Asymptotically Convergent Modified Recursive Least-Squares with Data-Dependent Updating and Forgetting Factor for Systems with Bounded Noise

SOURA DASGUPTA AND YIH-FANG HUANG, MEMBER, IEEE

Abstract—Continual updating of estimates required by most recursive estimation schemes often involves redundant usage of information and may result in system instabilities in the presence of bounded output disturbances. An algorithm which eliminates these difficulties is investigated. Based on a set theoretic assumption, the algorithm yields modified least-squares estimates with a forgetting factor. It updates the estimates selectively depending on whether the observed data contain sufficient information. The information evaluation required at each step involves very simple computations. In addition, the parameter estimates are shown to converge asymptotically, at an exponential rate, to a region around the true parameter.

I. Introduction

MANY SYSTEMS commonly found in communication and control theory can be modeled by autoregressive exogenous input (ARX) schemes of the form:

$$y_k = \sum_{i=1}^{n} a_i y_{k-i} + \sum_{j=0}^{m} b_j u_{k-j} + v_k. \tag{1.1}$$

Here $\{y_k\}$ and $\{u_k\}$ are the measurable output and input sequences, respectively, and $\{v_k\}$ is a sequence of uncorrelated disturbances corrupting the system. An important problem in both adaptive signal processing and control concerns the use of recursive least squares (RLS) and other estimation techniques for the identification of processes such as (1.1).

A feature of most recursive algorithms [1]–[5] is the continual update of parameter estimates without regard to the benefits provided. Thus even if a new measurement contains no fresh information and even if its use fails to result in any improvement in the quality of estimation, the update does not cease. In practice this may lead to significant redundancies, whose elimination could result in more efficient algorithms with fewer parameter estimate updates. Accordingly, one of the issues which this paper addresses is the formulation of adaptive algorithms having more discerning update strategies.

The second issue of interest relates to the case where a bound on the magnitude of v_k is available. Such a situation occurs frequently in both signal processing and control. In speech processing systems, for example, the disturbances in voice-band signals obey such a bound. Currently available recursive estimators result in prediction errors which eventually become less than or equal to the disturbance bound. However, the parameter estimates continue to be updated unless either the prediction error goes to zero or the update gain is asymptotically driven to zero [6]. While the former situation is necessarily rare, the latter removes any ability of tracking slow time variation. On the other hand, in most applications the asymptotic cessation of the update of parameter estimates is highly desirable. In adaptive control, for example, noncessation of updating could lead to system instability.

In this paper, we reformulate RLS estimation with the aforementioned issues in mind. Ours is similar to the set theoretic approach of [7] and [8] with the following important differences. Our algorithm, in the ideal case, is assured of convergence and the asymptotic cessation of updating, properties lacking in the formulation of [7], [8]. Further, in [7], [8] the condition which must be checked at each instant, to see if an update is required, entails greater computational complexity than does its counterpart in this paper. Finally, as simulations show, the use of a time-varying information-dependent forgetting factor equips the algorithm of this paper with an ability to track slow time variations in the unknown coefficients. The use of an information-dependent forgetting factor has also been made in a different context in [9]. A comparison of the strategy of [9] with the one employed here will be made after our algorithm is presented.

Several previous treatments of the bounded noise case appear in the literature [2], [10]–[13]. In some of these, e.g., [2], [13], the strategy has been to introduce a dead zone which causes the updates to be stopped when the prediction error becomes smaller than twice the assumed noise

Manuscript received April 29, 1985; revised November 8, 1985. This work was supported in part by the National Science Foundation under Grant ECS-8505218. This work was partially presented at the 24th Conference on Decision and Control, Fort Lauderdale, FL, December 11–13, 1985.

S. Dasgupta was with the University of Notre Dame, Notre Dame, IN. He is now with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242, USA.

Y. F. Huang is with the Department of Electrical and Computer Engineering, University of Notre Dame, Notre Dame, IN 46556, USA.

IEEE Log Number 8611422.

0018-9448/87/0500-0383$01.00

bound γ. The disadvantage here is that when γ is overestimated, the prediction error, in general, has limiting values no smaller than twice the assumed bound. For our algorithm, simulations show that even with up to 20 percent overestimation of γ, the prediction error approaches values smaller than the actual bound on the noise. In [10]–[12] other strategies are proposed in the adaptive control context to restrict the magnitude of the parameter estimates so as to prevent the information vector from becoming unbounded. In many of these, pointwise convergence of parameter estimates is not achieved, while in the others the same difficulty as in [2], [13] is present.

Section II of this paper is devoted to presenting the algorithm; the convergence problems are addressed in Section III. A key requirement for the convergence of any recursive estimator is that the inputs be sufficiently uncorrelated or persistently exciting so as to make the coefficients in (1.1) uniquely identifiable. Such a requirement is present here as well, and Section IV describes conditions for meeting it. Section V presents simulation results and Section VI makes concluding remarks. The appendices contain most of the proofs.

II. The Algorithm

Consider the estimation problem of (1.1), re-expressed as

$$y_k = \theta^{*T}x_k + v_k \tag{2.1}$$

where $\theta^{*T} \triangleq [a_1,\cdots,a_n,b_0,b_1,\cdots,b_m]$ and $x_k^T \triangleq [y_{k-1},\cdots,y_{k-n},u_k,\cdots,u_{k-m}]$. It is worth noting that the analysis in the sequel, except for that in Section IV, will apply to any system satisfying (2.1), i.e., any x_k, and not just to ARX processes. It is assumed that for each k, v_k is bounded in magnitude by γ, i.e.,

$$v_k^2 \le \gamma^2, \text{ for all } k. \tag{2.2}$$

Equations (2.1) and (2.2) together yield

$$\left(y_k - \theta^{*T}x_k\right)^2 \le \gamma^2. \tag{2.3}$$

Let S_k be a subset of R^{n+m+1} defined by

$$S_k = \left\{\theta\colon \left(y_k - \theta^T x_k\right)^2 \le \gamma^2, \theta \in R^{n+m+1}\right\}. \tag{2.4}$$

From a geometrical point of view, S_k is a convex polytope [14]. Thus with each measured value of (y_k, x_k), (2.1) and (2.2) together yield a convex polytope in the parameter space.

The fundamental concept of our approach is summarized in the following. Each S_k can be regarded as a degenerate ellipsoid in R^{n+m+1} [7], [8]. At any instant k, consider the intersection of the sequence of polytopes $S_1,\cdots,S_k$. It must contain the modeled parameter θ^* and so must any ellipsoid which bounds it. The recursive algorithm thus starts with a sufficiently large ellipsoid which covers all possible values of θ^*. After (y_1, x_1) is acquired, it finds an ellipsoid which bounds the intersection of the initial ellipsoid and S_1, and which is in a sense "optimal." Such an ellipsoid is denoted by E_1. By the same token, one can then obtain a sequence of optimal bounding ellipsoids (OBE) $\{E_k\}$. The estimate for θ^* at the kth instant is then defined to be the center of E_k.

