Scientia Iranica, Vol. 9, No. 4, pp 378–384
© Sharif University of Technology, October 2002

SCIENTIA
IRANICA

Research Note

NLMS Algorithm with Variable Step-Size Using Set-Membership Identification

A.M. Doost-Hoseini[1] and K. Shahtalebi*

In this paper, set-membership identification is used to derive a simple algorithm which is a sign version of the normalized least mean square algorithm. Convergence analysis is carried out. With some simulation examples, the performance of the algorithm, in the cases of slow and fast variations of a parameter, is compared with the modified Dasgupta-Huang optimal bounding ellipsoid algorithm. These examples show the performance of the proposed algorithm.

INTRODUCTION

Set-Membership (SM) identification, a technique that uses a priori assumptions about a parametric model in order to constrain the solutions to certain sets, has, in recent years, been the focus of extensive research efforts [1-3]. Its adaptive capabilities are receiving considerable attention and it is becoming increasingly popular around the world [4,5]. Lurking in one small, but very significant, corner of SM research, is a point of tangency with Least-Square Error (LSE) identification methods. Fundamentally, this common ground is manifested in a class of algorithms known as Optimal Bounding Ellipsoid (OBE) algorithms. The original version of OBE is attributable to Fogel and Huang [6]; Dasgupta, Kosut, Wahlberg et al. have also used it in robust adaptive control [7]. The Normalized Least Mean Square (NLMS) algorithm is, however, easy to implement and is a robust method for tracking slowly varying parameters of signals and systems [8]. There are many successful applications of the NLMS algorithm and there is a variety of modifications to fit it into specific application requirements. The NLMS algorithm uses normalized instantaneous estimates of the gradient of the mean square error to update parameter estimates. However, it suffers from slow convergence and requires heuristic adjustments of the step size that is a trade-off between excess estimation errors and convergence rate [9-11]. In this paper, the theory of SM identification is used to present a method for automatic adjustment of step-size to improve the convergence rate and reduce steady-state parameter estimation errors. The result is a simple algorithm in the form of sign NLMS. There will be also a review of the fundamentals of OBE algorithms [7]. Then the main idea is introduced, where the structure of the algorithm is described. In each iteration, an estimate of the parameter and a simple spheroid are obtained with its center at the estimate of parameter. Under natural conditions, similar to other OBE algorithms, only a small percentage of the data is used to update the estimates. Convergence analysis is, then, presented and by utilizing suitable simulations, the performance of the proposed algorithm is compared with that of the well-known Dasgupta-Huang Optimal Bounding Ellipsoid (DHOBE) algorithm [7].

SIMPLE STRUCTURE OF OBE ALGORITHM

OBE algorithms are used to identify a model of the general form:

$$y_n = W^T X_n + v_n, \tag{1}$$

in which $W^T = [w_1, \cdots, w_m]$ is the unknown parameter vector, $\{v_n\}$ is a disturbance, error, or input sequence and X_n is a measurable sequence of m-vectors. It is assumed that for each n, v_n is bounded in magnitude by γ, i.e.:

$$v_n^2 \leq \gamma^2, \quad \text{for all } n. \tag{2}$$

Equations 1 and 2 yield:

$$(y_n - W^T X_n)^2 \leq \gamma^2, \quad \text{for all } n. \tag{3}$$

1. *Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, I.R. Iran.*
*. *Corresponding Author, Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, I.R. Iran.*

Let S_n be a subset of R^m defined by:

$$S_n = \left\{ W : \left(y_n - W^T X_n \right)^2 \le \gamma^2,\ W \in R^m \right\}. \quad (4)$$

