ABSTRACT

Many strategies have recently been proposed to improve both the stationary and nonstationary performance of the LMS adaptive algorithm [1-5]. The algorithm presented here uses the power of the filtered gradient estimate to form the step-size parameter, $\mu$. Each weight coordinate has its own variable step-size parameter to more accurately update the weight vector in the direction of the minimum of the performance surface. Though somewhat similar to the algorithm put forth by Shan and Kailath [1], this work solves the problems mentioned in [2] and extends the discussion to include the effect on algorithm misadjustment due to using the gradient to derive the step-size parameter.

INTRODUCTION

A number of variable step-size LMS algorithms have been developed that either strive to maintain a constant sum of a posteriori squared error [3-4] or make use of information present in the gradient estimate [1-2][5] to derive the step-size parameter $\mu$. The major drawback of the constant error criterion is the inability to distinguish between observation noise (noise in the steady state solution) and weight vector lag (tracking error) which has conflicting requirements for the choice of $\mu$. When the observation noise increases, the a posteriori error increases. When the lag increases, the a posteriori error increases.

In the gradient-based category, a variation of the LMS algorithm using the power of the filtered gradient to obtain the step-size parameter is considered (Figure 1). This technique is able to differentiate between observation noise and weight vector lag. If the misadjustment is solely due to observation noise, a small $\mu$ leads to lower excess mean square error. The expected value of the gradient vector after convergence is zero and the value of $\mu$ based on this gradient will drop to zero. The case of weight vector lag requires a larger step-size to allow rapid weight tracking towards the minimum of the performance surface. As the weight vector lags the optimal weight vector solution, the expected value of the gradient grows, pointing in the direction of the optimal solution. The step-size grows with the extent of the weight lag, enabling rapid adjustment of the weights to minimize this lag.

The use of the gradient to derive a variable $\mu$ gives rise to an increase in misadjustment, primarily as a result of gradient noise. The misadjustment of this algorithm is directly proportional to the minimum mean square error, unlike the misadjustment of the classic LMS algorithm, and may be an undesirable characteristic of the various gradient-based algorithms.

The computational complexity of this algorithm is only 2-3 times that of the classic LMS algorithm and it is significantly less complex than the RLS algorithm. The performance gain for this modest increase in complexity makes it an attractive choice for many adaptive filtering applications.

A computer simulation of the noise cancellation problem was used to evaluate the stationary and nonstationary performance of this variable step-size algorithm, as compared to the classic LMS algorithm.

ALGORITHM DESCRIPTION

The gradient-based variable step-size algorithm is a variation of the classic LMS algorithm with the fixed, scalar $\mu$ replaced by a diagonal step-size matrix $\mathbf{M}_k$ [5] that is derived from the power of the gradient estimate.

$$
\hat{V}_k = \begin{bmatrix}
\frac{\partial^2 e_k}{\partial V_k^2} & 0 & \frac{\partial e_k}{\partial V_k} \\
0 & \frac{\partial^2 e_k}{\partial V_{k-1}^2} \\
\frac{\partial e_k}{\partial V_k} & \frac{\partial e_k}{\partial V_{k-1}} & -2e_kX_k
\end{bmatrix}
$$

To approximate the expected value of the gradient, the gradient vector of equation (1) is applied to a low-pass moving average (MA) filter. This type of filter was chosen for convenience and any suitable filter could be used to approximate the expected gradient. The resulting filtered gradient estimate is squared to obtain the gradient power in each weight coordinate. After multiplying by a gain parameter, $\gamma$, the resulting step-size matrix $\mathbf{M}_k$ for the 2-weight case is

$$
\mathbf{M}_k = \gamma \mathbf{J} \mathbf{V}_k \mathbf{J} = \gamma \begin{bmatrix}
\mathbf{J} \mathbf{V}_k & 0 \\
0 & \mathbf{J} \mathbf{V}_{k-1}
\end{bmatrix}
$$

$$
\mathbf{M}_k = \gamma \mathbf{J} \mathbf{V}_k \mathbf{J} = \gamma \begin{bmatrix}
|\mathbf{J} \mathbf{V}_k|^2 & 0 \\
0 & |\mathbf{J} \mathbf{V}_{k-1}|^2
\end{bmatrix}
$$