Suppose that E_{k-1}, at any instant $k-1$, is given by

$$E_{k-1} = \left\{\theta\colon (\theta-\theta_{k-1})^T P_{k-1}^{-1}(\theta-\theta_{k-1}) \le \sigma_{k-1}^2\right\} \tag{2.5}$$

for some positive definite matrix P_{k-1} and a nonzero scalar σ_{k-1}. Then given (y_k, x_k), an ellipsoid that bounds $E_{k-1} \cap S_k$ is given by

$$\Big\{\theta\colon (1-\lambda_k)(\theta-\theta_{k-1})^T P_{k-1}^{-1}(\theta-\theta_{k-1}) + \lambda_k\left(y_k-\theta^T x_k\right)^2 \le (1-\lambda_k)\sigma_{k-1}^2 + \lambda_k\gamma^2\Big\} \tag{2.6}$$

for any $0 \le \lambda_k < 1$. As Theorem 2.1 below shows, there exist P_k and σ_k such that (2.6) can be re-expressed as

$$\left\{\theta\colon (\theta-\theta_k)^T P_k^{-1}(\theta-\theta_k) \le \sigma_k^2\right\} \tag{2.7}$$

where the nonsingularity of P_k will be a subject of later elaboration. In the sequel, x_k and y_k shall be assumed to be bounded.

Theorem 2.1: Consider the inequality

$$(1-\lambda_k)(\theta-\theta_{k-1})^T P_{k-1}^{-1}(\theta-\theta_{k-1}) + \lambda_k\left(y_k-\theta^T x_k\right)^2 \le (1-\lambda_k)\sigma_{k-1}^2 + \lambda_k\gamma^2 \tag{2.8}$$

where P_{k-1} is an $N \times N$ positive definite symmetric matrix, x_k, θ, and θ_{k-1} are N dimensional vectors, and y_k, σ_{k-1}, γ, and λ_k are scalars with $0 \le \lambda_k < 1$. Then with

$$P_k^{-1} = (1-\lambda_k)P_{k-1}^{-1} + \lambda_k x_k x_k^T \tag{2.9a}$$

$$\theta_k = \theta_{k-1} + \lambda_k P_k x_k \delta_k \tag{2.9b}$$

$$\delta_k = y_k - x_k^T\theta_{k-1} \tag{2.9c}$$

$$\sigma_k^2 = (1-\lambda_k)\sigma_{k-1}^2 + \lambda_k\gamma^2 - \frac{\lambda_k(1-\lambda_k)\delta_k^2}{1-\lambda_k+\lambda_k G_k} \tag{2.9d}$$

$$G_k = x_k^T P_{k-1} x_k, \tag{2.9e}$$

(2.8) is equivalent to

$$(\theta-\theta_k)^T P_k^{-1}(\theta-\theta_k) \le \sigma_k^2. \tag{2.10}$$

Proof: For $0 \le \lambda_k < 1$, P_k must be positive definite symmetric as well. Thus from (2.9a) and the matrix inversion Lemma

$$P_k = \frac{1}{1-\lambda_k}\left[P_{k-1} - \frac{\lambda_k P_{k-1}x_k x_k^T P_{k-1}}{1-\lambda_k+\lambda_k G_k}\right] \tag{2.11}$$

whence

$$\begin{aligned} P_k\left[(1-\lambda_k)P_{k-1}^{-1}\theta_{k-1} + \lambda_k x_k y_k\right] &= \frac{1}{1-\lambda_k}\left[P_{k-1} - \frac{\lambda_k P_{k-1}x_k x_k^T P_{k-1}}{1-\lambda_k+\lambda_k G_k}\right] \cdot\left[(1-\lambda_k)P_{k-1}^{-1}\theta_{k-1} + \lambda_k x_k y_k\right] \\ &= \theta_{k-1} + \frac{\lambda_k P_{k-1}x_k\delta_k}{1-\lambda_k+\lambda_k G_k} \end{aligned} \tag{2.12}$$

where the last step follows by multiplying the terms in the previous equation and (2.9c). Moreover, by (2.9b) and (2.11)

$$\theta_k = \theta_{k-1} + \frac{\lambda_k}{1-\lambda_k}\left[P_{k-1} - \frac{\lambda_k P_{k-1}x_k x_k^T P_{k-1}}{1-\lambda_k+\lambda_k G_k}\right]x_k\delta_k$$

$$= \theta_{k-1} + \frac{\lambda_k P_{k-1}x_k\delta_k}{1-\lambda_k+\lambda_k G_k} \quad (2.13)$$

$$= P_k\left[(1-\lambda_k)P_{k-1}^{-1}\theta_{k-1} + \lambda_k x_k y_k\right], \quad (2.13a)$$

the last step arising from (2.12). Consider next the left-hand side of (2.8) which equals

$$(1-\lambda_k)\theta^T P_{k-1}^{-1}\theta + \lambda_k(\theta^T x_k)^2 - 2\theta^T \cdot\left[(1-\lambda_k)P_{k-1}^{-1}\theta_{k-1} + \lambda_k x_k y_k\right] + (1-\lambda_k)\theta_{k-1}^T P_{k-1}^{-1}\theta_{k-1} + \lambda_k y_k^2$$
$$= (\theta-\theta_k)^T P_k^{-1}(\theta-\theta_k) - \theta_k^T P_k^{-1}\theta_k + (1-\lambda_k)\theta_{k-1}^T P_{k-1}^{-1}\theta_{k-1} + \lambda_k y_k^2$$

which follows from (2.13a). Thus (2.8) becomes

$$(\theta-\theta_k)^T P_k^{-1}(\theta-\theta_k) \le (1-\lambda_k)\sigma_{k-1}^2 + \lambda_k\gamma^2 - \left[\lambda_k y_k^2 - \theta_k^T P_k^{-1}\theta_k + (1-\lambda_k)\theta_{k-1}^T P_{k-1}^{-1}\theta_{k-1}\right].$$

After some routine algebra, the result follows.

We have thus established that (2.7), with the quantities of interest defined in (2.9), is a bounding ellipsoid. There are such ellipsoids corresponding to every value of λ_k. We choose the OBE to be the one for which σ_k^2 in (2.7) is the smallest, since σ_k^2 is a bound on the estimation error. Also, from an analytical viewpoint, σ_k^2 is a natural bound on the Lyapunov function to be used in Section III. Thus minimizing σ_k^2 with respect to λ_k will facilitate convergence. More interestingly, this choice leads to an information evaluation criterion which is computationally easier than its counterparts in [7], [8]. Note that the notion of λ_k in (2.6), which introduces a forgetting factor $(1-\lambda_k)$, is also different from that in [7], [8]. The forgetting factor also aids in the convergence analysis. Let the optimum value of λ_k be denoted by λ_k^*, defined as follows.

Definition 2.1: The parameter λ_k^* is such that

1) $\lambda_k^* \in [0, \alpha]$ for some $\alpha < 1$.
2) $\sigma_k^2(\lambda_k^*) \le \sigma_k^2(\lambda_k)$ for all $\lambda_k \in [0, \alpha]$.

Here α is a design parameter smaller than one since $\lambda_k = 1$ implies that P_k is singular (2.9a). From (2.9d), $\sigma_k^2(0) = \sigma_{k-1}^2$ whence $\sigma_k^2(\lambda_k^*) \le \sigma_{k-1}^2$. Thus if $d\sigma_k^2/d\lambda_k \ge 0$ for every positive λ_k, then one concludes that the use of information available at the kth instant does not improve σ_k^2 and hence at that instant $\lambda_k^* = 0$ and no update is made. Lemma 2.1, proved in Appendix I, gives explicit expressions for calculating λ_k^*.

