A Variable Step Size Adaptive Algorithm with Simple Parameter Selection

Daniel G. Tiglea, Renato Candido, and Magno T. M. Silva, *Member, IEEE*

***Abstract*—We propose a normalized least mean squares algorithm with variable step size. Unlike other solutions, it has low computational cost, only three parameters that are simple to choose, and its steady-state performance can be easily predicted. Simulations show a competitive performance in comparison with other solutions, and validate our theoretical analysis.**

***Index Terms*—Variable step size, normalized least mean squares, adaptive filtering, steady-state analysis**

I. Introduction

Since their inception, adaptive filtering algorithms have been successfully employed in numerous signal processing applications such as equalization, active noise control, acoustic echo cancellation, biomedical engineering, among others [1], [2]. The normalized least-mean-square (ϵ-NLMS) algorithm is one of the most popular adaptive filters due to its simplicity and improved robustness in comparison with its non-normalized counterpart, the LMS algorithm [1], [2]. Its stability and performance are governed by a fixed step size μ, whose choice is tied to a compromise: lower step sizes lead to improved steady-state performances, but slow down the convergence rate [1]–[3]. Hence, several versions of the LMS and ϵ-NLMS algorithms with variable step-sizes (VSS) have been proposed in the literature [4]–[10]. They seek to implement time-dependent step sizes that remain high during the transient and decrease when the mean squared error (MSE) is sufficiently low, thus improving the steady-state performance. These algorithms introduce new parameters that control the evolution of the step size. However, in many cases it is difficult to determine how they should be chosen [6], with poor selections cutting the benefit of the VSS mechanism. A "non-parametric" VSS algorithm was proposed in [6], but in this case it is difficult to predict the value of the step size during steady state, and, consequently, the performance of the algorithm. In [10], a VSS algorithm was proposed that switches from a high step size to a lower one after a number of iterations. Although attractive for its simplicity, it is not robust to abrupt changes in the environment and is usually outperformed by more sophisticated solutions, as expected [10]. Another alternative consists in employing a convex combination of adaptive filters [11]–[13]. In this scheme, two or more filters with different step sizes run simultaneously, and the output of the algorithm is formed by combining the responses of each filter in an adaptive manner, so as to minimize the MSE. However, the computational cost of these approaches is at least twice as large as that of each individual filter, assuming that they run the same algorithm.

Inspired by an algorithm we proposed for the sampling and censoring of adaptive diffusion networks [14], in this letter we propose a low-cost VSS-NLMS algorithm with three parameters that are simple to choose based on the filter length and desired steady-state performance. It is possible to ensure that its step size does not decrease in the mean until the MSE drops sufficiently. Moreover, by assuming that the data are Gaussian and that the filter length is sufficiently long, its steady-state performance can be predicted with good accuracy. Simulations validate the analysis even when these assumptions do not hold, and show that its performance is competitive in comparison with other state-of-the art solutions. Lastly, its robustness to changes in the environment is also verified.

This letter is organized as follows. In Sec. II, the formulation of the problem is presented. In Sec. III, the proposed VSS-NLMS algorithm is introduced. In Sec. IV, we show how the parameters of the algorithm can be easily selected and obtain theoretical results for its steady-state performance. Finally, in Sec. V we provide simulation results and Sec. VI closes the paper with the main conclusions of our work.

Notation. We use normal font letters for scalars, boldface lowercase letters for vectors, and boldface uppercase letters for matrices. Moreover, $(\cdot)^{\mathrm{T}}$ denotes transposition, $\mathrm{E}\{\cdot\}$ the mathematical expectation, $\mathrm{Tr}[\cdot]$ the trace of a matrix, and $\|\cdot\|$ the Euclidean norm. To simplify the arguments, we assume real data throughout the letter.

II. Problem Formulation

Let us denote by $d(n)$ the desired response of the filter, which is modeled as $d(n) = \mathbf{u}^{\mathrm{T}}(n)\mathbf{w}_{\mathrm{o}} + v(n)$, where $\mathbf{u}(n) = [u(n)\ u(n-1)\ \cdots\ u(n-M+1)]^{\mathrm{T}}$ is the input regressor vector, $\mathbf{w}_{\mathrm{o}}$ is the optimal system, and $v(n)$ is the measurement noise, which is assumed to be independent from any other signal, and independent and identically distributed (iid) with zero mean and variance σ_v^2 [1], [2]. Denoting by $\mathbf{w}(n)$ the estimate of $\mathbf{w}_{\mathrm{o}}$ produced by the algorithm at time instant n, the adaptation of ϵ-NLMS is given by [1], [2]

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\mu}{\epsilon + \|\mathbf{u}(n)\|^2}\mathbf{u}(n)e(n), \qquad (1)$$

where $e(n) = d(n) - \mathbf{u}^{\mathrm{T}}(n)\mathbf{w}(n-1)$ is the estimation error, $\epsilon > 0$ is a regularization factor, and $0 < \mu < 2$ is a step size [1], [2].

