GENERALIZED GRADIENT ADAPTIVE STEP SIZES FOR STOCHASTIC GRADIENT ADAPTIVE FILTERS

S. C. Douglas
Department of Electrical Engineering
University of Utah
Salt Lake City, UT 84112 USA

ABSTRACT

In this paper, we derive new adaptive step size algorithms for two general classes of modified stochastic gradient adaptive filters that include the sign-error, sign-data, sign-sign, and normalized gradient adaptive filters as specific cases. These computationally-simple parameter adjustment algorithms are based on stochastic gradient approximations of steepest descent procedures for the unknown parameters. Analyses of the algorithms show that the stationary points of the steepest descent procedures yield the optimum step size values at each time instant as obtained from statistical analyses of the adaptive filter updates. Simulations verify the theoretical results and indicate that near-optimal tracking performance can be obtained from each of the adaptive step size algorithms without any knowledge of the rate of change of the unknown system.

1. INTRODUCTION

Least-mean-square (LMS) adaptive finite-impulse-response (FIR) filters have proven to be extremely useful in a number of signal processing tasks. However, for many types of real-world input data, LMS adaptive filters suffer from a slower rate of convergence for a given steady-state mean-square error as compared to the behavior of recursive least-squares adaptive filters. Consequently, several computationally-simple methods for improving the convergence properties of the LMS adaptive filter have been proposed [1, 2, 3, 4]. In general, these methods specify a procedure for adjusting the algorithm step size to obtain fast convergence when the error in the adaptive filter coefficients is large and to obtain a small mean-square error when the error in the adaptive filter coefficients is small. Most of these approaches seem reasonable from a heuristic standpoint but satisfy no specific optimality criteria. A notable exception are gradient adaptive step size algorithms that adjust the step size according to a stochastic gradient descent procedure on the squared output error [2, 3, 4]. It is shown in [3] that these algorithms can provide near-optimal tracking performance under a standard Markov model for the nonstationary unknown system. Interestingly, the algorithm was first developed to improve the convergence properties of the backpropagation algorithm for multilayer feedforward neural networks, where it has been termed the "delta-delta rule" [5].

In some situations, it is desirable to modify the LMS algorithm update, either to simplify its implementation in hardware [6, 7] or to improve its robustness and performance in the presence of non-Gaussian or time-varying signal statistics [8, 9, 10, 11]. Adaptive step size procedures for these modified algorithms remain largely unexplored.

In this paper, we present new adaptive step size procedures for the two families of modified stochastic gradient adaptive algorithms given by

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu_k f(e_k) G(\mathbf{X}_k) \quad (1)$$

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \frac{f(e_k) G(\mathbf{X}_k)}{\beta_k + ||\mathbf{X}_k||^q}, \quad (2)$$

where $\mathbf{W}_k = [w_{1,k} \cdots w_{L,k}]^T$ are the L coefficients of the adaptive filter at time k, $\mathbf{X}_k = [x_k \cdots x_{k-L+1}]^T$ are the data elements in current filter memory, $e_k = d_k - \mathbf{W}_k^T \mathbf{X}_k$ is the adaptive filter error, d_k is the desired response signal. μ_k and β_k are adaptive parameters, $||\cdot||$ denotes a suitable vector norm, $G(\cdot)$ is a vector-valued nonlinearity, and $f(\cdot)$ is an odd-symmetric scalar nonlinearity. By specifying the nonlinearities $f(\cdot)$ and $G(\cdot)$, the vector norm, and the value of q, equations (1) and (2) can be used to describe a number of useful adaptive filter algorithms, including the sign-error, sign-data, sign-sign, and other quantized state adaptive algorithms [6]; the normalized LMS (NLMS), sign-data NLMS, sign-error NLMS, and other normalized adaptive algorithms [9, 10, 11, 12]; and Newton-type algorithms. For the algorithm families in (1) and (2), we derive simple data-adaptive procedures for adjusting the step size parameter μ_k and parameter β_k, respectively, to achieve both fast tracking in nonstationary environments and a low excess mean-square error (MSE) in near-stationary environments. Our derivation is based on the following stochastic gradient descent procedure, as given here for μ_k in (1) [2, 3] (the expression for β_k is analogous):