Lemma 2.1: With P_k positive semidefinite and

$$\sigma_k^2 = (1-\lambda_k)\sigma_{k-1}^2 + \lambda_k\gamma^2 - \frac{\lambda_k(1-\lambda_k)\delta_k^2}{1-\lambda_k+\lambda_k G_k} \quad (2.9d)$$

consider λ_k^* of Definition 2.1 and define $\beta_k \triangleq (\gamma^2 - \sigma_{k-1}^2)/\delta_k^2$. Then the following is true:

1) if $\gamma^2 \ge \sigma_{k-1}^2 + \delta_k^2$, then $\lambda_k^* = 0$ (2.14)
2) otherwise,

$$\lambda_k^* = \min(\alpha, \nu_k) \quad (2.14a)$$

where

$$\nu_k = \begin{cases} \alpha, & \text{if } \delta_k^2 = 0 \quad (2.15a)\\ \dfrac{1-\beta_k}{2}, & \text{if } G_k = 1 \quad (2.15b)\\ \dfrac{1}{1-G_k}\left[1 - \sqrt{\dfrac{G_k}{1+\beta_k(G_k-1)}}\right], & \text{if } \beta_k(G_k-1)+1 > 0 \quad (2.15c)\\ \alpha, & \text{if } \beta_k(G_k-1)+1 \le 0. \quad (2.15d)\end{cases}$$

One can see that $\gamma^2 < \sigma_{k-1}^2 + \delta_k^2$ implies $\lambda_k^* > 0$; furthermore, if $\gamma^2 < \sigma_{k-1}^2$, then

$$\lambda_k^* \ge \min\left\{\alpha, \frac{1}{1+\sqrt{G_k}}\right\}. \quad (2.16)$$

Remark 2.1: A detailed study of computational aspects is postponed until later. It suffices to note for the moment that $\lambda_k^* = 0$ if (2.14) is satisfied. Thus to check if an update is required, only the prediction error δ_k need be found. If (2.14) is found to hold then the calculations in (2.15) are not required.

Remark 2.2: If $\delta_k^2 = 0$ and (2.14) does not hold, $\beta_k = -\infty$. Thus $\beta_k(G_k - 1) + 1 > 0$ implies $G_k < 1$, whence by (2.15c) and (2.14a) $\lambda_k^* = \alpha$, since $\nu_k \ge 1$. On the other hand, $\beta_k(G_k - 1) + 1 \le 0$ implies $G_k > 1$, whence $\nu_k = \alpha$. Thus (2.15a) is a special case of (2.15c) and (2.15d).

Having established a recursion for the OBE's $\{E_k\}$, we now state what E_0 is. It is given by

$$E_0 = \{\theta: \|\theta\|^2 \le 1/\epsilon\} \quad (2.17)$$

where $1/\epsilon$ is a suitably large number and $\|\theta\|^2 \triangleq \theta^T\theta$. In general, ϵ can be as small as one pleases and should be such that

$$\|\theta^*\|^2 \le 1/\epsilon.$$

It is evident that according to (2.7) and (2.17)

$$P_0 = I \qquad \theta_0 = 0 \qquad \sigma_0^2 = 1/\epsilon. \tag{2.18}$$

As far as computation of θ_k is concerned, the following equation, rather than (2.9a), needs to be implemented:

$$P_k = \frac{1}{1-\lambda_k^*}\left[P_{k-1} - \lambda_k^* P_{k-1}x_k x_k^T P_{k-1}/ (1-\lambda_k^* + \lambda_k^* G_k)\right]. \tag{2.19}$$

Of course, in the other equations of (2.9) λ_k^* from Lemma 2.1 should replace λ_k.

Computationally, the greatest complexity lies in (2.19) and finding $P_{k-1}x_k$ and G_k. If (2.14) holds, then none of these need be computed. The relevant condition to be checked only involves finding δ_k, the prediction error. If (2.14) is false, then (2.19) requires G_k in any case. Thus finding λ_k^* involves no additional complexities. Observe also that θ_k, the center of E_k, is the parameter estimate and that $1-\lambda_k^*$ can be viewed as an information-dependent forgetting factor which may vary with time. It is worth noting that, conventionally, λ_k^* rather than $1-\lambda_k^*$ is the notation for forgetting factors. However, λ_k^* plays a dual role here. While $1-\lambda_k^*$ acts as a forgetting factor, λ_k^* acts as an update gain in (2.9b) and (2.19).

As noted in the Introduction, Fortesque *et al.* have used variable forgetting factors in [9]. While our selection of these factors is based on the need to minimize the extent of the feasible parameter estimate set, the approach in [9] arises from a different consideration. There, an information measure related to the cumulative sum of squared prediction errors is proposed, and the forgetting factor is selected to maintain this measure at a constant value. Thus, if the prediction error, at any stage, becomes high, less reliance is placed on prior information.

A key difference between the two approaches is the absence of λ_k^* as a gain factor in [9]. Thus, whereas the algorithm of this paper will cease updating when the forgetting factor $(1-\lambda_k^*)$ becomes one, that is not the case in [9]. It is this feature which equips our algorithm with the ability to handle bounded disturbances.

III. Convergence Issues

In this section convergence properties associated with (2.9), (2.14), (2.15), and (2.19) are examined. Clearly, the infinite memory associated with (2.19) guarantees that P_k is always positive definite. However, some of the results presented here require that α_1, α_2 exist such that for all k,

$$0 < \alpha_1 I \le P_k \le \alpha_2 I < \infty. \tag{3.1}$$

In Section IV we shall show that (3.1) is satisfied if there exist N, α_3, and α_4 such that for all k,

$$0 < \alpha_3 I \le \sum_{i=k}^{k+N} x_i x_i^T \le \alpha_4 I < \infty. \tag{3.2}$$

With (3.1) holding, it is shown first that the parameter error converges exponentially to a region in which

$$\|\theta_k - \theta^*\|^2 \le \gamma^2/\alpha_5 \tag{3.3}$$

where

$$0 < \alpha_5 I \le P_k^{-1} \le \alpha_6 I. \tag{3.4}$$

Moreover,

$$\lim_{k\to\infty} \|\theta_{k+1} - \theta_k\| = 0 \tag{3.5}$$

$$\lim_{k\to\infty} \delta_k^2 \in [0, \gamma^2]. \tag{3.6}$$

With a further restriction on x_i, we show that

$$\lim_{k\to\infty} \lambda_k^* = 0. \tag{3.7}$$

Note throughout this section, expressions like (3.6) should not be taken to mean that $\lim_{k\to\infty}\delta_k^2$ exists but rather that δ_k^2 becomes asymptotically less than or equal to γ^2. These results require first the following lemma and the assumption that G_k and x_k are bounded.

Lemma 3.1: Consider (2.9), (2.14), (2.15), and (2.19). Then,

$$\lim_{k\to\infty} \sigma_k^2 \in [0, \gamma^2] \tag{3.8}$$

where the rate of convergence is exponential.

Proof: From (2.9d)

$$(\sigma_k^2 - \gamma^2) - (\sigma_{k-1}^2 - \gamma^2) \le -\lambda_k^*(\sigma_{k-1}^2 - \gamma^2). \tag{3.9}$$

Moreover, from Lemma 2.1, $\sigma_{k-1}^2 > \gamma^2$ implies $\lambda_k^* \ge \min\{\alpha, 1/(1+\sqrt{G_k})\}$. Thus the result follows.

We now prove (3.3) using Lyapunov theory.

Theorem 3.1: Consider (2.1), (2.9), (2.14), (2.15), and (2.19). Suppose $\theta^* \in E_0$. Then $\theta^* \in E_k$ for all subsequent k. Moreover, if (3.4) holds, then θ_k converges exponentially to a region where (3.3) holds.

