NOVEL VARIABLE STEP SIZE NLMS ALGORITHMS FOR ECHO CANCELLATION

Mohammad Asif Iqbal

Qualcomm Incorporated
5775 Morehouse Drive
San Diego, CA 92121
Email: asif.umr@gmail.com

Steven L. Grant

University of Missouri-Rolla
228 Emerson Electric Co. Hall
Rolla, MO 65409
Email: sgrant@umr.edu

ABSTRACT

In this paper we present two new variable step size (VSS) methods for adaptive filters. These VSS methods are so effective, they eliminate the need for a separate double-talk detection algorithm in echo cancellation applications. The key feature of both approaches is the introduction of a new near-end signal energy estimator (NESEE) that provides accurate and computationally efficient estimates even during double-talk and echo path change events. The first VSS algorithm applies the NESEE to the recently proposed Nonparametric VSS NLMS (NPVSS-NLMS) algorithm. The resulting algorithm has excellent convergence characteristics with an intrinsic immunity to double-talk. The second approach is somewhat more ad hoc. It is composed of a combination of an efficient echo path change detector and the NESEE. This VSS method also has excellent convergence, double talk immunity, and computational efficiency. Simulations demonstrate the efficacy of both proposed algorithms.

Index Terms— Double-talk detection, acoustic echo canceller, echo path change detection, and AEC.

1. INTRODUCTION

Adaptive algorithms are extensively used in signal processing applications [1]. The normalized least mean square (NLMS) algorithm is highly popular because of it robustness and simplicity. The stability and adaptation speed of this algorithm is governed by a step size parameter. The choice of this parameter reflects the tradeoff between fast convergence on one hand and poor steady state misalignment on the other. To address these conflicting requirements, numerous variable step size (VSS) algorithms have been proposed [2] [3] [4] [5] [6]. A key parameter in most VSS algorithms is the estimate of the energy of the near-end signal. Often, minimum statistics methods [7] are used, but these only estimate the energy of the background noise of the near-end signal, not the energy of the total signal – the background *and* the near-end talker. Other attempts have been made to estimate the near-end talker's energy, often based on deviations of a subset of the adaptive filter coefficients, but these methods are usually not robust.

Figure 1 shows the basic block diagram of an AEC. The far-end signal $\mathbf{x}(n)$ is filtered through the echo-path $\mathbf{h}$ to get the echo signal

$$y(n) = \mathbf{h}^{\mathbf{T}}\mathbf{x}(n) \tag{1}$$

where

$$\mathbf{h} = [h_0(n)\ \ h_1(n)\ \ ....\ ,\ \ h_{l-1}(n)]^T,$$

$$\mathbf{x}(n) = [x(n)\ \ x(n-1)\ \ ....\ ,\ \ x(n-l+1)]^T,$$

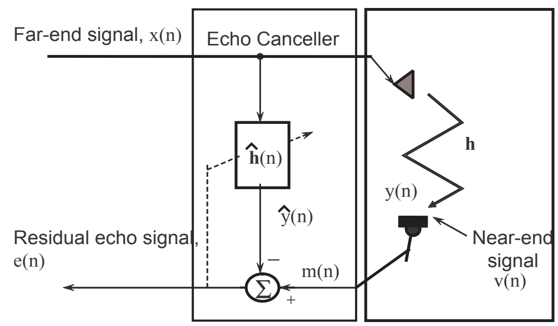

Fig. 1. Basic AEC Model

and l is the length of the adaptive filter. This echo signal is acoustically added to the near-end speech signal $v(n)$ to get the microphone signal:

$$m(n) = y(n) + v(n) + w(n) \tag{2}$$

where $w(n)$ is the additive background noise. We define the error signal at time n as

$$e(n) = m(n) - \hat{\mathbf{h}}^{\mathbf{T}}(n)\mathbf{x}(n) \tag{3}$$

It is used to adapt the l taps of the AEC's adaptive filter $\hat{\mathbf{h}}(n)$ to generate an estimate of the echo,

$$\hat{y}(n) = \hat{\mathbf{h}}^T(n)\mathbf{x}(n) \tag{4}$$

When the near-end talker, $v(n)$ is active or when the speech comes from both the far-end and near-end, identification of the echo-path becomes problematic as the adaptive filter coefficients diverge from the true echo-path if the adaptation is not halted. To prevent this, a double-talk detector is used to stop the AEC's filter adaptation during periods of near-end speech.

