ARMA Parameter Estimation Using a Novel Recursive Estimation Algorithm with Selective Updating

ASHOK K. RAO, YIH-FANG HUANG, MEMBER, IEEE, AND SOURA DASGUPTA

Abstract—This paper investigates an extension of a recursive estimation algorithm (the so-called OBE algorithm) [9]-[11], which features a discerning update strategy. In particular, an extension of the algorithm to ARMA parameter estimation is presented here along with convergence analysis. The extension is similar to the extended least-squares algorithm. However, the convergence analysis is complicated due to the discerning update strategy which incorporates an information-dependent updating factor. The virtues of such an update strategy are: 1) more efficient use of the input data in terms of information processing, and 2) a modular adaptive filter structure which would facilitate the development of a parallel-pipelined signal processing architecture. It is shown in this paper that if the input noise is bounded and the moving average parameters satisfy a certain magnitude bound, then the *a posteriori* prediction errors are uniformly bounded. With an additional persistence of excitation condition, the parameter estimates are shown to converge to a neighborhood of the true parameters, and the *a priori* prediction errors are shown to be asymptotically bounded. Simulation results show that the parameter estimation error for the EOBE algorithm is comparable to that for the ELS algorithm.

I. INTRODUCTION

IN many adaptive signal processing applications, such as speech processing, seismic data processing, and channel equalization, a signal $y(t)$ is often considered as the output of an IIR filter driven by unknown white noise $w(t)$ [1]. The signal $y(t)$ can therefore be modeled as an autoregressive moving average (ARMA) process of the form

$$y(t) = a_1 y(t-1) + \cdots + a_n y(t-n) + w(t)$$
$$+ c_1 w(t-1) + \cdots + c_r w(t-r). \quad (1.1)$$

Fitting this ARMA model to the measured data $y(t)$, $t = 1, 2, \cdots$, requires the estimation of the parameters a_1, $\cdots$, a_n, c_1, $\cdots$, c_r. Many methods for the estimation of ARMA parameters have been proposed in the literature, particularly from the spectral estimation viewpoint. Among the more recent are Cadzow's overdetermined ra-

tional equation method [2], the spectral matching technique of Friedlander and Porat [3], and the extended Yule–Walker method of Kaveh [4]. A common feature of these methods is the use of the sample autocorrelation sequence of the output process $y(t)$. In the context of system identification, the extended least-squares (ELS), the recursive maximum likelihood (RML), and multistage least-squares algorithms have been used to recursively estimate ARMA parameters [5], [6], [12]. The ELS algorithm uses the *a posteriori* prediction error $\epsilon(t)$, as an estimate of $w(t)$. The regressor vector is formed from $y(t-1), \cdots, y(t-n)$ and $\epsilon(t-1), \cdots, \epsilon(t-r)$. The standard recursive least-squares (RLS) algorithm is then employed to update the estimates. The algorithm is conceptually simple but restrictive in the sense that convergence of the algorithm can be assured only if the underlying transfer function $H(q^{-1}) = 1/C(q^{-1}) - 1/2$ is strictly positive real (SPR), with q^{-1} being the delay operator and

$$C(q^{-1}) = 1 + c_1 q^{-1} + c_2 q^{-2} + \cdots + c_r q^{-r}. \quad (1.2)$$

The RML algorithm, which uses a filtered version of the regressor vector used in the ELS algorithm, does not require $H(q^{-1})$ to be SPR. However, the estimates have to be monitored and projected into a stability region to ensure convergence [5].

In addition to the aforementioned least-squares based methods, there exists a different class of estimation algorithms that estimate membership sets of parameters which are consistent with the measurements and noise constraints [7]-[11]. These algorithms are particularly useful when the noise distribution is unknown but constraints in the form of bounds on the instantaneous values of the noise are available. To the best of our knowledge, none of the algorithms has been applied to the problem of ARMA parameter estimation. Among these algorithms based on membership sets, a group of seminal recursive algorithms are the so-called optimal bounding ellipsoid (OBE) algorithms [9]-[11]. The OBE algorithms have been developed using a set-theoretic formulation and are applicable to autoregressive with exogenous input (ARX) models with bounded noise. One of the main features of these temporally recursive algorithms is a discerning update strategy. This feature, obtained by the introduction

of an information dependent updating/forgetting factor, yields a modular structure thereby increasing the potential for concurrent and pipelined processing of signals. The presence of such a forgetting factor also gives the algorithms the ability to track slowly time varying parameters. One of the algorithms [11] has been shown to possess the advantageous feature of automatic asymptotic cessation of updates if the model is time invariant. If a loose upper bound on the noise magnitude is known, and if the input is persistently exciting and sufficiently uncorrelated with the noise, then it has been shown in [11] that the parameter estimates converge asymptotically to a neighborhood of the true parameter vector.

In this paper, we extend one of the OBE algorithms [11] to the ARMA case. For the ARMA parameter estimation problem, the OBE algorithm cannot be applied in its present form. However, by assuming that the input white noise is bounded in magnitude, the OBE algorithm can be extended in a manner similar to the ELS algorithm. Convergence analysis of the resulting algorithm is performed by imposing a bound on the sum of the magnitudes of the MA coefficients. This ensures that the true parameter vector is contained in all the optimal bounding ellipsoids. A uniform bound on the *a posteriori* prediction error can then be derived. In contrast, even though the *a posteriori* prediction errors are generated in a stable fashion in the ELS algorithm [5], it is difficult to obtain an expression for even the asymptotic bound, if such a bound exists. By imposing a persistence of excitation condition on the regressor vector, the *a priori* prediction error of the extended OBE algorithm is shown to be bounded and the parameter estimates are shown to converge to a neighborhood of the true parameter vector.

The paper is organized in the following manner. In Section II, a brief review of the OBE algorithm and its properties is presented. In Section III, the algorithm is extended to ARMA parameter estimation. Convergence analysis of the extended algorithm is performed in Section IV. The performance of the algorithm is compared to the ELS algorithm through simulation studies in Section V. Section VI concludes the paper.

