PERIODICA POLYTECHNICA SER. EL. ENG. VOL. 36, NO. 2, PP. 107–119 (1992)

A NEW ALGORITHM FOR ADAPTIVE IIR FILTERS

TH. N. MOHAMMED

Department of Measurement and Instrument Engineering
Technical University, of Budapest

Received: July 2, 1992.

Abstract

This paper introduces a Simplified Hyperstable SEquential Regression (SHSER) adaptive algorithm designed for use with Infinite Impulse Response (IIR) digital filters, which offers a reduced computational load since it avoids direct matrix inversions, fulfills the hyperstability condition under certain circumstances, and has a high convergence rate. The proposed SHSER algorithm is a combination of the Simplified Hyperstable Adaptive Recursive Filters (SHARF) algorithm and the recursive version of the SEquential Regression (SER) algorithm for IIR adaptive filters. Some important comments associated with the proposed algorithm including the SPR condition, and the step size are briefly given. Simulation results are included comparing the proposed SHSER algorithm to the SHARF algorithm and SER algorithm with respect to the convergence rate behaviour.

Keywords: adaptive IIR digital filters, hyperstable digital filters.

Introduction

The most widely used adaptive filtering algorithm is Widrow's (WIDROW, 1985) Least Mean Squared (LMS) algorithm, which is relevant to FIR filters. The related formulation involves a steepest descent approach which has been extended to recursive filters by Stearns (STEARNS, 1976). In the case of noisy conditions (i. e. the noisy gradient) the LMS/Newton method has been proposed, which is generally superior to the LMS algorithm (WIDROW, 1985). In this paper the recursive version of the LMS/Newton algorithm will be discussed in combination with the SHARF algorithm (LARIMORE, 1980). When the environment is nonstationary the inverse of the input correlation matrix needs to be calculated at each iteration in the coefficients updating equation. To decrease the computational efforts and to improve the convergence rate (see Section 1) the correlation matrix will be evaluated recursively, similarly as in the case of the SER algorithm (AHMED, 1977), and combined with the prefiltering error mechanism of the SHARF algorithm (LARIMORE, 1980), resulting in the Simplified Hyperstable SEquential Regression (SHSER) adaptive filter algorithm. In this SHSER algorithm the sequential regression formula is used to estimate sequentially the inverse of input correlation matrix and updating it at each iteration by using the inverse matrix lemma (GRAUPE, 1972) so the matrix inversions are avoided. The SHSER algorithm is an output error realization, thus it does not generate biased estimates (JOHNSON, 1979) in updating the feedback coefficients but the output error is clearly a nonlinear function of the coefficients. This type of realization is called in many references as a Pseudo linear regression (LJUNG, 1983), for which the solutions may be suboptimal (i. e. having local minima) unless a certain transfer function is Strictly Positive Real (SPR) (JOHNSON, 1979). These local minima can be avoided under certain conditions (SÖDERSTRÖM, 1982). The convergence to the global minimum is achieved by prefiltering the error (i. e. smoothed version of the output error) according to specific smoothing coefficients vector $\mathbf{C}$ (LARIMORE, 1980).

1. Formulation of the SHSER Algorithm

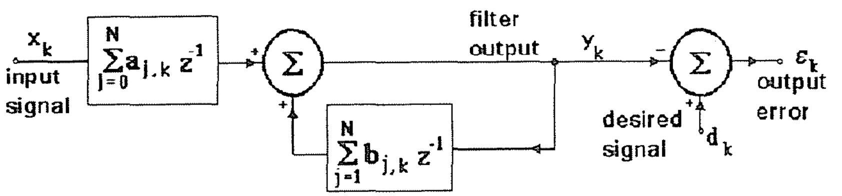

Fig. 1. The recursive filter structure

Consider the recursive digital filter structure in *Fig. 1*. The output signal of the filter can be expressed:

$$y_k = \sum_{j=1}^{N-1} b_{j,k} y_{k-j} + \sum_{j=0}^{N-1} a_{j,k} x_{k-j}, \tag{1}$$

Defining the vectors:
where, $\{\mathbf{b}_\mathbf{k}$and $\mathbf{a}_\mathbf{k}\}$ are the adjustable coefficients.

$$\mathbf{u}_\mathbf{k} = [x_k, x_{k-1}, \dots, x_{k-(N-1)}, y_{k-1}, y_{k-2}, \dots, y_{k-(N-1)}]^T$$

and

$$\mathbf{w}_\mathbf{k} = [a_{0,k}, a_{1,k}, \dots, a_{(N-1),k}, b_{1,k}, b_{2,k}, \dots, b_{(N-1),k}]^T$$

equation (1) can be rewritten:

$$y_k = \mathbf{w_k}^T \ \mathbf{u_k}. \tag{2}$$

The error signal is calculated:

$$\epsilon_k = d_k - y_k = d_k - \mathbf{w_k}^T \ \mathbf{u_k} \tag{3}$$

where d_k is the kth data point of desired output.