Instead of using an explicit double-talk detector, we propose using a variable step size NLMS that automatically addresses this adaptation control problem. More recently Benesty, et. al. [8] proposed a Nonparametric VSS-NLMS algorithm (NPVSS-NLMS) that is very easy to control in real world applications. We, derive a novel near-end energy detector and modify the NPVSS-NLMS algorithm to handle the double-talk adaptation control problem. A second, novel VSS-NLMS adaptive algorithm is also introduced. Simulations in the context of acoustic echo cancellation show that the proposed algorithms have better convergence at the same tracking rate as compared to the classical NLMS algorithm with maximum step size. More importantly, the proposed algorithms have good immunity to double-talk and no divergence is observed during periods of near-end speech. Furthermore, both of these variable step sizes have very low computational complexity.

This paper is structured as follows. We introduce the novel near-end energy estimator in Section 2. In Section 3, we develop two novel variable step size NLMS algorithms for echo cancellation. A comprehensive study on the proposed algorithms is done in Section 4 which is followed by a summary and conclusions in Section 5.

2. NOVEL NEAR-END ENERGY ESTIMATOR

In this section we derive the near-end signal energy estimator (NESEE) and an explicit convergence statistic, which subsequently will be used in deriving the variable step size algorithms. Referring to figure 1, the cross-correlation between the far-end vector $\mathbf{x}(n)$ and the residual error $e(n)$ is given by:

$$\begin{aligned}\mathbf{r}_{e\mathbf{x}}(n) &= E[\mathbf{x}(n)e(n)] \\ &= R_{\mathbf{xx}}\Delta\mathbf{h}(n-1)\end{aligned}$$

where $E[\ ]$ denotes the mathematical expectation, $\mathbf{R_{xx}} = E[\mathbf{xx^T}]$ and $\Delta\mathbf{h}(n-1) = \mathbf{h} - \hat{\mathbf{h}}(n-1)$. The variance of the cancellation error $e(n)$ is given by:

$$\begin{aligned}\sigma_e^2(n) &= E[e^2(n)] \\ &= \Delta\mathbf{h}^T(n-1)\mathbf{R_{xx}}\Delta\mathbf{h}(n-1) + \sigma_v^2(n) + \sigma_w^2(n)\end{aligned} \tag{5}$$

where $\sigma_v^2(n)$ is the power of the near-end signal and $\sigma_w^2(n)$ is the background noise power.

We define the NESEE as:

$$\gamma(n) = \sigma_e^2(n) - \frac{1}{\sigma_x^2(n)}\mathbf{r}_{e\mathbf{x}}(n)^T\mathbf{r}_{e\mathbf{x}}(n) \tag{6}$$

$$\approx \sigma_v^2(n) + \sigma_w^2(n) \tag{7}$$

where $\sigma_x^2(n)$ is the variance of the excitation signal. The values of $\sigma_e^2(n)$, $\sigma_x^2(n)$ and $\mathbf{r}_{e\mathbf{x}}(n)$ in equation (6) are exact and not available in practice. An easily computed estimate is given by:

$$\hat{\gamma}(n) = \hat{\sigma}_e^2(n) - \frac{1}{\hat{\sigma}_x^2(n)}\hat{\mathbf{r}}_{e\mathbf{x}}(n)^T\hat{\mathbf{r}}_{e\mathbf{x}}(n) \tag{8}$$

