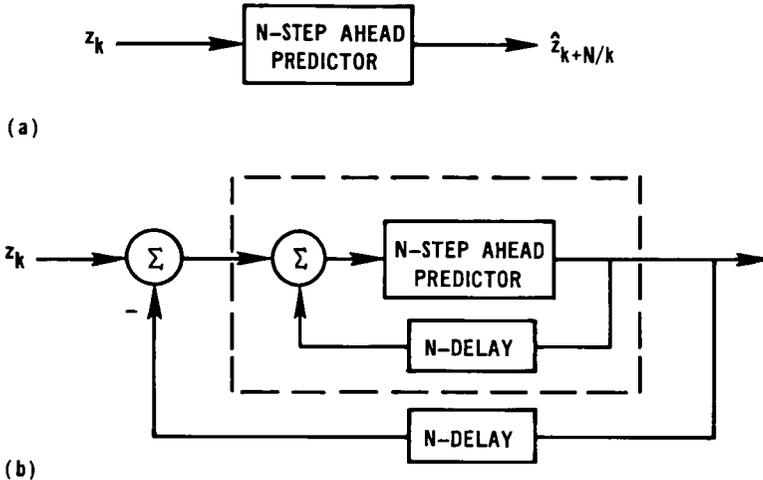
(a)

(b)

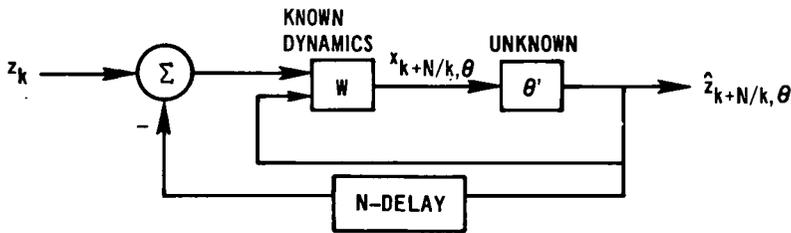Fig. 4.3. Predictor re-organized as a feedback structure.

Fig. 4.4. $N$-step-ahead predictor conditioned on $\theta$.

affect the convergence analysis. The derivations of the passivity conditions in terms of the parameter $\theta$ and $W$ is straightforward.

## Multivariable linear system identification

The schemes previously described under the heading "An output error algorithm" and "Extended least squares" are for scalar output system identification. A natural question to ask is to what extent can the ideas be extended to cover the

multivariable case. Of course, one can immediately replace all the scalar parameters $a_i, b_i, c_i$ by matrices $A_i, B_i, C_i$, and $\theta$ is then a matrix rather than a vector. The catch is that the models are not uniquely parameterized. There may be situations where it is important to work with a uniquely parameterized model so that $\hat{\theta}_k$ has
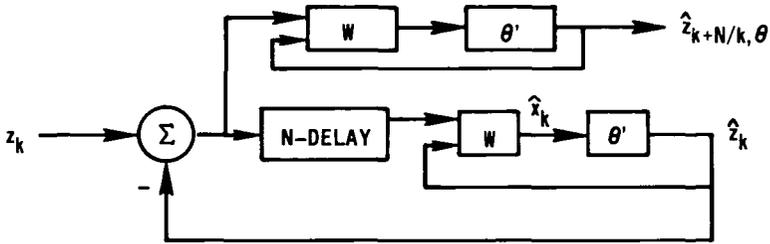


(a)
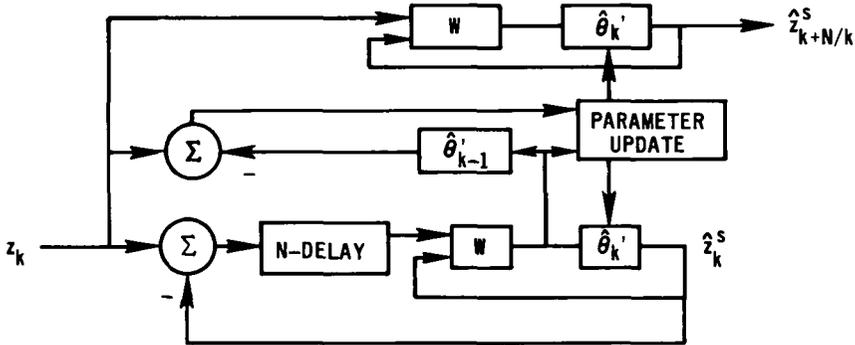


(b)

Fig. 4.5. (a) Conditional optimal predictor. (b) Self-tuning version of optimal predictor.

some significance and the difficulties sometimes associated with non-unique parameterization are avoided. (One such difficulty is that the matrix $P_k$ perhaps approaches a singular but non-zero matrix giving numerical problems.)