Consider first the LMS algorithm (WIDROW, 1985) to estimate the recursive gradient as:

$$\tilde{\nabla}_k = \partial(\epsilon_k)^2/\partial(\mathbf{w_k}) = -2 \cdot \epsilon_k \cdot (\partial(y_k)/\partial(\mathbf{w_k}))\,. \tag{4}$$

It has been shown that, because y_k is now a recursive function, it can be obtained recursively (STEARNS, 1976):

$$\Phi_{n,k} = \partial(y_k)/\partial(a_n) = y_{k-n} + \sum_{l=1}^{N-1} b_l \ \partial(y_{k-1})/\partial(a_n) = y_{k-n} + \sum_{l=1}^{N-1} \Phi_{n,k-1}, \tag{5a}$$

$$\beta_{n,k} = \partial(y_k)/\partial(b_n) = x_{k-n} + \sum_{l=1}^{N-1} b_l \ \partial(y_{k-1})/\partial(b_n) = x_{k-n} + \sum_{l=1}^{N-1} \beta_{n,k-1}. \tag{5b}$$

Thus, $\tilde{\nabla}_k$ in (4) can be expressed as:

$$\tilde{\nabla}_k = -2 \ \epsilon_k \cdot [\beta_{0,k}, \ldots, \beta_{(N-1),k}, \Phi_{1,k}, \ldots, \Phi_{(N-1),k}] \tag{6}$$

for convenience (6) can be rewritten as:

$$\tilde{\nabla}_k = -2 \ \epsilon_k \ \Omega_k, \tag{7}$$

where $\Omega_k = [\beta_{0,k}, \ldots, \beta_{(N-1),k}, \Phi_{1,k}, \ldots, \Phi_{(N-1),k}]$.

With this simple estimate of the gradient, the LMS algorithm for updating the filter coefficients is expressed as follows:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mathbf{M} \ \tilde{\nabla}_k, \tag{8}$$

where $\mathbf{M}$ is a diagonal step-size matrix (see Section 3).

If the input correlation matrix $\mathbf{R}$ is known, a more efficient step-size matrix can be composed. A widely used technique is the LMS/Newton algorithm (WIDROW, 1985) which is expressed as follows:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mathbf{M} \ \lambda_{av} \ \mathbf{R}^{-1} \ \tilde{\nabla}_k, \tag{9}$$

where λ_{av} is the mean of eigenvalues of $\mathbf{R}$ and $\mathbf{R}^{-1}$ is the inverse of the correlation matrix $\mathbf{R}$, which reduces to (8), if $\lambda_{av}\ \mathbf{R}^{-1}=\mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

AHMED (1977) has discussed the LMS/Newton algorithm (9) in terms of the SER algorithm, where $\mathbf{R}$ is expressed using a recursively calculated $\mathbf{Q}$ matrix as follows:

$$\mathbf{R}_k^{-1} = \Big((1-\psi^{k+1})/(1-\psi)\Big)\ \mathbf{Q}_k^{-1}, \tag{10}$$

where ψ is the so called forgetting factor, chosen such that the half life of exponential function is equal to the number of iterations over which the input is stationary. Typically, ψ would have the value between 0.9 and 0.99 corresponding to an effective memory between 10 samples and 100 samples, respectively (GRAUPE, 1972):

$$0.9 < \psi = 2^{-1/LSO} < 0.99,$$

where LSO is the stationarity length of the input signal.

