Exact Expectation Evaluation and Design of Variable Step-Size Adaptive Algorithms

Pedro Lara, Filipe Igreja, Luís D. T. J. Tarrataca, Diego B. Haddad, and Mariane R. Petraglia, *Senior Member, IEEE*

***Abstract*—The choice of a fixed step size in adaptive filtering algorithms implies a conflict between the convergence rate and the steady-state performance. In order to address this trade-off more effectively, variable step-size schemes have been proposed. The efficiency evaluation of such techniques requires comparisons of the resulting step size values with theoretical optimum values obtained from a statistical analysis of the adaptive algorithm convergence. The analysis generally employs statistical approximations, the most critical being the assumption of independence between the input signal and the filter coefficients. In this work it is argued that such a comparison can be misleading because the supposedly optimal step-size sequence sometimes induces divergence in the first phase of learning. This occurs most often when the input signal is colored and/or heavy-tailed. This instability trend can be explained by a convergence analysis that does not employ the independence hypothesis. In addition, the use of this exact analysis implies an optimal step-size sequence that can be significantly different from that obtained with standard analysis methods. This approach can be used to improve the design process of variable step size adaptive filtering algorithms.**

***Index Terms*—Adaptive Filtering, Variable Step Size, Exact Expectation Analysis.**

I. Introduction

An adaptive filtering scheme can be interpreted as a recursive and nonlinear estimator that stores in a vector $\boldsymbol{w}(k) \in \mathbb{R}^N$, where k is the iteration index, estimates of an optimal (and unknown) vector $\boldsymbol{w}^\star \in \mathbb{R}^N$. It is expected that the mean-square deviation (MSD), defined as

$$\xi(k) \triangleq \mathbb{E}\left[\|\boldsymbol{w}(k) - \boldsymbol{w}^\star\|^2\right], \tag{1}$$

where $\mathbb{E}[\cdot]$ is the expectation operator, will be progressively reduced in the transient phase until it attains the steady-state baseline. In supervised settings, which is the focus of this paper, a reference signal $d(k)$ is assumed to be available (sometimes by ingenious ways). It is also assumed that such a signal can be described by the affine-in-parameter data model

$$d(k) \triangleq \boldsymbol{x}^T(k)\boldsymbol{w}^\star + \nu(k), \tag{2}$$

This work was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico and in part by Fundação Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro. Pedro Lara, Filipe Igreja, Luís D. T. J. Tarrataca, and Diego B. Haddad are with Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Brazil (emails: pedro.lara@cefet-rj.br, figreja@gmail.com, luis.tarrataca@cefer-rj.br, diego.haddad@cefet-rj.br). M. R. Petraglia is with the Program of Electrical Engineering, Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Universidade Federal do Rio de Janeiro, 21945-970 Rio de Janeiro-RJ, Brasil (email: mariane@pads.ufrj.br).

where $\nu(k)$ accounts for measurement noise and/or modeling errors, and $\boldsymbol{x}(k) \triangleq \begin{bmatrix} x(k) & x(k-1) & \dots & x(k-N+1)\end{bmatrix}^T$ is the currently memory input data, assuming a tapped delay line filtering structure. The discrepancy between the output of the filter $y(k) \triangleq \boldsymbol{w}^T(k)\boldsymbol{x}(k)$ and the reference signal $d(k)$ is the error signal

$$e(k) \triangleq d(k) - y(k), \tag{3}$$

which normally feeds the estimation procedure.

One of the most popular adaptive algorithms is the least mean squares (LMS), which adopts a stochastic gradient of the mean squared error (MSE) function $\vartheta(k) \triangleq \mathbb{E}\left[e^2(k)\right]$, giving place to the recursive update equation

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta\boldsymbol{x}(k)e(k), \tag{4}$$

where β is the step size or learning factor. In general, small values of β produce good steady-state performance and small probability of divergence, while large step sizes lead to fast convergence and superior tracking ability [1]. Such a trade-off can be alleviated by a proper control policy that adjusts a time-variant step size $\beta(k)$ in a data-dependent manner [2]–[7]. In order to evaluate the robustness of a variable step size (VSS) scheme, it is necessary to compare the resulting $\beta(k)$ values to an optimal step-size sequence derived from a theoretical analysis [8]. The commonly employed convergence analysis uses the independence assumption (IA), which states that the deviation vector

$$\tilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}^\star - \boldsymbol{w}(k) \tag{5}$$

is statistically independent of $\boldsymbol{x}(k)$ [9], with the resulting step sizes denoted in this paper by $\beta_{\text{IA}}(k)$.