From a geometrical point of view, S_n is a convex polytope [12]. Thus, with each measured set (y_n, X_n), Equations 1 and 2 together yield a convex polytope in the parameter space. At any instant, n, the intersection of the sequence of polytopes $S_1, \cdots, S_n$ must be considered. It must contain the model parameter W and so must, also, any ellipsoid which bounds it. OBE algorithms starts with a sufficiently large ellipsoid which covers all possible values of W. After (y_1, X_1) is acquired, an ellipsoid is found which bounds the intersection of the initial ellipsoid and S_1. Each algorithm uses a specific optimization criteria and a particular method to find this ellipsoid, which is denoted by E_1, and optimizes it according to its criteria. By the same token, a sequence of optimal bounding ellipsoids E_n can then be obtained. The estimate for W at the nth instant is then defined to be the center of E_n. Suppose that E_{n-1}, at instant $n-1$, is given by:

$$E_{n-1} = \{ W : (W - W_{n-1})^T P_{n-1}^{-1} (W - W_{n-1}) \le \eta_{n-1}^2 \}, \quad (5)$$

for some positive definite matrix P_{n-1} and a nonzero scalar η_{n-1}. Observing (y_n, X_n), an ellipsoid that bounds $E_{n-1} \cap S_n$ is given by:

$$E_n = \{ W : (W - W_n)^T P_n^{-1} (W - W_n) \le \eta_n^2 \}, \quad (6)$$

where:

$$P_n^{-1} = (1 - \lambda_n) P_{n-1}^{-1} + \lambda_n X_n X_n^T, \quad (7)$$

or, equivalently, (using matrix inversion lemma)

$$P_n = \frac{1}{1 - \lambda_n} \left[P_{n-1} - \frac{\lambda_n P_{n-1} X_n X_n^T P_{n-1}}{1 - \lambda_n + \lambda_n X_n^T P_{n-1} X_n} \right], \quad (8)$$

$$e_n = y_n - X_n^T W_{n-1}, \quad (9)$$

$$W_n = W_{n-1} + \frac{\lambda_n P_{n-1} X_n}{1 - \lambda_n + \lambda_n X_n^T P_{n-1} X_n} e_n, \quad (10)$$

$$\eta_n^2 = (1 - \lambda_n) \eta_{n-1}^2 + \lambda_n \gamma^2 - \frac{\lambda_n (1 - \lambda_n) e_n^2}{1 - \lambda_n + \lambda_n X_n^T P_{n-1} X_n}, \quad (11)$$

and λ_n is some scalar in [0,1) [7].

As stated earlier, each OBE algorithm uses a specific criterion to find optimal value for λ_n in renewing ellipsoids. Minimizing η_n^2 det $[P_n], \eta_n^2$ trace $[P_n]$ and η_n^2 are three examples. Dasgupta and Huang chose the last one, because η_n^2 is a bound on the Lyapunov function used in the minimization at time n. Hence, the convergence of the Lyapunov function is used to prove the convergence of the algorithm. With this idea, the best value of λ_n is used in the set $[0, \alpha)$ where α is optional. In the next section, the above structure (Equations 7 to 11) is utilized to design a modified NLMS algorithm.

NLMS recursion can be abtained with step-size μ, i.e.:

$$W_n = W_{n-1} + \frac{\mu X_n}{\theta + \mu X_n^T X_n} e_n, \quad (12)$$

from Equation 10, simply by setting $P_{n-1} = \mu I$ and $\lambda_n = \frac{1}{1+\theta} = cte$. It is also instructive to note that NLMS can be regarded as the exact solution to a minimization problem, using criterion H^∞. However, using H^∞ norm in the design of robust algorithms has some disadvantages. For example, minimizing H^∞ norm may be regarded as minimizing the maximum energy gain from all disturbances to the error and it is obvious that "energy" is not a momentary quantity. It means that when H^∞ norm criterion is used, it is conceivable that in some iterations the estimator does not perform properly, but the overall maximum energy gain is minimized. A suitable solution to overcome this problem is variable step-size μ_n instead of μ. In the next section, it is aimed to find a recursive equation for μ_n using SM identification.