Consider the uniquely parameterized model for $z_k$ an $m$-vector

$$z_k = \sum_{i=1}^{n} a_i z_{k-i} + \sum_{i=1}^{p} C_i \nu_{k-i} + \nu_k,$$

where $a_i$ are scalar and $C_i$ matrices. In [20], least squares ideas have been applied

G

by re-organizing this equation in terms of an unknown parameter vector $\phi$ as

$$z_k = y_k + \nu_k, \quad y_k = X'_k,$$

$$\phi' = [\theta' \vdots c'_1 c'_2 \dots c'_m], \quad \theta' = [a_1 a_2 \dots a_n],$$

$$\bar{\theta}' = [C_1 C_2 \dots C_p] = [c_1 c_2 \dots c_m]',$$

$$X'_k = [\bar{X}'_k D_k], \quad D_k = \text{block diag}\,[x'_k, x'_k, \dots, x'_k],$$

$$\bar{X}'_k = [z_{k-1} z_{k-2} \dots z_{k-p}], \quad x'_k = [\nu'_{k-1} \nu'_{k-2} \dots \nu'_{k-p}].$$

Note that here $y_k$ is a bilinear function of $\theta$ and $x_k$ (and $z_i$). The feedforward system required to be passive by the convergence theory in this case is

$$W(z) = C^{-1}(z) - \tfrac{1}{2}I,$$

where $C(z) = I + C_1 z^{-1} + C_2 z^{-2} + \dots C_p z^{-p}$.

The adaptive estimation algorithms for estimating $\theta$ require inefficient manipulation of the sparse matrix $X_k$.

Here we present a novel formulation of the problem to achieve novel faster algorithms without the need to manipulate sparse matrices. First note the reorganization

$$z_k = \bar{X}'_k \theta + \bar{\theta}' x_k + \nu_k.$$

Mild variations of the extended least squares derivations lead to the "recursions"

$$S_k = S_{k-1} - S_{k-1} \hat{\bar{X}}_k (\hat{\bar{X}}'_k S_{k-1} \hat{\bar{X}}_k + I)^{-1} \hat{\bar{X}}'_k S_{k-1},$$

$$R_k = R_{k-1} - R_{k-1} \hat{x}_k (\hat{x}'_k R_{k-1} \hat{x}_k + 1)^{-1} \hat{x}'_k R_{k-1},$$

$$\hat{\theta}_k = \hat{\theta}_{k-1} - S_k \hat{\bar{X}}_k (\hat{\bar{X}}'_k \hat{\theta}_{k-1} + \hat{\bar{\theta}}'_k \hat{x}_k - z_k),$$

$$\hat{\bar{\theta}}_k = \hat{\bar{\theta}}_{k-1} - R_k \hat{x}_k (\hat{x}'_k \hat{\bar{\theta}}_{k-1} + \hat{\theta}'_k \hat{\bar{X}}_k - z'_k).$$

Observe that truly recursive calculations† can be obtained by substituting for $\bar{\theta}_k$ to yield

$$\hat{\theta}_k = (I - \hat{x}'_k R_k \hat{x}_k S_k \hat{\bar{X}}_k \hat{\bar{X}}'_k)^{-1} [(I - S_k \hat{\bar{X}}_k \hat{\bar{X}}'_k) \hat{\theta}_{k-1} + (1 - \hat{x}'_k R_k \hat{x}_k) S_k \hat{\bar{X}}_k (z_k - \hat{\bar{\theta}}'_k \hat{x}_k)].$$

---

† We are indebted to a student, Mr. T. H. Dinh, for this observation and the remarks to follow.

Of course we require that the inverses exist. From our simulation experience we do not see this as a significant limitation. As an example of the computational efficiency of these algorithms, for a fourth-order system with four outputs, the number of multiplications required in these faster algorithms is reduced by a factor of 13. There is a greater reduction with higher order systems.

## 5. Conclusions and discussions

1. The signal model and adaptive estimation schemes considered in the paper have been shown to specialize to a number of useful estimation schemes such as novel adaptive Kalman filters, novel adaptive predictors and novel adaptive parameter identifiers in uniquely parameterized multivariable signal models, as well as to the more familiar ARMA parameter identifiers and extended least squares identification schemes. They can also be mildly modified to treat the model reference identification schemes [4, 5] and the algorithms using instrumental variables [9, 10].