When analyzing the performance of adaptive filters, some of the most commonly adopted metrics are the MSE and

This work was supported in part by CAPES under Grant 88887.512247/2020-00 and Finance Code 001 and in part by the São Paulo Research Foundation (FAPESP) under Grant 2021/02063-6.
The authors are with the Electronic Systems Engineering Department, Escola Politécnica, University of São Paulo, São Paulo, SP, Brazil, e-mails:{dtiglea, renatocan, magno}@lps.usp.br, ph. +55-11-3091-5134.

the excess MSE (EMSE), which are respectively given by MSE$(n) \triangleq$ E$\{e^2(n)\}$ and EMSE$(n) \triangleq$ E$\{e_a^2(n)\}$, where $e_a(n) \triangleq \mathbf{u}^{\mathrm{T}}(n)[\mathbf{w}_{\mathrm{o}} - \mathbf{w}(n-1)]$ is the *a priori* estimation error [1], [2]. Since $v(n)$ is assumed independent from any other signal, the MSE and EMSE are related by [1], [2]

$$\text{MSE}(n) = \text{EMSE}(n) + \sigma_v^2 > \sigma_v^2. \tag{2}$$

Thus, the measurement noise power marks a limit in the performance of adaptive algorithms. Several theoretical expressions have been derived for the MSE and EMSE of most adaptive filtering solutions. For the ϵ-NLMS with sufficiently small ϵ and μ, its steady-state EMSE can be approximated by [1]

$$\text{EMSE}_{\epsilon-\text{NLMS}}(\infty) \approx \frac{\mu\sigma_v^2\rho}{2-\mu}, \tag{3}$$

where $\rho \triangleq \mathrm{E}\{1/\|\mathbf{u}(n)\|^2\}\mathrm{Tr}(\mathbf{R}_u)$, with $\mathbf{R}_u \triangleq \mathrm{E}\{\mathbf{u}(n)\mathbf{u}^{\mathrm{T}}(n)\}$. From (3) we see that the steady-state EMSE increases with μ. However, low step sizes affect the convergence rate of ϵ-NLMS [1], [2]. In fact, it can be shown that the fastest convergence rate occurs by adopting $\mu = 1$ [3]. To conciliate these conflicting goals, we propose in Sec. III a VSS algorithm that ensures a high step size as long as the MSE lies above a certain threshold, and gradually decreases it as the MSE drops.

III. The proposed VSS algorithm

In order to control the step size, we multiply it by an auxiliary variable $s(n) \in [0,1]$. Thus, (1) can be recast as

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\bar{\mu}(n)}{\epsilon + \|\mathbf{u}(n)\|^2}\mathbf{u}(n)e(n), \tag{4}$$

where $\bar{\mu}(n) = \mu s(n)$. We suggest adopting $\mu = 1$ for optimal convergence rate, but a lower μ can be used for better steady-state performance. Ideally, $s(n)$ should be equal to one in the transient, and then decrease during steady state. Inspired by [14], we adopt the following cost function for $s(n)$:

$$J_s(n) = s^2(n) \cdot \gamma\sigma_v^2 + \big[1 - s(n)\big]\, e^2(n), \tag{5}$$

where γ is a positive parameter that the filter designer must choose, and whose influence on the algorithm will be studied in Sec. IV. The rationale behind (5) is that when $e^2(n)$ is high, $J_s(n)$ is minimized by making $s(n)$ closer to one. In contrast, as $e^2(n)$ becomes smaller than $\gamma\sigma_v^2$, $J_s(n)$ is minimized by decreasing $s(n)$. Thus, in the mean, the step size should decrease when the MSE is sufficiently close to its minimum value σ_v^2 based on our choice for γ. The reason why we adopt $s^2(n) \cdot \gamma\sigma_v^2$ instead of simply $s(n) \cdot \gamma\sigma_v^2$ in (5) is because the latter option encourages the algorithm to make $s(n) = 0$ when the MSE is smaller than $\gamma\sigma_v^2$. Hence, the adaptation would eventually stop. In contrast, by adopting $s^2(n) \cdot \gamma\sigma_v^2$, this problem is avoided, as will become clear further ahead. Moreover, we show in Sec. IV that in this case we can ensure $s(n) > 0$ in the mean during steady state.