MODIFIED NLMS ALGORITHM

The basic idea in order to derive the Modified NLMS (MNLMS) algorithm, is to replace P_n in Equation 8 by a diagonal matrix $\mu_n I > P_n$ (where $A > B$ means $A - B$ is non negative definite) and use an expanded set:

$$\overline{E}_n = \left\{ W : \mu_n^{-1} (W - W_n)^T (W - W_n) \le \eta_n^2 \right\}, \quad (13)$$

which includes E_n. i.e.:

$$E_n \in \overline{E}_n. \quad (14)$$

To meet this need, suppose at time $n-1$, P_{n-1} is replaced by $\mu_{n-1} I$. Therefore, from Equation 7, non-negative definiteness of $X_n X_n^T$ and $\lambda_n > 0$, the following is obtained:

$$P_n^{-1} = (1 - \lambda_n) \mu_{n-1}^{-1} I + \lambda_n X_n X_n^T \ge (1 - \lambda_n) \mu_{n-1}^{-1} I, \quad (15)$$

hence,

$$P_n \le (1 - \lambda_n)^{-1} \mu_{n-1} I. \quad (16)$$

Comparing Equations 6, 13 and 16, a proper choice for $\overline{E}_n$ and μ_n will be:

$$\overline{E}_n=\left\{W:(1-\lambda_n)\mu_{n-1}^{-1}(W-W_n)^T(W-W_n)\leq\eta_n^2\right\}$$
$$=\left\{W:\frac{1-\lambda_n}{\mu_{n-1}\eta_n^2}(W-W_n)^T(W-W_n)\leq 1\right\}, \quad (17)$$

$$\mu_n=\frac{\mu_{n-1}}{1-\lambda_n}, \quad (18)$$

where:

$$\eta_n^2=(1-\lambda_n)\eta_{n-1}^2+\lambda_n\gamma^2-\frac{\lambda_n(1-\lambda_n)e_n^2}{1-\lambda_n+\lambda_n\mu_{n-1}X_n^TX_n}. \quad (19)$$

From Equation 17, minimizing $\frac{\eta_n^2}{1-\lambda_n}$ with respect to λ_n, there is a right step towards convergene of the algorithm. However, from Equation 19, with definition:

$$f_n(\lambda_n)=\frac{\eta_n^2}{1-\lambda_n}$$
$$=\eta_{n-1}^2+\frac{\lambda_n\gamma^2}{1-\lambda_n}-\frac{\lambda_n e_n^2}{1-\lambda_n+\lambda_n\mu_{n-1}X_n^TX_n}. \quad (20)$$

Differentiating with respect to λ_n and finding the root of the resultant expression, the following is obtained:

$$\lambda_n^*=\frac{||e_n||-\gamma}{||e_n||-\gamma+\gamma\mu_{n-1}X_n^TX_n} \quad ||e_n||>\gamma.$$

Under the condition, $||e_n||>\gamma$, $\frac{d^2f_n(\lambda_n)}{d\lambda_n^2}|\lambda_n=\lambda_n^*>0$ and λ_n^* minimizes $f_n(\lambda_n)$. It is not difficult to show that for the case $||e_n||<\gamma$, $\lambda_n^*=0$ minimizes $f_n(\lambda_n)$. Hence:

$$\lambda_n^*=\begin{cases}\frac{||e_n||-\gamma}{||e_n||-\gamma+\gamma\mu_{n-1}X_n^TX_n}, & ||e_n||>\gamma\\ 0, & ||e_n||<\gamma\end{cases}. \quad (21)$$

Substituting Equation 21 in Equation 19:

$$\eta_n^2=\begin{cases}\frac{\gamma\mu_{n-1}X_n^TX_n}{||e_n||-\gamma+\gamma\mu_{n-1}X_n^TX_n}\Big(\eta_{n-1}^2-\\ \quad\frac{1}{\mu_{n-1}X_n^TX_n}(||e_n||-\gamma)^2\Big) & ||e_n||>\gamma\\ \eta_{n-1}^2 & ||e_n||<\gamma\end{cases} \quad (22)$$

Also, using $\mu_{n-1}I$ instead of P_{n-1} and $\lambda_n=\lambda_n^*$ in Recursion 10 leads to (after some routine algebra):

$$W_n=W_{n-1}+\frac{\lambda_n^*\mu_{n-1}X_n}{1-\lambda_n^*+\lambda_n^*\mu_{n-1}X_n^TX_n}e_n$$
$$=\begin{cases}W_{n-1}, & ||e_n||\leq\gamma\\ W_{n-1}+\frac{||e_n||-\gamma}{X_n^TX_n}X_n\ \mathrm{sign}(e_n), & ||e_n||>\gamma\end{cases}, \quad (23)$$

where sign $(e_n)=\frac{e_n}{||e_n||}$. This is the foundation of the algorithm propounded here. As one can see, parameters γ_n^*,η_n, or μ_n do not have any direct role in executing the algorithm, which is similar to a sign version of the NLMS algorithm.