Proof: Consider the Lyapunov function

$$V_k = \Delta\theta_k^T P_k^{-1} \Delta\theta_k \tag{3.10}$$

with $\Delta\theta_k \triangleq \theta^* - \theta_k$. Using analysis similar to that in [15], we find that

$$V_k = (1-\lambda_k^*)V_{k-1} + \lambda_k^* v_k^2 - \frac{\lambda_k^*(1-\lambda_k^*)\delta_k^2}{1-\lambda_k^* + \lambda_k^* G_k}. \tag{3.11}$$

Thus

$$V_k - V_{k-1} \le -\lambda_k^*\left[V_{k-1} - \gamma^2 + \frac{(1-\lambda_k^*)\delta_k^2}{1-\lambda_k^* + \lambda_k^* G_k}\right]. \tag{3.12}$$

Also

$$V_k \le (1-\lambda_k^*)V_{k-1} + \left[\sigma_k^2 - (1-\lambda_k^*)\sigma_{k-1}^2\right],$$

i.e.,

$$V_k - \sigma_k^2 \le (1-\lambda_k^*)\left[V_{k-1} - \sigma_{k-1}^2\right]. \tag{3.13}$$

Note that

$$V_{k-1} \le \sigma_{k-1}^2 \quad \text{iff} \quad \theta^* \in E_{k-1}$$

which implies

$$V_k \le \sigma_k^2 \quad \text{whence} \quad \theta^* \in E_k.$$

Thus by (3.13), it can be seen that if $\theta^* \in E_0$, then $\theta^* \in E_k$ for all k. Moreover, (3.12) implies that

$$\left(V_k - \gamma^2\right) - \left(V_{k-1} - \gamma^2\right) \le -\lambda_k^*\left[V_{k-1} - \gamma^2\right]. \quad (3.14)$$

Thus $V_k - \gamma^2$ converges exponentially to a value smaller than 0 if λ_k^* is uniformly nonzero while $V_k > \gamma^2$. Now $V_k > \gamma^2$ implies $\sigma_k^2 > \gamma^2$, whence by Lemma 2.1

$$\lambda_k^* \ge \min\left\{\alpha, \frac{1}{1 + \sqrt{G_k}}\right\}.$$

Since G_k is bounded, the result follows from (3.4).

We shall comment on the case when $\theta^* \notin E_0$ later. The theorem below deals with the convergence of a range of other parameters. In particular, it shows that in the limit, parameter estimate variation decays to zero and that

$$\sigma_{k-1}^2 + \delta_k^2 \le \gamma^2.$$

Theorem 3.2: For bounded G_k (i.e., (3.1) holding) with (2.1), (2.9), (2.14), (2.15), and (2.19), the following holds with $\theta^* \in E_0$:

$$\lim_{k\to\infty} \|\theta_{k+1} - \theta_k\| = 0, \quad (3.5)$$

$$\lim_{k\to\infty} \left(\sigma_{k-1}^2 + \delta_k^2\right) \in \left[0, \gamma^2\right], \quad (3.15)$$

and

$$\lim_{k\to\infty} \delta_k^2 \in \left[0, \gamma^2\right]. \quad (3.6)$$

Proof: See Appendix II.

Remark 3.1: Equation (3.15) does not necessarily imply that λ_k^* goes to zero, as λ_k^* goes to zero iff β_k goes to one. If δ_k^2 approaches zero faster than σ_{k-1}^2 approaches γ^2, β_k may not approach one. Thus for λ_k^* to vanish we need some further conditions such as those used in Theorem 3.3.

Remark 3.2: The quantities σ_{k-1}^2 and δ_k^2, respectively, measure the parameter and prediction errors. Condition (3.15) states that their sum, in the limit, falls below γ^2. However, if $\gamma_1^2 < \gamma^2$ is such that $v_k^2 \le \gamma_1^2$, V_k would now be governed by (see (3.14))

$$V_k - V_{k-1} \le -\lambda_k^*\left[V_{k-1} - \gamma_1^2 + \frac{(1-\lambda_k^*)\delta_k^2}{1 - \lambda_k^* + \lambda_k^* G_k}\right]. \quad (3.16)$$

Thus unless λ_k^* vanishes before $\sigma_{k-1}^2 + \delta_k^2 \le \gamma_1^2$ occurs, the sum of V_{k-1} and δ_k^2 could still become smaller than γ_1^2. Now λ_k^* goes to zero only when $\sigma_{k-1}^2 + \delta_k^2 \le \gamma^2$. Thus provided V_0 is sufficiently smaller than σ_0^2, one can see from (3.13) that $V_{k-1} + \delta_k^2$ may well be close to γ_1^2 before updating ceases.

Remark 3.3: The approach of this paper is predicated on $\theta^* \in E_0$. If $\theta^* \notin E_0$, then the notion of bounding ellipsoids no longer applies. Yet θ_k remains a valid estimate of θ^*, V_k still decreases as long as $\lambda_k^* \ne 0$, and the stopping condition is still (2.14). However, too great a violation of $\theta^* \in E_0$ could cause σ_k^2 to become negative. Thus at the point of convergence δ_k^2 could exceed γ^2.

With slow time variations in θ^* the algorithm should still perform well, for as long as $\theta^*(k) \in E_k$, small changes in θ^* beyond initial convergence would result in large δ_k, violation of (2.14), and the resumption of tracking. Large time variations could cause difficulties similar to those explained for large violations of $\theta^* \in E_0$. A modification to the algorithm for handling large time variations will be the subject of a forthcoming paper.

Remark 3.4: A conceivable alternative to the λ_k^* selection strategy outlined above is to set $\lambda_k^* = 0$ as soon as δ_k^2 becomes smaller than γ^2. However, the strategy selected here often leads to even smaller eventual δ_k^2 and σ_k^2.

We now discuss the convergence of λ_k^*, which together with the results of Theorem 3.2 implies the convergence of the OBE E_k.

Theorem 3.3: Consider the system (2.1) and the (2.9), (2.14), (2.15), and (2.19). Suppose there exist α_7, α_8, and $N_1 > 0$ such that for all k,

$$\alpha_7 I \le \sum_{i=k}^{k+N_1} \begin{bmatrix} x_i \\ v_i \end{bmatrix} \left[x_i^T v_i\right] \le \alpha_8 I. \quad (3.17)$$

Then

$$\lim_{k\to\infty} \lambda_k^* = 0.$$

Proof: See Appendix II.

Remark 3.5: Condition (3.17) essentially demands that u_k and v_k be sufficiently uncorrelated with each other. The next section clarifies this issue further.

IV. Persistence of Excitation

In Section III we stated that the satisfaction of (3.1) and (3.17) is required for certain convergence results. As is the case in many other identifiers [4], [5], [15], this translates to a persistence of excitation condition on the input and a lack of correlation condition on u_k and v_k. In this section, these results are formalized. We show first that (3.1) is implied by (3.2) and then go on to suggest ways in which (3.2) can be satisfied.

Theorem 4.1: With P_k defined in (2.19) and x_k such that, for some positive α_3, α_4, and N, and all k

$$0 < \alpha_3 I \le \sum_{i=k}^{k+N} x_i x_i^T \le \alpha_4 I < \infty, \quad (3.2)$$

there exist $\alpha_1, \alpha_2 > 0$ such that

$$0 < \alpha_1 I \le P_k \le \alpha_2 I < \infty \quad (3.1)$$

as long as $0 \le \lambda_i^* < \alpha < 1$.

Proof: See Appendix III.

Remark 4.1: The result can be interpreted as follows. Looking at measurements over a finite interval is equivalent to looking at measurements over an arbitrarily long interval with infinite discounting factor on all but a finite subinterval.

The following result shows two conditions under which (3.2) can be satisfied. It is proved in Appendix III.