Rather than directly adjusting $s(n)$, we update an auxiliary variable $\alpha(n)$ deterministically related to $s(n)$ via [12]

$$s(n) = \frac{\mathrm{sgm}[\alpha(n)] - \mathrm{sgm}[-\alpha^+]}{\mathrm{sgm}[\alpha^+] - \mathrm{sgm}[-\alpha^+]}, \tag{6}$$

where $\mathrm{sgm}[x] = (1+e^{-x})^{-1}$ is the sigmoidal function, and α^+ is the maximum value that $\alpha(n)$ can assume. We should notice that $s(n)$ attains values 1 and 0 for $\alpha(n) = \alpha^+$ and $\alpha(n) = -\alpha^+$, respectively. A common value adopted in the literature is $\alpha^+ = 4$. We encourage the reader to consult, e.g., [11]–[15] and their references for an in-depth explanation. Finally, $\alpha(n)$ is truncated in the range $[-\alpha^+, \alpha^+]$.

By taking the derivative of (5) with respect to $\alpha(n)$, we obtain the following stochastic gradient descent rule for $n \geqslant 1$:

$$\alpha(n+1) = \alpha(n) + \mu_s s'(n)[e^2(n) - 2\gamma\sigma_v^2 s(n)], \tag{7}$$

where $\mu_s > 0$ is a step size whose selection will be discussed in Sec. IV, $\alpha(0) = \alpha^+$, and [12]

$$s'(n) \triangleq \frac{ds(n)}{d\alpha(n)} = \frac{\mathrm{sgm}[\alpha(n)]\{1 - \mathrm{sgm}[\alpha(n)]\}}{\mathrm{sgm}[\alpha^+] - \mathrm{sgm}[-\alpha^+]}. \tag{8}$$

Eqs. (4), (6), and (7) form the basis of the proposed algorithm, which we name γVSS-NLMS. Its operation in the mean can be interpreted as follows. When the MSE is high, $\bar{\mu}(n)$ is kept at its maximum. As the MSE drops, the term between brackets in (7) becomes negative, at which point $\alpha(n)$, $s(n)$, and $\bar{\mu}(n)$ drop as well, further decreasing the MSE. However, as $s(n)$ drops, so does the incentive for $\alpha(n)$ to continue decreasing, due to the presence of $s(n)$ in the second term between brackets in (7). Eventually, an equilibrium is reached and $s(n)$ stabilizes. Indeed, in Sec. IV, we show that choosing a finite value for γ leads to $\mathrm{E}\{s(\infty)\} > 0$, thus the adaptation never stops completely. Moreover, since $0 \leqslant s(n) \leqslant 1$ and $\bar{\mu}(n) = \mu s(n)$, the stability of the algorithm is ensured if $0 < \mu < 2$. If $\mu_s s'(n)$ and $s(n)$ are implemented by a lookup table, γVSS-NLMS requires $2M+6$ multiplications and $2M+4$ sums per iteration, only slightly more than the $2M+2$ multiplications and $2M+2$ sums of ϵ-NLMS.

IV. Theoretical Analysis

Introducing the quantity $\Delta\alpha(n) \triangleq \alpha(n+1) - \alpha(n)$ and taking expectations from both sides of (7), we obtain

$$\mathrm{E}\{\Delta\alpha(n)\} = \mu_s \mathrm{E}\{s'(n)[e^2(n) - 2\gamma\sigma_v^2 s(n)]\}. \tag{9}$$

To make the analysis more tractable, we assume that $s'(n)$ is independent from the term between brackets in (9). Since $s'(n)$ varies much more slowly with time than the term between brackets, their effects can be taken into account separately, according to the averaging principle [16]. Hence, we can write

$$\mathrm{E}\{\Delta\alpha(n)\} \approx \mu_s \mathrm{E}\{s'(n)\}\mathrm{E}\{e^2(n) - 2\gamma\sigma_v^2 s(n)\}. \tag{10}$$

During the initial convergence, since $\alpha(0) = \alpha^+$ and we truncate α in $[-\alpha^+, \alpha^+]$, we see from (10) that $\mathrm{E}\{\alpha(n+1)\} = \alpha^+$ as long as $\mu_s \mathrm{E}\left\{s'(n)\right\} \mathrm{E}\left\{e^2(n) - \gamma\sigma_v^2 s(n)\right\} > 0$. Since $\mu_s > 0$, $\mathrm{E}\{s(n)\} \in [0,1]$ and $s'(n) > 0$ for $\alpha \in [-\alpha^+, \alpha^+]$, we conclude that $\mathrm{E}\{\alpha(n+1)\} = \alpha^+$, $\mathrm{E}\{s(n+1)\} = 1$, and $\mathrm{E}\{\bar{\mu}(n+1)\} = \mu$ as long as

