Variable Step-Size NLMS and Affine Projection Algorithms

Hyun-Chool Shin, Ali H. Sayed, *Fellow, IEEE*, and Woo-Jin Song, *Member, IEEE*

***Abstract*—This letter proposes two new variable step-size algorithms for normalized least mean square and affine projection. The proposed schemes lead to faster convergence rate and lower misadjustment error.**

***Index Terms*—Adaptive filters, affine projection algorithm, normalized least mean square (NLMS), variable step-size.**

I. Introduction

COLORED input data tend to deteriorate the convergence performance of least mean square (LMS)-type adaptive filters [1]–[3]. To overcome this problem, Ozeki and Umeda proposed an affine projection algorithm (APA) [4] that is based on affine subspace projections. In contrast to NLMS, which updates the weight vector based only on the current input vector, APA updates the weight vector based on K input vectors. In both cases of normalized least mean square (NLMS) and APA, the step-size μ governs the rate of convergence and the steady-state excess mean-square error. To meet the conflicting requirements of fast convergence and low misadjusment, the step-size needs to be controlled. In standard LMS, various schemes for controlling the step-size have been proposed [5]–[8]. The performance of these schemes is determined by how accurately they can estimate how far the filter is from optimal performance. Various criteria have been developed for this purpose. Kwong and Johnston [5] used squared instantaneous errors. To improve noise immunity under Gaussian noise, Aboulnasr and Mayyas [6] used the squared autocorrelation of errors at adjacent time, and Pazaitis and Constantinides [7] adopted the fourth-order cumulant of instantaneous error. In [8] and in some of the references therein, the optimum step-size for NLMS is obtained by minimizing the mean-square derivation at each iteration. These criteria work effectively for LMS but are not directly applicable to APA. This is because the instantaneous error of APA is a vector, not a scalar quantity as in LMS.

Manuscript received January 7, 2003; revised March 6, 2003. This work was supported in part by the National Science Foundation under Grants ECS-9820765 and CCR-0208573. The work was performed while the first author was a visiting graduate student at the UCLA Adaptive Systems Laboratory. His visit was supported by the Brain Korea (BK) 21 Program funded by the Ministry of Education and in part by HY-SDR Reserch Center at Hanyang University under the ITRC Program of MIC, Korea. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Alle-Jan van der Veen.

H.-C. Shin and W.-J. Song are with Division of Electronics and Computer Engineering, Pohang University of Science and Technology (POSTECH), Pohang, Korea.

A. H. Sayed is with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA (e-mail: sayed@ee.ucla.edu).

Digital Object Identifier 10.1109/LSP.2003.821722

In this letter, we propose a criterion that provides a measure of the adaptive filter state, i.e., it indicates how close the adaptive filter is to optimal performance. Using this criterion, we develop a variable step-size APA, which has faster convergence rate and lower misadjustment error than existing schemes. We also develop, as a special case, a variable step-size NLMS algorithm. Throughout the letter, the following notations are adopted: $\|\cdot\|$ is the Euclidean norm of a vector; and $\mathrm{Tr}(\cdot)$ is the trace of a matrix.

II. Variable Step-Size APA

Consider data $\{d(i)\}$ that arise from the model

$$d(i) = \mathbf{u}_i \mathbf{w}^\circ + v(i) \tag{1}$$

where $\mathbf{w}^\circ$ is an unknown column vector that we wish to estimate, $v(i)$ accounts for measurement noise and $\mathbf{u}_i$ denotes $1 \times M$ row input (regressor) vectors. Let $\mathbf{w}_i$ be an estimate for $\mathbf{w}^\circ$ at iteration i. The affine projection algorithm computes $\mathbf{w}_i$ via

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu U_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i \tag{2}$$

where

$$U_i = \begin{bmatrix} \mathbf{u}_i \\ \mathbf{u}_{i-1} \\ \vdots \\ \mathbf{u}_{i-K+1} \end{bmatrix} \quad \mathbf{d}_i = \begin{bmatrix} d(i) \\ d(i-1) \\ \vdots \\ d(i-K+1) \end{bmatrix}$$

$\mathbf{e}_i = \mathbf{d}_i - U_i \mathbf{w}_{i-1}$, and μ is the step-size.