Remark 1

In replacing P_n by $\mu_n I$, the volume of ellipsoid containing W is expanded. Hence, the ambiguity in the parameter increases. It is the penalty that must be paid for the simplicity of the algorithm.

Remark 2

The recursive form of Equations 18 and 22 has an important role in the proposed approach. At first glance, Recursion 22 confirms that η_n^2 is nonincreasing. In the next section, further discussion is presented about $\eta_n^2\mu_n$.

Remark 3

In the algorithm proposed, W_n is not refreshed (i.e., $W_{n+1}=W_n$) when $||e_n||\leq\gamma$, while in DHOBE, refreshing ceases when $||e_n||^2+\eta_{n-1}^2\leq\gamma^2$. Hence, it seems that the latter uses the measurements more efficiently. This occurs for the same reason explained in Remark 1.

Remark 4

In general, in OBE algorithms, the checking procedure for the presence of acceptable innovation in the data requires $O(m^2)$ operations per sample, while in the proposed algorithm, only $O(m)$ operations are needed. The comparison holds for overall operations also.

CONVERGENCE ANALYSIS

With the aid of a useful theorem, the convergence properties of the algorithm are established in this section. Define:

$$\zeta_n^2=\eta_n^2\mu_n. \quad (24)$$

From the definition of $\overline{E}_n$ (Equation 13) it is found that:

$$||W-W_n||\leq\zeta_n. \quad (25)$$

The following theorem shows convergence of the algorithm.

Theorem 1

If $W\in\overline{E}_n$, then ζ_n^2 is a nonincreasing function of n (hence $\overline{E}_n$ has a nonincreasing volume). Also, for all n, ζ_n^2 is nonnegative.

Proof

For $||e_n||\leq\gamma$, $\zeta_n^2=\zeta_{n-1}^2$ and the consequence is trivial. For $||e_n||>\gamma$, using Equation 21 in Equation 18, it is found that:

$$\mu_n = \frac{||e_n|| - \gamma + \gamma\mu_{n-1}X_n^T X_n}{\gamma X_n^T X_n}. \tag{26}$$

Multiplying the Left Hand Side (LHS) of Equation 22 with the LHS of Equation 26 and the Right Hand Side (RHS) with the RHS of Equation 26 and using Equation 24 leads to:

$$\zeta_n^2 = \zeta_{n-1}^2 - \frac{(||e_n|| - \gamma)^2}{X_n^T X_n}, \tag{27}$$

which is a decreasing function of n. Hence, it is maximized when $||e_n|| = \gamma$(i.e., $\zeta_n^2 = \zeta_{n-1}^2$) and minimized for $\sup||e_n||$. But:

$$||e_n|| = ||X_n^T(W - W_{n-1}) + v_n||.$$

Therefore;

$$||e_n|| \leq ||X_n^T(W - W_{n-1})|| + ||v_n||.$$

Because $W \in \overline{E}_n$:

$$\begin{aligned}||X_n^T(W - W_{n-1})|| &= (X_n^T(W - W_{n-1})(W - W_{n-1})^T X_n)^{1/2} \\ &\leq \zeta_{n-1}(X_n^T X_n)^{1/2}.\end{aligned}$$

Hence:

$$\begin{aligned}||e_n|| &\leq \zeta_{n-1}(X_n^T X_n)^{1/2} + ||v_n|| \\ &\leq \zeta_{n-1}(X_n^T X_n)^{1/2} + \gamma.\end{aligned}$$

Using $||e_n|| = \zeta_{n-1}(X_n^T X_n)^{1/2} + \gamma$ in the Recursion 27 leads to:

$$\zeta_n^2 = \zeta_{n-1}^2 - \frac{\left(\zeta_{n-1}(X_n^T X_n)^{1/2} + \gamma - \gamma\right)^2}{X_n^T X_n} = 0,$$

and the proof is complete. Theorem 1 expresses that with any ζ_0 and W_0 satisfying:

$$||W - W_0|| \leq \zeta_0,$$

the algorithm does not diverge. Of course, this is true when W is time invariant and inequality $||v_n|| \leq \gamma$ is valid.