II. THE OBE ALGORITHM

Consider the ARX model described by

$$y(t) = a_1 y(t-1) + \cdots + a_n y(t-n) + b_0 u(t)$$
$$+ b_1 u(t-1) + \cdots + b_m u(t-m) + v(t)$$

where $y(t)$ is the output, $u(t)$ is the measurable input, and $v(t)$ represents the uncertainty or noise. The above equation can be recast as

$$y(t) = \theta^{*T} \Phi(t) + v(t) \qquad (2.1)$$

where

$$\theta^* = [a_1, a_2, \cdots, a_n, \quad b_0, b_1, \cdots, b_m]^T$$

is the vector of true parameters and

$$\Phi(t) = [y(t-1), y(t-2), \cdots, y(t-n),$$
$$u(t), u(t-1), \cdots, u(t-m)]^T$$

is the regressor vector. A key assumption here is that the noise is bounded in magnitude, i.e., there exists a $\gamma_0 \geq 0$, such that

$$v^2(t) \leq \gamma_0^2, \qquad \text{for all } t, \text{ hence,}$$
$$\left(y(t) - \theta^{*T} \Phi(t)\right)^2 \leq \gamma_0^2.$$

Let S_t be a subset of the euclidean space R^{n+m+1}, defined by

$$S_t = \left\{ \theta \colon \left(y(t) - \theta^T \Phi(t)\right)^2 \leq \gamma_0^2, \quad \theta \in R^{n+m+1} \right\}.$$

From a geometric point of view, S_t is a convex polytope in the parameter space and contains the vector of true parameters. The OBE algorithm starts off with a large ellipsoid, E_0, in R^{n+m+1} which contains all possible values of the modeled parameter θ^*. After the first observation $y(1)$ is acquired, an ellipsoid is found which bounds the intersection of E_0 and the convex polytope S_1. This ellipsoid must be optimal in some sense, say minimum volume [9], [10] or by any other criterion [9], [11], to hasten convergence. Denoting the optimal ellipsoid by E_1, one can proceed exactly as before with the future observations and obtain a sequence of optimal bounding ellipsoids $\{E_t\}$. The center of the ellipsoid E_t can be taken as the parameter estimate at the tth instant and is denoted by $\theta(t)$. If at a particular time instant i, the resulting optimal bounding ellipsoid would be of a "smaller size," thereby implying that the data point $y(i)$ conveys some fresh "information" regarding the parameter estimates, then the parameters are updated. Otherwise, E_i is set equal to E_{i-1} and the parameters are not updated. It can also be shown [11] that all the ellipsoids $\{E_t, t = 1, 2, \cdots\}$ contain the true parameter θ^*, provided that E_0 does.

Let the ellipsoid E_{t-1} at the $(t-1)$th instant be formulated by

$$E_{t-1} = \left\{ \theta \colon \left(\theta - \theta(t-1)\right)^T P^{-1}(t-1) \right.$$
$$\left. \cdot \left(\theta - \theta(t-1)\right) \leq \sigma^2(t-1) \right\} \quad (2.2)$$

for some positive definite matrix $P(t-1)$ and a nonnegative scalar $\sigma^2(t-1)$. Then, given $y(t)$, an ellipsoid which bounds $E_{t-1} \cap S_t$ "tightly" is

$$\left\{ \theta \colon (1 - \lambda_t)\left(\theta - \theta(t-1)\right)^T P^{-1}(t-1) \right.$$
$$\cdot \left(\theta - \theta(t-1)\right) + \lambda_t \left(y(t) - \theta^T \Phi(t)\right)^2$$
$$\left. \leq (1 - \lambda_t)\sigma^2(t-1) + \lambda_t \gamma_0^2 \right\} \qquad (2.3)$$

where the forgetting factor $\lambda(t)$ satisfies $0 \leq \lambda(t) < 1$. The size of the bounding ellipsoid is related to the scalar $\sigma^2(t-1)$ and the eigenvalues of $P(t-1)$. The update equations for $\theta(t)$, $P(t)$, and $\sigma^2(t)$ are derived in [11]. The optimal ellipsoid which bounds the intersection of

E_{t-1} and S_t is defined in terms of an optimal value of λ_t. For the OBE algorithm of [11], the optimum value λ_t^* is determined by minimizing $\sigma^2(t)$ with respect to λ_t at every time instant. The minimization procedure results in λ_t^* being set equal to zero (no update) if

$$\sigma^2(t-1) + \delta^2(t) \leq \gamma_0^2. \qquad (2.4)$$

If (2.4) is not satisfied, then the optimal value of λ_t is computed. The parameter estimation procedure is depicted in Fig. 1. An outgrowth of this modular recursive estimation procedure is a parallel-pipelined networking structure [13]. The algorithm is such that the computational complexity of the information evaluation (IE) procedure is much less than that of the updating procedure (UPD). Since, in general, a good number of data samples would be rejected by the IE, both the IE and the UPD would involve significant amounts of idle time. A viable scheme then is to configure a parallel-pipelined network comprising of such modular estimators to process signals from multiple channels. Apart from reducing hardware costs, such a scheme would offer increased reliability since the failure of one UPD processor would not cause any of the channels to fail, in contrast to a system with a dedicated UPD processor for each channel.

III. EXTENSION TO ARMA MODELS

The ARMA model described by (1.1) can be rewritten as

$$w(t) = y(t) - \theta^{*T}\Phi'(t) \qquad (3.1)$$

where θ^*, the vector of true parameters, and $\Phi'(t)$ are defined by

$$\theta^* = [a_1, a_2, \cdots, a_n, \quad c_1, c_2, \cdots, c_r]^T$$

$$\Phi'(t) = [y(t-1), \cdots, y(t-n),$$
$$w(t-1), \cdots, w(t-r)]^T.$$

Here again, $w(t)$ is assumed to be bounded in magnitude, i.e., there exists positive γ_0^2 such that

$$w^2(t) \leq \gamma_0^2. \qquad (3.2)$$

Since the values of the noise sequence $\{w(t)\}$ are not available, the regressor vector $\Phi'(t)$ is not known exactly. If, however, at time t, an estimate of θ^*,

$$\theta(t) = [a_1(t), \cdots, a_n(t) \quad c_1(t), \cdots, c_r(t)]^T \qquad (3.3)$$

is available, $w(t)$ could be estimated by the *a posteriori* prediction error

$$\epsilon(t) = y(t) - \theta^T(t)\Phi(t) \qquad (3.4)$$

where

$$\Phi(t) = [y(t-1), \cdots, y(t-n),$$
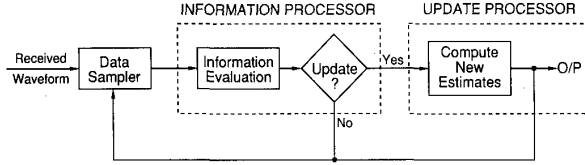$$\epsilon(t-1), \cdots, \epsilon(t-r)]^T. \qquad (3.5)$$

Fig. 1. A modular recursive estimator.