A. Optimal Variable Step-Size

The update recursion (2) can be written in terms of the weight-error vector, $\tilde{\mathbf{w}}_i = \mathbf{w}^\circ - \mathbf{w}_i$, as

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \mu U_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i. \tag{3}$$

Squaring both sides and taking expectations, we find that the mean-square deviation (MSD) satisfies

$$\begin{aligned} E\|\tilde{\mathbf{w}}_i\|^2 &= E\|\tilde{\mathbf{w}}_{i-1}\|^2 - 2\mu \,\mathrm{Re}\left(E\left[\mathbf{e}_i^* \left(U_i U_i^*\right)^{-1} U_i \tilde{\mathbf{w}}_{i-1}\right]\right) \\ &\quad + \mu^2 E\left[\mathbf{e}_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i\right] \\ &\triangleq E\|\tilde{\mathbf{w}}_{i-1}\|^2 - \Delta(\mu). \end{aligned} \tag{4}$$

If we choose μ such that $\Delta(\mu)$ is maximized, then this choice guarantees that the MSD will undergo the largest decrease from iteration $(i-1)$ to iteration i. Maximizing

$$\Delta(\mu) = 2\mu \,\mathrm{Re}\left(E\left[\mathbf{e}_i^* \left(U_i U_i^*\right)^{-1} U_i \tilde{\mathbf{w}}_{i-1}\right]\right) - \mu^2 E\left[\mathbf{e}_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i\right]$$

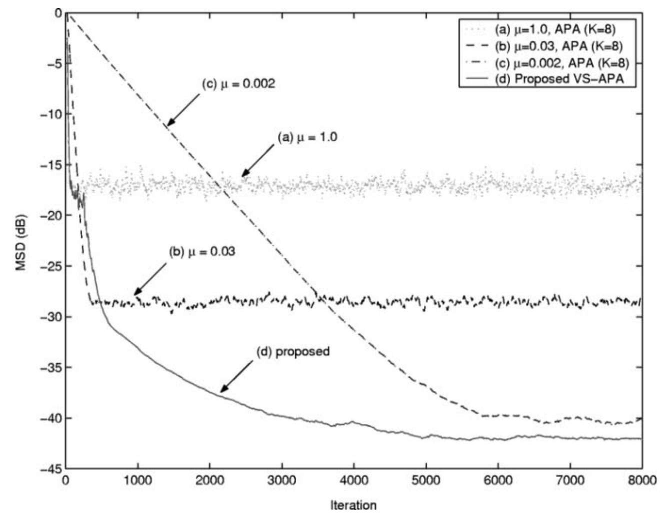

Fig. 1. Plot of the MSD for VS-APA and standard APA ($K = 8, C = 0.15$, Input: Gaussian AR(1), pole at 0.9).

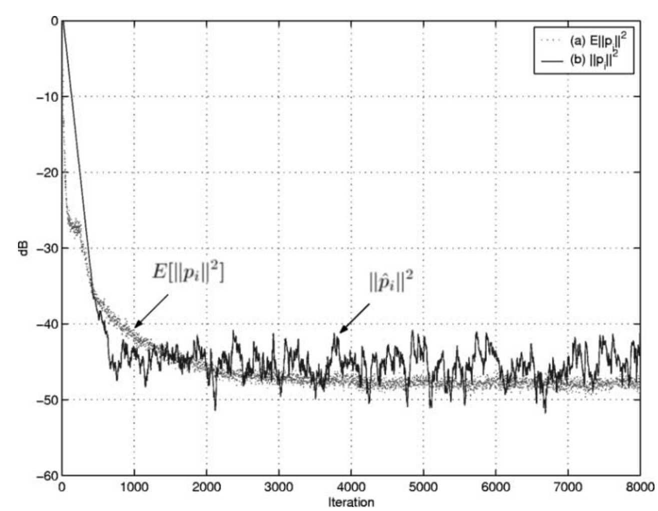

Fig. 2. Comparison of the estimate $\|\hat{\mathbf{p}}_i\|^2$ with $E\|\mathbf{p}_i\|^2$ ($K = 4, C = 0.01$, Input: Gaussian AR(1), pole at 0.9).