Theorem 4.2: Consider the system (1.1), (2.1). Assume that $z^n + \sum_{i=0}^{n-1} a_{n-i} z^i$ and $\sum_{j=0}^{m} b_{m-j} z^j$ are coprime and (1.1) is stable. Define

$$W_0(k) = [u_k, u_{k-1}, \cdots, u_{k-n-m}]^T \tag{4.1}$$

and

$$W_1(k) = \left[W_0^T(k), v_{k-1}, \cdots, v_{k-n}\right]^T. \tag{4.2}$$

Then if there exist $\beta_1, \beta_2 > 0$ such that for all k

$$\beta_1 I \le \sum_{i=k+n}^{k+N} W_1(i) W_1^T(i) \le \beta_2 I, \tag{4.3}$$

then (3.2) is satisfied. Alternatively, if there exist $\beta_3, \beta_4 > 0$ such that

$$\beta_3 I + n\gamma^2 I \le \sum_{i=k+n}^{k+N} W_0(i) W_0^T(i) \le \beta_4 I \tag{4.4}$$

for all k, then (3.2) is satisfied.

Remark 4.2: Equation (4.3) states that the inputs u_k should be sufficiently rich in frequency and must be uncorrelated with the noise. Equation (4.4), on the other hand, states that the input should be rich enough to overcome the effect of the noise. In practice the noise sequence is usually uncorrelated with the input sequence, thus (4.3) is easier to satisfy.

Remark 4.3: The above theorem states conditions under which all of the convergence results in the previous section, except $\lim_{k\to\infty} \lambda_k^* = 0$, are satisfied. Of course, even if λ_k^* does not go to zero, $\|\theta_{k+1} - \theta_k\|$ still may vanish in the limit.

Below, we show how condition (3.17), sufficient for $\lim_{k\to\infty} \lambda_k^* = 0$, can be satisfied. The proof follows in the same vein as that of the previous theorem and is omitted.

Theorem 4.3: Under the assumptions of Theorem 4.2 define

$$W_2(k) = \left[W_0^T(k), v_k, \cdots, v_{k-n}\right]^T.$$

Then (3.17) is satisfied if there exist $\beta_5, \beta_6 > 0$ such that

$$\beta_5 I \le \sum_{i=k+n}^{k+N} W_2(k) W_2^T(k) \le \beta_6 I \tag{4.5}$$

for all k.

Remark 4.4: Observe that (4.5) implies (4.3). In fact, (4.5) and (4.3) are almost the same, and it is highly unlikely that (4.3) is satisfied yet (4.5) is not.

V. Simulations

Consider the system

$$y_k = 0.3y_{k-1} - 0.28y_{k-2} + 0.46y_{k-3} - 0.1y_{k-4} + v_k$$

where v_k is a zero mean uniformly distributed white noise sequence, bounded in magnitude by one. Suppose that each of the four actual parameters undergoes a ten-percent step change in magnitude at every 200 sampling points. Then Figs. 1–4, respectively, show the trajectories of 1) actual parameters, 2) the RLS estimates, and 3) the estimates generated by the algorithm of this paper with $\alpha = 0.9$. The superior tracking ability of this algorithm over that of RLS is evident. Moreover, in the 2000-sample

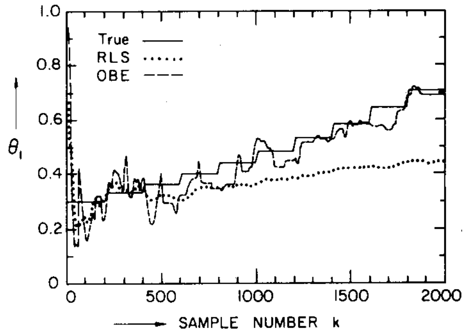

Fig. 1. Tracking of parameter θ_1 (starting value = 0.3).

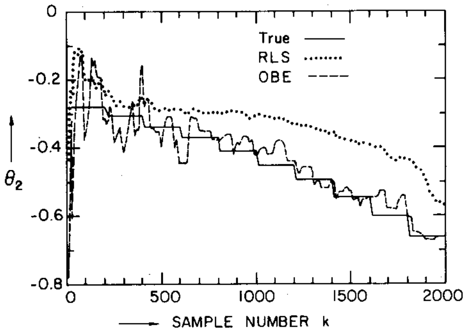

Fig. 2. Tracking of parameter θ_2 (starting value = −0.28).

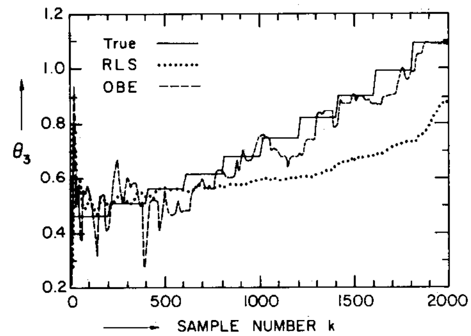

Fig. 3. Tracking of parameter θ_3 (starting value = 0.46).

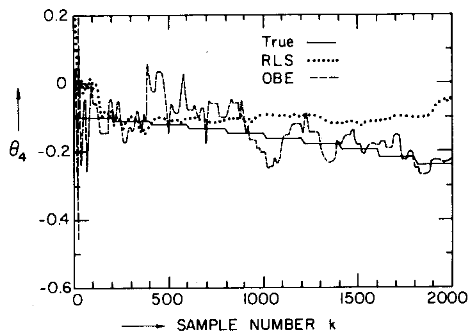

Fig. 4. Tracking of parameter θ_4 (starting value = −0.1).

interval the number of updates is only 209, and the final prediction error is $\delta_k^2 = 0.6 < 1$.

In all the examples we tried, with or without time variation, the number of updates did not exceed 15 percent of the number of samples, representing a significant computational saving. Moreover, even when the noise bound γ was over estimated by 20 percent of its actual value, the resulting prediction errors were smaller than the *actual* bound. The implication here is that, should the modeler be uncertain about the value of γ, a conservative estimate of γ could yet result in $|\delta_k|$ less than the *actual* γ.

From the example given, it appears that the initial behavior of the OBE algorithm is inferior to RLS when time variations are absent. This is not surprising partly due to the smoother transients of RLS. The OBE does not update as often as RLS, and when updates are made they turn out to be more substantial. Also, without time variations the need for having weighted information in the initial stages is less compelling, as redundancies in information are less frequent. At the same time, other advantages of the OBE, particularly the computational saving due to infrequent updates, amply justify its use.

VI. Conclusion

A reformulation of RLS estimation based on a bounded noise assumption has been shown to yield an algorithm whose updates are information-dependent. A Lyapunov approach has been used to prove the asymptotic convergence of the estimates. There are several key features of the algorithm. 1) By eliminating redundant updates of the parameter estimates, computational complexity can be expected to improve. 2) In the face of bounded output disturbances, asymptotic cessation of updating is still ensured once the sum of the prediction error and a certain bound on the estimation error becomes smaller than the disturbance bound. 3) The convergence of the estimation error to a region determined by the degree of excitation and the measurement disturbance bound is exponential. This is a property which strengthens the robustness characteristics of the algorithm. 4) Finally, the algorithm can cope with modest departures from idealistic assumptions. Thus even if the system has slow time variation or the disturbance sequence does not strictly obey the imposed magnitude bound, the algorithm can still be expected to perform adequately.

Acknowledgment

The authors thank Mr. Ashok K. Rao for his help with the simulations and the anonymous reviewers for their comments.