It follows from further calculations (GRAUPE, 1972; AHMED, 1979) that $\mathbf{Q}_k^{-1}$ can be computed recursively:

$$\mathbf{Q}_{k+1}^{-1} = (1/\psi)\ \left\{\mathbf{Q}_k^{-1} - \left[\Big((\mathbf{Q}_k^{-1}\mathbf{u}_k)(\mathbf{Q}_k^{-1}\mathbf{u}_k)^T\Big)\ /\ \Big(\psi + \mathbf{u}_k^T(\mathbf{Q}_k^{-1}\mathbf{u}_k)\Big)\right]\right\}. \tag{11}$$

Note that the vector $\mathbf{s}_k = \mathbf{Q}_k^{-1}\cdot\mathbf{u}_k$ is used three times in (11), and would be computed first in the algorithm and also the denominator term in (11) is scalar therefore would be computed separately (see *Table 1*).

In (LARIMORE, 1980) the recursive calculation of the derivative in (6) has been avoided by approximation Ω_k with u_k, and a smoothed version of ϵ_k is used in calculating the gradient:

$$\gamma_k = \sum_{n=0}^{P-1} c_n \epsilon_{k-n}, \tag{12}$$

where the vector $\mathbf{C}$ has been defined as: $\mathbf{C} = [c_0, c_1, \ldots, c_{p-1}]$ with c_0 equal to one, vector $\mathbf{C}$ is chosen such that the SPR condition is fulfilled (see Section 2).

In our experience it is worth combine into the SER and SHARF algorithms. This combination means that (9) will be expressed as:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mathbf{M}\lambda_{av}\Big((1-\psi^{k+1})/(1-\psi)\Big)\mathbf{Q}_k^{-1}\tilde{\nabla}_k, \tag{13}$$

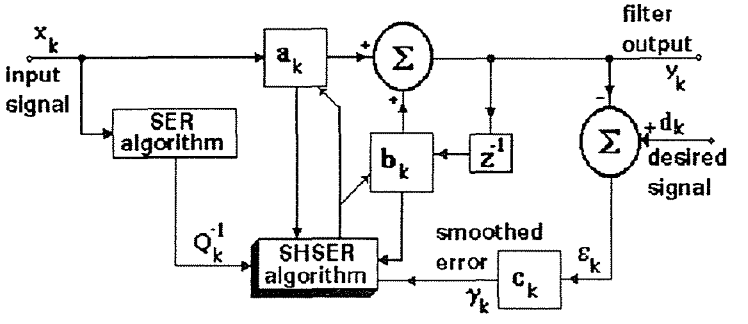

Fig. 2. The structure of the SHSER algorithm

where $\tilde{\nabla}_k$ in (13) takes the form:

$$\tilde{\nabla}_k = -2 \ \gamma_k \ \mathbf{u}_k. \tag{14}$$

We refer to the equations (11) and (13) as the SHSER algorithm (see *Table 1*), where, $\mathbf{Q}_k^{-1}$ is computed recursively in (11), the initial value $\mathbf{Q}_0^{-1}$ is necessary to be observed before determining $\mathbf{Q}_k^{-1}$ in (11), $\mathbf{Q}_0^{-1}$ typically equal to the identity matrix $\mathbf{I}$, λ_{av} in (13) can be estimated from actual data, and kept fixed during the adaptation process with a value typically between 0.05 and .1 when signal statistics are unknown. The structure of the SHSER algorithm is shown in *Fig. 2*.

2. Hyperstability and SPR Condition

The Hyperstable Adaptive Recursive Filter algorithm (HARF) has been introduced by Johnson (JOHNSON, 1979). It was shown by Larimore (LARIMORE, 1980), that the HARF algorithm may not converge to a minimum unless the transfer function is Strictly Positive Real (SPR). Treichler (TREICHLER, 1978) has introduced the SHARF algorithm, which is hyperstable only for slow rate of adaptation, and the output error can be filtered by a simple moving average technique.

The SHARF algorithm has been applied in the proposed SHSER algorithm. The SHARF algorithm requires that the transfer function of a specific linear time-invariant system:

$$\mathbf{G}(z) = \mathbf{C}(z)/\mathbf{B}(z) = \left(1 + \sum_{l=1}^{P-1} c_l \ z^{-1}\right) / \left(1 - \sum_{l=1}^{N-1} b_l \ z^{-1}\right) \tag{15}$$

fulfills the SPR condition, i. e. $Re\{\mathbf{G}(z)\} > 0$ for all $|z| = 1$. Note that (15) applies only to points on the unit circle. In the SHSER algorithm in modelling applications the denominator polynomial of $\mathbf{G}(z)$ is $\mathbf{B}(z)$. The components of $\mathbf{C}$ are selected to meet the SPR condition. This corresponds to the filtering of ϵ_k in (12), however, the choice of $\mathbf{C}$ given by (15) already assumes a great deal of a priori knowledge of model parameters. Unfortunately, this is not always possible to achieve in practice since the coefficients of the system are unknown in applications such as modelling configuration. An alternate approach has been applied on the SHSER algorithm, which allows recursive estimation of the $\mathbf{C}$ coefficients (PARIKH, 1979). This problem is still a subject of research in IIR adaptive filter theory (WIDROW, 1985).