CH2958-3/91/0000-1083/01.00$15991IEEE
Each element of $\mathbf{M}$ is limited to $\mu_{\text{max}}$ for stability and is applied to the weight update equation to obtain

$$
\mathbf{w}_{k+1} = \mathbf{w}_k - \mathbf{M}_k \mathbf{v}_k
$$

(3)

**Misadjustment of the Variable Step-Size Algorithm**

The misadjustment of the variable step-size algorithm will be considered next. Following (6), when the weight vector has converged to a solution approximately equal to $\mathbf{w}_{\text{opt}}$, the gradient noise is given by

$$
\mathbf{N}_k = \hat{\mathbf{v}}_k - 2\mathbf{c}_k \mathbf{x}_k
$$

(4)

Assuming that the mean-square error, $e^2$, is uncorrelated with the signal vector, the gradient noise power is the covariance of $\mathbf{N}_k$

$$
\text{cov} [\mathbf{N}_k] = E [\mathbf{N}_k \mathbf{N}_k^H] = 4E [e^2 \mathbf{x}_k \mathbf{x}_k^H] = 4 \mu_{\text{min}} \mathbf{R}
$$

(5)

which can be expressed in terms of the weight principal-axis coordinate system as

$$
\text{cov} [\mathbf{N}_k] = 4 \mu_{\text{min}} \mathbf{A}
$$

(6)

The variable step-size matrix $\mathbf{M}_k$ is the output of the MA filter with the gradient noise power as the input, multiplied by a gain constant $\gamma$.

$$
\mathbf{M}_k = \frac{4 \mu_{\text{min}} \mathbf{A}}{L}
$$

(7)

The average excess mean-square error (MSE) due to gradient noise is

$$
J_{\text{excess}} = \frac{\mu_{\text{min}} \text{tr} [\mathbf{A}] + \mu \text{tr} [\mu \mathbf{A}]}{L}
$$

(8)

Substituting $\mathbf{M}_k$ for $\mu$ in equation (8), the excess MSE becomes

$$
J_{\text{excess}} = \frac{4 \mu_{\text{min}} \text{tr} [\mathbf{A}]}{L}
$$

(9)

when the expected value of the gradient is approximated by an $L$-tap low-pass MA filter. Recalling that the misadjustment is the ratio of the excess MSE to the minimum MSE, we obtain the following expression for the misadjustment

$$
M = \frac{4 \mu_{\text{min}} \sum_{i=0}^{L-1} \lambda_i^2}{L}
$$

(10)

The misadjustment of the gradient-based, variable step-size LMS algorithm is directly proportional to the minimum MSE, unlike the misadjustment of the classic LMS algorithm. This may be an undesirable feature of the gradient-based algorithms.

**Computational Complexity**

The classic LMS algorithm has, as its major advantage, a low computational cost. If we look at the basic equations that make up the LMS algorithm,

$$
\mathbf{w}_{k+1} = \mathbf{w}_k - \mathbf{M}_k \mathbf{v}_k
$$

(11)

we see that filtering an $M$ point data record with a $N$-tap adaptive filter requires $3MN$ multiplies and $2MN$ additions. Note that the constant multiplier typically seen in the LMS calculations has been wrapped up in the fixed $\bar{a}$.

The variable step-size LMS algorithm utilizes the same equations for the taps weight updates and for computation of the error $\mathbf{e}_k$ as does the classic LMS algorithm. The variable step-size parameter for each weight coordinate is given by

$$
\mu_k = \gamma \frac{1}{E [\hat{\mathbf{v}}_k]}
$$

(12)

where the expectation of the gradient is approximated by filtering the gradient with a low-pass MA filter. After filling the filter with gradient samples, the equation for the expected gradient is given by

$$
E [\hat{\mathbf{v}}_{k+1}] = \frac{1}{L} \left( \mathbf{e}_k \mathbf{x}_{k+1} - \hat{\mathbf{v}}_{k+1} \right)
$$

(13)

The total computational cost for the variable step-size LMS algorithm for an $M$ point data record is $7MN$ multiplies and $4MN$ additions, using the low-pass MA filter. Although the variable step-size LMS algorithm is 2-3 times as complex as the classic LMS algorithm, its anticipated increase in performance over the classic LMS algorithm should make it an attractive alternative.

**SIMULATION RESULTS**

**Description of the Simulation**

Algorithm testing was conducted using a 2-tap adaptive filter, to solve the noise cancellation problem (Figure 2). Both the classic LMS algorithm and the gradient-based variable step-size algorithm presented in this paper were simulated to allow head-to-head comparison.