where the estimates denoted by hat are obtained using the exponential recursive weighting algorithm, [9] [10]:

$$\begin{aligned}\hat{\mathbf{r}}_{e\mathbf{x}}(n) &= \lambda\hat{\mathbf{r}}_{e\mathbf{x}}(n-1) + (1-\lambda)\mathbf{x}(n)e(n) \\ \hat{\sigma}_x^2(n) &= \lambda\hat{\sigma}_x^2(n-1) + (1-\lambda)x^2(n) \\ \hat{\sigma}_e^2(n) &= \lambda\hat{\sigma}_e^2(n-1) + (1-\lambda)e^2(n)\end{aligned} \tag{9}$$

where $\lambda = 0.97$ is the exponential weighting factor. We plot $\gamma(n)$ and the true near-end signal energy at a near-end to far-end ratio (NFR) of -5dB and at a signal to noise ratio (SNR) of 12 dB in Figure 2. These simulations demonstrate the efficacy of the proposed statistic, the estimate is almost always equivalent to the true energy.

Next, we look at the echo-path change detection statistic proposed in [11] which is a direct measure of the adaptive filter's convergence. Referring to figure 1, the cross-correlation between the microphone signal $m(n)$, and the cancellation error $e(n)$ is given by:

$$\begin{aligned}r_{em}(n) &= E[m(n)e(n)] \\ &= \mathbf{h^T}\mathbf{R_{xx}}\Delta\mathbf{h}(n-1) + \sigma_v^2(n)\end{aligned} \tag{10}$$

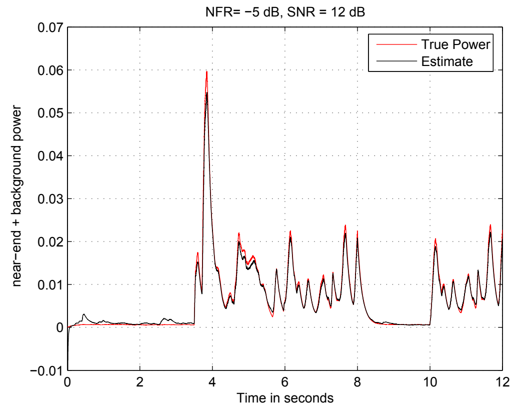

Fig. 2. Tracking near-end energy + background noise at a SNR of 12 dB and NFR of -5 dB.

The variance of the microphone signal is given by:

$$\begin{aligned}\sigma_m^2(n) &= E[m^2(n)] \\ &= \mathbf{h^T}\mathbf{R_{xx}}\mathbf{h} + \sigma_v^2(n).\end{aligned} \tag{11}$$

and the variance of the cancellation error $e(n)$ is given in equation 5.

We define our convergence statistic to be

$$\xi(n) = \left|\frac{r_{em}(n) - \sigma_e^2(n)}{\sigma_m^2(n) - r_{em}(n)}\right| \tag{12}$$

substituting equations 10 , 11 and 5 in 12 we get:

$$\xi(n) = \left|\frac{\Delta\mathbf{h}^T(n-1)\mathbf{R_{xx}}\hat{\mathbf{h}}(n-1)}{\mathbf{h^T}\mathbf{R_{xx}}\hat{\mathbf{h}}(n-1)}\right| \tag{13}$$

We observe from equation 13, for $\mathbf{h} \approx \hat{\mathbf{h}}(n-1)$, $\xi(n) \approx 0$ and for $\mathbf{h} \neq \hat{\mathbf{h}}(n-1)$, $\xi(n) > 0$. Thus the proposed statistic is a good measure of the adaptive filter's convergence. It has been shown in [11], that the proposed statistic accurately tracks the residual echo energy due to echo-path variations.

3. VSS-NLMS ALGORITHMS

In this section, we introduce two different novel variable step size (VSS) NLMS algorithms for effective double-talk control.