$$\mu_k = \mu_{k-1} - \rho \frac{\partial \phi(e_k)}{\partial \mu_{k-1}}, \quad (3)$$

where $\phi(\cdot)$ denotes the relevant cost function to be minimized and ρ is a convergence parameter. These step size algorithms are found to be extremely simple to implement. Moreover, by suitable choice of the cost function $\phi(\cdot)$, the adaptive step size algorithm is robust to large errors in the desired response signal, a situation for which the algorithm in [2, 3] is known to perform poorly [4].

2. DERIVATION

In this section, we derive the step size update for μ_k explicitly; the derivation for the update for β_k is similar. Consider the step size update given by (3). Evaluating the partial derivative in (3) using the chain rule gives

$$\frac{\partial \phi(e_k)}{\partial \mu_{k-1}} = \frac{\partial \phi(e_k)}{\partial e_k} \frac{\partial e_k}{\partial \mu_{k-1}}. \quad (4)$$

0-7803-2431-5/95 $4.00 © 1995 IEEE

Following the notation of [8], we define $f(e_k) = \partial\phi(e_k)/\partial e_k$. We now substitute $\mathbf{W}_k = \mathbf{W}_{k-1} + \mu_{k-1} f(e_{k-1}) G(\mathbf{X}_{k-1})$ into the expression for e_k, which gives

$$e_k = d_k - \mathbf{W}_{k-1}^T \mathbf{X}_k - \mu_{k-1} f(e_{k-1}) G(\mathbf{X}_{k-1})^T \mathbf{X}_k. \quad (5)$$

Thus, the second partial derivative in (4) is

$$\frac{\partial e_k}{\partial \mu_{k-1}} = -f(e_{k-1}) G(\mathbf{X}_{k-1})^T \mathbf{X}_k. \quad (6)$$

Combining (3), (4), and (6) yields the step size update as

$$\mu_k = \mu_{k-1} + \rho f(e_k) f(e_{k-1}) G(\mathbf{X}_{k-1})^T \mathbf{X}_k, \quad (7)$$

where ρ is a convergence parameter for the step size.

Remark 1: As is the case for the LMS adaptive filter step size update [3], determining the range of ρ to guarantee stability of the coupled system of equations in (1) and (7), even when the statistics of the system are completely known, is a difficult task. However, it is possible to limit the range of values that μ_k can have to a set of values that are known to guarantee convergence of the algorithm. For example, the mean-square error of the sign-error adaptive filter with fixed step size will remain bounded for a wide range of statistical conditions if $0 < \mu < \infty$ [7]. Thus, if μ_k produced by the step size update is negative, we conjecture that the sign-error adaptive algorithm with adaptive step size can be stabilized by setting the step size equal to a value $\epsilon > 0$ whenever the update in (7) causes $\mu_k < 0$. Our experience has shown that the upper bound check required for the LMS adaptive filter with gradient adaptive step size is not necessary for the sign-error adaptive filter version. This issue is currently under investigation.

Remark 2: As in the case of the adaptive step size algorithm for the LMS adaptive filter, there exist input signals that will cause $G(\mathbf{X}_{k-1})^T \mathbf{X}_k$ to be zero for all k [3]. In such cases, the step size will never be updated. However, for most real-world input signals, the probability of such a situation occurring is extremely small.

Remark 3: In some situations, it may be beneficial for implementation reasons to approximate the adaptive step size update in (7) as

$$\mu_k = \mu_{k-1} + \rho f(e_k) f(e_{k-1}) G(\mathbf{X}_{k-1})^T G(\mathbf{X}_k). \quad (8)$$

For example, in the sign-sign algorithm for which $f(e_k) = \mathrm{sgn}(e_k)$ and $G(\mathbf{X}_k) = \mathrm{sgn}(\mathbf{X}_k)$, the update in (8) can be implemented using a simple up-down counter, as the value of $\mathrm{sgn}(\mathbf{X}_{k-1})^T \mathrm{sgn}(\mathbf{X}_k)$ is always an integer.