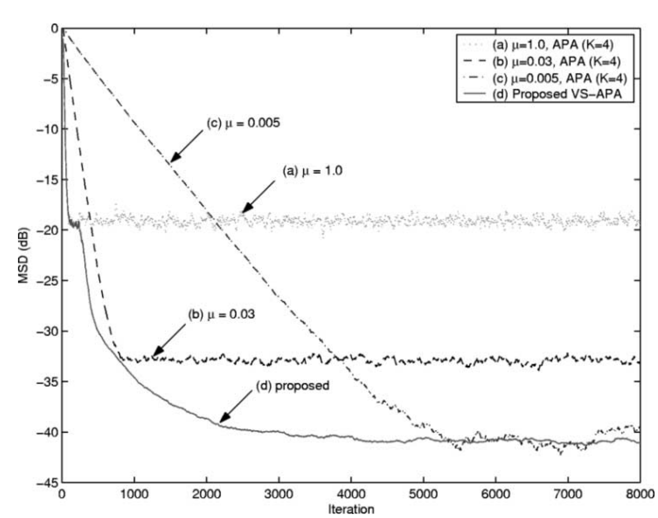

Fig. 3. Plot of the MSD for VS-APA and standard APA ($K = 4, C = 0.01$, Input: Gaussian AR(1), pole at 0.9).

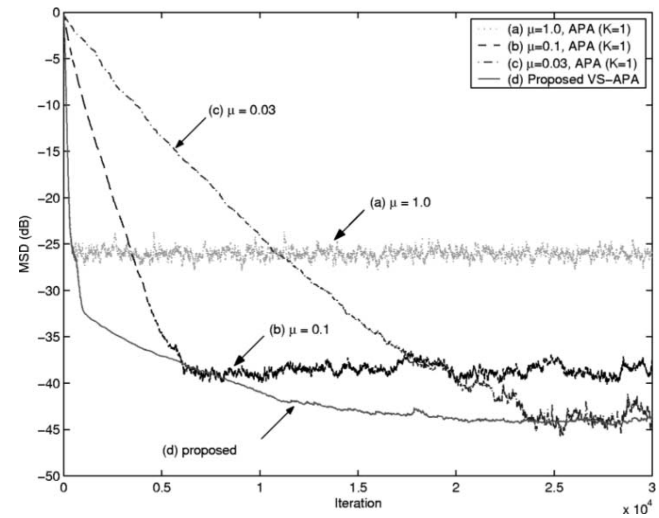

Fig. 4. Plot of the MSD for VS-NLMS and standard NLMS ($K = 1, C = 0.0001$, Input: Gaussian AR(1), pole at 0.9).

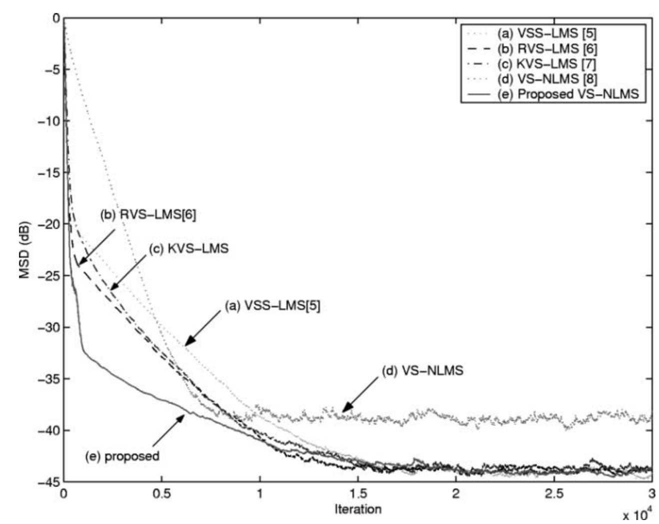

Fig. 5. Plot of the MSD for the proposed and other VS algorithms ($K = 1, C = 0.0001$, Input: Gaussian AR(1), pole at 0.9).