$$\text{MSE}(n) > 2\gamma\sigma_v^2. \tag{11}$$

Thus, we have a clear understanding of when $\bar{\mu}(n)$ begins to drop based on γ. Moreover, we can estimate the expected steady-state value of $s(n)$. Since in steady state we should have $\mathrm{E}\{\Delta\alpha(\infty)\}=0$, we conclude from (10) that

$$\mathrm{E}\{e^2(\infty)-2\gamma\sigma_v^2 s(\infty)\}\approx 0. \tag{12}$$

Using (2), (12) can be recast as

$$\mathrm{EMSE}_\gamma(\infty)+\sigma_v^2-2\gamma\sigma_v^2\mathrm{E}\{s(\infty)\}\approx 0, \tag{13}$$

where $\mathrm{EMSE}_\gamma(\infty)$ is the steady-state EMSE of the proposed algorithm. Since in steady state γVSS-NLMS should operate as the ϵ-NLMS with a lower step size, we estimate $\mathrm{EMSE}_\gamma(\infty)$ by replacing μ with $\mathrm{E}\{\bar{\mu}(\infty)\}=\mu\mathrm{E}\{s(\infty)\}$ in (3), obtaining

$$\mathrm{EMSE}_\gamma(\infty)\approx\frac{\mu\mathrm{E}\{s(\infty)\}\sigma_v^2\rho}{2-\mu\mathrm{E}\{s(\infty)\}}. \tag{14}$$

Replacing (14) in (13) and taking into account that we must have $\mathrm{E}\{s(\infty)\}\in[0,1]$ at all times, after some algebra we get

$$\mathrm{E}\{s(\infty)\}\approx\frac{4\gamma+\mu(1-\rho)-\sqrt{[4\gamma+\mu(1-\rho)]^2-16\gamma\mu}}{4\gamma\mu} \tag{15}$$

if the right-hand side (rhs) is real and lesser than or equal to one. Otherwise, we assume that $\mathrm{E}\{s(\infty)\}=1$. To obtain the minimum value $\gamma_{\min}$ for γ, we must ensure that the rhs of (15) is simultaneously: i) real, and ii) less than one. Condition i) can be satisfied by adopting $\gamma>\mu\left(1+\rho+2\sqrt{\rho}\right)/4$ for any μ. For $\mu\geqslant\zeta\triangleq 2(1+\sqrt{\rho})^{-1}$, this also enforces Condition ii). For $\mu<\zeta$, we must adopt the stricter rule $\gamma>2+(\rho-1)\mu]/[2(2-\mu)]$ to enforce Condition ii) (see the supplementary material for a detailed explanation). We thus get

$$\gamma_{\min}=\begin{cases}\mu\left(1+\rho+2\sqrt{\rho}\right)/4, & \text{if } \zeta\leqslant\mu<2\\ [2+(\rho-1)\mu]/[2(2-\mu)], & \text{otherwise}\end{cases}. \tag{16}$$

These results can be simplified assuming that the data are Gaussian and that the filter length M is sufficiently long. Under these assumptions, we can write [17], [18]

$$\rho\approx M/(M-2). \tag{17}$$

As shown in Sec. V, the results thus obtained hold fairly even if these conditions are not met. Replacing (17) in (16) yields

$$\gamma_{\min}=\begin{cases}\dfrac{\mu}{2}\left(\dfrac{M-1}{M-2}+\sqrt{\dfrac{M}{M-2}}\right), & \text{if } \zeta\leqslant\mu<2\\ \dfrac{[(M-2)+\mu]}{(M-2)(2-\mu)}, & \text{otherwise}\end{cases}. \tag{18}$$

From (17) we see that typically $\rho\approx 1$ and thus $\zeta\approx 1$ for large M. Moreover, for large γ or small μ we may assume that $\mu\mathrm{E}\{s(\infty)\}\ll 2$, enabling us to recast (14) as $\mathrm{EMSE}_\gamma(\infty)\approx\mu\mathrm{E}\{s(\infty)\}\sigma_v^2\rho/2$. In this case, we get from (13) and (17)

$$\mathrm{E}\{s(\infty)\}\approx\frac{2(M-2)}{4(M-2)\gamma-\mu M}. \tag{19}$$

Replacing (19) in (14) and using (17), we conclude that

$$\mathrm{EMSE}_\gamma(\infty)\approx\frac{\mu\sigma_v^2 M}{4(M-2)\gamma-2\mu(M-1)} \tag{20}$$

for large γ or small μ. Otherwise, a more precise result can be obtained by replacing (15) and (17) in (14). From (20), we can see that γVSS-NLMS achieves a lower EMSE than ϵ-NLMS for large γ. Moreover, μ_s does not influence the steady-state performance of γVSS-NLMS, although it can affect its convergence. If μ_s is too small, $\bar{\mu}(n)$ decreases very slowly, which is undesirable. If μ_s is too large, $\bar{\mu}(n)$ can drop suddenly after (11) ceases to hold, harming the convergence rate. Simulations suggest that, for $0<\mu\leqslant 1$, the transient performance of γVSS-NLMS is preserved by making