3. Step-size Matrix

In non-recursive adaptive filter the LMS algorithm is updated the coefficients as (WIDROW, 1985):

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu \ \tilde{\nabla}_k, \tag{16}$$

which is the same as in (8), but replace $\mathbf{u}_k$ by $\mathbf{x}_k$, and $\mathbf{M}$ by μ (see Section 1). The parameter μ in (16) is a step size constant that governs stability and the rate of convergence. In the recursive adaptive filter of (8), for updating the coefficients, the constant μ has been replaced by the following diagonal matrix

$$\mathbf{M} = \mathrm{diag}[\mu, \mu, \cdots, \mu, v_1, v_2, \cdots, v_{N-1}]. \tag{17}$$

Thus, due to the nonquadratic error (WIDROW, 1985) surface, a common convergence parameter μ is obtained for each numerator coefficient and different convergence factors are obtained for the denominator coefficients (see *Table 1*). The chosen value of μ is the same as in the non-recursive case (16) (WIDROW, 1985), and from practical experience for choosing $v_1 \cdots v_{N-1}$ in applications such as system identification and in inverse filtering (i. e. equalizing), the following hints are given: the ratios between v_i(i.e. $v_1/v_2 \quad v_2/v_3 \quad v_{N-2}/v_{N-1}$) are equal to the ratio of the denominator plant coefficients (i.e. $b_2/b_1 \cdots b_N/b_{N-1}$), respectively. This choice improves the speed of the convergence but, unfortunately, there is no general way yet to prove this.

Since, the SER algorithm (AHMED, 1979) is applied in the proposed algorithm to estimate the correlation matrix $\mathbf{R}$ recursively equation (11), this allows us to increase the step-size matrix $\mathbf{M}$ in equation (13) (PARIKH, 1978). In our experience, to increase the convergence rate the large positive value of $\mathbf{M}$ is selected and slightly reduced through the adaptation process by dividing the matrix $\mathbf{M}$ by a factor $\alpha > 1$, so $\mathbf{M}$ is constrained to be between the initial step-size matrix $\mathbf{M}_{ini}$ and the final step-size matrix $\mathbf{M}_f$, the filter in this case remains stable under the condition that $\mathbf{R}$ always be positive definite. This experience must have justification, the convergency of the SHSER algorithm is under investigation.

4. Simulation Results

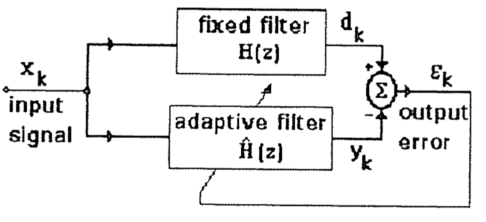

Fig. 3. The system identification structure

A simple system identification problem shown in *Fig. 3* is considered, where the fixed filter is denoted by $H(z)$ and the adaptive filter is denoted by $\widehat{H}(z)$. The adaptive algorithm applied on this configuration is one of the following: 1) SER algorithm; 2) SHARF algorithm; 3) SHSER algorithm. The smoothing error coefficients $\mathbf{C}$, the step-size matrix $\mathbf{M}$, and the matrix reduction α are set due to Sections 2 and 3. In examples 1 and 2 the comparison is made between the SHARF and SHSER algorithms and also between SER and SHSER algorithms in terms of parameter convergence, and Mean Square Error (MSE) convergence. In example 3 the modelling system is driven by the special type of nonstationary signal which is a single sinusoid with sinusoidally varying frequency, for which the SHARF and SER algorithms have failed to converge either to a local minimum or the global minimum while the SHSER does.

