EXACT EXPECTATION ANALYSIS OF THE SIGN-DATA LMS ALGORITHM
FOR I.I.D. INPUT DATA

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

ABSTRACT

Previous statistical analyses of the sign-data least-mean-square (LMS) and other nonlinearly-modified data adaptive algorithms assume that the filter coefficients are statistically independent of the input data samples currently in filter memory. This assumption is incorrect for a tapped-delay-line filter; thus, the analysis using this assumption is only an approximate description of the adaptive filter’s convergence behavior. In this paper, we present an automated method of deriving an exact description of the convergence behavior of a class of nonlinearly-modified data adaptive algorithms for system identification modelling with independent, identically-distributed (i.i.d.) samples as input data. Using our method, we can identify a set of linear equations that exactly describes a nonlinear data algorithm’s stochastic behavior at each time step. Moreover, we can obtain precise bounds upon the step size to guarantee convergence of the algorithm in the mean and in mean square. Simulations indicate that the equations produced by the exact method are much more accurate than previous analyses in predicting convergence behavior of the sign-data LMS adaptive algorithm, particularly in fast adaptation situations.

1. INTRODUCTION

Stochastic gradient adaptive filtering algorithms such as the popular least-mean-square (LMS) algorithm often employ nonlinearities within the coefficient update, either to simplify the implementation of the adaptive filter in hardware or to improve the performance of the adaptive filter. One algorithm that modifies the data vector within the update to allow a simple hardware implementation is the sign-data LMS algorithm, given by

\[ W_{k+1} = W_k + \mu (d_k - W_k^T X_k) \text{sgn}(X_k), \tag{1} \]

where \( W_k = [w_{1,k} \ w_{2,k} \ldots w_{N,k}]^T \) and \( X_k = [x_k \ x_{k-1} \ldots x_{k-N+1}]^T \) are the weight and data vectors, respectively, \( d_k \) is the desired response, \( \mu \) is a user-specified convergence parameter, and the \( \text{sgn}(\cdot) \) nonlinearity operates individually on each element of the data vector \( x_k \).

Previous analyses of this algorithm and other nonlinearly-modified data adaptive algorithms have assumed that the weight vector \( W_k \) is statistically independent of the data vector \( X_k \) \([1, 2, 3, 4]\), an assumption which is clearly not true because of the shift-input nature of the elements of the data vector. Recently, an exact analysis of a two-tap LMS adaptive filter operating on independent, identically-distributed (i.i.d.) Gaussian input data samples has been presented \([5]\). This exact analysis predicts the convergence behavior of the LMS adaptive filter more accurately than the independence assumption analysis for i.i.d. Gaussian inputs, giving a precise bound upon the step size to guarantee convergence of the filter coefficients in mean-square. A formal method for developing an exact description of the convergence behavior of LMS adaptive filters with an arbitrary number of taps operating upon arbitrarily-distributed i.i.d. input data is presented in \([6]\). The equations produced by the method are shown to accurately predict the mean-square performance of LMS adaptive filters with two and three coefficients.

In this paper, we extend the method presented in \([6]\) to develop a set of update equations for describing without approximation the convergence behavior of nonlinearly-modified data adaptive algorithms of the form

\[ W_{k+1} = W_k + \mu (d_k - W_k^T X_k) F(X_k) \tag{2} \]

where \( F(\cdot) \) is a vector-valued function of the input data vector such that \( [F(X_k)] = f(x_{k-1}) = -f(-x_{k-1}) \). Equation (2) is a generalization of the sign-data LMS algorithm in (1). The analysis method relies upon a system identification desired response model, in which \( d_k \) is described as

\[ d_k = W^* X_k + n_k, \tag{3} \]

where \( W^* \) is a vector of filter coefficients for the unknown system and \( \{n_k\} \) is an i.i.d. noise sequence with variance \( \sigma_n^2 \) that is independent of the input data sequence. For the analysis presented in this paper, we assume that the input data and noise processes are both zero-mean and independent from sample to sample; however, the method for generating the exact analysis can be applied to finite-time-correlated data and noise models as well. By maintaining within the analysis the interdependence of weight vector elements and data vector elements that are possibly statistically related, we can analyze and quantitatively the effects that this interdependence has upon the adaptation behavior of the nonlinear data algorithm. Moreover, we can provide exact bounds upon the step size parameter to guarantee convergence in mean and in mean-square for any given input and noise distributions.