Section II raises some questions about the optimality of the sequence $\beta_{\text{IA}}(k)$, especially due to the fact that sometimes divergence is observed in practice. In Section III, such a phenomenon is theoretically explained by an *exact analysis* (EA), which does not employ IA [10]–[12]. Section IV devises a method to avoid divergence by providing an alternative step-size optimal sequence $\beta_{\text{EA}}(k)$. Section V displays the ensemble-average learning curves (EALCs) obtained with $\beta_{\text{IA}}(k)$ and $\beta_{\text{EA}}(k)$, showing that the EA step-size sequence leads to better performance than what is obtained with the IA one, and that such fact has implications on the design of new VSS adaptive algorithms. Section VI presents concluding remarks.

II. Classical Theoretical Step-size Sequence

Let us assume that the input signal $x(k)$ is stationary, so that the autocorrelation matrices of the input and weight vectors are $\boldsymbol{R_x} \triangleq \mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]$ and $\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) \triangleq \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right]$, respectively. Thus, the MSD and MSE can be written as

$$\xi(k) = \text{Tr}\left[\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right], \tag{6}$$

$$\vartheta(k) \approx \sigma_\nu^2 + \text{Tr}\left[\boldsymbol{R_x}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right]. \tag{7}$$

where Tr($\boldsymbol{A}$) denotes the trace of matrix $\boldsymbol{A}$. The above approximation employs IA and a new hypothesis, named noise independence assumption (NIA), which assumes that the measurement noise $\nu(k)$ is an independent and identically distributed (i.i.d.) random sequence with zero mean and statistically independent of the input signal $x(k)$.

Eqs. (6) and (7) explain the relevance of determining $\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)$ in most analyses. Normally, the adopted approach intends to establish a recursion for $\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k+1)$ in terms of $\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)$. Depending on the employed assumptions, different recursions can be obtained. For example, for a zero-mean Gaussian input and under IA and NIA conditions, such a recursion for a fixed step size can be written as [8]

$$\begin{aligned}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k+1) &= \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) - \beta\left[\boldsymbol{R_x}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) + \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\boldsymbol{R_x}\right] \\ &+\beta^2\left\{\boldsymbol{R_x}\text{Tr}\left[\boldsymbol{R_x}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right] + 2\boldsymbol{R_x}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\boldsymbol{R_x} + \sigma_\nu^2\boldsymbol{R_x}\right\},\end{aligned} \tag{8}$$

where σ_ν^2 is the variance of the measurement noise, which is assumed white and Gaussian.

Depending on the chosen assumptions and the metric to be optimized (MSD or MSE), different theoretical step sizes can be derived. Employing (8) and minimizing $\vartheta(k+1)$ yields [9]

$$\beta_{\text{IA}}(k) = \frac{\text{Tr}\left[\boldsymbol{R}_{\boldsymbol{x}}^2\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right]}{\text{Tr}\left[\boldsymbol{R}_{\boldsymbol{x}}^2\right]\text{Tr}\left[\boldsymbol{R_x}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right] + 2\text{Tr}\left[\boldsymbol{R}_{\boldsymbol{x}}^3\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right] + \sigma_\nu^2\text{Tr}\left[\boldsymbol{R}_{\boldsymbol{x}}^2\right]}, \tag{9}$$