$$\mu_s\approx\frac{1}{\gamma\sigma_v^2}\cdot\frac{\mu^2}{3\ln(M)}. \tag{21}$$

This is a heuristic result that can be interpreted as follows. The first term cancels the effect of γ and σ_v^2 on the convergence rate of $\alpha(n)$ in (7). The second term is related to the convergence of $\mathbf{w}(n)$, which slows down with the decrease of μ or the increase of M. However, it is important to mention that simulation results show that the performance of γVSS-NLMS is not very sensitive to small variations in μ_s.

We remark that an online estimator for σ_v^2 can be used (see, e.g., [7], [19] and their references) to cut the need for prior knowledge of the noise power. In this case, one may simply replace σ_v^2 with $\hat{\sigma}_v^2(n)$ in (7) and use $\mu_s(n)=\theta/[\epsilon+\hat{\sigma}_v^2(n)]$, where $\hat{\sigma}_v^2(n)$ is the estimate of σ_v^2 at iteration n and $\theta\triangleq\mu^2/[3\gamma\ln(M)]$, further simplifying the usage of the algorithm.

V. Simulation Results

The following simulations were obtained over an average of 10^3 independent realizations with $6\cdot 10^4$ iterations each in a system identification setup. The filter length is equal to that of $\mathbf{w}_\mathrm{o}$. The coefficients of $\mathbf{w}_\mathrm{o}$ are generated randomly following a uniform distribution in the range $[-1,1]$, and later normalized so that $\mathbf{w}_\mathrm{o}$ has unit norm. Unless stated otherwise, we consider $M=128$ and a Gaussian distribution for $u(n)$ with zero mean and unit variance. Lastly, we consider a Gaussian distribution for $v(n)$ with zero mean. The noise power σ_v^2 is set to ensure a signal-noise ratio of 20dB, and we adopt $\epsilon=10^{-5}$.

In the simulations of Fig. 1, we compare the EMSE curves of γVSS-NLMS with those of other VSS algorithms and of a convex combinations of two adaptive filters with coefficient transfer from the fast filter to the slow one [11]. In the simulations, the algorithm of [9] was outperformed by the other solutions. Since this algorithm was designed for acoustic echo cancellation with speech signals, we opted to not include it in Fig. 1 since this comparison could be unfair. Moreover, for the VSS algorithms of the LMS type, we adopted normalized step sizes to enable the comparison with solutions of the ϵ-NLMS kind. The parameters of every algorithm were adjusted to achieve roughly the same steady-state performance. For every algorithm, we set the initial step size to $\mu=1$. For γVSS-NLMS we adopt $\gamma=12.5$, which satisfies (18), and consider two versions: i) σ_v^2 known *a priori*, and ii) estimated online using the algorithm of [19]. Since the solutions of [4]–[7] achieved very similar results, we opted to only depict the EMSE curve of the algorithm of [7] in Fig. 1, as its performance is marginally better than most others, and it is the most recent solution among them. Otherwise, the figure would be very polluted. In Tab. I we show the values adopted for each parameter of the algorithms considered, keeping the