This paper is organized as follows. In the next section, we present a method for generating the linear equations necessary to represent the statistical behavior of nonlinear-data adaptive algorithms, and we give the resulting equations for the two-tap case with i.i.d. input data and noise signals. In Section 3, we specialize the results to the sign-data LMS algorithm, and we compare the accuracy of the exact analysis to that of the independence assumption analysis through Monte Carlo simulations. In Section 4, we give details of a computer program written in the symbolic-manipulation software package MAPLE\(^1\) that automatically derives the exact analysis equations for the sign-data LMS

\(^{1}\)MAPLE is a registered trademark of Waterloo Maple Software.
algorithm for an arbitrary number of filter coefficients. Section 5 presents our conclusions and issues for future research.

2. DERIVATION OF THE EXACT ANALYSIS

The method for deriving an exact statistical description of the behavior of algorithms of the form in (2) requires that the input data is correlated only for a finite set of time lags. In other words, for some positive integer $M$, we require that

$$E[x_{m+k}] = 0, \quad m > M. \tag{4}$$

With this constraint, we can develop a set of linear difference equations that exactly describes the convergence of the filter coefficient equations and the resulting mean-square error (MSE) at the filter output. In this paper, we will take the case of $M = 1$, which corresponds to i.i.d. input data samples, although in general these techniques can be applied to the correlated data case as well. As an example, we shall illustrate our method using a two-tap filter; Section 3 discusses the extension of this method to filters with more coefficients.

To derive the exact linear difference equations, we first identify a quantity of interest, such as the excess MSE at the filter output due to statistical fluctuations in the filter coefficients. For a two-tap filter, we can express the excess MSE as

$$\zeta_k = E[(V_k^2 X_k)] \tag{5}$$

$$= E[E[v_{k+1}^2] E[v_{k+1}] + E[v_{k+1}^2] E[v_{k+1}^2], \tag{6}$$

where the weight error vector $V_k$ is defined as $V_k = W_k - W^*$. Not e that $x_k$ is zero-mean and independent of both $v_{1:k}$ and $v_{2:k}$ because the input data is i.i.d.; however, $x_{k-1}$ is not independent of $v_{1:k}$ and $v_{2:k}$ because $x_{k-1}$ is used in the previous time step to update the filter coefficients. We would like to develop linear difference equations for the two quantities $E[v_k^2]$ and $E[v_k^2 v_{k-1}^2]$ that will compute the values of these quantities at time $k+1$ given the values of these and possibly other quantities at time $k$. Using equation (3), we write the weight error vector update for the algorithm in (2) as

$$V_{k+1} = (I - \mu F(X_k) X_k) V_k + \mu v_k F(X_k). \tag{7}$$

To find the update for $E[v_k^2]$, we square the first element for the weight error vector update in (7) and take expectations. Since all quantities that depend upon $x_k$ are independent of $v_{1:k}$, $v_{2:k}$, and $x_{k-1}$, we can separate these expectations, which gives

$$E[v_k^2] = (1 - \mu f_{1:k} + \mu f_{2:k}) E[v_k^2] + \mu f_{2:k} E[v_k^2 v_{k-1}^2] - \mu f_{2:k} E[v_k^2 v_{k-1}^2], \tag{8}$$

where we have defined the doubly-indexed variable $f_{mn}$ as

$$f_{mn} = E[x_m x_n]. \tag{9}$$

where the expectation is taken over the probability density function of the input data. Considering the term $E[v_k^2 v_{k-1}^2]$, we can develop a similar difference equation in the same fashion, which yields

$$E[v_{k+1}^2 v_{k+1}^2] = f_{1:k} E[v_{k+1}^2] + \mu f_{2:k} E[v_{k+1}^2 v_{k-1}^2 f_{2:k-1}] + \mu f_{2:k} E[v_{k+1}^2 v_{k-1}^2 f_{2:k-1}] - 2\mu f_{2:k} E[v_{k+1}^2 v_{k-1}^2 f_{2:k-1}]. \tag{10}$$