where it can be observed that the learning factor $\beta_{\text{IA}}(k)$ tends to attain large values in the transient response, in order to achieve a high convergence rate [13]. Such a fact should raise concerns, because the IA is accurate solely when the step size is small [14], [15]. Additionally, the assessment of VSS techniques using EALCs could be misleading, since the results depend strongly on the number of independent repeated experiments, here denoted as K. This argument can be clarified through an example case. Let $\sigma_\nu^2 = 10^{-6}$ and $\boldsymbol{w}^\star$ be an ideal vector filled with 1's. Fig. 1 presents some EALCs with the IA theoretic sequence[1] $\beta_{\text{IA}}(k)$ designed for the minimization of $\xi(k)$. Note that EALCs evaluated with small values of K, e.g., $K = 10$, present a reasonable learning behavior, which may hinder the fact that divergence in the first phase of the learning process indeed occurs in some experiments (see the curves with $K = 1$, whose experiments were chosen in order to illustrate this issue). The impact of such a divergence, which occurs with low probability, is captured by EALCs for large values of K, e.g., $K = 10^4$. The importance of employing a large number of simulations in order to reasonably assess the mean-square behavior of adaptive filtering algorithms with constant step sizes is highlighted in [16]. Fig. 1 shows that such a concern is also relevant to VSS schemes.

[1]Such a sequence was computed using the statistical properties of the input signal, e.g., autocorrelation function and probability distribution.

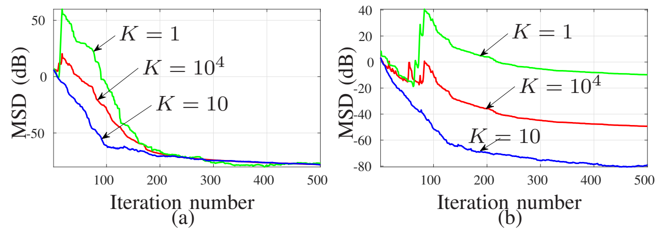

Fig. 1. MSD evolution of the LMS algorithm using the time-variant step-size sequence $\beta_{\text{IA}}(k)$ for K independent experiments (for $K = 1$, an experiment with divergence trend was selected): (a) unitary variance white Laplacian input signals ($N = 5$); (b) colored Gaussian input signals, obtained by filtering a unitary variance white Gaussian signal using $H(z) = 1 + 0.8z^{-1} + 0.2z^{-2}$ ($N = 3$).

III. Exact Expectation Analysis

In order to accurately analyze the behavior of a VSS adaptive filter that employs IA-based optimal step-size sequences, it is necessary to avoid the independence assumption. This is an important feature of the so-called exact expectation analysis [11], [12], which consists of a recursive and systematic procedure that creates a state vector $\boldsymbol{y}^{(\text{EA})}(k)$ containing the necessary information to derive the mean-square deviation evolution of the adaptive coefficients. The exact analysis approach assumes the following M-length FIR filter data model for $x(k)$ [12]:

$$x(k) = \sum_{m=0}^{M-1} b_m u(k-m), \tag{10}$$

where $u(k)$ is an i.i.d. zero-mean sequence, whose statistical moments are given by $\gamma_n \triangleq \mathbb{E}\left[u^n(k)\right]$. By a straightforward manipulation of Eq. (4) with variable step size $\beta(k)$, the recursion for each deviation error coefficient in terms of $u(k)$ can be written as

$$\begin{aligned}\tilde{w}_i(k+1) = \tilde{w}_i(k) - \beta(k)\sum_{m=0}^{M-1} b_m u(k-i-m)\,\cdot \\ \cdot\left\{\nu(k) + \sum_{j=0}^{N-1}\left[\sum_{n=0}^{M-1} b_n u(k-j-n)\right]\tilde{w}_j(k)\right\},\end{aligned} \tag{11}$$

where $\tilde{w}_i(k)$, for $i \in \{0, 1, \ldots, N-1\}$, is the ith element of $\tilde{\boldsymbol{w}}(k)$. Eq. (11) is an exact difference equation that should be squared for each value of i, before the application of the expectation operator $\mathbb{E}\left[\cdot\right]$, in order to provide second order statistical information about the deviations. If IA is employed, terms such as $\mathbb{E}\left[u^2(k-1)\tilde{w}_0(k)\tilde{w}_1(k)\right]$ can be approximated as

$$\mathbb{E}\left[u^2(k-1)\tilde{w}_0(k)\tilde{w}_1(k)\right] \approx \gamma_2\mathbb{E}\left[\tilde{w}_0(k)\tilde{w}_1(k)\right], \tag{12}$$

whereas in the case of exact analysis new recursive equations are required. For example, the evaluation of Eq. (12) requires multiplying $\tilde{w}_0(k+1)$ by $\tilde{w}_1(k+1)$, using Eq. (11), followed by the multiplication of the resulting terms by $u^2(k)$ before the application of the operator $\mathbb{E}[\cdot]$. The improved recursive equations may provide new terms that will in their turn require new equations. The recursion procedure eventually