Appendix I
Proof of Lemma 2.1

By the definition of λ_k^* and (2.9d) we have that

$$\sigma_k^2(\lambda_k^*) \le \sigma_k^2(0) = \sigma_{k-1}^2. \tag{A.1}$$

Thus if $d\sigma_k^2/d\lambda_k \ge 0$ everywhere on $\lambda_k \in [0, \alpha]$, then $\lambda_k^* = 0$. From (2.9d)

$$\frac{d\sigma_k^2}{d\lambda_k} = \gamma^2 - \sigma_{k-1}^2 - \delta_k^2 \frac{(1-\lambda_k)^2 - \lambda_k^2 G_k}{(1-\lambda_k+\lambda_k G_k)^2} \tag{A.2}$$

and

$$\frac{d^2\sigma_k^2}{d\lambda_k^2} = \frac{2\delta_k^2 G_k}{(1-\lambda_k+\lambda_k G_k)^3}. \tag{A.3}$$

If $\delta_k^2 G_k \neq 0$, the positive definiteness of P_{k-1} implies that $d^2\sigma_k^2/d\lambda_k^2$ has the same sign as $(1-\lambda_k+\lambda_k G_k)$, which for any $\lambda_k \in [0,1)$ is positive. Let us prove Lemma 2.1 case by case.

Case I: $\delta_k^2 = 0$. From (A.2), $d\sigma_k^2/d\lambda_k < 0$ if and only if $\gamma^2 < \sigma_{k-1}^2$. Thus

$$\lambda_k^* = \begin{cases} 0, & \text{if } \gamma^2 \ge \sigma_{k-1}^2 \\ \alpha, & \text{if } \gamma^2 < \sigma_{k-1}^2 \end{cases}.$$

Note that in this case both (2.15a) and (2.16) are satisfied. Now, for subsequent cases, it is assumed that $\delta_k \neq 0$.

Case II: $G_k = 1$,

$$\frac{d\sigma_k^2}{d\lambda_k} = \delta_k^2[\beta_k - 1 + 2\lambda_k] \tag{A.4}$$

with β_k defined in the statement of the Lemma. Also, $d^2\sigma_k^2/d\lambda_k^2 \ge 0$ for any $\lambda_k \ge 0$. Thus σ_k^2 is minimized when

$$\lambda_k = \frac{1-\beta_k}{2}, \qquad \beta_k < 1.$$

If $\beta_k \ge 1$, $(1-\beta_k)/2$ is nonpositive and $\lambda_k^* = 0$. Note that $\beta_k \ge 1$ is equivalent to $\gamma^2 \ge \sigma_{k-1}^2 + \delta_k^2$, (2.14), provided that $\delta_k \neq 0$. Thus both (2.15b) and (2.16) are satisfied.

Case III: $\beta_k(G_k - 1) + 1 > 0$. By (A.2),

$$\frac{d\sigma_k^2}{d\lambda_k} = 0 \quad \text{iff } \lambda_k = \frac{1}{1-G_k}\left\{1 \pm \sqrt{\frac{G_k}{1+\beta_k(G_k-1)}}\right\}. \tag{A.5}$$

Since $1 + \beta_k(G_k - 1) > 0$, λ_k is real. It is easy to show that only

$$\lambda_k = \frac{1}{1-G_k}\left\{1 - \sqrt{\frac{G_k}{1+\beta_k(G_k-1)}}\right\} \tag{A.6}$$

corresponds to a minimum. Moreover, in (A.6)

$$\lambda_k > 0 \Leftrightarrow \beta_k < 1 \Leftrightarrow \gamma^2 < \sigma_{k-1}^2 + \delta_k^2$$

and

$$\beta_k \le 0 \Rightarrow \lambda_k \ge \frac{1-\sqrt{G_k}}{1-G_k} = \frac{1}{1+\sqrt{G_k}}.$$

Further, if λ_k in (A.6) is greater than α, it is easy to see that

$$\frac{d\sigma_k^2}{d\lambda_k} < 0$$

for all $\lambda_k \in [0, \alpha]$. Thus λ_k^* is as given by (2.14a) and (2.15c). In addition, (2.16) clearly is satisfied for $G_k > 0$. If $G_k = 0$, then $\lambda_k = 1$ and $\beta_k < 1$. Thus (2.14a), (2.15c), and (2.16) hold.

Case IV: $\beta_k(G_k - 1) + 1 \le 0$. Suppose the equality holds. Then

$$\frac{d\sigma_k^2}{d\lambda_k} = \delta_k^2\left[\frac{1}{1-G_k} - \frac{(1-\lambda_k)^2 - \lambda_k^2 G_k}{(1-\lambda_k+\lambda_k G_k)^2}\right] = \frac{G_k\delta_k^2}{(1-G_k)(1-\lambda_k+\lambda_k G_k)^2}.$$

With the fact that $0 \le G_k$ and $\beta_k = 1/(1-G_k)$ we have $\beta_k \ge 1$ if and only if $G_k < 1$ and $\beta_k < 0$ if and only if $G_k > 1$. Further, $d\sigma_k^2/d\lambda_k$ has the same sign as $(1 - G_k)$. Thus λ_k^* equals 0 if $\beta_k \ge 1$ and equals α if $\beta_k < 0$. Note that $\beta_k < 1$ is not possible for this case. If $\beta_k(G_k - 1) + 1 < 0$, then (A.5) is complex and $d\sigma_k^2/d\lambda_k$ has the same sign everywhere. Now,

$$\left.\frac{d\sigma_k^2}{d\lambda_k}\right|_{\lambda_k=0} = \delta_k^2[\beta_k - 1].$$

Thus $\lambda_k^* = \alpha$ if $\beta_k \le 1$ and $\lambda_k^* = 0$ otherwise. Hence (2.14a), (2.15d), and (2.16) are satisfied.

Appendix II
Proof of Theorems 3.2 and 3.3

Proof of Theorem 3.2

By Theorem 3.1

$$\theta^* \in E_0 \Rightarrow \theta^* \in E_k \Rightarrow \sigma_k^2 \ge 0 \qquad \forall k. \tag{B.1}$$

Also by (A.2) if $\lambda_k^* > 0$, then

$$\left.\frac{d\sigma_k^2}{d\lambda_k}\right|_{\lambda_k=\lambda_k^*} \le 0$$

$$\Leftrightarrow \gamma^2 - \sigma_{k-1}^2 - \frac{(1-\lambda_k^*)\delta_k^2}{1-\lambda_k^*+\lambda_k^* G_k} \le -\frac{\lambda_k^*\delta_k^2 G_k}{(1-\lambda_k^*+\lambda_k^* G_k)^2}.$$

Thus

$$\sigma_k^2 \le \sigma_{k-1}^2 - \frac{\lambda_k^{*2}\delta_k^2 G_k}{(1-\lambda_k^*+\lambda_k^* G_k)^2}. \tag{B.2}$$

Of course if, in the limit, $\lambda_k^* = 0$, then $\theta_{k+1} = \theta_k$ and by Lemma 2.1, both (3.5) and (3.15) are satisfied. Equations (B.1) and (B.2) imply

$$\lim_{k\to\infty} \lambda_k^{*2}\delta_k^2 G_k = 0. \tag{B.3}$$

To show (3.5), we need to show that

$$\lim_{k\to\infty} \lambda_k^{*2}\delta_k^2 = 0. \tag{B.4}$$

Now (B.3) implies that for all $\epsilon > 0$ there exists N, such that for all $k > N$,

$$\lambda_k^{*2}\delta_k^2 G_k < \epsilon. \tag{B.5}$$

Suppose for some k, $\lambda_k^{*2}\delta_k^2 > a > 0$. Then

$$G_k < \epsilon/a. \tag{B.6}$$

So

$$\sigma_k^2 - \sigma_{k-1}^2 = \delta_k^2\lambda_k^*\left[\beta_k - \frac{1-\lambda_k^*}{1-\lambda_k^*+\lambda_k^* G_k}\right] = \delta_k^2\lambda_k^*\left[\beta_k - \frac{1}{1+(\lambda_k^*/1-\lambda_k^*)G_k}\right] \le \delta_k^2\lambda_k^*[\beta_k - 1 + 0(\epsilon)]. \tag{B.7}$$

Thus if (B.4) is violated and (B.1) holds,

$$\lim_{k\to\infty} \beta_k \in [1,\infty) \Rightarrow \lim_{k\to\infty} \sigma_{k-1}^2 + \delta_k^2 \in [0,\gamma^2] \Rightarrow \lim_{k\to\infty} \delta_k^2 \in [0,\gamma^2] \text{ whence } \lim_{k\to\infty} \lambda_k^* = 0,$$

which contradicts (B.4). On the other hand, if (B.4) holds, then (3.5) is automatically satisfied.