Note that the right-hand size of (10) contains four new expectation terms requiring their own linear difference updates. We can proceed in this fashion, creating the necessary equations, until all terms on the right-hand-side of these equations have updates. This process produces the following five additional equations:

$$E[v_{k+1}^2] = E[v_k^2] + \mu f_{2:k} E[v_k^2 f_{2:k-1}] + \mu f_{2:k} E[v_k^2 f_{2:k-2}] + \mu f_{2:k} f_{2:k-1} + 2\mu f_{2:k} E[v_k^2 f_{2:k-2}]. \tag{11}$$

$$E[v_{k+1}^2 f_{2:k}] = (f_{1:k} - \mu f_{1:k} + \mu f_{2:k}) E[v_k^2] + \mu f_{2:k} E[v_k^2 f_{2:k-1}] + \mu f_{2:k} E[v_k^2 f_{2:k-2}] + \mu f_{2:k} f_{2:k-1} - 2\mu f_{2:k} E[v_k^2 f_{2:k-2}]. \tag{12}$$

$$E[v_{k+1}^2 v_{k-1}^2] = f_{2:k} E[v_k^2] + \mu f_{2:k} f_{2:k-1} + \mu f_{2:k} E[v_k^2 f_{2:k-1}] + \mu f_{2:k} f_{2:k-1} + 2\mu f_{2:k} E[v_k^2 f_{2:k-2}]. \tag{13}$$

$$E[v_{k+1}^2 v_{k-2}^2 f_{2:k}] = f_{2:k} E[v_k^2] + \mu f_{2:k} f_{2:k-1} + \mu f_{2:k} E[v_k^2 f_{2:k-1}] + \mu f_{2:k} f_{2:k-1} - 2\mu f_{2:k} E[v_k^2 f_{2:k-2}]. \tag{14}$$

$$E[v_{k+1}^2 v_{k-2}^2] = f_{2:k} E[v_k^2] + \mu f_{2:k} f_{2:k-1} + \mu f_{2:k} E[v_k^2 f_{2:k-1}] + \mu f_{2:k} f_{2:k-1} - 2\mu f_{2:k} E[v_k^2 f_{2:k-2}]. \tag{15}$$

Because the data is correlated only to a finite number of time lags, the equation-generating process we have illustrated will eventually stop for a general $N$-tap filter after we have accounted for all terms on the right-hand sides of all of the equations. In this example, the seven equations in (8)-(15) are linear in the seven state variables $E[v_k^2], E[v_k^2 v_{k-1}^2], E[v_k^2 f_{2:k-1}], E[v_k^2 f_{2:k}], E[v_k^2 v_{k-1}^2 f_{2:k}], E[v_k^2 v_{k-2}^2 f_{2:k}],$ and $E[v_k^2 v_{k-2}^2 f_{2:k-2}]$. Thus, they can be written in the form

$$Y_{k+1} = A Y_k + B, \tag{16}$$

where $Y_k$ is a seven-dimensional vector of states variables, and the entries of the transition matrix $A$ and vector $B$ are determined from the equations above and depend upon the $f_{mn}$, $\sigma_{mn}^2$, and the step size parameter $\mu$. For a general $N$-tap filter, the resulting equations will also be linear in the state variables as thus can be written in the same form as in (16).