The derivation of the update for β_k in (2) is similar to the above derivation for μ_k. The resulting update is

$$\beta_k = \beta_{k-1} - \rho \frac{f(e_k) f(e_{k-1}) G(\mathbf{X}_{k-1})^T \mathbf{X}_k}{(\beta_{k-1} + \|\mathbf{X}_{k-1}\|^q)^2}. \quad (9)$$

For coefficient stability, the valid range of β_k depends on the algorithm form. For the NLMS algorithm, stability can be guaranteed for $\beta_k \geq 0$.

3. IMPLEMENTATION

We now consider the implementation costs of the parameter updates in (7), (8), and (9) in the context of FIR filtering. Because of the shift-input nature of the input data vector, these updates can be easily calculated in most cases using only a few mathematical operations. For example, consider the case in which $G(\mathbf{X}_k) = [g(x_k) \cdots g(x_{k-L+1})]^T$, where $g(\cdot)$ is a scalar nonlinearity. Then, the scalar quantity $b_k = G(\mathbf{X}_{k-1})^T \mathbf{X}_k$ can be calculated from its previous value as

$$b_k = b_{k-1} + x_k g(x_{k-1}) - x_{k-L} g(x_{k-L-1}). \quad (10)$$

Equation (10) is critically-stable but can be reinitialized every L iterations without any additional multiplies. The update for $G(\mathbf{X}_{k-1})^T G(\mathbf{X}_k)$ used in (8) is similar. For the update of β_k, consider the case where $\|\cdot\|$ denotes the L_q norm. Then, it is natural to define $a_k = \beta_k + c_k$ and update β_k and c_k according to

$$\beta_k = \beta_{k-1} - \rho \frac{f(e_k) f(e_{k-1}) b_k}{a_{k-1}^2} \quad (11)$$

$$c_k = c_{k-1} + |x_k|^q - |x_{k-L}|^q. \quad (12)$$

Since the quantity $1/a_k$ is used in the filter coefficient update, no additional divides are needed to implement the adaptive parameter update.

Table 1 lists several adaptive algorithms and compares the number of multiplies, adds, and divides necessary to implement both fixed and adaptive parameter versions. As can be seen, the additional operations required per iteration to implement the adaptive parameter versions are few and do not depend on the filter length L, making them amenable to real-time implementation.

Algorithm	*Fixed Param.*		*Adaptive Param.*	
	# Mults.	*# Adds*	*# Mults.*	*# Adds*
sign-error	$2L$	$2L$	$2L+3$	$2L+3$
sign-data	$L+1$	$2L$	$L+4$	$2L+3$
sign-sign	L	$2L$	$L+1$	$2L+1$
NLMS	$2L+2$ ($1\div$)	$2L+2$	$2L+9$ ($1\div$)	$2L+6$
sign-error NLMS	$2L$ ($1\div$)	$2L+2$	$2L+6$ ($1\div$)	$2L+6$
sign-data NLMS	L ($1\div$)	$2L+2$	$L+6$ ($1\div$)	$2L+6$

Table 1: Computational complexity of adaptive algorithms.

4. ANALYSIS

The adaptive step size algorithms described above are based on an approximate gradient descent procedure. What optimality properties, if any, do these algorithms possess? To answer this question, we first determine the optimum step size sequences for these algorithms using the statistical analyses presented in [8, 9, 13]. We also assume that the desired response signal is generated from a nonstationary model such that

$$d_k = \mathbf{W}_{opt,k}^T \mathbf{X}_k + n_k \quad (13)$$

$$\mathbf{W}_{opt,k+1} = \mathbf{W}_{opt,k} + \mathbf{M}_k, \quad (14)$$

where n_k is an i.i.d. noise sequence with zero mean and variance σ_n^2 and $\mathbf{M}_k$ is an i.i.d vector sequence with zero mean and covariance matrix $\sigma_m^2 I$.