3.1. Extension of the Nonparametric VSS-NLMS Algorithm

In this section, we extend the idea of the recently proposed Nonparametric VSS-NLMS algorithm [8], for effective double-talk control. The NPVSS-NLMS update equations are given by:

$$e(n) = m(n) - \hat{\mathbf{h}}^{\mathbf{T}}(n-1)\mathbf{x}(n) \tag{14}$$

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mu_{NPVSS}(n)\frac{\mathbf{x}(n)e(n)}{\mathbf{x}^T(n)\mathbf{x}(n) + \delta}$$

where

$$\mu_{NPVSS}(n) = 1 - \frac{\sigma_w}{\sigma_e(n)} \tag{15}$$

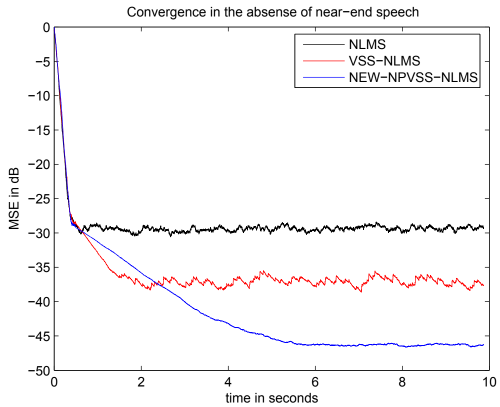

Fig. 3. Misalignment of the adaptive filter. (a) NLMS, (b) VSS-NLMS and (c) NEW-NPVSS-NLMS. The excitation signal is the white Gaussian noise, $L = 512$, and SNR is 30 dB.

where $\sigma_e^2(n) = E[e^2(n)]$ is the power of the error signal and σ_w^2 is the background noise level, which was assumed to be known. Instead of using the background noise level σ_w, we propose using the near-end + background noise estimate $\gamma(n)$ proposed in Section 2 i.e.

$$\mu_{NEW-NPVSS}(n) = \begin{cases} 1 - \frac{\hat{\gamma}(n)}{\hat{\sigma}_e(n)} & \text{if } \xi < \epsilon \\ 1 & \text{Otherwise.} \end{cases} \quad (16)$$

where ϵ is a small positive quantity ($\epsilon > 0$).

When the filter is converged ($\xi_{EP}(n) < \epsilon$) and double-talk is introduced, we have, $\gamma(n) \approx \sigma_e(n) \approx \sigma_v(n)$ and hence the adaptation step size is $\mu_{NEW-NPVSS}(n) \approx 0$. Further, during echo-path variations (in the absence of near-end speech), we have, $\mu_{NEW-NPVSS}(n) = 1$ since $\xi_{EP}(n) > \epsilon$. This way, we force very slow adaptation during periods of near-end speech and fast adaptation during echo-path variations.

3.2. Novel VSS-NLMS Algorithm

In this section, we introduce a second variable step size NLMS algorithm. The update equations are given by:

$$\begin{aligned} e(n) &= m(n) - \hat{\mathbf{h}}^{\mathbf{T}}(n-1)\mathbf{x}(n) \quad (17) \\ \hat{\mathbf{h}}(n) &= \hat{\mathbf{h}}(n-1) + \mu_{VSS}(n)\frac{\mathbf{x}(n)e(n)}{\mathbf{x}^T(n)\mathbf{x}(n)+\delta} \end{aligned}$$

We define our new variable step size to be:

$$\mu_{VSS}(n) = \frac{\xi(n)}{\xi(n)+\gamma(n)} \quad (18)$$

where $\xi(n)$ and $\gamma(n)$ are as defined in equations 13 and 7 respectively. When the filter is converged (i.e. $\xi(n) < \epsilon$) and double-talk is introduced, we have:

$$\begin{aligned} \mu_{VSS}(n) &\approx \frac{\epsilon}{\epsilon + \sigma_v(n)} \\ &\approx 0 \end{aligned} \quad (19)$$

thus, the filter adaptation is almost inhibited. During echo-path variations $\xi(n) >> \epsilon$ (in the absence of near-end speech), we have,