Example 1

The identification system shown in *Fig. 3* is considered consisting of the fixed filter:

$$H(z) = 1/(1 - 1.2z^{-1} + 0.6z^{-2})$$

and the adaptive IIR filter:

$$\widehat{H}(z) = a_0/(1 - b_1 z^{-1} - b_2 z^{-2}),$$

where $\widehat{H}(z)$ is used to minimize the MSE. The input signal is nonstationary sinusoidal, its magnitude and phase change randomly. The forgetting factor ψ has been set according to Section 1 as 0.95. *Fig. 4a* shows the comparison of coefficients and MSE convergence between the SER algorithm (PARIKH, 1978) and the proposed algorithm SHSER, while *Fig. 4b* shows the above comparison but between the SHARF algorithm (JOHNSON, 1979) and the proposed algorithm SHSER.

Example 2

The problem in (PARIKH, 1978) is considered where the fixed filter as:

$$H(z) = \frac{.34444 - .34444z^{-2}}{1 + .55697z^{-1} + .31121z^{-2}}$$

and the adaptive IIR filter has the form of

$$\widehat{H}(z) = \frac{a_0 - a_2 z^{-2}}{1 - b_1 z^{-1} - b_2 z^{-2}}.$$

The system is driven by white noise, ψ is set to 0.93. *Fig. 5a* shows the same comparison in Example 1 between the SER algorithm and SHSER algorithm, while *Fig. 5b* shows the comparison between the SHARF algorithm and the SHSER algorithm.

Example 3

We consider again the identification system in Example 1 with the same $H(z)$ and $\widehat{H}(z)$ but we apply in this example a special nonstationary input signal which is a single sinusoid with sinusoidally varying frequency as :

