2012 IX International Symposium on Telecommunications (BIHTEL)
October 25-27, 2012, Sarajevo, Bosnia and Herzegovina

A zero-attracting variable step-size LMS algorithm for sparse system identification

Mohammad Shukri Salman[1], Mohammad N. S. Jahromi[2], Aykut Hocanin[2] and Osman Kukrer[2]

[1]Electrical and Electronic Engineering Dept.
Mevlana (Rumi) University, Selcuklu
Konya, Turkey
Email: mssalman@mevlana.edu.tr

[2]Electrical and Electronic Engineering Dept.
Eastern Mediterranean University, Famagusta
TRNC, Mersin 10, Turkey
{mohammad.sabet,aykut.hocanin,osman.kukrer}@emu.edu.tr

***Abstract*—In this paper, new adaptive algorithms are proposed to improve the performance of the variable step-size LMS (VSSLMS) algorithm when the system is sparse. The first proposed algorithm is the zero-attracting (ZA) VSSLMS. This algorithm outperforms the standard VSSLMS if the system is highly sparse. However, the performance of the ZA-VSSLMS algorithm deteriorates when the sparsity of the system decreases. To further improve the performance of the ZA-VSSLMS filter, the weighted zero-attracting (WZA)-VSSLMS algorithm is introduced. The algorithm performs the same or better than the ZA-VSSLMS if the system is highly sparse. On the other hand, when the sparsity of the system decreases, it performs better than the ZA-VSSLMS and better or the same as the standard VSSLMS algorithm. Also, both proposed algorithms have the same order of computational complexity as that of the VSSLMS algorithm ($O(N)$). For a system identification setting, the results indicate the high performance of the proposed algorithms in convergence speed and/or steady-state error under sparsity condition compared with the standard VSSLMS algorithm.**

I. Introduction

Adaptive filtering has application in many real-world systems [1], [2], such as system identification. Conventional adaptive algorithms, such as the least-mean-square (LMS), the normalized LMS (NLMS) algorithms and most of their variants [1], [3], severely suffer from being sensitive to highly correlated inputs and hence low convergence speed to the steady state mean-square-error (MSE) or even divergence. This is very crucial with a kind of filters with long impulse response where only a small percentage of coefficients are active and most of the others are zeros; i.e. the impulse response of the system is sparse. Variable step-size (VSS) methods [4], [5] aim to improve the convergence speed of the LMS algorithm, while preserving the steady-state performance. However, they still do not exploit the sparsity of the system.

In the past years, a number of adaptive algorithms that address sparsity are proposed based on applying a subset selection scheme during the filtering process [6]. These algorithms, basically, rely on the standard LMS (has bad performance if the input signal is correlated or the filter length is relatively large), i.e., the zero-attarcting LMS algorithm [7], a modified variable step-size NLMS algorithm [8] and a transform domain LMS algorithm for system identification [9], or the recursive-least-squares (RLS) algorithm (has high computational complexity).

In this paper, with the same order of computational complexity as that of the LMS algorithm ($O(N)$) but with much better performance, we propose a new approach called the zero-attracting variable step-size LMS (ZA-VSSLMA). This algorithm outperforms the standard VSSLMS algorithm [5] if the system is highly sparse. This is because of an introduced shrinkage term to the update equation of the algorithm that gives the ZA-VSSLMS algorithm the ability of attracting zeros and improves its performance when the system is highly sparse. However, the performance of the ZA-VSSLMS algorithm deteriorates when the sparsity of the system decreases. This is because of the fact that the shrinkage term in the ZA-VSSLMS does not distinguish between zero taps and non-zero taps. To further improve the performance of the ZA-VSSLMS filter, with a slight cost in the number of computations,the weighted zero-attracting (WZA)-VSSLMS algorithm is introduced by adding a log-sum penalty to the cost function of the VSSLMS algorithm. The algorithm performs similar or better than the ZA-VSSLMS if the system is highly sparse. On the other hand, if the sparsity of the system, it performs better than the ZA-VSSLMS and better or the same as the standard VSSLMS algorithm. Also, both proposed algorithms should be stable since the only introduced term is an ℓ_1 norm dependent (please check the ℓ_1 stability lemma [10]).