From the linear update in (16), we can accurately predict the mean-square convergence behavior of the nonlinear algorithm by specifying the initial values of the state variable quantities in $Y_0$ and iterating upon the relationship. The excess MSE at any time step can then be determined from equation (6). To find the steady-state excess MSE, we can solve for the stationary point of the iteration,

$$\lim_{k \to \infty} Y_k = (I - A)^{-1} B, \tag{17}$$

from which the excess MSE can be computed from (6). Perhaps most importantly, a precise range of $\mu$ can be defined which guarantees that all the eigenvalues of the matrix $A$ are of magnitude less than one, thus guaranteeing mean-square stability of the adaptive algorithm. Thus, no approximate bounds based upon the independence assumption 4 or upon other conservative bounding techniques are necessary if the correlation structure of the data is known. In the next section we specialize these results to
the sign-data LMS algorithm and explore the accuracy of the exact analysis as compared to the independence assumption analysis under fast adaptation conditions.

3. SIMULATIONS: SIGN-DATA ADAPTATION, TWO-TAP CASE

Moschner studied adaptation algorithms with clipped input data [1], where the data were quantized to a single bit in the adaptation algorithm. His analysis assumed Gaussian input data and invoked the independence assumption. The convergence of the excess MSE is given by, for the independence assumption with i.i.d. input data,

$$\xi_{k+1} = (1 - 2\mu \gamma_1 + \mu^2 \gamma_2) \xi_k + \mu^2 \gamma_2$$

where \( \gamma_m \) is the \( m \)th absolute-value moment of the input data, defined as

$$\gamma_m = E[|z|^m].$$

Using the equations for the exact updates in (8)-(15), it can be shown that for the \( \text{sgn}(\cdot) \) nonlinearity with Gaussian data, the exact equations that describe the excess MSE convergence behavior are given by (16), with \( A \), \( Y \), and \( B \) given by

$$A = \begin{bmatrix} 1 - 2\mu \gamma_2 & \mu^2 \gamma_2 & 0 & 0 & \mu^2 \\ \mu^2 \gamma_2 & 2\mu \gamma_2 & 1 & -2\mu & \mu^2 \\ 2\mu \gamma_2 & 2\mu \gamma_2 & \gamma_2 & -2\mu \gamma_2 & \mu^2 \gamma_2 \\ 3\mu \gamma_2^2 & 3\mu \gamma_2^2 & \gamma_2 & -2\mu \gamma_2 & \mu^2 \gamma_2 \end{bmatrix}$$

$$Y = \begin{bmatrix} E[v_1^2] \\ E[v_2^2] \\ E[v_3^2|z_{k-1}] \\ E[v_4^2|z_{k-1}] \end{bmatrix}$$

$$B = \begin{bmatrix} \mu^2 \sigma_x^2 \\ \mu^2 \sigma_x^2 \\ \mu^2 \sigma_x^2 \\ \mu^2 \sigma_x^2 \end{bmatrix}.$$

Note that only four state variables are required for the two-tap sign-data LMS algorithm, because \( E[v_1^2] = E[v_2^2] = E[v_3^2|z_{k-1}] = E[v_4^2|z_{k-1}] \) for the \( \text{sgn}(\cdot) \) nonlinearity. In our simulations, the coefficients of the unknown system to be identified were chosen arbitrarily as \( W^0 = [1 1]^T \), and the initial weight values for the adaptive filter were both set to 2. The variance of the input data was chosen to be unity, and the variance of the interfering noise was chosen to be \( \sigma_n^2 = 0.01 \). In order to get highly-accurate simulation convergence curves for comparison with the analyses, 10000 simulation experiments were run and the results averaged. Steady-state quantities were estimated by averaging 100 time steps in these simulations for a total of one million averaged iterations.

Figure 1 shows the convergence of the excess MSE of a two-tap filter adapted using the sign-data algorithm for a step size of \( \mu = 0.48 \). This step size is much larger than that typically chosen for the sign-data LMS algorithm in other studies. As can be seen, the actual convergence rate of the excess MSE is slower than that predicted by the independence assumption analysis, and the final excess MSE in steady-state is greater than that predicted by the independence assumption analysis. An average of 10000 simulations indicates the accuracy of the exact analysis, whereas the predicted convergence curve from the independence assumption analysis is clearly biased for this large step size.