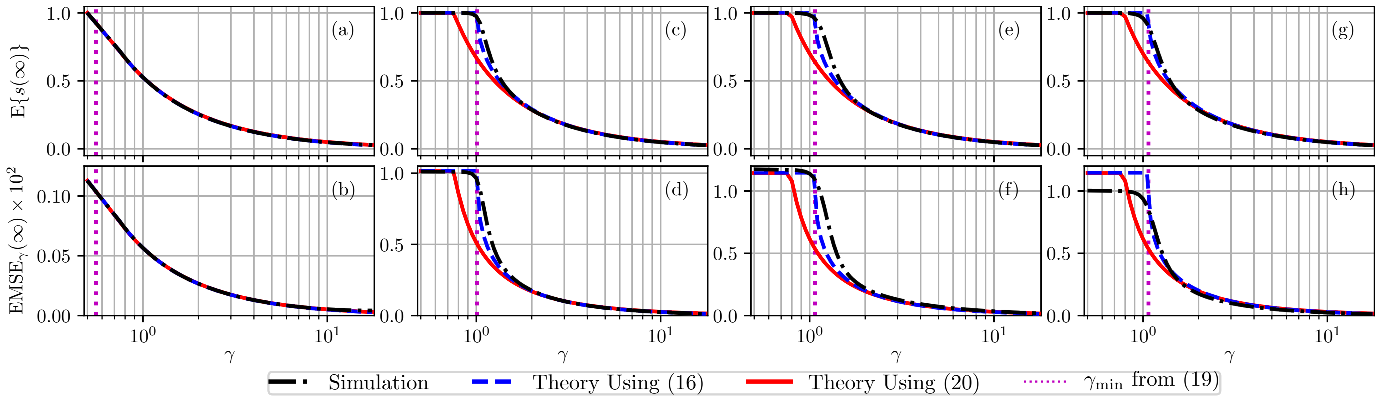

Fig. 2: Comparison between simulation results and the analysis from Sec. IV in terms of $\mathrm{E}\{s(\infty)\}$ (top row) and $\mathrm{EMSE}_\gamma(\infty)$ (bottom row) for $0.5\leqslant\gamma\leqslant 17.5$. (a) and (b): $\mu=0.2$, M=128, and $u(n)\sim\mathcal{N}(0,1)$. (c) and (d): μ=1, M=128 and $u(n)\sim\mathcal{N}(0,1)$. (e) and (f): μ=1, M=16 and $u(n)=r(n)-0.8u(n-1)$, with $r(n)\sim\mathcal{N}(0,1)$. (g) and (h): μ=1, M=16 and $u(n)\in\{-1,1\}$.

notation of the original references. We also present the results obtained with ϵ-NLMS with $\mu=1$ as well as $\mu=0.408$, which corresponds to the steady-state step size of γVSS-NLMS, and indicate the theoretical steady-state EMSE yielded by (20). To simulate an abrupt change in the environment, in the middle of each realization we divide $\mathbf{w}_\mathrm{o}$ by 2. We can see that γVSS-NLMS converged as fast as the solution of [7] in both transients, and faster than the other algorithms. Hence, its performance is competitive with other solutions. We also see that the steady-state EMSE matches the results yielded by (20) closely, and that the use of an online estimator for σ_v^2 had almost no impact on the performance.

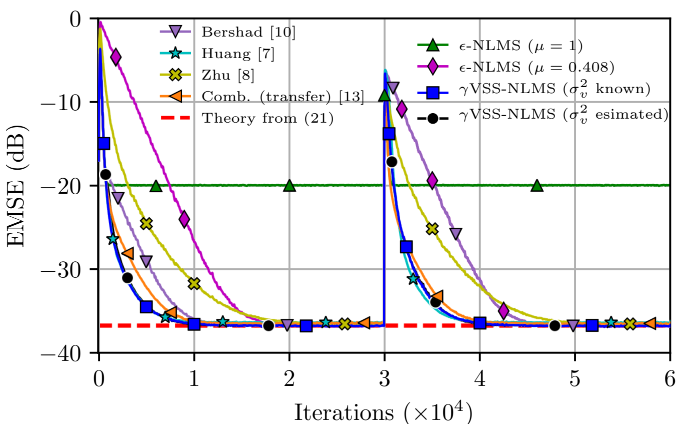

Fig. 1: Comparison between γVSS-NLMS and the VSS algorithms from [7], [8], [10], as well as convex combinations of ϵ-NLMS filters with different step sizes [11]–[13]. The curves are filtered by a moving-average filter with 64 coefficients.

To validate the analysis of Sec. IV, in Fig. 2 we compare simulation results to the theoretical values for $\mathrm{E}\{s(\infty)\}$ and for the steady-state EMSE of γVSS-NLMS based on (15) and (19). For this, we consider an ensemble average of the last 6000 iterations of each realization, with σ_v^2 known *a priori* and several values for $0.5\leqslant\gamma\leqslant 17.5$ in four scenarios. In Figs. 2(a) and (b), we adopt $\mu=0.2$, whereas in Figs. 2(c) and (d), we use $\mu=1$. In Figs. 2(e) and (f), we adopt $\mu=1$, $M=16$ and a

TABLE I: Parameters of the VSS algorithms and convex combinations of adaptive filters used in the simulations of Fig. 1.

Solution	Parameters
γVSS-NLMS (σ_v^2 known)	$\gamma=12.5$, $\mu_s=0.5496$, $\alpha^+=4$
γVSS-NLMS (σ_v^2 estimated online using [19])	$\gamma=12.5$, $\theta=5.4960\cdot 10^{-3}$, $\alpha^+=4$, $w_s=0.9998$, $w_m=0.9995$, and $w_f=0.9984$
Huang [7]	$\mu_{\max}$=1, $\mu_{\min}$=10^{-5}, α=0.998, $\beta=25$, ζ=0.35
Zhu [8]	$m=1$, $A=2$, $B=0.7$, $\lambda=0.99$
Bershad [10]	$\mu_1=0.892$, $\mu_2=0.408$, $n_s=607$
Combination (with coefficient transfer) [13]	$\mu_1=1$, $\mu_2=0.408$, $\mu_a=1.5$ $\eta=0.98$, $\lambda_0=0.8$, $\ell=0.95$