Due to space limitations, we consider general error criteria adaptive algorithms of the form [8]

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu_{opt,k} f(e_k) \mathbf{X}_k, \quad (15)$$

where $f(\cdot)$ is an odd-symmetric nonlinearity and $\{\mu_{opt,k}\}$ is an optimum step size sequence to be determined. It can be

shown (cf. [8], eq. (48)) that the excess MSE at time $k+1$ is given by

$$\xi_{ex,k+1} = \xi_{ex,k} - 2\mu_{opt,k}E[f'(e_k)]E[\mathbf{V}_k^T R^2 \mathbf{V}_k] + \mu_{opt,k}^2 E[f^2(e_k)\mathbf{X}_k^T R \mathbf{X}_k] + \sigma_m^2 \mathrm{tr}[R]. \quad (16)$$

where $\mathbf{V}_k = \mathbf{W}_k - \mathbf{W}_{opt,k}$ and $R = E[\mathbf{X}_k\mathbf{X}_k^T]$.

To derive the optimum step size $\mu_{opt,k}$, we can follow the method of derivation presented in [9]. From (16), we can view the quantity $\xi_{ex,k+1}$ as a quadratic function of the step size $\mu_{opt,k}$. To choose an optimum step size to minimize this quantity, we take $\partial\xi_{ex,k+1}/\partial\mu_{opt,k}$ and set the result equal to zero. This operation gives $\mu_{opt,k}$ as

$$\mu_{opt,k} = \frac{E[f'(e_k)]E[\mathbf{V}_k^T R^2 \mathbf{V}_k]}{E[f^2(e_k)\mathbf{X}_k^T R \mathbf{X}_k]}. \quad (17)$$

Now, consider the adaptive step size procedure given in (7) with $G(\mathbf{X}_{k-1}) = \mathbf{X}_{k-1}$. It can be shown that e_k can be written as

$$e_k = \widehat{d}_k - \mu_{k-1}\widehat{x}_k \quad (18)$$

$$\widehat{d}_k = n_k - \mathbf{V}_{k-1}^T\mathbf{X}_k - \mathbf{M}_{k-1}^T\mathbf{X}_k \quad (19)$$

$$\widehat{x}_k = f(e_{k-1})\mathbf{X}_{k-1}^T\mathbf{X}_k. \quad (20)$$

Substituting (18) and (20) into (7), we get

$$\mu_k = \mu_{k-1} + \rho f(\widehat{d}_k - \mu_{k-1}\widehat{x}_k)\widehat{x}_k. \quad (21)$$

Examining equation (21), we see that it is in an explicit stochastic gradient descent form, in which μ_{k-1} is the parameter to be adjusted so that $\mu_{k-1}\widehat{x}_k$ follows $\widehat{d}_k$.

As in the LMS algorithm, we can relate the update in (21) to a steepest descent procedure for the parameter $\mu_{SD,k}$, that is simply the expectation of (21) over the probability densities of $\mathbf{V}_{k-1}$, $\mathbf{X}_k$, $\mathbf{X}_{k-1}$, $\mathbf{M}_{k-1}$, and n_k. This procedure is

$$\mu_{SD,k} = \mu_{SD,k-1} + \rho E[f(\widehat{d}_k - \mu_{SD,k-1}\widehat{x}_k)\widehat{x}_k]. \quad (22)$$

Despite the apparent difficulty of evaluating the right-hand-side of (22), it is relatively straightforward to determine the stationary point $\mu_{SD}^{(k)}$ of this steepest descent procedure, which is simply given by

$$\mu_{SD}^{(k)} = \frac{E[\widehat{d}_k\widehat{x}_k]}{E[\widehat{x}_k^2]}. \quad (23)$$

Considering the numerator of (23) and using a result from Appendix A of [8], we evaluate it using our assumptions as

$$E[\widehat{d}_k\widehat{x}_k] \approx E[\mathbf{V}_{k-1}^T R^2 \mathbf{V}_{k-1}]E[f'(e_{k-1})]. \quad (24)$$