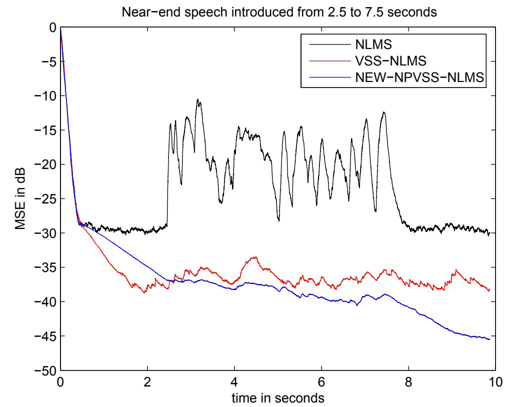

Fig. 4. Misalignment of the adaptive filter during double-talk. (a) NLMS, (b) VSS-NLMS and (c) NEW-NPVSS-NLMS. The excitation signal is the white Gaussian noise, $L = 512$, and SNR is 30 dB.

$\gamma(n) \approx \sigma_w(n) \approx \epsilon$, hence,

$$\begin{aligned} \mu_{VSS}(n) &\approx \frac{\xi(n)}{\xi(n)+\epsilon} \\ &\approx 1 \end{aligned} \quad (20)$$

Thus, the filter adapts to these new variations.

4. SIMULATION RESULTS

In our simulations, we use 512 taps of a measured room response of a $10' \times 10' \times 8'$ room as the echo-path. The same length is used for the adaptive filters. We use white Gaussian signal and speech as the excitation signal. These signals are sampled at 8 kHz. An independent white Gaussian noise signal is added to the microphone signal $m(n)$ at a SNR of 30dB. We use the convergence of the normalized misalignment (in dB), $20\log_{10}(\|\mathbf{h} - \hat{\mathbf{h}}(n)\|_2/\|\mathbf{h}\|_2)$, as a measure of performance. Figure 3 compares the convergence of the NLMS algorithm (with step size = 1), with the convergence of the proposed NEW-NPVSS NLMS and VSS-NLMS algorithms. It can be observed from this figure that when the convergence rate is same for all the algorithms, NEW-NPVSS NLMS algorithm has a better misalignment by almost 18 dB. It can also be observed that the VSS-NLMS is 10dB better than the NLMS.

Next, we study the affects of double-talk on misalignment. Here, again we use white Gaussian signal as the excitation signal and introduce near-end speech after 2.5 seconds for a period of 5 seconds. We observe in Figure 4, that the NLMS algorithm diverges during double-talk whereas the proposed algorithms are immune. We observe no divergence during periods of near-end speech for VSS-NLMS and NEW-NPVSS NLMS algorithms. With these algorithms we do not need an explicit double-talk detector to inhibit adaptation during periods of near-end speech. This is very desirable and is computationally very attractive.

Tracking is another important aspect in acoustic echo cancellation, echo-paths vary randomly and rapidly. It is important that an adaptive filter tracks fast to avoid annoying echo. Figure 5, compares these algorithms during echo-path variations. To create these variations, the room response is shifted to right by 10 samples after 5 seconds. According to the simulation, the convergence rate is

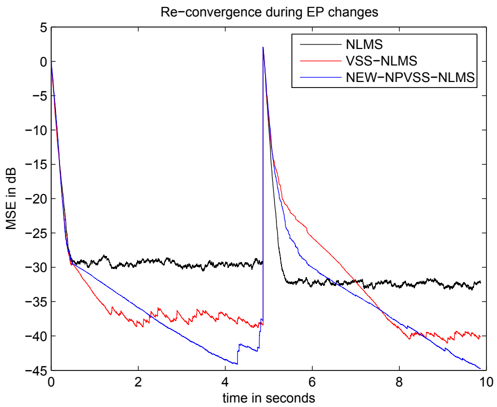

Fig. 5. Misalignment of the adaptive filter during echo-path variations. (a) NLMS, (b) VSS-NLMS and (c) NEW-NPVSS-NLMS. The excitation signal is the white Gaussian noise, $L = 512$, and SNR is 30 dB.