Now just as in the ARX case, define for some suitable γ^2 the convex polytope

$$S_t = \left\{ \theta: (y(t) - \theta^T\Phi(t))^2 \leq \gamma^2, \quad \theta \in R^{n+r} \right\}$$

and the bounding ellipsoid

$$E_t = \left\{ \theta \in R^{n+r}: (\theta - \theta(t))^T P^{-1}(t)(\theta - \theta(t)) \right.$$
$$\left. \leq \sigma^2(t) \right\}.$$

The update equations for $\theta(t)$, $P(t)$, and $\sigma^2(t)$, which then follow directly from [11], are as in the ARX case, with the only difference being that the regressor vector is now given by (3.5):

$$P^{-1}(t) = (1 - \lambda_t)P^{-1}(t-1) + \lambda_t\Phi(t)\Phi^T(t) \qquad (3.6a)$$

$$\theta(t) = \theta(t-1) + \lambda_t P(t)\Phi(t)\delta(t) \qquad (3.6b)$$

$$\delta(t) = y(t) - \theta^T(t-1)\Phi(t) \qquad (3.6c)$$

$$\sigma^2(t) = (1 - \lambda_t)\sigma^2(t-1) + \lambda_t\gamma^2$$
$$- \frac{\lambda_t(1 - \lambda_t)\delta^2(t)}{1 - \lambda_t + \lambda_t G(t)} \qquad (3.6d)$$

where

$$G(t) = \Phi^T(t)P(t-1)\Phi(t). \qquad (3.6e)$$

The matrix inversion lemma can be used in (3.6a) to obtain the following recursion for $P(t)$:

$$P(t) = \frac{1}{1 - \lambda_t}\left[P(t-1) - \frac{\lambda_t P(t-1)\Phi(t)\Phi^T(t)P(t-1)}{1 - \lambda_t + \lambda_t G(t)} \right]. \qquad (3.6f)$$

As in the OBE algorithm, the bounding ellipsoids are optimized by choosing λ_t^* to minimize $\sigma^2(t)$. In order to facilitate the subsequent analysis, the initial conditions are modified to

$$P(0) = MI_{n+r}, \quad \theta(0) = 0, \quad \text{and } \sigma^2(0) = \gamma^2 - \epsilon \qquad (3.7)$$

where $M \gg 1$, $\epsilon \ll 1$, and I_{n+r} is the identity matrix of dimension $n + r$. This choice of initial conditions ensures that the initial ellipsoid E_0 will contain the true parameter vector θ^* and, more importantly, as shown in Appendix A, simplifies the optimum forgetting factor

determination formula to

$$\text{If } \sigma^2(t - 1) + \delta^2(t) \le \gamma^2$$

$$\text{then } \lambda_t^* = 0, \tag{3.8}$$

otherwise

$$\lambda_t^* = \frac{1 - \beta(t)}{2}$$

$$\text{if } G(t) = 1 \tag{3.9a}$$

$$\frac{1}{1 - G(t)}\left[1 - \sqrt{\frac{G(t)}{1 + \beta(t)(G(t) - 1)}}\right]$$

$$\text{if } G(t) \ne 1 \tag{3.9b}$$

where

$$\beta(t) \triangleq \frac{\gamma^2 - \sigma^2(t - 1)}{\delta^2(t)}. \tag{3.9c}$$

Remarks:

1) It is shown in Appendix A that if $\sigma^2(t - 1) + \delta^2(t) > \gamma^2$, then λ_t^* given by (3.9) satisfies

$$\left.\frac{d\sigma^2(t)}{d\lambda_t}\right|_{\lambda_t = \lambda_t^*} = 0$$

and furthermore, $0 < \lambda_t^* < 1$. Thus, unlike [11], no upper bound need be imposed on the forgetting factor.

2) Since $\sigma^2(t) = \sigma^2(t - 1)$ if $\lambda_t^* = 0$, any nonzero value of λ_t^* which minimizes $\sigma^2(t)$ will cause $\sigma^2(t) < \sigma^2(t - 1)$. Thus, choosing λ_t^* to minimize $\sigma^2(t)$ causes $\{\sigma^2(t)\}$ to be a nonincreasing sequence.

The recursive relations (3.6), the initial conditions (3.7), the selective update strategy (3.8), and the forgetting factor determination formula (3.9) form the Extended Optimal Bounding Ellipsoid (EOBE) estimation algorithm [14]. The choice of the threshold γ^2 will become clear from the analysis below. The algorithm retains the discerning update strategy and the modular adaptive filter structure of the OBE algorithm [11], [13].

IV. Analysis of the EOBE Algorithm

The main difficulty in the analysis of the EOBE algorithm arises from the presence of the *a posteriori* prediction errors in the regressor vector. Unlike the OBE algorithm, in this case, boundedness of $w(t)$ does not guarantee that all the convex polytopes S_t, $t = 1, 2, \cdots$, will contain θ^*. The first step in the analysis is to find conditions under which this happens. The minimization of $\sigma^2(t)$, at every time instant, and the choice of initial conditions (3.7), facilitate the characterization of the behavior of the *a posteriori* prediction errors.

Lemma 1: For the EOBE algorithm of Section III, if $\sigma^2(t - 1) + \delta^2(t) > \gamma^2$, i.e., if an update occurs at time instant t, then

i) $\sigma^2(t) + \epsilon^2(t) = \gamma^2,$ (4.1)

ii) $\epsilon^2(k) \le \epsilon^2(t)$ for all time instants $k < t$,   (4.2)

and if $t + j$ is the time instant at which the next update occurs, then

iii) $\epsilon^2(k) \le \epsilon^2(t)$ for all $k < t + j$. (4.3)

Proof:

i) It has been shown in Appendix A that if $\sigma^2(t - 1) + \delta^2(t) > \gamma^2$, then the optimum forgetting factor λ_t^* satisfies

$$\left.\frac{d\sigma^2(t)}{d\lambda_t}\right|_{\lambda_t = \lambda_t^*} = 0. \tag{4.4}$$

Taking the derivative in (3.6d) and using (4.4) yields

$$\gamma^2 - \sigma^2(t - 1) - \frac{(1 - \lambda_t)\delta^2(t)}{1 - \lambda_t + \lambda_t G(t)}$$

$$= -\frac{\lambda_t \delta^2(t)\, G(t)}{\left(1 - \lambda_t + \lambda_t G(t)\right)^2} \tag{4.5a}$$

which can be rewritten in the form

$$\gamma^2 - \sigma^2(t - 1) = \delta^2(t)\frac{(1 - \lambda_t)^2 - \lambda_t^2 G(t)}{\left(1 - \lambda_t + \lambda_t G(t)\right)^2}. \tag{4.5b}$$

In (4.5) and in the remainder of the paper, when there is no risk of confusion, the optimum forgetting factor λ_t^* will be denoted by λ_t. It is also easily shown from (3.6b), (3.6c), and (3.6f) that the *a posteriori* and *a priori* prediction errors are related by

$$\epsilon(t) = \frac{1 - \lambda_t}{1 - \lambda_t + \lambda_t G(t)}\delta(t). \tag{4.6}$$