The choice of a proper bounding level, γ, for noise, is critical. Over-bounding only increases the estimation error, but under-bounding is riskier as it can cause divergence. The value of ζ_n^2 at each time instant helps in discovering this situation. When ζ_n^2 goes negative, either an error in the maximum level of noise or a variation in the true parameter W has occured and proper values must be chosen for ζ_n and γ. These will be considered in other papers. From Theorem 1 it is obvious that in order to find an upper bound for $||W - W_{n-1}||$, the following is obtained:

$$\lim_{n\to\infty} ||e_n|| \in [0, \gamma], \tag{28}$$

$$||e_n|| = ||X_n^T(W - W_{n-1}) + v_n||.$$

Suppose at $n = n_0$, the sequence $\{v_n\}_n^\infty$ chooses those values in the set $[0, \gamma]$ that yields:

$$||e_n|| \leq \gamma.$$

Hence, for all $n > n_0$:

$$\zeta_n = \zeta_{n_0}, \quad W_n = W_{n_0},$$

and:

$$||e_n|| = ||X_n^T(W - W_{n_0}) + v_n|| \leq \gamma,$$

Because $||v_n|| \leq \gamma$,

$$||X_n^T(W - W_{n_0})|| \leq 2\gamma. \tag{29}$$

In addition, suppose there exist $M, \alpha_1, \alpha_2 > 0$, so that for every n_0:

$$M\alpha_1 I \leq \sum_{n=n_0}^{n_0+M} X_n X_n^T \leq M\alpha_2 I. \tag{30}$$

From Equation 29:

$$(W - W_{n_0})^T \left(\sum_{n=n_0}^{n_0+M} X_n X_n^T\right)(W - W_{n_0}) \leq 4M\gamma^2. \tag{31}$$

Therefore:

$$(W - W_{n_0})(W - W_{n_0})^T \leq 4M\gamma^2 \left(\sum_{n=n_0}^{n_0+M} X_n X_n^T\right)^{-1}.$$

Hence, for all $n \geq n_0$

$$||W - W_n||^2 \leq 4\gamma^2/\alpha_1. \tag{32}$$

SIMULATION EXAMPLES

In practice, adaptive filters are used in time-varying environments. It is, thus, important to investigate the performance of these algorithms, allowing the system-model parameters to vary with time. In case of time-varying systems, it is important to ensure that the time-varying parameters remain inside the bounding ellipsoid $\overline{E}_n$. In this section, with suitable simulation examples, the proposed algorithm is compared with the well-known DHOBE (with the rescue procedure considered in [13]). The tracking properties of these two algorithms are studied for an ARX(1, 1) model:

$$y_n = ay_{n-1} + bu_n + v_n.$$

The nominal values for the parameters are $a = -0.5$ and $b = 1$. The sequence v_n and u_n is pseudorandom noise with uniform distribution in $[-1, 1]$. For the DHOBE algorithm, $\alpha = 0.2$, $\eta_0^2 = 100$ are chosen (see previous section). Obviously, the maximum level of noise will be $\gamma = 1$. The parameters were varied as follows.

Case 1: Slow Variation in the Parameter Vector

The parameters a and b were varied by one percent for every 10 samples, starting from first sample and the output data, $\{d_n\}$, were generated for $n = 1, 2, \cdots, 1000$. The parameter estimates, i.e., the centers of the OBE, are plotted against the true parameter in Figure 1.