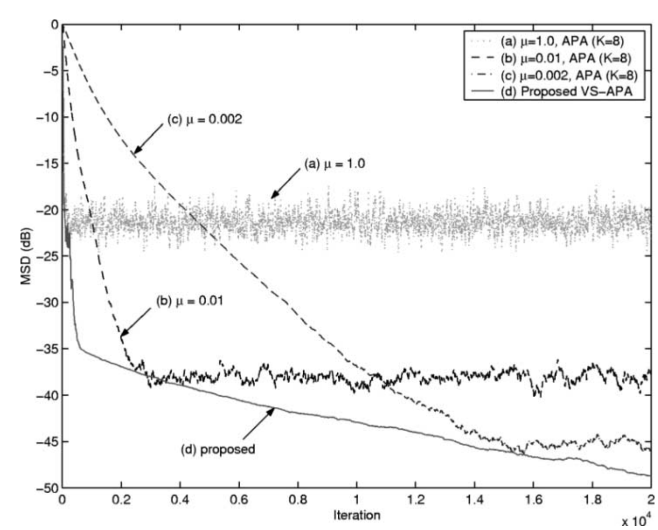

Fig. 6. Plot of the MSD for VS-APA and standard APA [$K = 8, C = 0.15$, Input: Gaussian ARMA(4,2)].

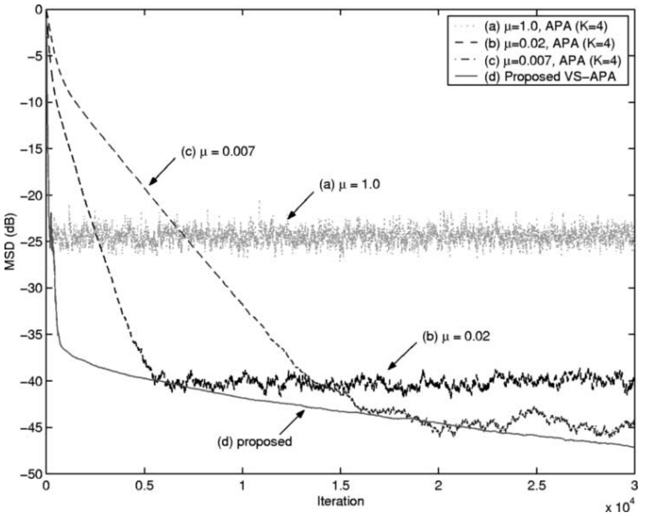

Fig. 7. Plot of the MSD for VS-APA and standard APA [$K = 4, C = 0.01$, Input: Gaussian ARMA(4,2)].

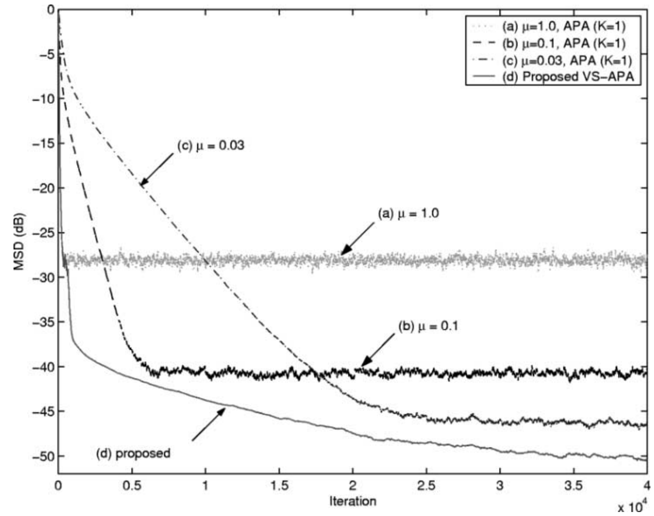

Fig. 8. Plot of the MSD for VS-NLMS and standard NLMS [$K = 1, C = 0.0001$, Input: Gaussian ARMA(4,2)].

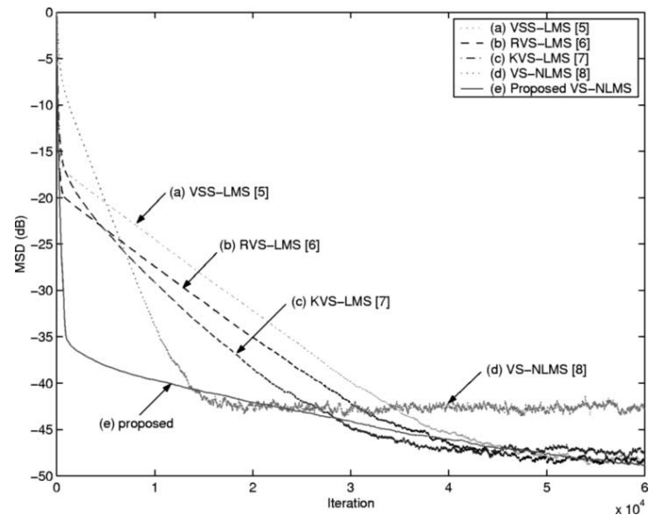

Fig. 9. Plot of the MSD for the proposed and other VS algorithms [$K = 1$, $C = 0.0001$, Input: Gaussian ARMA(4,2)].