The paper is organized as follows. In section II, a brief review of the VSSLMS algorithm is introduced. In section III, the proposed ZA-VSSLMS algorithm is derived. In section IV, the proposed WZA-VSSLMS algorithm is derived. In section V, a discussion on the performance of the algorithms and their behavior is presented. In section VI, simulation results that compare the performance of the proposed algorithms with that of the standard VSSLMS algorithm in sparse systems, are

978-1-4673-4876-8/12/$31.00 ©2012 IEEE

shown. Finally, in the last section, conclusions are drawn.

II. Review of the Variable Step-Size LMS Algorithm

In the standard LMS algorithm, the cost function $J(k)$ is defined as

$$J(k) = \frac{1}{2}e^2(k) \tag{1}$$

where $e(k)$ is the error signal given by

$$e(k) = y(k) - \mathbf{w}^T(k)\mathbf{x}(k) \tag{2}$$

where $\mathbf{w}(k)$ is the coefficient weight vector of the adaptive algorithm with length N, $\mathbf{x}(k)$ is the system input tap vector and $y(k)$ is a sample of an observation of output signal given as

$$y(k) = \mathbf{w}^T(k)\mathbf{x}(k) + \eta(k) \tag{3}$$

where $\eta(k)$ is the additive observation noise which is independent from $x(k)$.

By the steepest descent method, the filter coefficient vector is updated by

$$\begin{aligned} \mathbf{w}(k+1) &= \mathbf{w}(k) - \frac{\mu}{2}\nabla J(k) \\ &= \mathbf{w}(k) + \mu e(k)\mathbf{x}(k) \end{aligned} \tag{4}$$

where μ is the adaptation step-size which controls the convergence and steady-state behavior of the LMS algorithm.

In [5], a variable step-size LMS (VSSLMS) algorithm is proposed. The algorithm uses a variable step-size which is estimated as

$$\mu'(k+1) = \alpha\mu'(k) + \gamma e^2(k) \tag{5}$$

with $0 < \alpha < 1$ and $\gamma > 0$. Then,

$$\mu(k) = \begin{cases} \mu_{max} & \text{if } \mu'(k+1) > \mu_{max} \\ \mu_{min} & \text{if } \mu'(k+1) < \mu_{min} \\ \mu'(k+1) & \text{otherwise} \end{cases} \tag{6}$$

where $0 < \mu_{min} < \mu_{max}$ and $\mu'(0)$ has no restriction (a good choice could be μ_{max}, [5]). As seen from (5), the step-size is always positive and controlled by α, γ and the prediction error $e(k)$. Intuitively speaking, a large prediction error at the beginning provides a large step-size which, in turn, leads to a faster tracking. When the prediction error starts decreasing, the step-size decreases and hence the misadjustment reduced. μ_{max} is chosen in a way to guarantee bounded mean-square-error (mse) [5]. A sufficient condition for that is

$$\mu_{max} \leq \frac{2}{3tr(\mathbf{R})} \tag{7}$$

where $\mathbf{R}$ is the expected value of the autocorrelation matrix of the input tap vector.

III. Zero-Attracting Variable Step-Size LMS Algorithm

A. Proposed Algorithm

Modifying the cost function given in (1) by adding the ℓ_1 norm penalty of the coefficient vector to the square of the prediction error gives [7]

$$J_1(k) = \frac{1}{2}e^2(k) + \lambda\|\mathbf{w}(k)\|_1 \tag{8}$$

where λ is a positive constant. By the gradient method, the update equation becomes

$$\begin{aligned} \mathbf{w}(k+1) &= \mathbf{w}(k) - \frac{\mu(k)}{2}\nabla J_1(k) \\ &= \mathbf{w}(k) + \mu(k)e(k)\mathbf{x}(k) - \rho(k)f(\mathbf{w}(k)) \end{aligned} \tag{9}$$

where $\rho(k) = \lambda\mu(k)$ and $f(\mathbf{w}(k))$ is the sign function ($f(\mathbf{w}(k)) = sgn(\mathbf{w}(k))$).