From (20), we can express $E[\widehat{x}_k^2]$ as

$$E[\widehat{x}_k^2] = E[f^2(e_{k-1})\mathbf{X}_{k-1}^T\mathbf{X}_k\mathbf{X}_k^T\mathbf{X}_{k-1}] \quad (25)$$

$$= E[f^2(e_{k-1})\mathbf{X}_{k-1}^T R\mathbf{X}_{k-1}], \quad (26)$$

where we have used the independence of $\mathbf{X}_k$ from all other quantities in the expression.

Substituting (24) and (26) into (23) yields the necessary expression for the stationary point as

$$\mu_{SD}^{(k)} = \frac{E[\mathbf{V}_{k-1}^T R^2 \mathbf{V}_{k-1}]E[f'(e_{k-1})]}{E[f^2(e_{k-1})\mathbf{X}_{k-1}^T R\mathbf{X}_{k-1}]}. \quad (27)$$

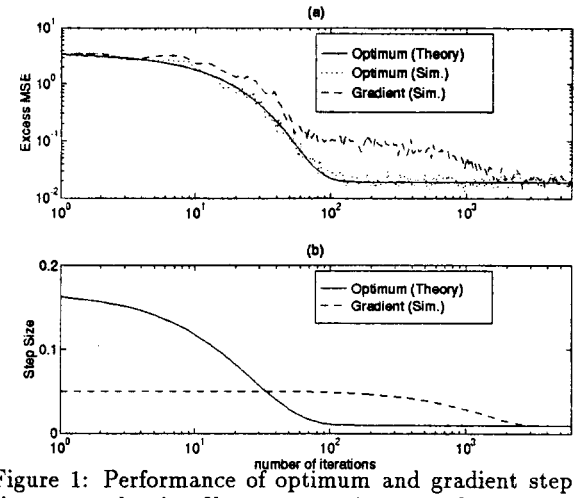

Figure 1: Performance of optimum and gradient step size sign-error adaptive filters, nonstationary unknown system.

Comparing equations (27) and (17), we see that $\mu_{opt,k-1} = \mu_{SD}^{(k)}$. This implies that the optimum step size as found from analysis is *exactly the same* as the stationary point of the steepest descent iteration on which the gradient adaptive step size algorithm is based[1]. In other words, for the kth iteration of the adaptive filter update, if the gradient adaptive step size algorithm converges to to the minimum point of its associated steepest descent cost function, then the resulting value of μ_k would be the optimum value $\mu_{opt,k-1}$.

In practice, the step size μ_k will not converge to $\mu_{opt,k-1}$, for the following reasons. 1) Because (7) is a stochastic gradient approximation to a steepest descent procedure, the step size can only converge to its optimum value in expectation, and its actual value will continue to fluctuate about its optimum value. 2) During the transient stages of the adaptation of the adaptive filter, the level of excess MSE is constantly changing, and therefore the optimum value $\mu_{opt,k}$ is constantly changing. Thus, the adaptive step size algorithm must attempt to track these changes in the step size, which results in a lag error. However, if the level of excess MSE stays nearly fixed, as in the case of tracking a slowly-moving unknown system, then the adaptive step size can achieve near-optimal performance.

We have verified this form of optimality for (7) for the case $\{f(e_k) = e_k, G(\mathbf{X}_k) = [g(x_k) \cdots g(x_{k-L+1})]^T\}$ (nonlinear data algorithm) when x_k is an i.i.d. input sequence, using analytic results from [13]. Thus, these adaptive step size procedures attempt to solve a well-posed optimization problem, in contrast to other heuristic procedures.