with respect to μ, leads to the optimum step-size

$$\mu^\circ(i) = \frac{\text{Re}\, E\left(\left[\mathbf{e}_i^* \left(U_i U_i^*\right)^{-1} U_i \tilde{\mathbf{w}}_{i-1}\right]\right)}{E\left[\mathbf{e}_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i\right]}. \tag{5}$$

Assuming the noise sequence $v(i)$ is identically and independently distributed and statistically independent of the regression data $\{U_i\}$, and neglecting the dependency of $\tilde{\mathbf{w}}_{i-1}$ on past noises, $\mu^\circ(i)$ is approximated as

$$\mu^\circ(i) \approx \frac{E\|\tilde{\mathbf{w}}_{i-1}\|_\Sigma^2}{E\|\tilde{\mathbf{w}}_{i-1}\|_\Sigma^2 + \sigma_v^2 \text{Tr}\left\{E\left[\left(U_i U_i^*\right)^{-1}\right]\right\}} \tag{6}$$

where $E\|\tilde{\mathbf{w}}_{i-1}\|_\Sigma^2 = E[\tilde{\mathbf{w}}_{i-1}^* U_i^* (U_i U_i^*)^{-1} U_i \tilde{\mathbf{w}}_{i-1}]$.

Observe that $U_i^*(U_i U_i^*)^{-1} U_i$ is a projection matrix onto $\mathcal{R}(U_i^*)$, the range space of U_i^*. Let $\mathbf{p}_i \triangleq U_i^* \left(U_i U_i^*\right)^{-1} U_i \tilde{\mathbf{w}}_{i-1}$, which is the projection of $\tilde{\mathbf{w}}_{i-1}$ onto $\mathcal{R}(U_i^*)$. Since $\|\mathbf{p}_i\|^2 = \tilde{\mathbf{w}}_{i-1}^* U_i^* \left(U_i U_i^*\right)^{-1} U_i \tilde{\mathbf{w}}_{i-1}$, the optimum step-size in (6) becomes

$$\mu^\circ(i) = \frac{E\|\mathbf{p}_i\|^2}{E\|\mathbf{p}_i\|^2 + \sigma_v^2 \text{Tr}\left\{E\left[\left(U_i U_i^*\right)^{-1}\right]\right\}} \tag{7}$$

In calculating this $\mu^\circ(i)$, however, the major obstacle is that $\mathbf{p}_i$ is not available during adaptation, since $\mathbf{w}^\circ$ is unknown.

B. Variable Step-Size APA

However, note that when $v(i) = 0$, $\mathbf{p}_i = U_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i$ and even with noise, it holds under expectation that

$$E[\mathbf{p}_i] = E\left[U_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i\right].$$

Motivated by these facts, we propose to estimate $\mathbf{p}_i$ by time-averaging as follows:

$$\hat{\mathbf{p}}_i = \alpha \hat{\mathbf{p}}_{i-1} + (1 - \alpha) U_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i \tag{8}$$

with a smoothing factor α $(0 \leq \alpha < 1)$.

Using $\|\hat{\mathbf{p}}_i\|^2$ instead of $E\|\mathbf{p}_i\|^2$ in (7), the proposed variable step-size (VS) APA becomes

$$\begin{aligned} \mathbf{w}_i &= \mathbf{w}_{i-1} + \mu(i) U_i^* \left(U_i U_i^*\right)^{-1} \mathbf{e}_i \\ \mu(i) &= \mu_{\max} \cdot \frac{\|\hat{\mathbf{p}}_i\|^2}{\|\hat{\mathbf{p}}_i\|^2 + C} \end{aligned} \tag{9}$$