The data set for the simulation is a sinusoidal signal combined with colored Gaussian noise. The eigenvalue spread (performance surface asymmetry) is obtained by passing white Gaussian noise through a one-pole recursive filter. The filter output is colored Gaussian noise with eigenvalue spread determined by the noise coloring filter. A nonstationary data set is obtained by creating a step change in the signal to noise ratio of the input data while still using the same noise power for the

Figure 2 - Simulation of noise cancellation problem.
noise cancellation filter. The nonstationary testing was more limited in scope and is meant to provide a qualitative comparison of the algorithms simulated.

Algorithm Performance

An example of this algorithm's performance for the noise cancellation problem is shown in Figure 3. The input signal (a) is a sinusoid summed with Gaussian noise \( n_k \), having an eigenvalue spread of 5 and a signal-to-noise ratio of 1. This input signal is given by

\[
d_k = \sin(2\pi k/30) + n_k
\]  

(14)

The output waveform (b) shows the results of the adaptive filtering on the noisy sinusoid. From (c) we can see that the variable step-size starts out at \( \mu_{\text{max}} \) and decreases to \( \mu_{\text{min}} \) as the adaptive filter converges. The step-size parameter shown here is for a single trial and shows how it is effected by short-term noise statistics.

![Graphs and plots related to algorithm performance.](image)

Figure 3 - Plots of variable step-size LMS algorithm performance.

- (a) Noisy input signal;
- (b) Filtered output signal;
- (c) Adaptation of the step-size parameter \( \mu \);
- (d) Weight tracks.

The weight tracking plot (d) shows the search for the Wiener optimal solution for an eigenvalue spread of 5. The weights quickly pull into one axis of the principal-axis coordinate system and then walk in to the final solution with a small amount of rattling. The MSE learning curves of Figures 4 - 5 show the convergence characteristics of this algorithm for various eigenvalue spreads and signal-to-noise ratios.

![MSE learning curves](image)

Figure 4 - Gradient-based variable step-size LMS algorithm. MSE learning curves for various eigenvalue spreads and with S/N = 1.
The weight tracks of Figure 8 (a-b) provide a comparison of these algorithm's misadjustment and their extent of tap weight rattling. Their parameters were chosen to yield similar convergence rates.

From these weight tracks, we can see that the variable step-size algorithm converges with much less rattling than the LMS algorithm. The comparison of the misadjustment of the two algorithms for various signal-to-noise ratios and eigenvalue spreads is shown below in Table 1.

For the signal-to-noise ratios and eigenvalue spreads shown, the gradient-based variable
step-size LMS algorithm converged with significantly less misadjustment than the classic LMS algorithm.

**Nonstationary Performance Testing**

The nonstationary performance of the two algorithms was compared by applying a step change in signal-to-noise ratio to the adaptive filter input, with the noise power at the tap weight inputs held constant. This input signal may be expressed as

\[ d_k = \sin(2\pi k/3) + \frac{1}{\sqrt{S/N}} x_k, \quad S/N = 16 \text{ for } k = 0-749, \]
\[ S/N = 1 \text{ for } k = 750-1500 \]

and the noise at the weight inputs is \( x_k \). The parameters of the two algorithms were selected to yield approximately the same speed of convergence. The resulting misadjustment for the gradient-based, variable step-size LMS algorithm is less than that of the classic LMS algorithm. The main benefit of the gradient-based variable step-size LMS algorithm is that the value of \( \mu_{\text{max}} \) used to obtain this weight convergence and tracking performance was significantly larger than for the LMS algorithm. This larger \( \mu_{\text{max}} \) will allow faster convergence for stationary signals or signals that vary slowly with time.

![Graph](image)

**Figure 9 - Nonstationary learning curves for eigenvalue spread = 5 and S/N = 16 for the first 750 samples, taking a step change to S/N = 1 for the last 750 samples.**

**CONCLUSIONS**

The performance results show that the gradient-based variable step-size algorithm is capable of faster convergence with less misadjustment than the classic LMS algorithm. This performance improvement holds in general for various signal-to-noise ratios, eigenvalue spreads and for both stationary and nonstationary signals. The complexity is only 2-3 times that of the LMS algorithm, making it an attractive choice for many adaptive filtering applications. A drawback of this algorithm is that the misadjustment is directly proportional to the minimum mean-square error, making this algorithm dependent upon the mean value of the signals to be filtered.

**REFERENCES**