Note that the nonnegativeness of $G(t)$ implies that $\epsilon^2(t) \le \delta^2(t)$. Substituting (4.6) in (4.5b) and rearranging terms yields

$$(1 - \lambda_t)\gamma^2 - (1 - \lambda_t)\sigma^2(t - 1)$$

$$= (1 - \lambda_t)\epsilon^2(t) - \frac{\lambda_t^2 G(t)\,\epsilon^2(t)}{1 - \lambda_t}. \tag{4.7}$$

Now using (4.6) in (3.6d) gives

$$\sigma^2(t) = (1 - \lambda_t)\sigma^2(t - 1) + \lambda_t\gamma^2$$

$$- \lambda_t\epsilon^2(t) - \frac{\lambda_t^2 G(t)}{1 - \lambda_t}\epsilon^2(t). \tag{4.8}$$

Finally, subtracting (4.8) from (4.7) gives (4.1).

ii) *Case 1:* If $k < t$ is an updating instant, then (4.1) gives

$$\sigma^2(k) + \epsilon^2(k) = \gamma^2. \tag{4.9}$$

But since $\{\sigma^2(t)\}$ is a nonincreasing sequence, (4.9) and (4.1) together would imply that

$$\epsilon^2(k) \le \epsilon^2(t).$$

Case 2: If $k < t$ is a nonupdating instant, then $\epsilon^2(k)$ $= \delta^2(k)$, and so by (3.8), $\sigma^2(k-1) + \epsilon^2(k) \leq \gamma^2$, and since $\sigma^2(t)$ is nonincreasing, $\epsilon^2(k) \leq \epsilon^2(t)$.

iii) Since λ_k, $k = t+1, t+2, \cdots, t+j-1$, are all zero, $\sigma^2(k) = \sigma^2(t)$, for all $t < k < t+j$. And because k is a nonupdating instant, $\sigma^2(k-1) + \epsilon^2(k) = \sigma^2(t) + \epsilon^2(k) \leq \gamma^2$, and so (4.3) follows.

We can now derive sufficient conditions under which the convex polytopes S_t and E_t will contain θ^*.

Theorem 1: The convex polytopes S_t and consequently the ellipsoids E_t, $t = 1, 2, \cdots$, will contain the true parameter, if

i) E_0 contains θ^*,   (4.10a)

ii) the true moving average coefficients satisfy

$$\left[\sum_{i=1}^{r} |c_i| \right]^2 < 0.5, \qquad (4.10b)$$

iii) the threshold γ^2 satisfies

$$\gamma^2 \geq \left[2\gamma_0^2 \frac{\left[1 + \sum_{i=1}^{r} |c_i| \right]^2}{1 - 2\left[\sum_{i=1}^{r} |c_i| \right]^2} \right]. \qquad (4.10c)$$

Proof: Let the induction hypothesis be $\theta^* \in E_{t-1}$. Then defining

$$V(t) = \left(\theta(t) - \theta^* \right)^T P^{-1}(t) \left(\theta(t) - \theta^* \right) \quad (4.11)$$

and recalling the definition of E_{t-1} yields

$$V(t-1) \leq \sigma^2(t-1) \qquad (4.12)$$

and since $P^{-1}(t)$ is positive definite for all t, $\sigma^2(t-1) > 0$.

Now using (3.1) and (3.5)

$$\left(y(t) - \theta^{*T}\Phi(t) \right)^2$$
$$= \left(C(q^{-1})[w(t)] - \left(C(q^{-1}) - 1 \right)[\epsilon(t)] \right)^2$$

where the operator $C(q^{-1})$ has been defined in (1.2). Defining $n(t) = C(q^{-1})[w(t)]$, and recalling an elementary algebraic inequality

$$(a - b)^2 \leq 2a^2 + 2b^2$$

yields

$$\left(y(t) - \theta^T\Phi(t) \right)^2 \leq 2n^2(t) + 2\big(c_1\epsilon(t-1)$$
$$+ c_2\epsilon(t-2) \cdots + c_r\epsilon(t-r)\big)^2. \qquad (4.13)$$

But

$$n^2(t) \leq \tfrac{1}{2}\gamma'^2, \qquad \text{for all } t$$

where

$$\gamma'^2 \triangleq 2\gamma_0^2 \left(1 + \sum_{i=1}^{r} |c_i| \right)^2. \qquad (4.14)$$

Hence,

$$\left(y(t) - \theta^{*T}\Phi(t) \right)^2 \leq \gamma'^2 + 2\big(|c_1|\,|\epsilon(t-1)|$$
$$+ |c_2|\,|\epsilon(t-2)| + \cdots + |c_r|\,|\epsilon(t-r)| \big)^2. \qquad (4.15)$$

But by Lemma 1, if $t - j$ is the updating instant immediately preceding time instant t, then

$$\left| \epsilon(t-i) \right| \leq \left| \epsilon(t-j) \right| \qquad \text{for } 1 \leq i \leq r.$$

Thus

$$\left(y(t) - \theta^{*T}\Phi(t) \right)^2$$
$$\leq \gamma'^2 + 2\left(\sum_{i=1}^{r} |c_i| \right)^2 \epsilon^2(t-j)$$
$$\leq \gamma'^2 + 2\left(\sum_{i=1}^{r} |c_i| \right)^2 \left(\gamma^2 - \sigma^2(t-1) \right).$$

Since $\epsilon^2(t-j) = \gamma^2 - \sigma^2(t-j) = \gamma^2 - \sigma^2(t-1)$. Now by the induction hypothesis, $\sigma^2(t-1) \geq 0$. Hence,

$$\left(y(t) - \theta^{*T}\Phi(t) \right)^2 \leq \gamma'^2 + 2\left(\sum_{i=1}^{r} |c_i| \right)^2 \gamma^2. \quad (4.16)$$

So the convex polytope S_t will contain θ^* if

$$\gamma'^2 + 2\left(\sum_{i=1}^{r} |c_i| \right)^2 \gamma^2 \leq \gamma^2. \qquad (4.17)$$

The inequality (4.17) will hold iff (4.10b) and (4.10c) are true. Assuming (4.10b) and (4.10c) thus guarantees that for all time instants t

$$\left(y(t) - \theta^{*T}\Phi(t) \right)^2 \leq \gamma^2. \qquad (4.18)$$

Using (3.6) and (4.11), it can be shown that

$$V(t) - \sigma^2(t) \leq (1 - \lambda_t)\big(V(t-1) - \sigma^2(t-1) \big)$$
$$+ \lambda_t\big[\left(y(t) - \theta^{*T}\Phi(t) \right)^2 - \gamma^2 \big] \qquad (4.19)$$

and so from (4.18) it follows that

$$V(t) - \sigma^2(t) \leq (1 - \lambda_t)\big(V(t-1) - \sigma^2(t-1) \big). \qquad (4.20)$$

Finally, by (4.12), it follows that

$$V(t) - \sigma^2(t) \leq 0, \qquad (4.21)$$

i.e., E_t contains θ^*, and $\sigma^2(t)$ is nonnegative for all t.