$$x(t) = \sin(2\pi(0.25 + 0.02\sin(0.01t))t)$$

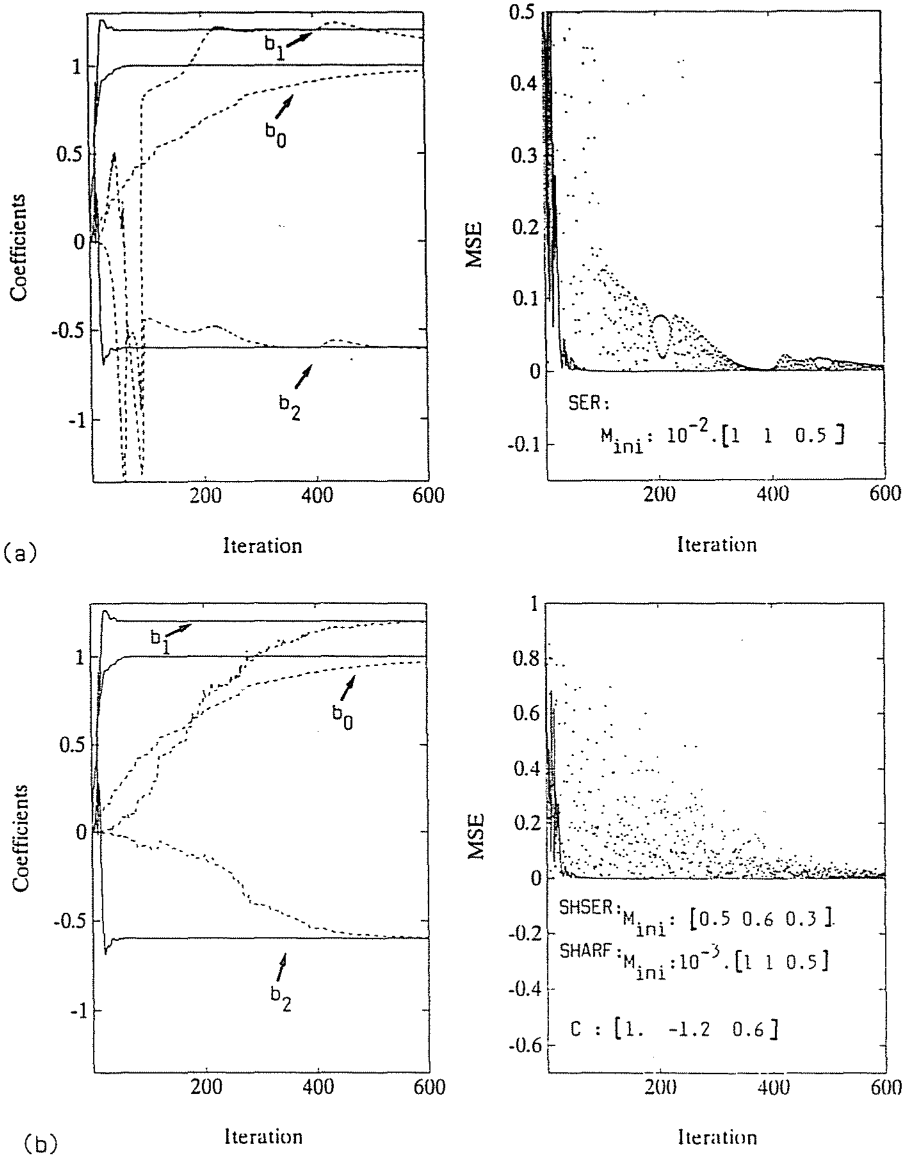

Fig. 4. (a) Comparison of the convergence between SER and SHSER algorithms
(b) The same comparison between SHARF and SHSER algorithms
continuous lines:SHSER
dotted lines:SER

and ψ is set to 0.99. **C** contains the same values as in Example 1. The convergence is failed when we apply this signal to the SHARF and SER