colored input given by $u(n)=r(n)-0.8u(n-1)$, where $r(n)$ is a Gaussian noise with zero mean and unit variance. Lastly, in Figs. 2(g) and (h) we consider $\mu=1$, $M=16$ and a binary input $u(n)$ that can assume the values -1 or 1 with equal probability. In each plot, we also indicate the $\gamma_{\min}$ yielded by (18) by vertical dotted lines. For intermediate values of γ we see that (15) provides a more precise model than (19), albeit more complex. However, the theoretical values yielded by both models match the simulation results closely for large γ, i.e. $\gamma\geqslant 2$, in all four scenarios. We can also see that Figs. 2(a), (c), (e), and (f) validate (18). For $\mu=0.2$ (Figs. 2(a) and (b)), the results match closely for $\gamma\geqslant 0.5$ in general, which makes sense since (3) is a better approximation for small μ. Lastly, we see from Figs. 2(e) to (h) that the main results of Sec. IV hold even when the assumptions of Gaussian data and long filter lengths do not.

VI. Conclusions

Unlike most other solutions, the proposed γVSS-NLMS algorithm has low computational cost and only three parameters (α^+, γ, and μ_s) that are easy to choose. Moreover, we obtained theoretical results for its steady-state performance based on the value of its parameters, further aiding their selection. Simulations show that the performance of γVSS-NLMS is competitive in comparison with the other solutions and validate the theoretical analysis. In future works we intend to analyze the transient behavior of γVSS-NLMS, to derive more precise results for the selection of μ_s, and to test the algorithm in different scenarios.

REFERENCES

[1] A. H. Sayed, *Adaptive Filters*, John Wiley & Sons, NJ, 2008.
[2] S. Haykin, *Adaptive Filter Theory*, Pearson, Upper Saddle River, 5th edition, 2014.
[3] D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Trans. Signal Process.*, vol. 41, pp. 2811–2825, Sep. 1993.
[4] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Trans. Signal Process.*, vol. 40, pp. 1633–1642, Jul. 1992.
[5] T. Aboulnasr and K. Mayyas, "A robust variable step-size lms-type algorithm: analysis and simulations," *IEEE Transactions on signal processing*, vol. 45, no. 3, pp. 631–639, Mar. 1997.
[6] J. Benesty, H. Rey, L. R. Vega, and S. Tressens, "A nonparametric VSS NLMS algorithm," *IEEE Signal Processing Letters*, vol. 13, no. 10, pp. 581–584, Oct. 2006.
[7] H.-C. Huang and J. Lee, "A new variable step-size NLMS algorithm and its performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 2055–2060, Dec. 2011.
[8] Y.-G. Zhu, Y.-G. Li, S.-Y. Guan, and Q.-S. Chen, "A novel variable step-size NLMS algorithm and its analysis," *Procedia Engineering*, vol. 29, pp. 1181–1185, Feb. 2012.
[9] M. Hamidia and A. Amrouche, "Improved variable step-size NLMS adaptive filtering algorithm for acoustic echo cancellation," *Digital Signal Processing*, vol. 49, pp. 44–55, Feb. 2016.
[10] N. J. Bershad and J. C. M. Bermudez, "A switched variable step size NLMS adaptive filter," *Digital Signal Processing*, vol. 101, pp. 102730, Jun. 2020.
[11] J. Arenas-Garcia, L. A. Azpicueta-Ruiz, M. T. M. Silva, V. H. Nascimento, and A. H. Sayed, "Combinations of adaptive filters: Performance and convergence properties," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 120–140, Jan. 2016.
[12] M. Lázaro-Gredilla, L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-Garcia, "Adaptively biasing the weights of adaptive filters," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3890–3895, Jul. 2010.
[13] J. Arenas-García, M. Martínez-Ramón, A. Navia-Vázquez, and A. R. Figueiras-Vidal, "Plant identification via adaptive combination of transversal filters," *Signal Processing*, vol. 86, pp. 2430–2438, Sep. 2006.
[14] D. G. Tiglea, R. Candido, and M. T. M. Silva, "A low-cost algorithm for adaptive sampling and censoring in diffusion networks," *IEEE Transactions on Signal Processing*, vol. 69, pp. 58–72, Jan. 2021.
[15] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 54, pp. 1078–1090, Mar. 2006.
[16] C. Samson and V. U. Reddy, "Fixed poit error analysis of the normalized ladder algorithms," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, pp. 1177–1191, Oct. 1983.
[17] V. H. Nascimento and M. T. M. Silva, "Adaptive filters," in *Academic Press Library in Signal Processing: Signal Processing Theory and Machine Learning*, R. Chellapa and S. Theodoridis, Eds., vol. 1, chapter 12, pp. 619–761. Academic Press, Chennai, 2014.
[18] M. C. Costa and J. C. M. Bermudez, "An improved model for the normalized LMS algorithm with Gaussian inputs and large number of coefficients," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Orlando, FL, 2002, vol. II, pp. 1385–1388.
[19] T. Strutz, "Estimation of measurement-noise variance for variable-step-size nlms filters," in *Proc. EUSIPCO*, 2019, pp. 1–5.