Comparing (4) and (9), in (9) there is an extra term $(-\rho(k)f(\mathbf{w}(k)))$. This term always attracts the tap coefficients to zero. In other words, if the majority of the coefficients of $\mathbf{w}$ is zero, the zero-attractor will speed up the convergence behavior. $\rho(k)$ controls the strength of the zero-attractor. Due to this, we call the algorithm the zero-attracting VSSLMS (ZA-VSSLMS).

B. Discussion on the Convergence Behavior of the ZA-VSSLMS

From (9), it is known that $-\rho(k)sgn(\mathbf{w}(k))$ is bounded between $-\rho(k)$ and $\rho(k)$ or $-\lambda\mu_{max}$ and $-\lambda\mu_{min}$. Hence the convergence criterion of the ZA-VSSLMS algorithm is the same as that of the VSSLMS algorithm [5]. However, the ZA-VSSLMS algorithm provides a biased estimate of the coefficient vector of the VSSLMS and yields a lower mse if the system sparse.

IV. Weighted Zero-Attracting Variable Step-Size LMS Algorithm

The shrinkage in the ZA-VSSLMS algorithm does not distinguish between the zero taps and the non-zero taps. Hence, its performance would deteriorate for less sparse systems. So weighting the the ZA term in (9) will enhance its performance if the system is less sparse [7].

Redefining the cost function in (1) as

$$J_2(k) = \frac{1}{2}e^2(k) + \lambda' \sum_{i=1}^{N} log\left(1 + \frac{|w_i|}{\zeta'}\right) \tag{10}$$

where λ' and ζ' are positive constants. Then, the same as before, by applying the gradient method we get

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k)e(k)\mathbf{x}(k) - \rho(k)\frac{sgn\left[\mathbf{w}(k)\right]}{1+\zeta|\mathbf{w}(k)|} \tag{11}$$

where $\rho(k) = \frac{\mu(k)\lambda'}{\zeta'}$, $\zeta = \frac{1}{\zeta'}$ and $|\mathbf{w}(k)| = \sqrt{\sum_{i=1}^{N}(w_i^2)}$.

The weighted zero-attracting effect appears only on the taps that have magnitude comparable to $\frac{1}{\zeta}$ and there is a little shrinkage exerted on the taps whose magnitude is much greater than $\frac{1}{\zeta}$ [7]. As a result of that, the bias of the weighted zero-attracting VSSLMS (WZA-VSSLMS) algorithm can be reduced.

V. Discussion

In the VSSLMS algorithm, the update equation can be expressed as

$$\begin{pmatrix} \text{new tap} \\ \text{weights} \end{pmatrix} = \begin{pmatrix} \text{old tap} \\ \text{weights} \end{pmatrix} + \begin{pmatrix} \text{gradient} \\ \text{correction} \end{pmatrix} \qquad (12)$$

where the filter coefficients are updated along the negative gradient direction. Also, the update equations of the ZA-VSSLMS and WZA-VSSLMS algorithms can be expressed as [11]

$$\begin{pmatrix} \text{new tap} \\ \text{weights} \end{pmatrix} = \begin{pmatrix} \text{old tap} \\ \text{weights} \end{pmatrix} + \begin{pmatrix} \text{gradient} \\ \text{correction} \end{pmatrix} + \begin{pmatrix} \text{zero} \\ \text{attraction} \end{pmatrix} \qquad (13)$$

where the zero-attracting term in (13) (equivalent to $-\rho(k) f(\mathbf{w}(k))$ and $-\rho(k)\frac{sgn[\mathbf{w}(k)]}{1+\zeta|\mathbf{w}(k)|}$ in (9) and (11), respectively) imposes an attraction to zero on small coefficients. Particularly, if the filter weight coefficient is positive, it will decrease and if it is negative, it will increase.

In (9) and (11), $\rho(k)$ results in a trade off between adaptation speed and adaptation quality. A large value of $\rho(k)$ results in a faster convergence since the intensity of attraction increases as $\rho(k)$ increases. On the other hand, the steady-state misalignment also increases as $\rho(k)$ increases. After the adaptation reaches steady state, most filter weights are near to zero due to the sparsity. So those near-zero coefficients $w_i(k)$ will move randomly in the small neighborhood of zero, driven by the attraction term as well as the gradient noise term. Therefore, a large $\rho(k)$ results in a large steady-state misalignment.