where C is a positive constant. From (7) and (9), we know that C is related to $\sigma_v^2 \text{Tr}\{E[(U_i U_i^*)^{-1}]\}$, and this quantity can be approximated as K/SNR. So C is proportional to K and inversely proportional to SNR. When $\|\hat{\mathbf{p}}_i\|^2$ is large, $\mu(i)$ tends to $\mu_{\max}$. On the other hand, when $\|\hat{\mathbf{p}}_i\|^2$ is small, the step-size is small. Thus depending on $\|\hat{\mathbf{p}}_i\|^2$, $\mu(i)$ varies between 0 and $\mu_{\max}$. To guarantee filter stability, $\mu_{\max}$ is chosen less than 2.

C. Variable Step-Size NLMS

A special case of (9) is a variable step-size NLMS algorithm obtained by setting $K = 1$. Recall that standard NLMS computes $\mathbf{w}_i$ via

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i) \frac{\mathbf{u}_i^*}{\|\mathbf{u}_i\|^2} e(i) \tag{10}$$

TABLE I
PARAMETERS FOR VARIABLE STEP-SIZE NLMS ALGORITHMS

VSS-LMS	RVS-LMS	KVS-LMS	VS-NLMS
$\mu_{\max} = 1$	$\mu_{\max} = 1$	$\mu_{\max} = 1$	$\alpha_1 = 0.93$
$\mu_{\min} = 0.02$	$\mu_{\min} = 0.02$	$\alpha = 10$	$\alpha_2 = 0.93$
$\alpha = 0.995$	$\alpha = 0.995$	$\beta = 0.99$	$\gamma = 0.92$
$\gamma = 0.015$	$\beta = 0.99$		
	$\gamma = 0.013$		

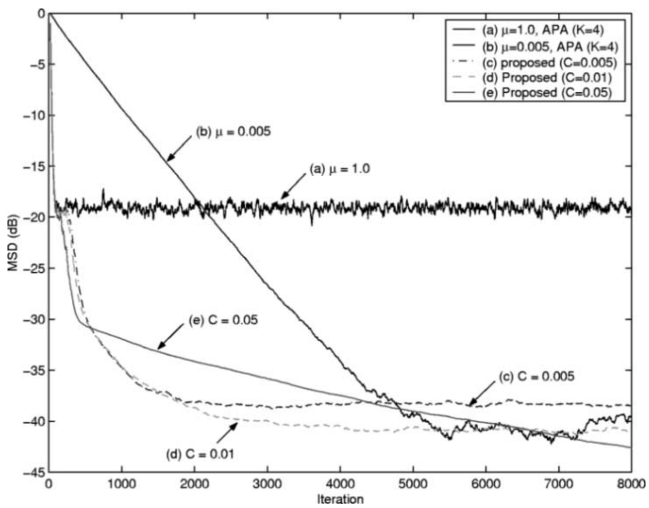

Fig. 10. Variation of the rate of convergence of the proposed VS-APA as a function of C.

where $e(i) = d(i) - \mathbf{u}_i\mathbf{w}_{i-1}$. Then from (9), the step-size is calculated as $\mu(i) = \mu_{\max} \cdot \|\hat{\mathbf{p}}_i\|^2/(\|\hat{\mathbf{p}}_i\|^2 + C)$ where now $\hat{\mathbf{p}}_i = \alpha\hat{\mathbf{p}}_{i-1} + (1-\alpha)\mathbf{u}_i e(i)/\|\mathbf{u}_i\|^2$.

III. SIMULATION RESULTS

We illustrate the performance of the proposed algorithms by carrying out computer simulations in a channel estimation scenario. The unknown channel $H(z)$ is represented by a moving average model with 16 taps. The adaptive filter and the unknown channel are assumed to have the same number of taps. Two Gaussian distributed signals are used for the input signal. The input signals are obtained by filtering a white, zero-mean, Gaussian random sequence through a first-order system $G_1(z) = 1/1 - 0.9z^{-1}$ or a second-order system

$$G_2(z) = \frac{1 + 0.9z^{-1} + 0.6z^{-2} + 0.81z^{-3} - 0.329z^{-4}}{1 + 1.0z^{-1} + 0.21z^{-2}}.$$