TABLE I
LENGTH OF STATE VECTORS $\boldsymbol{y}^{(\mathrm{IA})}(k)$ AND $\boldsymbol{y}^{(\mathrm{EA})}(k)$ FOR SEVERAL COMBINATIONS OF N AND M.

N	M	L (IA)	L (EA)
1	1	1	**1**
1	2	1	**3**
1	3	1	**19**
1	4	1	**152**
1	5	1	**1341**
1	6	1	12546
1	7	1	122213
2	1	2	**5**
2	2	3	**48**
2	3	3	**394**
2	4	3	**3517**
2	5	3	33130
3	1	3	**37**
3	2	6	**698**
3	3	6	**6409**
3	4	6	60957
4	1	4	**330**
4	2	10	**9578**
4	3	10	94697
5	1	5	**3046**
5	2	15	127638
6	1	6	**28181**
7	1	7	262134
8	1	8	2438009

halts, since the correlation between $x(k_1)$ and $x(k_2)$ is zero for $|k_1 - k_2| > M$, which is guaranteed by the model in Eq. (10). Note that the exact analysis still employs NAI, which allows the cancellation of terms such as $\mathbb{E}\left[\nu(k)u^2(k-1)\tilde{w}_1^2(k)\right]$.

The recursive procedure generates terms (like the one in Eq. (12)) that are progressively stored in a state vector $\boldsymbol{y}^{(\mathrm{EA})}(k) \in \mathbb{R}^L$, where L is a function of N and M, which can be updated by the linear state equation system

$$\boldsymbol{y}^{(\mathrm{EA})}(k+1) = \boldsymbol{A}^{(\mathrm{EA})}(k)\boldsymbol{y}^{(\mathrm{EA})}(k) + \boldsymbol{b}^{(\mathrm{EA})}(k). \quad (13)$$

The elements $a_{i,j}^{(\mathrm{EA})}(k)$ of $\boldsymbol{A}^{(\mathrm{EA})}(k) \in \mathbb{R}^{L\times L}$ and $b_i^{(\mathrm{EA})}(k)$ of $\boldsymbol{b}^{(\mathrm{EA})}(k) \in \mathbb{R}^L$ are functions of $\beta(k)$, γ_n, b_m and σ_ν^2. Originally, $\boldsymbol{A}^{(\mathrm{EA})}(k)$ and $\boldsymbol{b}^{(\mathrm{EA})}(k)$ were time invariant [11], [12], which is not the case in this work due to the use of variable step size.

The application of IA allows the simplification of terms such as Eq. (12), which in general yields a simpler model, that is,

$$\boldsymbol{y}^{(\mathrm{IA})}(k+1) = \boldsymbol{A}^{(\mathrm{IA})}(k)\boldsymbol{y}^{(\mathrm{IA})}(k) + \boldsymbol{b}^{(\mathrm{IA})}(k), \quad (14)$$

with $\boldsymbol{A}^{(\mathrm{IA})}(k)$ and $\boldsymbol{b}^{(\mathrm{IA})}(k)$ having sizes much smaller than $\boldsymbol{A}^{(\mathrm{EA})}(k)$ and $\boldsymbol{b}^{(\mathrm{EA})}(k)$.

A C++ program was developed to execute symbolic operations, in order to efficiently derive the algebraic equations of models in Eqs. (13) and (14) and evaluate numerically the transition matrices $\boldsymbol{A}^{(\mathrm{EA})}(k)$ and $\boldsymbol{A}^{(\mathrm{IA})}(k)$ for every deterministic step-size sequence $\beta(k)$. Table I displays, for different combinations of N and M, the respective values of L obtained with the C++ code for both IA and EA analyses[2].