Case 2: Jump in the MA Parameter at $n = 500$

b was changed by 100 percent at the five-hundredth sample and a was kept constant at its nominal value at all times. The parameter estimates are plotted against the true parameter in Figure 2. To have a better comparison in this case, consider:

$$W = [a, b]^T,$$

(for simplicity, the time–dependence of W, a and b is not shown) and:

$$f_d(n) = \frac{(W - W_n)^T P_n^{-1} (W - W_n)}{\eta_n^2},$$

where P_n and η_n^2 are defined in Equations 8 and 11, respectively. When $W \in E_n$, the above fraction is less than one. Also consider:

$$f_m(n) = \frac{(W - W_n)^T (W - W_n)}{\zeta_n^2},$$

which is less than one, when $W \in \bar{E}_n$ in the proposed algorithm. These two fractions are plotted against each other in Figure 3 for Case 2. It is observed that at $n = 500$ there are great jumps in $f_d(n)$ and $f_m(n)$. However, the level of $f_m(n)$ is, obviously, less than $f_d(n)$ for most of the time.

CONCLUSION

A simple form of a recursive SM parameter estimation algorithm has been proposed and its convergence analysis is carried out. Simulation results show that the tracking performance of this algorithm is comparable to that of the modified DHOBE algorithm.

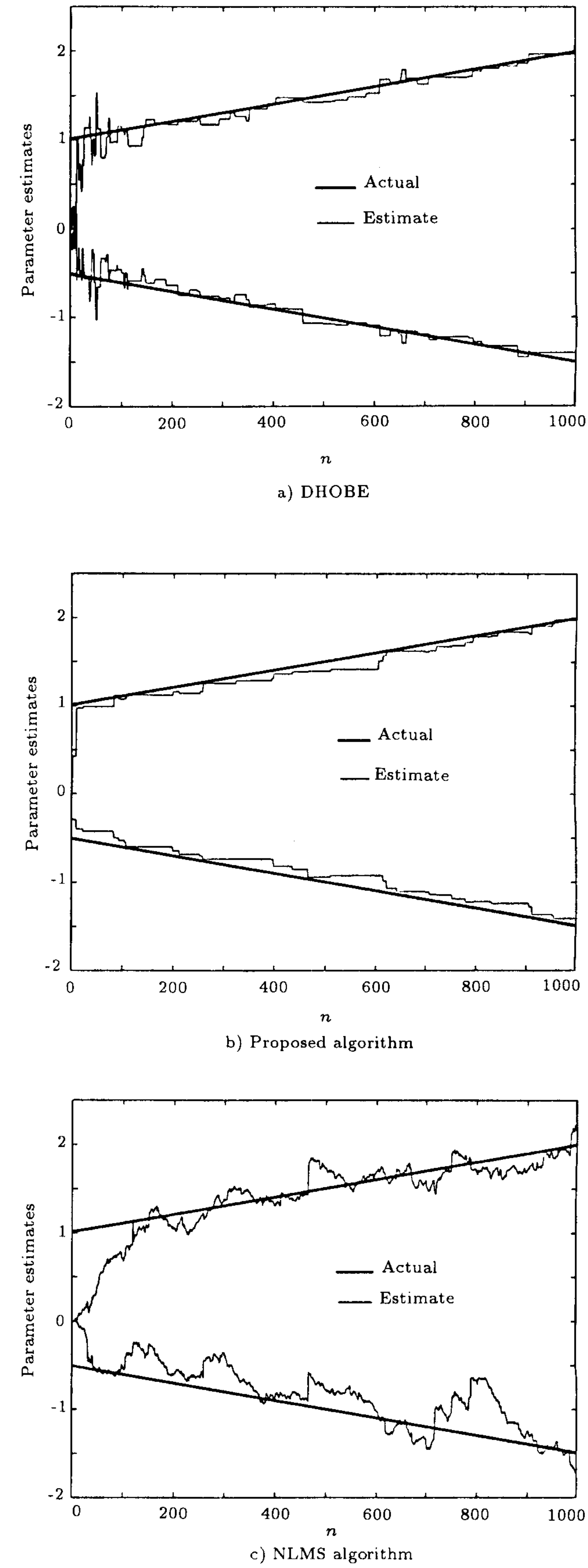

Figure 1. Parameter estimates for the case of slow variation in the true parameter from $n = 1$.