Further, (B.4) implies, for arbitrary $\epsilon > 0$, there exists N such that for any $k \ge N$

$$\lambda_k^*\delta_k^2 \le \epsilon^2. \tag{B.8}$$

Suppose (3.6) is not true. Then $\lim_{k\to\infty}\delta_k^2 \ne 0$. Suppose $\delta_k^2 > \gamma^2$. Then

$$\lambda_k^{*2} \le \epsilon^2/\gamma^2. \tag{B.9}$$

We shall show that

$$\beta_k \ge 1 - 0(\epsilon). \tag{B.10}$$

Consider the three cases of (2.15) applicable to this situation.

Case I: $G_k = 1$,

$$\lambda_k^* = \frac{1-\beta_k}{2} \le \epsilon/\gamma \Rightarrow \beta_k \ge 1 - 2\epsilon/\gamma.$$

Case II: $\beta_k(G_k - 1) + 1 \le 0$. If $\epsilon^2/\gamma^2 < \alpha$, then $\beta_k \ge 1$.

Case III: $\beta_k(G_k - 1) + 1 > 0$. For small enough ϵ,

$$\lambda_k^* = \frac{1}{1-G_k}\left[1 - \sqrt{\frac{G_k}{\beta_k(G_k-1)+1}}\right]$$

$$\Leftrightarrow \frac{G_k}{\beta_k(G_k-1)+1} = \{\lambda_k^*(G_k-1)+1\}^2$$

$$\Leftrightarrow \beta_k = \frac{1}{G_k-1}\left\{\frac{G_k}{[\lambda_k^*(G_k-1)+1]^2} - 1\right\}.$$

$$= \frac{1}{G_k-1}\left[\frac{G_k - 1 - \lambda_k^{*2}(G_k-1)^2 - 2\lambda_k^*(G_k-1)}{\{\lambda_k^*(G_k-1)+1\}^2}\right]$$

$$= \frac{1}{\{\lambda_k^*(G_k-1)+1\}^2} - \frac{\lambda_k^{*2}(G_k-1)}{\{\lambda_k^*(G_k-1)+1\}^2} - \frac{2\lambda_k^*}{\{\lambda_k^*(G_k-1)+1\}^2}$$

$$\ge 1 - 0(\epsilon).$$

Thus (B.10) holds. Hence

$$\gamma^2 \ge \sigma_{k-1}^2 + \delta_k^2 - 0(\epsilon)$$

and (3.6) and (3.15) are satisfied.

Proof of Theorem 3.3

From the proof Theorem 3.2 one can see that

$$\lim_{k\to\infty} \lambda_k^{*2}\delta_k^2 = 0 \tag{B.11}$$

and

$$\lim_{k\to\infty} \|\theta_k - \theta_{k-1}\| = 0. \tag{B.12}$$

Now

$$\delta_k = \Delta\theta_{k-1}^T x_k + v_k.$$

From (3.17) and (B.12), over any interval of length N_1, δ_k cannot be arbitrarily small. Thus at least one l_i exists in every interval of length N_1 such that for some a_2, $\delta_{l_i}^2 \geq a_2 > 0$. Now by Theorem 3.2, for all ϵ there exists N_2 such that for all $i \geq N_2$ and $k = l_i$,

$$\sigma_{k-1}^2 - \gamma^2 + \delta_k^2 \leq \epsilon$$

and so

$$\sigma_{k-1}^2 \leq \gamma^2 - a_2 + \epsilon$$

whence for small enough ϵ

$$\beta_k > 0.$$

Now σ_k^2 is nonincreasing. Thus for all $k \geq l_{N_2}$,

$$\beta_k = \frac{\gamma^2 - \sigma_{k-1}^2}{\delta_k^2} \geq \frac{a_2}{\delta_k^2}. \tag{B.13}$$

From (B.11) for any $\epsilon > 0$ there exists N_3 such that for all $k \geq N_3$,

$$\lambda_k^{*2} \delta_k^2 \leq \epsilon.$$

Thus either $\lambda_k^{*2} \leq 0(\epsilon)$ or $\delta_k^2 \leq 0(\epsilon)$. In the latter case, by (B.13), $\beta_k \geq a_2/0(\epsilon) > 1$ for sufficiently small ϵ, whence $\lambda_k^* = 0$. This completes the proof.

Appendix III
Proof of Theorems 4.1 and 4.2

Proof of Theorem 4.1

We first show that (3.4) holds, so (3.1) follows. The upper bound follows from the boundedness condition in (3.2), which implies for any unit vector η

$$\sum_{i=k}^{k+N} \left(\eta^T x_i\right)^2 \geq \alpha_3. \tag{C.0a}$$

From (2.19)

$$P_k^{-1} = \left\{\prod_{i=1}^{k} (1 - \lambda_{k-i})\right\} I + \sum_{j=1}^{k} \left(\prod_{i=j+1}^{k} (1 - \lambda_i)\right) \lambda_j x_j x_j^T.$$

Thus

$$J = \eta^T P_k^{-1} \eta = \prod_{i=1}^{k} (1 - \lambda_i) + \sum_{j=1}^{k} \left(\prod_{i=j+1}^{k} (1 - \lambda_i)\right) \lambda_j \left(x_j^T \eta\right)^2$$

where $0 \leq \lambda_i \leq \alpha < 1$. Consider the stationary points of J with respect to λ_i:

$$\frac{\partial J}{\partial \lambda_l} = -\prod_{i=1 i \neq l}^{k} (1 - \lambda_i) - \sum_{j=1}^{l-1} \prod_{i=j+1 i \neq l}^{k} (1 - \lambda_i) \lambda_j \left(x_j^T \eta\right)^2 + \left\{\prod_{i=l+1}^{k} (1 - \lambda_i)\right\} \left(x_l^T \eta\right)^2 = 0. \tag{C.0}$$

For $l = 1$

$$\frac{\partial J}{\partial \lambda_1} = -\prod_{i=2}^{k} (1 - \lambda_i) + \left\{\prod_{i=2}^{k} (1 - \lambda_i)\right\} \left(x_1^T \eta\right)^2 = 0.$$