VI. Simulation Results

In this section, the performances of the WZA-VSSLMS and ZA-VSSLMS algorithms are compared with that of the standard VSSLMS algorithm in a sparse system identification setting. All the experiments are implemented with 200 independent Monte-Carlo runs. All the parameters used were selected by extensive simulations to give optimal performance for each algorithm.

In the first experiment, in order to exploit the sparsity of the system, we use a filter of 10 coefficients in the time varying system. Initially, the one random tap of the unknown system is set to 1 and the others to zero; to have a sparsity of 1/10. After 500 iterations, all the 5 random taps are set to 1 and the rest kept to be zero, i.e., a sparsity of 5/10. Finally, after 1000 iterations all the taps are set with values of -1 and 1 randomly, leaving a completely non-sparse system.

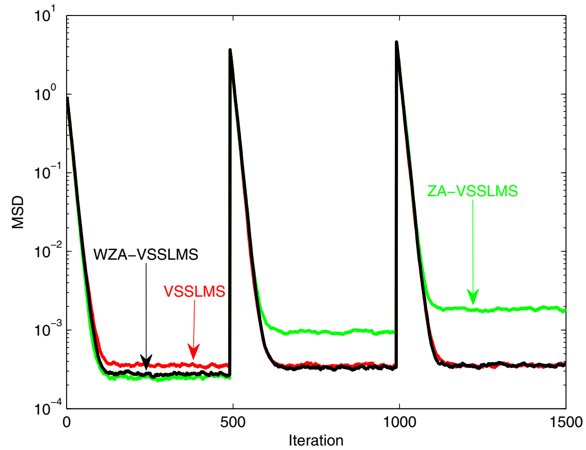

Fig. 1. Tracking and steady-state behaviors of a 10 tap adaptive filter driven by a white input signal.

The input signal and the observed noise are both assumed to be white Gaussian random sequences with variances 1 and 10^{-3}, respectively, in order to have a 30dB signal-to-noise ratio (SNR). The performance measure used is the mean-square deviation (MSD) defined as $MSD = E\|\mathbf{h} - \mathbf{w}(n)\|^2$. Simulations are done with the following parameters: For the VSSLMS: $\mu_{min} = 0.05$, $\mu_{max} = 0.07$, $\gamma = 0.048$ and $\alpha = 0.97$. For the ZA-VSSLMS: $\mu_{min} = 0.05$, $\mu_{max} = 0.07$, $\gamma = 0.048$, $\alpha = 0.97$ and $\rho = 1 \times 10^{-3}$. For the WZA-VSSLMS: $\mu_{min} = 0.05$, $\mu_{max} = 0.07$, $\gamma = 0.048$, $\alpha = 0.97$ and $\rho = 0.55 \times 10^{-3}$. 1 shows the average MSD estimate of all algorithms. As seen from the MSD results, when the system is very sparse (before the 500^{th} iteration), the ZA-VSSLMS and the WZA-VSSLMS algorithms converge at the same rate of the VSSLMS algorithm but with better steady-state MSD (1dB better). After the 500^{th} iteration, as the number of non-zero taps increases, we see that the performance of the ZA-VSSLMS algorithm deteriorates since the shrinkage in the ZA-VSSLMS algorithm does not distinguish between the zero taps and non-zero taps. However, the WZA-VSSLMS algorithm converges at the same rate to same MSD as that of the VSSLMS algorithm even if the system is non-sparse.