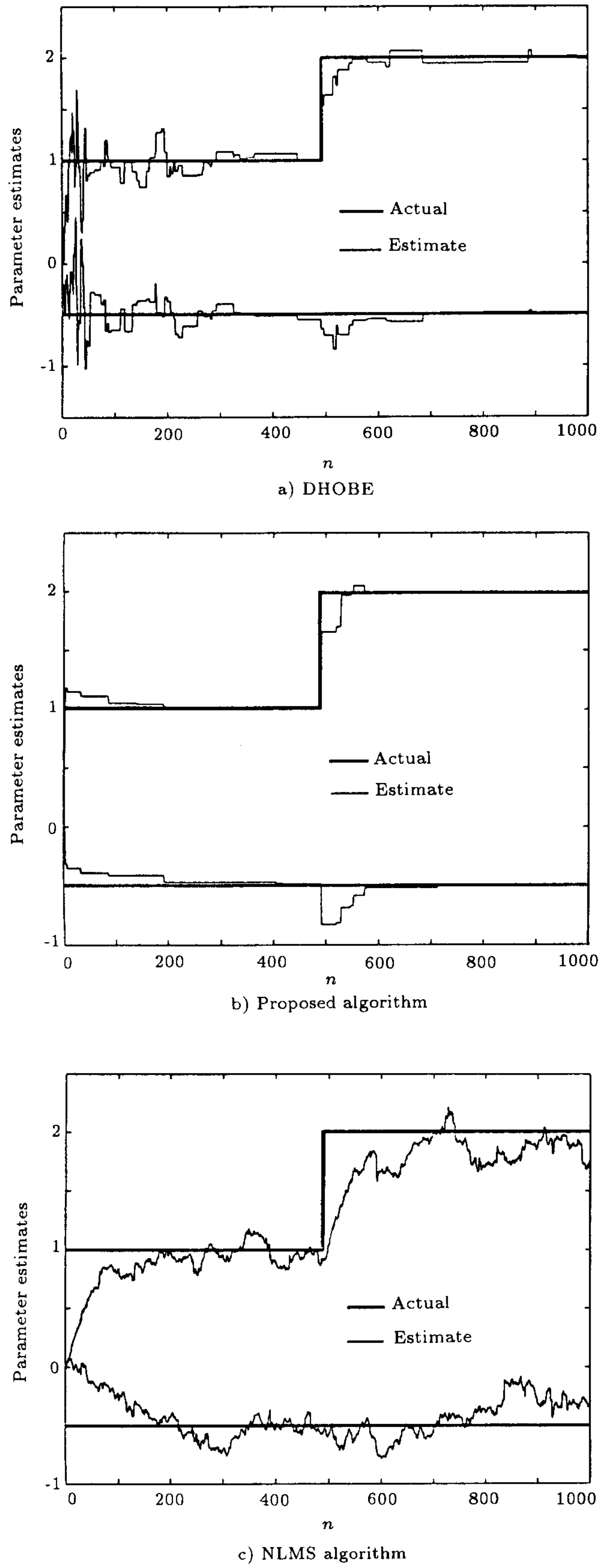

Figure 2. Parameter estimates for the case of a jump in the MA parameter at $n = 500$.