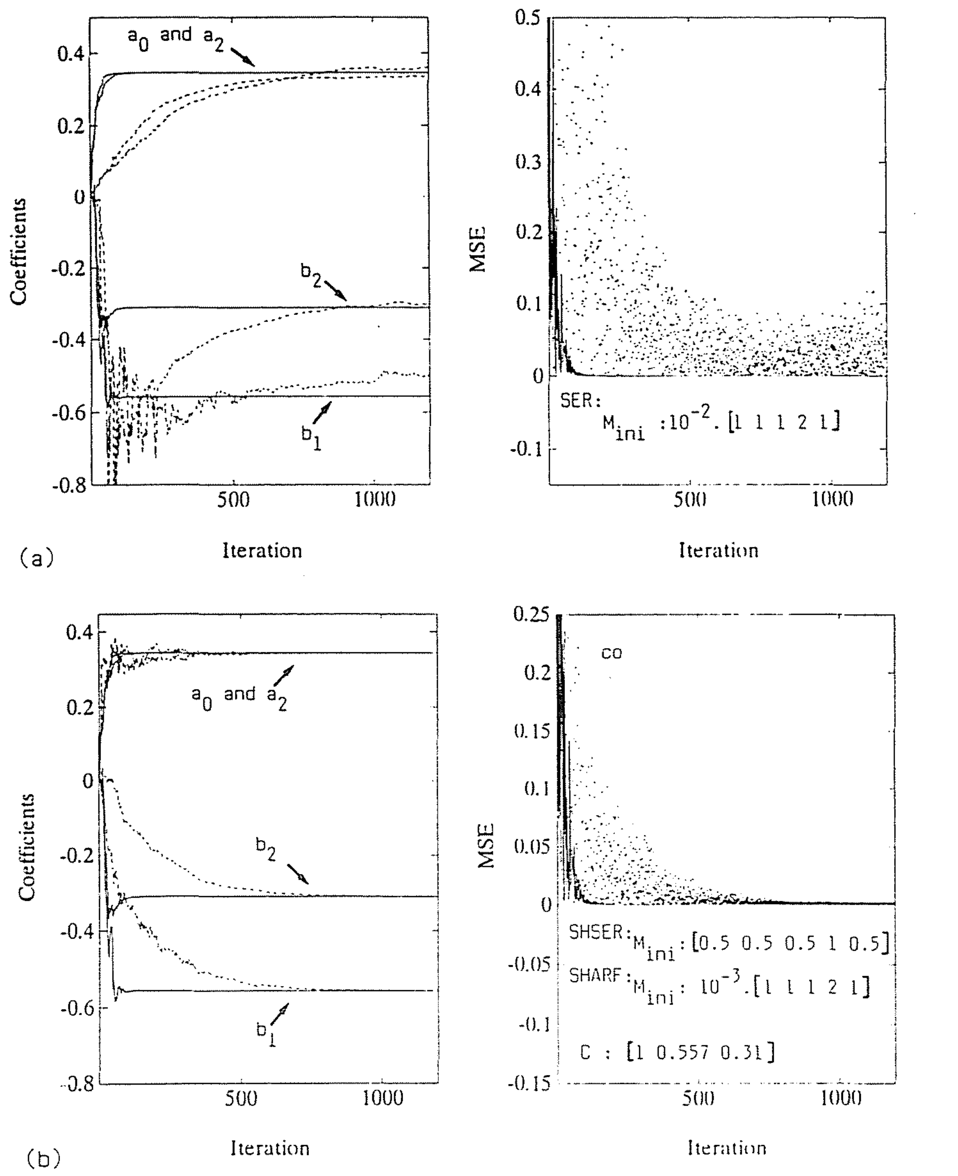

Fig. 5. (a) Comparison of the convergence between SER and SHSER algorithms
(b) The same comparison between SHARF and SHSER algorithms
continuous lines:SHSER
dotted lines:SER

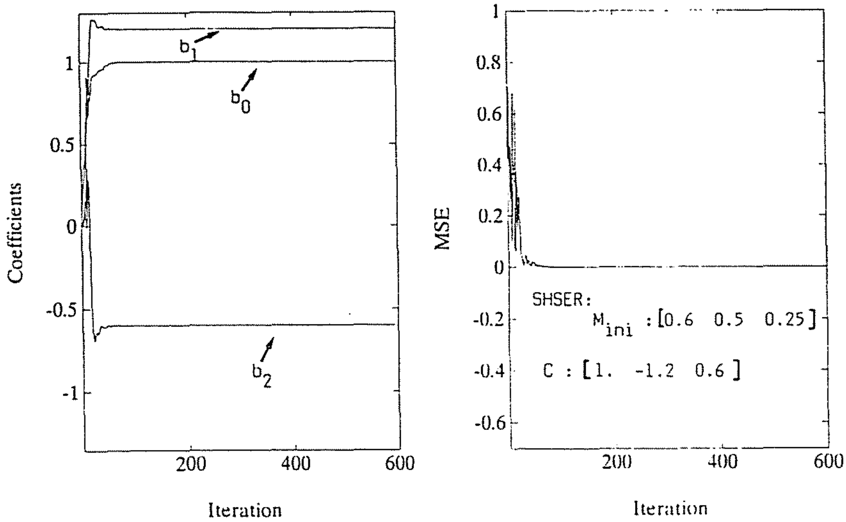

Fig. 6. Convergence of the coefficients and MSE in Example 3 continuous lines:SHSER

algorithms, while the proposed SHSER algorithm well converged. *Fig. 6* shows the convergence of coefficients of the adaptive filter and the convergence of the MSE by the SHSER algorithm.

It is evident from these simulations that the SHSER algorithm yielded the superior rate of convergence under the conditions previously described. Maintaining stability during the adaptive process was found to be the primary limiting factor for both SER and SHARF if the step-size matrix **M** is not small enough (WIDROW, 1985; LARIMORE, 1980). Further experimental results obtained, related to this work indicate that the SHSER algorithm is promising candidate for adaptive filtering in a nonstationary environment.

Table 1
The SHSER algorithm

Initialization	
$\mathbf{Q}_0^{-1}$	$= \delta \cdot \mathbf{I}$
$a_{j,0}$	$= b_{j,0} \quad = 0$
$x_{k-(N-1)}$	$= y_{k-(N-1)} \quad = 0$
$\mathbf{w}_0$	= starting weight vector=0
Vector and Scaler Definition	
$\mathbf{u}_k$	$= [x_k, x_{k-1}, \cdots, x_{k-(N-1)}, y_{k-1}, y_{k-2}, \cdots, y_{k-(N-1)}]^T$
$\mathbf{w}_k$	$= [a_{0,k}, a_{1,k}, \ldots, a_{(N-1),k}, b_{1,k}, b_{2,k}, \ldots, b_{(N-1),k}]^T$
$\mathbf{M}$	$= \text{diag}\,[\mu, \cdots, \mu, v_1, v_2, \cdots, v_{N-1}]$
ψ	$= 0.9 < 2^{-1/LSO} < 0.99$
LSO	= Length of stationary input signal
d_k	= kth data point of desired output
For Iteration $K \geq 1$	
y_k	$= \mathbf{w}_k^T \; \mathbf{u}_k$
$\mathbf{s}_k$	$= \mathbf{Q}_k^{-1} \; \mathbf{u}_k$
ζ	$= \psi + \mathbf{u}_k^T \; \mathbf{s}_k$
$\mathbf{Q}_{k+1}^{-1}$	$= (1/\psi)\; (\mathbf{Q}_k^{-1} - (\mathbf{s}_k \; \mathbf{s}_k^T/\zeta))$
ϵ_k	$= d_k - y_k$
γ_k	$= \sum_{n=0}^{P-1} c_n \; \epsilon_{k-n}$
$\tilde{\nabla}_k$	$= -2 \; \gamma_k \; \mathbf{u}_k$
$\mathbf{w}_{k+1}$	$= \mathbf{w}_k - \mathbf{M} \; \lambda_{av}[(1-\psi^{k+1})/(1-\psi)] \; \mathbf{Q}_k^{-1} \; \tilde{\nabla}_k$
$\mathbf{M}$	$= \mathbf{M}/\alpha$, where $\alpha > 1$ and $\mathbf{M}_f < \mathbf{M} < \mathbf{M}_{ini}$