Figure 2 shows the convergence of the first and second weight error variance for the sign-data LMS algorithm. As can be seen, the variance in the first filter tap is greater than that predicted by the independence assumption analysis. This behavior is similar to that observed for the two-tap LMS adaptive filter under fast adaptation conditions [6]. The exact analysis matches the convergence behavior in simulation extremely well and is much more accurate than the independence assumption analysis.

Table 1 presents the percent differences between the predicted
final values of various mean-square quantities and the actual values of these quantities observed through an average of one hundred time steps of an average of 10,000 simulations for different step sizes. The predictions from the exact analysis in this situation are very accurate, comparable to those of the exact analysis in LMS adaptation. The independence assumption analysis, however, is inaccurate in predicting the final quantities above the smallest step size tabulated, showing tendencies to underestimate the variance of the first filter tap and to overestimate the variance of the second filter tap. Note that the exact analysis is accurate for step sizes close to the stability bound of $\mu_{\text{max}} = 0.6564$ for the sign-data LMS algorithm with unity-variance white Gaussian input.

Using the exact analysis, we can predict an exact upper bound upon the step size of the sign-data LMS algorithm to guarantee mean-square convergence of the adaptive filter. Because this bound depends upon the maximum absolute eigenvalue of the transition matrix $A$, it is a complex function of the input data moments. For $N = 2$, it can be shown that the $A$ matrix in (20) has one zero eigenvalue; hence, the remaining three eigenvalues can be solved for analytically using the cubic equation (7). The resulting expression is too long to show here; instead, we evaluate it numerically. It can be shown that the maximum step size to guarantee mean square convergence for a two-tap filter with white Gaussian input of variance $\sigma_w^2$ is

$$\mu_{\text{max}} = \frac{0.656444616}{\sigma_w} \quad (23)$$

to eight decimal places. Note that this exact bound scales with $\sigma_w$ and is more stringent than that predicted by the independence bound, $\mu_{\text{max, ind}} = (2\pi)^{1/2}\sigma_w^2 = 0.79788456\sigma_w^2$.

4. AUTOMATIC GENERATION OF EXACT UPDATES

As mentioned in Section 2, the exact analysis can be extended to filters of arbitrary order. As might be surmised by the two-tap example, performing the derivation for the exact analysis for longer filters involves the creation of more state variables and their corresponding equations. However, the method for generating these equations is algorithmic and thus can be automated. We have written a simple computer program in the symbolic-manipulation software package MAPLE that automatically derives these equations for any length sign-data LMS adaptive filter. The program is less than one hundred lines of code in length and is included in its entirety in an appendix at the end of this paper.

We can describe the operation of this program by the following steps:

- Generate algebraic expressions for both the error $e_k$ and the elements of $V_{k+1}$ in terms of the elements of $V_k$, $X_k$, and $n_k$.
- Take the expectation of $(e_k - n_k)^2$ to find terms which constitute the excess MSE.
- Create a set of terms to be evaluated from these excess MSE terms.
- While this set is not empty,
  1. take the first term of the terms remaining in this set and determine its update equation;
  2. add this term to a set of evaluated terms;
  3. add any new terms appearing in the update equation to the set of terms to be evaluated; and
  4. remove all evaluated terms from the set of terms to be evaluated
- until the set of terms to be evaluated is empty.

We have used this program to generate the necessary system of equations for two, three, four, and five-tap sign-data LMS adaptive filters with i.i.d. input data. Moreover, we have verified the accuracy of the three-tap filter equations through extensive simulations similar to those in the last section, and in every case, the exact analysis is more accurate than the independence assumption analysis in predicting mean-square performance. The main difficulty in using the exact analysis equations in their raw form is their sheer number for even modest filter lengths. For example, the number of equations necessary to represent mean-square performance of the sign-data LMS adaptive filter for filter lengths of $N = \{2, 3, 4, 5\}$ taps is $\{4, 31, 300, 3065\}$ equations, respectively. However, these equations can still be a powerful tool in determining exact stability bounds upon the algorithm step size. In particular, only the maximum absolute eigenvalue of the transition matrix $A$ in (16) is important for stability considerations. As efficient search techniques for finding the maximum eigenvalue of a sparse matrix exist [8], such techniques could be incorporated within a general computer program that would take as input the filter length and the moments of the input data and produce as output the maximum step size to guarantee convergence in mean square. Such a program would be a powerful tool for the users of this as well as other stochastic-gradient adaptive algorithms.