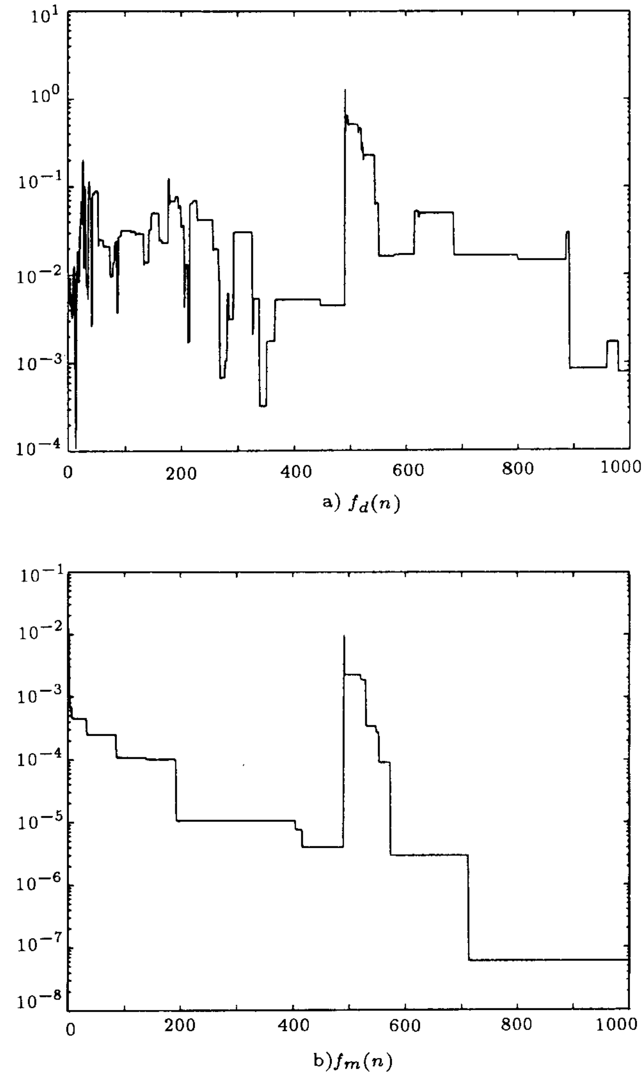

Figure 3. Values of $f_d(n)$ and $f_m(n)$ for the case of a jump in the MA parameter at $n = 500$.

REFERENCES

1. Deller, J.R. "Set membership identification in digital signal processing", *IEEE ASSP Magazine*, **6**, pp 4-20 (1990).
2. Milansese, M. and Vicino, A. "Optimal estimation theory for dynamic systems with set membership uncertainty: An overview", *Automatica*, **27**(6), pp 997-1009 (1991).
3. Kosut, R.L., Lau, M.K. and Boyd, S.P. "Set-membership identification of systems with parametric and nonparametric uncertainty", *IEEE Transaction on Automatic Control*, **AC-37**, pp 929-941 (July 1992).
4. Gollamudi, S. et al. "Set-membership adaptive equalization and an updator-shared implementation for multiple channel communications Systems", *IEEE Transactions on Signal Processing*, **SP-46**(9) (Sept. 1998).
5. Gollamudi, S. et al. "Set-membership filtering and a set-membership normalized LMS algorithm with an

adaptive step size", *IEEE Signal Processing Letters*, **5** (May 1998).

6. Fogel, E. and Huang, Y.F. "On the value of information in system identification-bounded noise case", *Automatica*, **20**, pp 229-238 (1982).
7. Dasgupta, S. and Huang, Y.F. "Asymptotically convergent modified recursive least-squares with data dependent updating and forgetting factor for systems with bounded noise", *IEEE Trans. Inform. Theory*, **IT-33**(3), pp 383-392 (May 1987).
8. Hassibi, B., Sayed, A.H. and Kailath, T. "Optimality of the LMS algorithm", *IEEE Trans. on Signal Processing*, **44**(2) (Feb. 1996).
9. Goodwin, G.C. and Sin, K.S., *Adaptive Filtering, Prediction and Control.*, Prentice-Hall, Englewood, Cliffs, NJ (1982).
10. Haykin, S., *Adaptive Filter Theory.*, 3rd Ed., Prentice-Hall, Englewood Cliffs, NJ (1996).
11. Farhang-Borojeny, B., *Adaptive Filters: Theory and Applications*, John-Wiley and Sons (1998).
12. Luenberger, D.G., *Optimization by Vector Space Methods*, Wiley, NewYork, USA (1969).
13. Rao, A.K. and Huang, Y.F. "Tracking characteristics of an OBE parameter estimation algorithm", *IEEE Trans. on Signal Processing*, **41**(3) (March 1993).