Conclusion

A simplified hyperstable adaptive filter algorithm based on the sequential regression formula is proposed. The algorithm has fast convergence rate, computationally simple, hyperstable under certain circumstances, and suitable for filtering applications. In this paper we have some propositions in choosing the smoothing error coefficients and the step-size matrix which improves the rate of convergence under the stability condition.

According to the examples, substantial improvement in convergence has been achieved in comparison with SER algorithm and with SHARF algorithm. The experimental results presented in this paper indicate that the SHSER algorithm under nonstationary environment is promising.

Acknowledgement

The author would like to thank Prof. Gábor Péceli and Dr. Ferenc Nagy for their helpful comments to improve this paper.

References

AHMED, N. — SOLDAN, D. — HUMMELS, D. — PARIKH, D. (1977): Sequential Regression Considerations of Adaptive Filtering, *Electronics Lett.*, Vol. 13, No. 15, pp. 446–448.

WIDROW, B. — STEARNS, S. (1985): Adaptive Signal Processing, Prentice-Hall, Englewood Cliffs, N. J.

STEARNS, S. — ELLIOTT, G. — AHMED, N. (1976): Adaptive Recursive Filtering, *Proc. of the 10th Asiolomar Conf. on Circuits, Syst., Comput.*, CA, pp. 5–10.

LARIMORE, M. — TREICHLER, J. — JOHNSON, C. (1980): SHARF: an Algorithm for Adapting IIR Digital Filters, *IEEE Trans. on ASSP*, Vol. ASSP–28, pp. 428–440.

GRAUPE, D. (1972): Identification of System, Van Nostrand, Princeton, NJ.

TREICHLER, J. — LARIMORE, M. — JOHNSON, C. (1978): Simple Adaptive IIR Filtering, *Proc. of ICASSP* 1978, pp. 118–122.

AHMED, N. — HUMMELS, D. — MICHAEL, L. — SOLDAN, D. (1979): A Short-term Sequential Regression Algorithm, *IEEE Trans. on ASSP*, Vol. ASSP–27, No. 5, pp. 453–457.

JOHNSON, C. — TREICHLER, J. — LARIMORE, M. (1979): Remarks, on The Use of SHARF as an Output Error Identifier, *Proc. 17th IEEE Conference on Decision and Control*, San Diego, CA, pp. 1094–1095.

TANG, K. — ROHRS, C. (1987): The Use of Large Adaptation Gains to Remove the SPR Condition form Recursive Adaptive Filter Algorithms, *IEEE Proc. of ICASSP*, Dallas, Texas, pp. 129–132.

PARIKH, D. — AHMED, N. (1978): A Sequential Regression Algorithm for Recursive Filters, *Electronics Lett.*, Vol. 14, No. 9, pp. 266–268.

LJUNG, L. — SÖDERSTRÖM, T. (1983): Theory and Practice of Recursive Identification, MIT press, Cambridge

SÖDERSTRÖM, T. — STOCIA, P. (1982): Some Properties of the Output Error Method, *Automatica*, Vol. 18, No. 1, pp. 93–99.

PARIKH, D. — SINHA, S. C. — AHMED, N. (1979): On a Modifications of the SHARF Algorithm, *Proc. of the 22nd Midwest Symposium on Circuits and Systems*

Address:

Thaier N. MOHAMMED
Department of Measurement and Instrument Engineering
Technical University
H–1521 Budapest, Hungary