Remarks:

1) The assumption (4.10b) says that the noise sequence $n(t) = C(q^{-1})[w(t)]$ should not be "too colored." This condition is analogous to the Strictly Positive Real (SPR) condition which appears in the ELS algorithm (cf. Section I). It is not very difficult to show that for the SPR condition to hold, it is necessary that

$$\sum_{i=1}^{r} |c_i|^2 < 1. \qquad (4.22)$$

It can also be seen that condition (4.10b) is a stricter form of the Strictly Dominant Passive (SDP) condition [15] which appears in the analysis of some signed LMS algorithms, and from [15], it follows that (4.10b) is sufficient for the SPR condition to hold and hence is more restrictive than the SPR condition.

2) Selection of the right "noise bound" γ^2 is made possible by (4.10c). The user would, however, need to have some knowledge of the magnitude of the true moving average coefficients. Simulation results show that overestimation of γ^2 has very little effect on the parameter estimates (centers of the bounding ellipsoids), although it may have an adverse effect on the size of the bounding ellipsoids.

3) The conditions (4.10b) and (4.10c) are not necessary conditions, and the algorithm has been observed to perform well in several examples where these conditions were violated.

The following result follows straightforwardly from Lemma 1 and Theorem 1.

Corollary 1: If the conditions of Theorem 1 hold then

a) $\displaystyle \lim_{t_j \to \infty} \epsilon^2(t_j)$ exists   (4.23a)

where $\{ t_j \}$ is the subsequence of updating instants of the EOBE algorithm, and

b) uniformly bounded *a posteriori* prediction errors

$$\epsilon^2(t) \le \gamma^2, \quad \text{for all time instants } t. \quad (4.23b)$$

Boundedness of $\delta^2(t)$, the *a priori* prediction error, and convergence of the parameter estimates to a neighborhood of the true parameter can be assured by requiring the regressor vector to be persistently exciting. The next lemma relates the positive definiteness of $P^{-1}(t)$ to the richness of the regressor vector $\Phi(t)$.

Lemma 2: If there exist positive α_3 and N such that, for all t

$$\sum_{i=t}^{t+N} \Phi(i) \Phi^T(i) \ge \alpha_3 I > 0, \quad (4.24a)$$

then there exists a positive α_4 such that

$$P^{-1}(t) \ge \alpha_4 I > 0. \quad (4.24b)$$

Proof of the lemma is the same as that of Theorem 4.1 of [11], it is thus omitted here.

Remark: The positive definiteness of $P^{-1}(t)$ implies that the eigenvalues of $P(t)$ are upper bounded.

Theorem 2: If the assumptions of Theorem 1 are satisfied and (4.24a) holds, then the EOBE algorithm ensures the following.

a) Parameter difference convergence

$$\lim_{t \to \infty} \| \theta(t) - \theta(t - k) \| = 0$$

for any finite k.   (4.25)

b) Asymptotically bounded parameter estimation errors

$$\| \theta(t) - \theta* \|^2 \to \left[0, 2\gamma_0^2 \left(1 + \sum_i |c_i| \right)^2 / \alpha_4 \right] \quad (4.26)$$

where γ_0^2 and α_4 are as in (3.2) and (4.24b), respectively.

c) If, in addition, the process (1.1) is stable, then the algorithm yields asymptotically bounded *a priori* prediction errors

$$\delta^2(t) \to [0, \gamma^2]. \quad (4.27)$$

Proof:
a) From (3.6b) and (3.6f)

$$\| \theta(t) - \theta(t - 1) \|^2$$
$$= \frac{\lambda_t^2 \Phi^T(t) P^2(t - 1) \Phi(t) \delta^2(t)}{\left(1 - \lambda_t + \lambda_t G(t) \right)^2} \quad (4.28)$$

$$\le e_{\max} \{ P(t - 1) \} \frac{\lambda_t^2 G(t) \delta^2(t)}{\left(1 - \lambda_t + \lambda_t G(t) \right)^2} \quad (4.29)$$

where $e_{\max} \{ P(t - 1) \}$ is the maximum eigenvalue of $P(t - 1)$, and $\| \cdot \|$ denotes the Euclidean norm. Using (3.6d) in (4.5a) yields

$$\sigma^2(t) = \sigma^2(t - 1) - \frac{\lambda_t^2 \delta^2(t) G(t)}{\left(1 - \lambda_t + \lambda_t G(t) \right)^2}. \quad (4.30)$$

The nonnegativity of $\sigma^2(t)$ therefore implies

$$\sum_{i=1}^{t} \frac{\lambda_i^2 \delta^2(i) G(i)}{\left(1 - \lambda_i + \lambda_i G(i) \right)^2} = \sigma^2(0) - \sigma^2(t) < \infty. \quad (4.31)$$

Hence,

$$\lim_{t \to \infty} \frac{\lambda_t^2 \delta^2(t) G(t)}{\left(1 - \lambda_t + \lambda_t G(t) \right)^2} = 0. \quad (4.32)$$

If (4.24a) holds, then by Lemma 2, $e_{\max} \{ P(t - 1)$, the maximum eigenvalue of $P(t - 1)$, is bounded for all t, and hence (4.29) and (4.32) yield

$$\| \theta(t) - \theta(t - 1) \| \to 0. \quad (4.33)$$

Applying the Minkowski inequality to $\| \theta(t) - \theta(t - k) \|$ and using (4.33) completes the proof of (4.25).

b) Using (3.6), (4.11), and (4.6), an expression similar to (4.19) can be derived as

$$V(t) = (1 - \lambda_t) V(t - 1) + \lambda_t \left[\left(C(q^{-1})[w(t)] \right. \right.$$
$$- \left. \left(C(q^{-1}) - 1 \right) [\epsilon(t)] \right)^2$$
$$- \left. \frac{1 - \lambda_t + \lambda_t G(t)}{1 - \lambda_t} \epsilon^2(t) \right]. \quad (4.34)$$

Just as in the proof of Theorem 1, (4.34) can be expressed as

$$V(t) \le (1 - \lambda_t)V(t - 1) + \lambda_t\gamma'^2$$

$$+ \lambda_t\left[2\left(\sum_{i=1}^{r} |c_i|\right)^2 \epsilon^2(t - j)\right.$$

$$\left. - \frac{1 - \lambda_t + \lambda_t G(t)}{1 - \lambda_t} \epsilon^2(t)\right] \quad (4.35)$$

where γ'^2 is as in (4.14), and $t - j$ is the updating instant immediately preceding time instant t. Assume t is an updating instant. Then (4.10b), (4.2), and the nonnegativity of $G(t)$ would imply that the term in square brackets on the right-hand side of (4.35) is not positive, and so

$$V(t) \le (1 - \lambda_t)V(t - 1) + \lambda_t\gamma'^2. \quad (4.36)$$

It is obvious that if t is not an updating instant, then (4.36) would still follow from (4.35). A nonrecursive form for (4.36) can be obtained as

$$V(t) \le \prod_{i=1}^{t} (1 - \lambda_i)V(0) + \gamma'^2 \sum_{i=1}^{t} q_{it} \quad (4.37)$$

where

$$q_{it} = \begin{cases} \lambda_i(1 - \lambda_{i+1}) \cdots (1 - \lambda_t), & i < t \\ \lambda_t & i = t. \end{cases} \quad (4.38)$$