The second experiment simulates a general sparse system. The driving signal and the observed noise are the same as the first experiment. The unknown system is selected as a general sparse system of 256-taps length with 28 randomly distributed non-zero coefficients. Simulations are done with the following parameters: For the VSSLMS: $\mu_{min} = 0.003$, $\mu_{max} = 0.006$, $\gamma = 0.0048$ and $\alpha = 0.97$. For the ZA-VSSLMS: $\mu_{min} = 0.003$, $\mu_{max} = 0.006$, $\gamma = 0.0048$, $\alpha = 0.97$ and $\rho = 2 \times 10^{-4}$. For the WZA-VSSLMS: $\mu_{min} = 0.003$, $\mu_{max} = 0.006$, $\gamma = 0.0048$, $\alpha = 0.97$ and $\rho = 3 \times 10^{-4}$. Fig. 2 shows the average MSD estimate of all algorithms. As shown, the ZA-VSSLMS algorithm converges to 1.5dB lower MSD and 1000 iterations faster convergence rate than that of the VSSLMS algorithm. The WZA-VSSLMS algorithm converges with the same rate as that of the ZA-VSSLMS but with 1dB lower

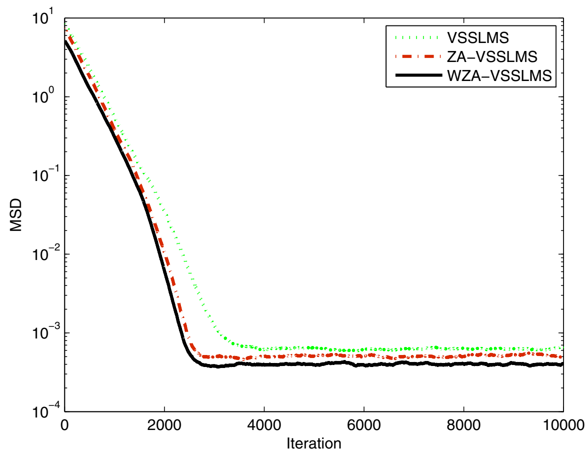

Fig. 2. Tracking and steady-state behaviors of a 256 tap acoustic echo canceller driven by a white input signal.

MSD (2.5dB better than that of the VSSLMS algorithm).

VII. Conclusions

In this paper, a novel adaptive ZA-VSSLMS and WZA-VSSLMS algorithms for sparse system identification are proposed. The proposed ZA-VSSLMS algorithm incorporates the ℓ_1 norm penalty of the coefficients into its cost function, which results in a shrinkage in the update formula. This shrinkage improves the performance of the adaptive filter when the majority of coefficients are zero. However, this shrinkage makes the performance of the algorithm deteriorate when the majority of system coefficients are non-zero. For this, a WZA-VSSLMS algorithm is proposed. The WZA-VSSLMS algorithm incorporates a log-sum penalty of the coefficients into its cost function, which improves the performance of the adaptive VSSLMS algorithm when the system is sparse. The WZA-VSSLMS algorithm is superior to the standard VSSLMS algorithm in both convergence rate and steady-state behaviors when the system is sparse, and it performs the same as the VSSLMS when the system is non-sparse.

References

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice Hall, Eaglewood Cliffs, NJ, 1985.

[2] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, NJ, 2002.

[3] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley, New York, NY, USA, 2003.

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

[5] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1633-1642, 1992.

[6] S. Kawamura and M. Hatori, "A tap selection algorithm for adaptive filters," *Proceedings of ICASSP*, vol. 11, pp. 2979-2982, 1986.

[7] Y. Chen, Y. Gu and A. O. Hero, "Sparse LMS for system identification," *IEEE International Conference on Acoustic, Speech, and Signal Processing* (ICASSP2009), pp. 3125-3128, 2009.

[8] S. Zhao, Z. Man and S. Khoo, "A class of modified variable step-size NLMS algorithms for system identification," *The* 4^{th} *IEEE Conference on Industrial Electronics and Applications* (ICIEA 2009), pp. 2987-2991, May 2009.

[9] K. Shi and X. Ma, "Transform domain LMS algorithms for sparse system identification," *IEEE International Conference on Acoustic, Speech, and Signal Processing* (ICASSP2010), pp. 3714-3717, 2010.

[10] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289-1306, 2006.

[11] Y. Gu, J. Jin and S. Mei, "ℓ_0 norm constraint LMS algorithm for sparse system identification," *IEEE Signal Processing Letter*, vol. 16, pp. 774-777, Sept. 2009.