The SNR is calculated by

$$\mathrm{SNR} = 10\log_{10}\left(\frac{E\left[y^2(i)\right]}{E\left[v^2(i)\right]}\right)$$

where $y(i) = \mathbf{u}_i\mathbf{w}^\circ$. The measurement noise $v(i)$ is added to $y(i)$ such that SNR $=$ 30 dB. The simulation results are obtained by ensemble averaging over 100 independent trials. We use the input signals generated by $G_1(z)$ and $G_2(z)$ for Figs. 1–5 and Figs. 6–9, respectively. In Fig. 1, we show the MSD $(E\|\tilde{\mathbf{w}}_i\|^2)$ for $K = 8$, $\alpha = 0.99$, $C = 0.15$, and $\mu_{\max} = 1.0$. Dashed lines indicate the results of APA with fixed step-sizes where we choose $\mu = 0.002, 0.03$, and 1.0. As can be seen, the proposed VS-APA converges faster and has lower mis-adjustment error. In Figs. 2 and 3, we choose $K = 4$, $\alpha = 0.99$, $C = 0.01$, and $\mu_{\max} = 1.0$. Fig. 2 shows how accurately $\|\hat{\mathbf{p}}_i\|^2$ estimates $E\|\mathbf{p}_i\|^2$. A similar result to Fig. 1 is observed in Fig. 3. Figs. 4 and 5 show the performance of the proposed variable step-size NLMS. For comparison purposes, the following variable step-size schemes are applied to (10); although the first three schemes have been originally developed for standard LMS

$$\begin{aligned}
\text{VSS-LMS [5]}: \quad & \mu(i) = \alpha\mu(i-1) + \gamma e^2(i)\\
\text{RVS-LMS [6]}: \quad & \mu(i) = \alpha\mu(i-1) + \gamma p^2(i)\\
& p(i) = \beta p(i-1) + (1-\beta)e(i)e(i-1)\\
\text{KVS-LMS [7]}: \quad & \mu(i) = \mu_{\max}\left(1 - e^{-\alpha C_4^e(i)}\right)\\
& C_4^e(i) = f(i) - 3p^2(i)\\
& f(i) = \beta f(i-1) + (1-\beta)e^4(i)\\
& p(i) = \beta p(i-1) + (1-\beta)e^2(i)\\
\text{VS-NLMS [8]}: \quad & \mu(i) = \frac{\sigma_u^2(i)}{\sigma_e^2(i)}\beta(i)\\
& \sigma_u^2(i) = \alpha_1\sigma_u^2(i-1) + (1-\alpha_1)u^2(i)\\
& \sigma_e^2(i) = \alpha_2\sigma_e^2(i-1) + (1-\alpha_2)e^2(i)\\
& \beta(i) = \gamma\beta(i-1) + (1-\gamma)\frac{e^2(i)}{u^2(i)}
\end{aligned}$$

The parameters used in Figs. 4 and 5 are shown in Table I. Figs. 6–9 are the simulation results with the different input signal generated by $G_2(z)$.

Finally, as indicated in Fig. 10, the rate of convergence of the proposed algorithm is not highly sensitive to the choice of C. In the figure, the value of C is varied by one order of magnitude.

IV. CONCLUSION

We have presented two variable step-size NLMS and APA schemes. The norm of the projected weighted error vector is used as a criterion to determine how close the adaptive filter is to optimum performance. The algorithms show improved filter performance.

REFERENCES

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
[2] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 1996.
[3] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.
[4] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electron. Commun. Jpn.*, vol. 67-A, no. 5, pp. 19–27, 1984.
[5] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Trans. Signal Processing*, vol. 40, pp. 1633–1642, July 1992.
[6] T. Aboulnasr and K. Mayyas, "A robust variable step-size LMS-type algorithm: Analysis and simulations," *IEEE Trans. Signal Processing*, vol. 45, pp. 631–639, Mar. 1997.
[7] D. I. Pazaitis and A. G. Constantinides, "A novel kurtosis driven variable step-size adaptive algorithm," *IEEE Trans. Signal Processing*, vol. 47, pp. 864–872, Mar. 1999.
[8] A. Mader, H. Puder, and G. U. Schmidt, "Step-size control for acoustic echo cancellation filters—An overview," *Signal Process.*, vol. 80, pp. 1697–1719, Sept. 2000.