For large t, the first term on the right-hand side of (4.37) can be neglected. In Appendix B, it is shown that

$$\sum_{i=1}^{t} q_{it} < 1. \quad (4.39)$$

Hence, for large enough t

$$V(t) = \big(\theta(t) - \theta^*\big)^T P^{-1}(t)\big(\theta(t) - \theta^*\big) \le \gamma'^2. \quad (4.40)$$

And so (4.26) follows from Lemma 2 and (4.14).

c) Stability of the process (1.1) and the boundedness of $w(t)$ implies that the outputs $y(t)$ are bounded. Hence, from (3.6e), (4.23b), and Lemma 2, it follows that

$$G(t) \le e_{\max}\big\{P(t - 1)\big\}\left[r\gamma^2 + n \max_{t-n \le i \le t-1} y^2(i)\right]$$

$$< \infty \quad (4.41)$$

where n is the order of the AR process and r is the order of the MA process. It can now be shown, just as in Theorem 3.2 of [11], that the a priori prediction errors satisfy (4.27).

Remarks:

1) The results of Theorem 1, and the results (4.25), (4.26) of Theorem 2, do not require the process to be stable. However, if the process is unstable, then on account of finite precision effects, the matrix $P(t)$ may not stay positive definite, thus invalidating the notion of bounding ellipsoids and causing the algorithm to fail. In this situation, the ELS algorithm will fail, too.

2) Theorems 1 and 2 do not impose any statistical properties on the input noise sequence $\{w(t)\}$. However, our simulation experience has been that the parameter estimates are usually not close to the true parameters if the noise is not white. Of course, such is also the case for the ELS algorithm.

V. SIMULATION RESULTS

Simulations have been performed to investigate the performance of the EOBE algorithm vis à vis the ELS algorithm. In this paper, we present simulation results for two examples—a broad-band ARMA (3, 3) process and a narrow-band ARMA (2, 2) process where the indexes n, r in an ARMA (n, r) process refer to the orders of the $A(q^{-1})$ and $C(q^{-1})$ polynomials, respectively.

Example 1—Broad-band ARMA (3, 3) Process: The output data $\{y(t)\}$ are generated by the following difference equation:

$$y(t) = -0.4y(t - 1) + 0.2y(t - 2) + 0.6y(t - 3)$$

$$+ w(t) - 0.22w(t - 1) + 0.17w(t - 2)$$

$$- 0.1w(t - 3).$$

The noise sequence $\{w(t)\}$ is generated by a pseudo-random number generator with a uniform probability distribution in $[-1.0, 1.0]$. The upper bound γ^2 was set equal to 25. The parameter estimates were obtained by applying the EOBE algorithm to 1000 point data sequences. Twenty-five runs of the algorithm were performed on the same model but with different input noise sequences. The average squared parameter error $L_1(t)$ is computed for the AR coefficients according to the formula

$$L(t) = \frac{1}{25}\sum_{j=1}^{25} l_j(t)$$

where $l_j(t)$, the squared AR parameter error at time t for the jth run, is defined by

$$l_j(t) = \sum_{i=1}^{n} \big(a_i(t) - a_i\big)^2$$

with a_i and $a_i(t)$ being defined by (1.1) and (3.3), respectively. The average squared parameter error $L_2(t)$ for the MA coefficients is defined analogously. Figs. 2 and 3 display the average squared estimation errors for AR and MA parameters using both the EOBE and the ELS algorithms. The curves show that the performance of the two algorithms is comparable. The average number of updates for the EOBE algorithm was 160 for 1000 point data sequences. Thus, only 16% of the samples are used for updates, as compared to the ELS algorithm which updates at every sampling instant.
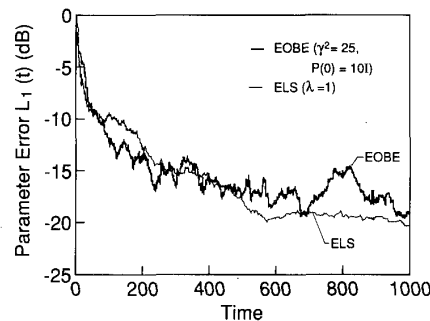
Fig. 2. Average squared AR parameter estimation error for the EOBE and
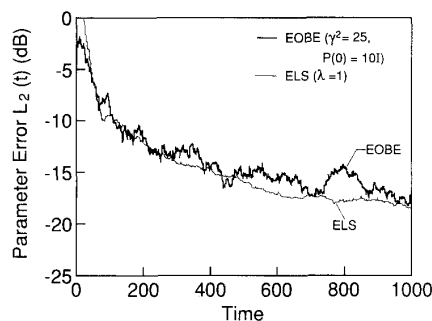ELS algorithms—Example 1.

Fig. 3. Average squared MA parameter estimation error for the EOBE and
ELS algorithms—Example 1.

TABLE I

Upper Bound γ^2	Average Tap Error	Average Number of Updates	Total Number of Times θ^* is Out of S_t	Total Number of Times θ^* is Out of E_t	Average Final Volume	Average Final Sum of Axes
0.5	0.031	160	7309	23952	—	—
1.0	0.031	160	315	0	0.22	10.46
2.0	0.031	160	0	0	2.6×10^4	74
5.0	0.031	154	0	0	5.4×10^7	265
25.0	0.031	153	0	0	2.1×10^{12}	1537
100.0	0.0308	156	0	0	1.0×10^{16}	6303

The effect of different choices for the upper bound γ^2 on the performance has also been studied. For each value of γ^2, the asymptotic average squared parameter error T was computed over 25 runs of the algorithm, according to the formula

$$T = \frac{1}{25} \sum_{j=1}^{25} \left\| \theta_j(1000) - \theta^* \right\|^2$$

where $\theta_j(1000)$ is the parameter estimate at the 1000th iteration in the jth run. The lower bound on γ^2 as calculated from (4.10c) is $\gamma^2 \geq 8.54$. The second column of Table I lists the different values of T obtained when γ^2 is varied from 0.5 to 100. It is clear that the centers of the bounding ellipsoids are insensitive to the value of γ^2, since the tap error is almost constant. However, the final size of the ellipsoids does depend on γ^2. The negative

volume obtained when $\gamma^2 = 0.5$ is an indication of the fact that $\sigma^2(t)$ is no longer positive and so bounding ellipsoids cannot be constructed.