5. SIMULATIONS

We now examine the the behavior of the adaptive step size algorithms through simulation of a nine-coefficient system identification problem with $\mathbf{W}_{opt,0}^T = [1\ 2\ 3\ 4\ 5\ 4\ 3\ 2\ 1]/5$. The signals $\{x_k\}$ and $\{n_k\}$ were chosen to be zero-mean Gaussian-distributed with $E[x_k x_{k-j}] = a^{|j|}$ and $\sigma_n^2 = 0.01$. The coefficients were allowed to vary according to the model in (14) where $\mathbf{M}_k$ is Gaussian-distributed with covariance matrix $\sigma_m^2 I$, as specified. One hundred simulations were run and the results averaged in each case.

[1] Because the steepest descent algorithm is based on an *a priori* error criterion, the sequence $\mu_{SD}^{(k)}$ lags $\mu_{opt,k-1}$ by one time step. The effect of this lag on the algorithm's performance is minimal compared to other factors.

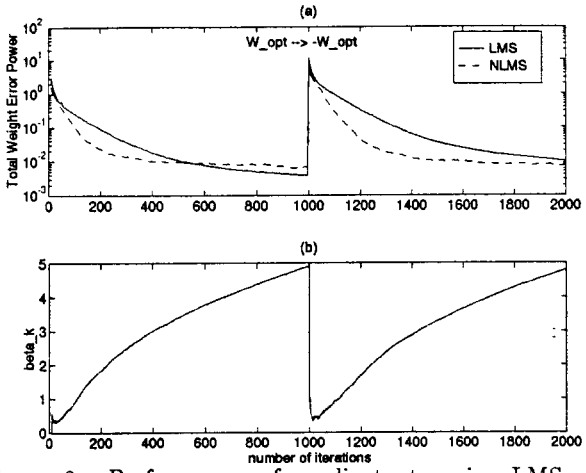

Figure 2: Performance of gradient step size LMS and NLMS algorithms for a sign change in coefficients, highly-correlated input data.

Figure 1a shows the convergence of the excess MSE for the gradient step size sign-error adaptive filter with $a = 0$ and $\sigma_m^2 = 0.0001$. For comparison, we also show the theoretical and simulated performance of the algorithm in (15) with the optimum step size sequence $\mu_{opt,k}$. As can be seen, the optimum step size algorithm gives the fastest initial convergence, as expected. Both algorithms achieve nearly the same steady-state error power while tracking. From the simulations, the steady-state excess MSE for the gradient step size algorithm is $\xi_{ss} = 0.0224$, about 15% greater than the optimum step size algorithm. Figure 1b shows the associated step size values for each algorithm. We see that the gradient step size algorithm does not follow the optimum step size sequence very closely initially. However, both algorithms achieve the optimum step size for tracking, given by $\lim_{k\to\infty} \mu_{opt,k} = \sigma_m^2/\sigma_n^2 = 0.01$ for this situation [7]. Note that the optimum step size algorithm requires *complete* knowledge of the signal and system statistics and thus cannot be implemented in most practical situations. The adaptive step size algorithm achieves similar steady-state performance without this knowledge.

We now compare the performance of the gradient adaptive step size LMS and NLMS algorithms when the unknown system coefficients undergo a sign change at iteration $k = 1000$. Figure 2a plots the total weight error power, given by $E[\mathbf{V}_k^T\mathbf{V}_k]$, for the two algorithms, where $a = 0.95$, $\sigma_m^2 = 0$, $\mu_0 = 0.01$ and $\rho = 0.0005$ for the LMS algorithm, and $\beta_0 = 0$ and $\rho = 20$ for the NLMS algorithm. To maintain stability of the LMS algorithm, the algorithm's step size was limited to the range $[0, 0.03]$ as determined by trial-and-error for this data. As can be seen, the adaptive step size NLMS algorithm provides faster convergence in this case because the LMS algorithm's step size must be limited to a very small range in order to maintain stability for this data. Figure 2b shows the average value of β_k for the NLMS algorithm. The algorithm's behavior is intuitively pleasing, as the small value of β_k yields fast convergence both initially and after the coefficient values change.