2. The convergence conditions for parameter identification in the stochastic case consist of some reasonable restrictions on the noise, noise-free convergence conditions, and additional conditions which are but mildly more restrictive than simply requiring that the matrix $P_k$ (a coefficient matrix in the parameter update equations) approaches zero as $k \to \infty$. Such restrictions are also required in the alternative ordinary differential equation (ODE) approach of [14], but in addition [14] includes stability restrictions. The ODE approach of [14] appeals to theorems which are not so simple in derivation as the martingale convergence theorem referred to in this paper.

The noise-free convergence condition, noted above, is that a system directly related to the signal model be passive (or positive real). That such a condition is required was first observed in [5]. There is also evidence that the condition is a necessary one [13, 14].

3. Continuous-time versions of the results in this paper are readily worked out. There is very little variation required in the technique. It appears that the ideas of [12–14] can also be extended to the continuous time case.

4. The less general estimation schemes of the earlier paper [11] are for very closely related parameter update algorithms. The noise-free convergence conditions in [11] are not as clear as in this paper and the convergence analysis of [11], although using the martingale convergence approach, yields weaker convergence results than in the present paper.

5. The adaptive schemes of this paper can be applied to yield adaptive controller designs for both minimum variance regulators and optimal state feedback regulators.

## Acknowledgement

### References

[1] K. J. Astrom and P. Eykhoff, "System identification—a survey", *Automatica* 7 (1971), 123–162.
[2] G. N. Saridis, "Stochastic approximation methods for identification and control", *IEEE Trans. Auto. Control AC*-19 (1974), 798–809.
[3] P. Kudva and K. S. Narendra, "An identification procedure for discrete multivariable systems", *IEEE Trans. Auto. Control AC*-19 (1974), 549–552.
[4] P. Kudva and K. S. Narendra, "The discrete adaptive observer", *Proc. of Decision and Control Conference* (1974), 307–312.
[5] I. D. Landau, "Unbiased recursive identification using model reference adaptive techniques", *IEEE Trans. Auto. Control AC*-21 (1976), 194–203. Also an addendum submitted for publication.
[6] B. D. O. Anderson, "An approach to multivariable identification", *Automatica* 13 (1977), 401–408.
[7] D. W. Clarke, "Generalized least squares estimation of the parameters of a dynamic model", *1st IFAC Symposium on Identification in Automatic Control Systems* (Prague, 1967).
[8] T. Soderstrom, "Convergence properties of the generalized least squares identification method", *Automatica* 10 (1974), 617–626.
[9] B. M. Finigan and I. H. Rowe, "Strongly consistent parameter estimation by the introduction of strong instrumental variables", *IEEE Trans. Auto. Control AC*-19 (1974), 825–831.
[10] P. D. Young, "An instrumental variable method for real-time identification of a noisy process", *Automatica* 6, 271–287.
[11] G. Ledwich and J. B. Moore, "Multivariable self-tuning filters", *Differential Games and Control Theory II*, (Lecture notes in pure and applied mathematics, Dekker, 1977), pp. 345–376.
[12] L. Ljung, "Analysis of recursive, stochastic algorithms", *IEEE Trans. Auto. Control AC*-22 (1977), 551–575.
[13] L. Ljung, "On positive real transfer function and the convergence of some recursive schemes", *IEEE Trans. Auto Control AC*-22 (1977), 539–551.
[14] L. Ljung, "Convergence of an adaptive filter algorithm", *International Control* 27 (1978), 673–694.
[15] (a) J. B. Moore, Martin C. Clark and B. D. O. Anderson, "On martingales and least squares linear system identification", *IFAC Symposium of Identification and System Parameter Estimation* (USSR, 1976). (b) Also J. B. Moore, "On strong consistency of least squares identification algorithms", *Automatica* 14 (1978), 505–509.
[16] B. D. O. Anderson and J. B. Moore, *Linear optimal control* (Prentice-Hall, Englewood Cliffs, N.J., 1970).

[17] C. Desoer and M. Vidyasaga, *Feedback systems: input–output properties* (Academic Press, New York, 1975).

[18] J. L. Doob, *Stochastic processes* (John Wiley and Sons, New York, 1953).

[19] B. Wittenmark, "A self-tuning predictor", *IEEE Trans. Auto. Control AC*-19 (1974), 848–851.

[20] G. C. Goodwin and R. L. Payne, *Dynamic System identification* (Academic Press, New York, 1977).

Electrical Engineering Department
University of Newcastle
Newcastle, N.S.W. 2308
and
Electrical Engineering Department
University of Queensland,
St Lucia, Qld 4067