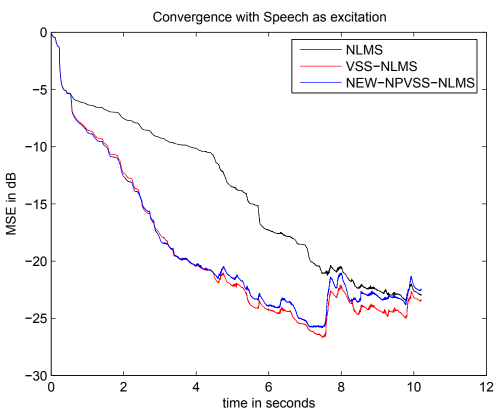

Fig. 6. Misalignment of the adaptive filter. (a) NLMS, (b) VSS-NLMS and (c) NEW-NPVSS-NLMS. The excitation signal is speech, $L = 512$, and SNR is 30 dB.

marginally decreased for the proposed algorithms as compared to the NLMS with its maximum step size. This is the price paid for total immunity towards double-talk.

Finally we compare our algorithms with speech signals as inputs. It can be observed from Figure 6, that proposed algorithms have similar performance and are substantially better than the NLMS algorithm with maximum step size.

5. CONCLUSIONS

In any adaptive algorithm, we need to find a compromise between convergence and steady state misalignment. However, in many applications this compromise may not be satisfactory. To address this problem, we have proposed two different novel variable step size NLMS algorithms. The step size can be easily computed in both the cases, simulations in the context of acoustic echo cancellation have shown better performance as compared to the classical NLMS algorithm. Also the proposed algorithms are immune to double-talk, which is highly desirable and computationally very attractive.

6. REFERENCES

[1] Simon Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, 1991.

[2] R.W. Harris, D.M. Chabries, and F.A. Bishop, "A variable step size adaptive filter algorithm," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. 34, pp. 309–316, April 1986.

[3] R.H. Kwong and E.W. Johnston, "A variable step size lms algorithm," *IEEE Transactions on Signal Processing*, vol. 40, pp. 1633–1642, July 1992.

[4] V.J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Transactions on Signal Processing*, vol. 41, pp. 2075–2087, June 1993.

[5] A. Mader, H. Puder, and G.U. Schmidt, "Step-size control for acoustic echo cancellation filters-an overview," *Signal Processing*, vol. 47, pp. 864–872, March 1999.

[6] H.C. Chin, A.H. Sayed, and W.J. Song, "Variable step size nlms and affine projection algorithms," *IEEE Signal Processing Letter*, vol. 11, pp. 132–135, Feb 2004.

[7] R. Martin, "Spectral subtraction based on minimum statistics," in *Proceedings of the 7th European Signal Processing Conference*, Edinburgh, Scotland, September 1994, pp. 1182–1185.

[8] Jacob Benesty, Hernan Rey, Leonardo Rey Vega, and Sara Tressens, "A nonparametric vss nlms algorithm," *IEEE Signal Processing Letters*, vol. 13, pp. 581–584, October 2006.

[9] B. Porat, "Second-order equvalence of rectangular and exponential windows in least-squares estimation of autoregressive processes," *IEEE Transactions on Acoustics,Speech and Signal Processing*, 1985.

[10] Y. Hua, *Adaptive Filter Theory and applications*, Ph.D. thesis, South-East University, Tiangsu, China, March 1989.

[11] Mohammad Asif Iqbal and Steven L. Grant, "A novel normalized cross-correlation based echo-path change detector," in *Proceedings of the IEEE Region 5 Conference*, Arkansas, April 2007.