The sign-error algorithm's robustness properties are well-known [8, 10, 11]. Figure 3 shows the convergence of the gradient adaptive step size sign-error and LMS algorithms for the case $a = 0.9$, $\sigma_m^2 = 0$, $\mu_0 = 0.01$, $\rho = 0.00002$, in which dropouts in the desired response occur with probability 0.01. The sign-error adaptive algorithm clearly outperforms the LMS algorithm, due to the former algorithm's

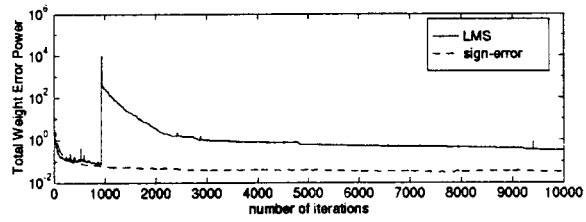

Figure 3: Performance of gradient step size sign-error and LMS adaptive filters, dropouts in desired response signal.

immunity to impulsive noises.

6. CONCLUSIONS

In this paper, we have presented new gradient adaptive step size procedures for several types of adaptive filters. These systems exhibit enhanced performance in situations where the LMS adaptive algorithm is known to perform poorly, such as with highly-correlated input data and in the presence of impulsive disturbances. Theoretical analysis and simulation show that the algorithms can achieve near-optimum tracking performance without knowledge of the speed of variation of the unknown system.

REFERENCES

[1] R.W. Harris, D.M. Chabries, and F.A. Bishop, "A variable step (VS) adaptive filter algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 2, pp. 309-316, April 1986.

[2] Y.K. Shin and J.G. Lee, "A study on the fast convergence algorithm for the LMS adaptive filter design," *Proc. KIEE*, vol. 19 , no. 5, pp. 12-19, October 1985.

[3] V.J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Trans. Signal Processing*, vol. SP-41, no. 6, pp. 2075-2087, June 1993.

[4] A. Sugiyama, "Stochastic gradient algorithms with a gradient-adaptive and limited step-size," *IEICE Trans. Fund. Elect., Comm., Comp. Sci.*, vol. E77-A, no. 3, pp. 534-538, March 1994.

[5] A.G. Barto and R.S. Sutton, "Goal seeking components for adaptive intelligence: An initial assessment," Air Force Tech. Rep. AFWAL-TR-81-1070, Wright-Patterson AFB, OH, 1981.

[6] D.L. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, no. 4, pp. 578-586, August 1982.

[7] S.H. Cho and V.J. Mathews, "Tracking analysis of the sign algorithm in nonstationary environments," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-38, no. 12, pp. 2046-2057, December 1990.

[8] S.C. Douglas and T.H.-Y. Meng, "Stochastic gradient adaptation under general error criteria," *IEEE Trans. Signal Processing*, vol. SP-42, no. 6, pp. 1335-1351, June 1994.

[9] S.C. Douglas and T.H.-Y. Meng, "Normalized data nonlinearities for LMS adaptation," *IEEE Trans. Signal Processing*, vol. SP-42, no. 6, pp. 1352-1365, June 1994.

[10] N.L. Freire and S.C. Douglas, "Adaptive cancellation of geomagnetic background noise using a sign-error normalized LMS algorithm," *Proc. IEEE Int'l Conf. on Acoust., Speech, Signal Processing*, Minneapolis, MN, v. III, pp. 523-526, April 1993.

[11] O. Arikan, A.E. Çetin, and E. Erzin, "Adaptive filtering for non-Gaussian stable processes," *IEEE Signal Processing Letters*, vol. SPL-1, no. 11, pp. 163-165, November 1994.

[12] S.C. Douglas, "A family of normalized LMS algorithms," *IEEE Signal Processing Letters*, vol. SPL-1, no. 3, pp. 49-51, March 1994.

[13] S.C. Douglas and T.H.-Y. Meng, "The optimum scalar data nonlinearity in LMS adaptation for arbitrary i.i.d. inputs," *IEEE Trans. Signal Processing*, vol. SP-40, no. 6, pp. 1566-1570, June 1992.