The performance of the algorithm, when the noise sequence $\{w(t)\}$ has a Gaussian distribution, was evaluated in a similar fashion. A constant value of $\gamma^2 = 25$ was used and the standard deviation of the noise was varied. The results for 25 runs of the algorithm are shown in Table II. It is clear that the unbounded noise has marginal effect on the parameter estimates.

Finally, the tracking capability of the EOBE algorithm was compared to that of the ELS algorithm (with forgetting factor = 0.99). The same model was used to generate 400 data points. The parameters were then changed by 150% and the next 400 points were generated. Finally, the last 200 points were generated by using the original parameters. The average squared parameter error was

TABLE II

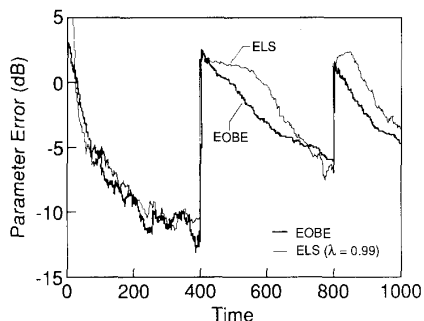| Standard Deviation of Noise	Average Tap Error	Average Number of Updates	Total Number of Times θ^* is Out of S_t	Total Number of Times θ^* is Out of E_t	Average Final Volume	Average Final Sum of Semiaxes
0.5	0.29	93	0	0	5.6×10^{12}	1575
1.0	0.317	108	0	0	4.89×10^8	333
2.0	0.313	117	323	0	1.47×10^2	20
3.0	0.32	122	2439	19787	—	—

Fig. 4. Tracking performance of the EOBE and ELS algorithms—average squared parameter estimation error.
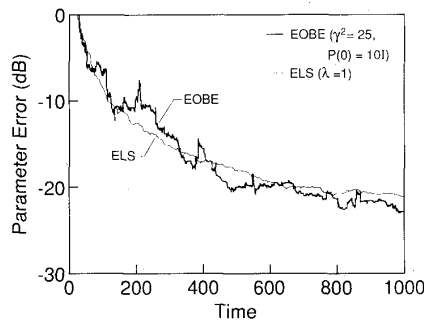

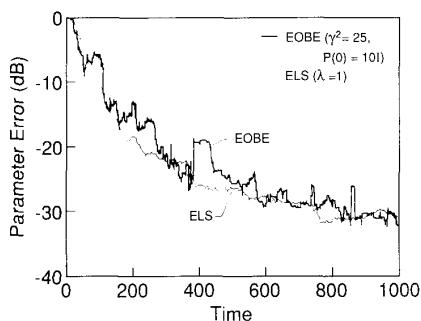Fig. 6. Average squared MA parameter estimation error for the EOBE and ELS algorithms—Example 2.

Fig. 5. Average squared AR parameter estimation error for the EOBE and ELS algorithms—Example 2.

evaluated over 25 runs and is shown in Fig. 4. Even though the formulation of bounding ellipsoids is based on the assumption that the parameters are constant, the simulation results show that the algorithm is able to accommodate changes in model parameters. Analysis of the tracking ability of the algorithm is currently under investigation.

Example 2—Narrow-band ARMA (2, 2) Process: The output data $\{y(t)\}$ are generated by the following difference equation:

$$y(t) = 1.4y(t-1) - 0.95y(t-2) + w(t)$$
$$- 0.86w(t-1) + 0.431w(t-2).$$

Note that, in this case, condition (4.10b) of Theorem 1 is violated. The noise sequence is uniformly distributed in $[-1.0, 1.0]$, as in the first example. The upper bound γ^2 was set equal to 25. The average squared AR and MA parameter estimation errors are calculated over twenty five

runs and plotted in Figs. 5 and 6, respectively. The average number of updates was 78 for 1000 point data sequences.

For this example too, different values of the upper bound γ^2 were used and no significant difference in the quality of estimates, number of updates or convergence rate was observed. Thus, it is verified once again that a precise knowledge of the upper bound is not a prerequisite for satisfactory performance of the algorithm.

VI. CONCLUSION

A recursive parameter estimation algorithm has been extended for ARMA parameter estimation. The main features of the algorithm are a membership set theoretic formulation and a discerning update strategy. Convergence analysis of the algorithm has been performed under the assumption that the noise is bounded. The main results of the analysis are that all the bounding ellipsoids will contain the true parameter, provided the true moving average coefficients satisfy a condition, which is analogous to the SPR condition of the ELS algorithm. In addition, the algorithm yields uniformly bounded *a posteriori* prediction errors. With a persistence of excitation condition on the regressor vector, boundedness of the *a priori* prediction errors can then be established and the parameter estimates are shown to converge to a neighborhood of the true parameters. Simulation results show that the performance of the algorithm is comparable to the ELS algorithm while requiring far fewer updates.

APPENDIX A

Proof of (3.8) and (3.9): The proof is along the lines of the proof of Lemma 2.1 in [11].

Since λ_t^* minimizes $\sigma^2(t)$

$$\sigma^2(t, \lambda_t^*) \le \sigma^2(t, 0) = \sigma^2(t - 1) \qquad (A.1)$$

and

$$\frac{d\sigma^2(t)}{d\lambda_t} = \gamma^2 - \sigma^2(t - 1)$$

$$- \delta^2(t) \frac{(1 - \lambda_t)^2 - \lambda_t^2 G(t)}{\left(1 - \lambda_t + \lambda_t G(t)\right)^2} \quad (A.2)$$

and

$$\frac{d^2\sigma^2(t)}{d\lambda_t^2} = \frac{2\delta^2(t) G(t)}{\left(1 - \lambda_t + \lambda_t G(t)\right)^3}. \qquad (A.3)$$

Thus, $d^2\sigma^2(t)/d\lambda_t^2 > 0$, unless $\delta^2(t) = 0$ or $G(t) = 0$. Since $P(t - 1)$ is positive definite, $G(t) = 0$ iff $\Phi(t) = 0$. The algorithm can be modified to detect the occurrence of a null $\Phi(t)$ and set it to a small nonzero value, prior to the calculation of $G(t)$. Thus, it can be assumed that $G(t) \ne 0$ for all t. If $\delta^2(t) = 0$ ($\sigma^2(t - 1) + \delta^2(t) < \gamma^2$ in this case), then, since $\sigma^2(0) < \gamma^2$ by (3.7), and since $\sigma^2(t)$ is nonincreasing, therefore, by (A.2) $d\sigma^2(t)/d\lambda_t$ is positive, and hence $\sigma^2(t)$ is minimized if $\lambda_t^* = 0$. Now, for the sequel, the second derivative of $\sigma^2(t)$ can be assumed to be positive, and hence the unique minimum occurs at $d\sigma^2(t)/d\lambda_t = 0$. From (A.2), if $G(t) = 1$, $\sigma^2(t)$ is minimized if

$$\lambda_t^* = \left(1 - \beta(t)\right)/2. \qquad (A.4)$$

Otherwise, if $G(t) \ne 1$, $\sigma^2(t)$ is minimized if

$$\lambda_t^* = \frac{1}{1 - G(t)} \left[1 - \sqrt{\frac{G(t)}{1 + \beta(t)\left(G(t) - 1\right)}} \right]. \qquad (A.5)$$