VII. Supplementary Material

Let us begin by analyzing the right-hand side of (15). Condition i) can be enforced by making

$$[4\gamma + \mu(1-\rho)]^2 - 16\gamma\mu > 0, \tag{22}$$

which can be recast as

$$16\gamma^2 - 8\gamma\mu(1+\rho) + \mu^2(1-\rho)^2 > 0. \tag{23}$$

Solving for γ and taking into account that this parameter should be positive, we obtain

$$\gamma > \frac{\mu\left(1+\rho+2\sqrt{\rho}\right)}{4} = \frac{\mu(1+\sqrt{\rho})^2}{4}, \tag{24}$$

which corresponds to the first case in (16).

To enforce Condition ii), let us firstly assume that the right-hand side of (15) is real, i.e. Condition i) is satisfied. Then, we simply need to enforce

$$\frac{4\gamma+\mu(1-\rho)-\sqrt{[4\gamma+\mu(1-\rho)]^2 - 16\gamma\mu}}{4\gamma\mu} < 1. \tag{25}$$

Solving for γ, we obtain after some algebraic manipulations

$$\gamma > \frac{2+(\rho-1)\mu}{2(2-\mu)}, \tag{26}$$

which corresponds to the second case in (16).

Inequality (26) represents a stricter rule for the selection of γ than (24). Intuitively, this is reasonable, since we assumed that Condition i) was met – which Inequality (24) enforces – in order to obtain (24). Furthermore, there is a single value of μ for which the Rules (24) and (26) coincide. Making

$$\frac{\mu(1+\sqrt{\rho})^2}{4} = \frac{2+(\rho-1)\mu}{2(2-\mu)} \tag{27}$$

and solving for μ yields

$$\mu = \zeta \triangleq \frac{2}{(1+\sqrt{\rho})}, \tag{28}$$

as we defined in Sec. IV.

As stated in Sec. IV, the relevance of the parameter ζ stems from the fact that, for $\mu \geqslant \zeta$, Inequality (24) already enforces Condition ii) as well as Condition i). Therefore, in these cases we can relax the selection of γ by adopting (24) instead of (26). Taking this into account, we obtain (16) as a whole.

To verify this, let us consider the special case $\mu = \zeta$. In this case, Inequality (24) yields $\gamma > \frac{1+\sqrt{\rho}}{2}$. If we make

$$\gamma = \frac{1+\sqrt{\rho}}{2},$$

this leads to

$$\gamma\mu = 1$$

and

$$4\gamma+\mu(1-\rho) = 4.$$

Replacing these results in (15) yields $\mathrm{E}\{s(\infty)\} = 1$. Increasing γ by just a little bit, $\mathrm{E}\{s(\infty)\}$ decreases, so we can indeed enforce Condition ii) by adopting (24) in this case. Moreover, if we increase μ further while enforcing (24), $\mathrm{E}\{s(\infty)\}$ decreases. For example, if we select $\mu = a\zeta$, where $1 < a < 1+\sqrt{\rho}$ is a constant (the condition that $a < 1+\sqrt{\rho}$ stems from the fact that we should have $\mu < 2$), (24) yields $\gamma > a(1+\sqrt{\rho})$. If we adopt $\gamma = a(1+\sqrt{\rho})$, we conclude that, in this case,

$$\gamma\mu = a^2.$$

Furthermore,

$$2\gamma + \mu(1-\rho) = 4a.$$

Replacing these results in (15), we obtain

$$\mathrm{E}\{s(\infty)\} = \frac{4a - \sqrt{(4a)^2 - 16\cdot a^2}}{4a^2} = \frac{1}{a} < 1,$$

since $a > 1$. Therefore, we can see that for $\mu > \zeta$, (24) is sufficient to enforce Condition ii). On the other hand, from this argument we can also see that if $a < 1$, i.e. $\mu < \zeta$, Eq. (16) would yield $\mathrm{E}\{s(\infty)\} > 1$. Thus, Condition ii) is not satisfied. In this case, we need to adopt (26) for the selection of γ.