This implies that either

$$\left(x_1^T \eta\right) = 1 \tag{C.1}$$

or

$$\prod_{i=2}^{k} (1 - \lambda_i) = 0. \tag{C.2}$$

Since $0 \leq \lambda_i \leq \alpha < 1$, (C.2) cannot hold and so (C.1) holds. For $l = 2$,

$$\frac{\partial J}{\partial \lambda_2} = -\prod_{i=1 i \neq 2}^{k} (1 - \lambda_i) - \sum_{j=1}^{1} \prod_{i=j+1 i \neq 2}^{k} (1 - \lambda_i) \lambda_j \left(x_j^T \eta\right)^2 + \left\{\prod_{i=3}^{k} (1 - \lambda_i)\right\} \left(x_2^T \eta\right)^2 = 0$$

$$\Leftrightarrow -\prod_{i=1 i \neq 2}^{k} (1 - \lambda_i) - \lambda_1 \prod_{i=3}^{k} (1 - \lambda_i) + \left\{\prod_{i=3}^{k} (1 - \lambda_i)\right\} \left(x_2^T \eta\right) = 0$$

$$\Leftrightarrow -\prod_{i=3}^{k} (1 - \lambda_i)\left[1 - \lambda_1 + \lambda_1 - x_2^T \eta\right] = 0$$

$$\Leftrightarrow \quad x_2^T \eta = 1.$$

Continuing this sequence, we find either $x_i^T \eta = 1$ or the minimum is at one of the extremities. If $x_i^T \eta = 1$, then J is clearly 1, no matter what the value of the λ_i is. If $x_i^T \eta \neq 1$, then we need consider either $\lambda_i = 0$ or $\lambda_i = \alpha$. In the former case $J = 1$, while in the latter

$$J = (1 - \alpha)^k + \alpha \sum_{j=1}^{k} (1 - \alpha)^{k-j} \left(\eta^T x_j\right)^2.$$

Thus for $k \leq N$,

$$\eta^T P_k^{-1} \eta \geq (1 - \alpha)^N.$$

Now suppose there does not exist an α_5 such that the lower bound of (3.4) holds for all k. Then in view of (C.3), for an arbitrary $\epsilon > 0$ there exists $k > N$ and a unit vector η such that

$$(1 - \alpha)^k + \alpha \sum_{j=1}^{k} (1 - \alpha)^{k-j} \left(\eta^T x_j\right)^2 \leq \epsilon.$$

Then for any finite N

$$\alpha \sum_{j=k-N}^{k} (1 - \alpha)^{k-j} \left(\eta^T x_j\right)^2 \leq \epsilon$$

so

$$\sum_{j=k-N}^{k} \left(\eta^T x_j\right)^2 \leq \frac{\epsilon}{\alpha(1 - \alpha)^N}$$

and (C.0a) is violated. Thus the lower bound of (3.2) implies that of (3.4).

Proof of Theorem 4.2

The approach used here is similar to that in [15]. Define d as the unit delay operator. Then (1.1) can be re-expressed as

$$A(d) y_k = B(d) u_k + v_k$$

where

$$A(d) = 1 - \sum_{i=1}^{n} a_i d^i$$

$$B(d) = \sum_{j=0}^{m} b_j d^j.$$

Suppose the lower bound of (3.2) is violated. Then for all $\epsilon > 0$, there exist a unit vector $\xi \triangleq [\gamma_1, \cdots, \gamma_n, \eta_0, \cdots, \eta_m]^T$ and a k

such that for any $i \in [k, k+N]$

$$|\xi^T x_i| < \epsilon$$

$$\Rightarrow \left| \sum_{j=1}^{n} \gamma_j y_{i-j} + \sum_{j=0}^{m} \eta_j u_{i-j} \right| < \epsilon, \qquad \forall i \in [k, k+N]$$

$$\Rightarrow \left| \sum_{i=1}^{n} \gamma_i d^i y_i + \sum_{j=0}^{m} \eta_j d^j u_i \right| < \epsilon, \qquad \forall i \in [k, k+N].$$

Define

$$\sum_{i=1}^{n} \gamma_i d^i = \gamma(d)$$

and

$$\sum_{j=0}^{m} \eta_j d^j = \eta(d).$$

Thus

$$\begin{aligned}
&|\gamma(d) y_i + \eta(d) u_i| < \epsilon, && \forall i \in [k, k+N] \\
\Rightarrow\ &|\gamma(d) a_j y_{i-j} + \eta(d) a_j u_{i-j}| < \epsilon |a_j|, && \forall i \in [k-j, k+N] \\
\Rightarrow\ &|\gamma(d) A(d) y_i + \eta(d) A(d) u_i| < 0(\epsilon), && \forall i \in [k-n, k+N]. \\
\Rightarrow\ &|\gamma(d) B(d) u_i + \eta(d) A(d) u_i + \gamma(d) v_i| < 0(\epsilon), && \forall i \in [k-n, k+N] \\
\Rightarrow\ &|\{\gamma(d) B(d) + \eta(d) A(d)\} u_i + \gamma(d) v_i| < 0(\epsilon), && \forall i \in [k-n, k+N]
\end{aligned} \tag{C.4}$$

Now $\gamma(d)B(d) + \eta(d)A(d) \neq 0$ as otherwise,

$$\frac{B(d^{-1})}{A(d^{-1})} = \frac{\sum_{j=0}^{m} \eta_{m-j} (d^{-1})^j}{\sum_{j=0}^{n-1} \gamma_{n-j} (d^{-1})^j}$$

which violates the assumption that $B(d^{-1})$ and $A(d^{-1})$ are coprime since the degree of $A(d^{-1})$ is n and that of $\sum_{j=0}^{n-1} \gamma_{n-j} (d^{-1})^j$ is $n-1$.

Thus there exists a χ, bounded away from zero such that

$$|\chi^T W_1(i)| < 0(\epsilon), \qquad \forall i \in [k-n, k+N]$$

so (4.3) is violated; hence (4.3) implies the desired result. Moreover, by (C.4)

$$|\{\gamma(d) B(d) + \eta(d) A(d)\} u_i| < 0(\epsilon) + \sqrt{\sum_{i=1}^{n} v_i^2} \leq 0(\epsilon) + \sqrt{n}\gamma.$$

Thus (4.4) is violated. Note that the upper bounds follow easily from our boundedness assumptions.

References

[1] Y. D. Landau, *Adaptive Control: The Model Reference Approach*. New York: Marcel Dekker, 1979.

[2] G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1982.

[3] C. R. Johnson, Jr., and B. D. O. Anderson, "On reduced order adaptive output error identification and adaptive IIR filtering," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 927–933, 1982.

[4] C. R. Johnson, Jr., "A convergence proof for a hyperstable adaptive recursive filter," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 746–749, 1979.

[5] B. D. O. Anderson and C. R. Johnson, Jr., "Exponential convergence of adaptive identification and control algorithms," *Automatica*, vol. 18, pp. 1–13, 1982.

[6] L. Ljung and T. Söderstrom, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.

[7] E. Fogel and Y. F. Huang, "On the value of information in system identification-bounded noise case," *Automatica*, vol. 20, pp. 229–238, 1982.

[8] Y. F. Huang, "A recursive estimation algorithm using selective updating for spectral analysis and adaptive signal processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1331–1334, 1986.

[9] T. R. Fortesque, L. S. Kershenbaum, and B. E. Ydstie, "Implementation of self tuning regulators with variable forgetting factors," *Automatica*, vol. 19, pp. 831–835, 1981.

[10] B. Egardt, *Stability of Adaptive Controllers*. New York: Springer-Verlag, 1979, p. 20.

[11] ——, "Global stability analysis of adaptive control systems with disturbances," *Proc. JACC*, 1980.

[12] C. Samson, "Stability analysis of adaptively controlled systems subject to bounded disturbances," *Automatica*, vol. 19, pp. 81–86, 1983.

[13] B. B. Peterson and K. S. Narendra, "Bounded error adaptive control," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 1161–1168, 1982.

[14] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.

[15] R. M. Johnstone and B. D. O. Anderson, "Exponential convergence of recursive least-squares with exponential forgetting factor-adaptive control," *Systems and Control Letters*, vol. 2, pp. 69–76, 1982.