Moreover, in (A.4) and (A.5)

$$\lambda_t^* > 0 \Leftrightarrow \beta(t) < 1 \Leftrightarrow \sigma^2(t - 1) + \delta^2(t) > \gamma^2. \qquad (A.6)$$

It is easy to show that $1 + \beta(t)(G(t) - 1)$ is always positive. Since $\sigma^2(0) < \gamma^2$ and $\sigma^2(t)$ is nonincreasing, therefore, $\beta(t) > 0$. From (A.6), $\beta(t) < 1$, hence $1 - 1/\beta(t) < 0$. Then

$$1 + \beta(t)\left(G(t) - 1\right) \le 0 \Rightarrow G(t)$$

$$\le 1 - 1/\beta(t) \Rightarrow G(t) < 0$$

which is a contradiction. Thus, (A.5) would always yield real λ_t^*. It is now shown that (A.4) and (A.5) yield values of λ_t^* which are upper bounded by unity. If $G(t) = 1$, then since $\beta(t) > 0$, (A.4) yields $\lambda_t^* < 1$. If $G(t) < 1$, then $\lambda_t^* \ge 1 \Leftrightarrow 1 - [G(t)/(1 + \beta(t)(G(t) - 1))]^{1/2} \ge 1 - G(t)$

$$\Leftrightarrow G(t)\left(1 + \beta(t)\left(G(t) - 1\right)\right) \ge 1. \qquad (A.7)$$

But $G(t) < 1$ and $\beta(t) > 0$ contradict (A.7). Hence, if $G(t) < 1$, then $\lambda_t^* < 1$. It can be shown in exactly the

same way that $G(t) > 1$ would imply that $\lambda_t^* < 1$. Thus, unlike the case in [11], no upper bound has to be imposed on the forgetting factor.

APPENDIX B

Proof of (4.39) (by Induction): Let

$$R(t) = \sum_{i=1}^{t} q_{it}. \qquad (B.1)$$

Then

$$R(t) = (1 - \lambda_t)R(t - 1) + \lambda_t \qquad (B.2)$$

and

$$R(1) = \lambda_1 < 1.$$

Assume

$$R(t - 1) < 1.$$

Then by (B.2)

$$R(t) < (1 - \lambda_t) \cdot 1 + \lambda_t,$$

i.e.,

$$R(t) < 1.$$

REFERENCES

[1] B. Friedlander, "System identification techniques for adaptive signal processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 240–246, Apr. 1982.
[2] J. A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, pp. 907–939, Sept. 1982.
[3] B. Friedlander and B. Porat, "A spectral matching technique for ARMA parameter estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 338–343, Apr. 1984.
[4] M. Kaveh, "High resolution spectral estimation for noisy signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 286–287, June 1979.
[5] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification.* Cambridge, MA: M.I.T. Press, 1983.
[6] D. Q. Mayne *et al.*, "Linear identification of ARMA processes," *Automatica* vol. 18, no. 4, pp. 461–466, July 1982.
[7] F. C. Schweppe, "Recursive state estimation: Unknown but bounded errors and system inputs," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 22–28, Feb. 1968.
[8] M. Milanese and G. Belaforte, "Estimation theory and uncertainty intervals," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 408–414, Apr. 1982.
[9] E. Fogel and Y. F. Huang, "On the value of information in system identification—bounded noise case," *Automatica*, vol. 18, no. 2, pp. 229–238, Mar. 1982.
[10] 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. 1131–1134, Oct. 1986.
[11] S. Dasgupta and Y. F. Huang, "Asymptotically convergent modified recursive least-squares," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 383–392, May 1987.
[12] G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction and Control.* Englewood Cliffs, NJ: Prentice-Hall, 1982.
[13] Y. F. Huang and R. W. Liu, "A high level parallel-pipelined network architecture for adaptive signal processing," in *Proc. 26th IEEE Conf. Decision Contr.*, vol. 1, Dec. 1987, pp. 662–666.
[14] A. K. Rao, "Applications and extensions of a novel recursive estimation algorithm with selective updating," M.S. thesis, Dep. Elec. Comput. Eng., Univ. Notre Dame, Notre Dame, IN, Aug. 1987.
[15] S. Dasgupta, J. S. Garnett, and C. R. Johnson, "Convergence of an adaptive filter with signed error filtering," in *Proc. 1988 Amer. Contr. Conf.*, Atlanta, vol. 1, pp. 400–405.
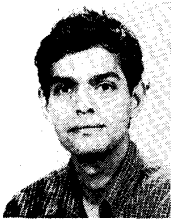
Ashok K. Rao was born in New Delhi, India, on January 12, 1962. He received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, India, in 1984, the M.S. degree in electrical engineering, and the Ph.D. degree, both from the University of Notre Dame in 1987 and 1989, respectively.

He is currently a Member of the Technical Staff at COMSAT Laboratories. His main research interests are in parameter estimation, adaptive filtering, communication theory and image coding.

Dr. Huang is currently Associate Editor on Neural Networks and Signal Processing for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS.

Yih-Fang Huang (S'80–M'82) received the B.S.E.E. degree from National Taiwan University, Taipei, Taiwan, in 1976, the M.S.E.E. degree in electrical engineering from University of Notre Dame, Notre Dame, IN, in 1979, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, 1982.

In 1982 he joined the Department of Electrical and Computer Engineering at University of Notre Dame. In May 1987 he became an Associate Professor at Notre Dame. His research interests are in statistical signal processing and artificial neural networks.

Soura Dasgupta was born in Calcutta, India, in April 1959. He received the B.E. degree in electrical engineering from the University of Queensland, Australia, in 1980. He received the Ph.D. degree in systems engineering from the Australian National University, Canberra, in 1984.

In 1981 he was a Research Fellow at the Electronics and Communication Sciences Unit in the Indian Statistical Institute, Calcutta. In the academic year of 1984–1985 he was a Visiting Assistant Professor in the Department of Electrical Engineering at the University of Notre Dame. He is currently with the Department of Electrical and Computer Engineering at the University of Iowa. His research interests are in adaptive systems theory, robust stability, and neural networks.