5. CONCLUSION

We have described a method for producing an exact description of nonlinear data adaptive algorithms such as the sign-data LMS algorithm with i.i.d. input. Simulations with both two- and three-tap adaptive filters operating on white Gaussian input data verify that the exact description is much more accurate than other approximate analyses in determining mean-square performance and stability of the sign-data LMS adaptive algorithm. We have also presented a computer program that generates this exact description for a filter with an arbitrary number of taps. Future work shall focus upon 1) incorporating correlated models within the data.

<table>
<thead>
<tr>
<th>step size</th>
<th>$%$ Error in Predicting Final Values, Indep.</th>
<th>$%$ Error in Predicting Final Values, Exact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E[V_k]$</td>
<td>$E[V_{k+1}]$</td>
</tr>
<tr>
<td>0.08</td>
<td>0.387</td>
<td>-0.665</td>
</tr>
<tr>
<td>0.15</td>
<td>0.849</td>
<td>-2.204</td>
</tr>
<tr>
<td>0.24</td>
<td>1.170</td>
<td>-4.651</td>
</tr>
<tr>
<td>0.32</td>
<td>4.053</td>
<td>-6.938</td>
</tr>
<tr>
<td>0.40</td>
<td>8.787</td>
<td>-8.411</td>
</tr>
<tr>
<td>0.48</td>
<td>21.100</td>
<td>-6.692</td>
</tr>
<tr>
<td>0.56</td>
<td>33.772</td>
<td>7.076</td>
</tr>
<tr>
<td>0.64</td>
<td>38.54</td>
<td>14.73</td>
</tr>
</tbody>
</table>

Table 1: Simulation results, sign-data algorithm, Gaussian input, $N = 2$. 

569
analysis and 2) providing an automated program for determining
an exact step size bound to guarantee mean-square convergence
of stochastic-gradient adaptive filters for given data correlation
statistics.

APPENDIX

This computer program derives the exact mean-square analysis
equations for a general N-tap sign-data LMS adaptive filter. It is
for use with the MAPLE software package. @Scott C. Douglas,
October, 1992

createequations:sgnLMS := proc(numcoeff);
local out, temp, tempset, tempterm;
tempterm := term;
for i from 1 to numcoeff do
  tempterm := subs(signum(xkm.i), tempterm);
end;
local out, temp, tempset, tempterm;
tempterm := term;
for i from 1 to numcoeff do
  tempterm := subs(signum(xkm.i), tempterm);
end;

modifyset := proc(expression, set, evalset, numcoeff)
local out, tempable, tempset, tempset2, i, N;
tempable := [op(expression)];
for i from 1 to (numcoeff) do
  tempable := subs(set[i]=tempable);
end;
tempset := subs(tempable);
tempset2 := tempset;
tempset := tempset minus evalset;
out := tempset minus [0];
end;
cardinality := proc(set)
local localout, tempout, i;
i := 0;
tempset := set;
while evalb(tempset <> [ ]) do
  i := i + 1;
tempset := tempset minus [set[i]];
end;
out := i;
end;

REFERENCES

Stanford University, Stanford, CA, June 1970.

February 1986.

LMS adaptation for arbitrary I.D. inputs," IEEE Trans. Signal Processing,


a tapped delay line (two-dimensional case)," IEEE Trans. Acoust., Speech,

adaptive filter without the independence assumption," Proc. International
Conf. Acoust., Speech, Signal Processing, vol. IV, pp. 61-64, San Francisco,


[8] G.H. Golub and C.F. Van Loan, Matrix Computations, 2nd. ed. (Baltimore:

570