From the respective theoretical analysis, convergence of the VSS adaptive algorithm is predicted if all the eigenvalues of the transition matrices have magnitudes less than unity [17], [18]. Therefore, only the largest magnitude eigenvalue $\left|\lambda_{\max}^{(\mathrm{EA})}[\beta(k)]\right|$ or $\left|\lambda_{\max}^{(\mathrm{IA})}[\beta(k)]\right|$ of $\boldsymbol{A}^{(\mathrm{EA})}(k)$ or $\boldsymbol{A}^{(\mathrm{IA})}(k)$ needs to be calculated, to enable one to decide whether the use of a deterministic step-size sequence $\beta(k)$ may cause instability. The largest magnitude eigenvalue of a matrix can be computed using the power method [19].

Fig. 2 presents the evolutions of $\left|\lambda_{\max}^{(\mathrm{EA})}[\beta_{\mathrm{IA}}(k)]\right|$ and $\left|\lambda_{\max}^{(\mathrm{IA})}[\beta_{\mathrm{IA}}(k)]\right|$ over the iterations for the configurations depicted in Fig. 1. The IA analysis predicts that the learning process is stable in all iterations, which is not in agreement with the experimental results presented in Fig. 1, where divergence is observed in the transient phase. Such a phenomenon is explained by the exact analysis, which indicates that the stability bound of the step size is indeed violated by $\beta_{\mathrm{IA}}(k)$ in the first learning phase. The low probability of divergence is not contradictory to the instability predicted by the exact analysis, due to the fact that mean-square instability may imply that the deviation in some experiments assume very large values before converging to a small steady-state baseline [20].

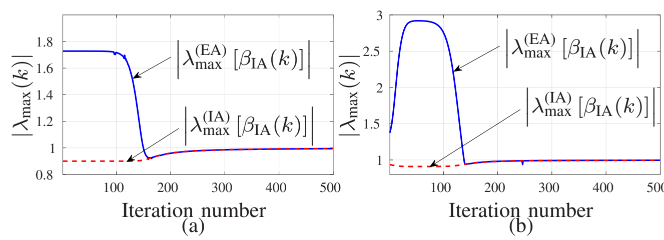

Fig. 2. Evolutions of the largest magnitude eigenvalues of the transition matrices $\boldsymbol{A}^{(\mathrm{EA})}$ and $\boldsymbol{A}^{(\mathrm{IA})}$ for step-size sequence $\beta_{\mathrm{IA}}(k)$: (a) configuration of Fig. 1(a); (b) configuration of Fig. 1(b).

IV. Exact Expectation Step-size Sequence

In the previous sections the risks involved in employing the optimum step-size sequence $\beta_{\mathrm{IA}}(k)$ derived from the IA analysis were described. In this section, a design procedure is proposed for the construction of an optimal step-size sequence that is aware of the correlations between the adaptive coefficients and past input samples. This awareness is guaranteed by using the exact analysis model in Eq. (13) for the evaluation of the MSD[3]. Note that $\xi(k) = \sum_{i=0}^{N-1}\mathbb{E}\left[\tilde{w}_i^2(k)\right]$ involves only a subset of elements of the $\boldsymbol{y}^{(\mathrm{EA})}(k+1)$ vector, that is, $y_j^{(\mathrm{EA})}(k+1)$ with $j \in \mathcal{I}$, where $\mathcal{I} = \{i_0, i_1, \ldots, i_{N-1}\}$ contains the indexes of such elements. From Eqs. (3), (4), (11) and (13), we obtain

$$\xi(k+1) = \sum_{j=0}^{N-1}\left[\sum_{l=0}^{L-1} a_{i_j,l}^{(\mathrm{EA})}(k)y_l^{(\mathrm{EA})}(k) + b_{i_j}^{(\mathrm{EA})}(k)\right], \quad (15)$$

[2] For the exact analysis, the values of L which have already been reported in the literature are shown in boldface. In [12], the reported number of equations L for $(N, M) = (3, 3)$ was 6449. We believe that the difference from our value presented in Table I is due to a typo.

[3] Similar reasoning can be employed for the MSE.

where $a_{i_j,l}^{(\mathrm{EA})}(k)$ and $b_{i_j}^{(\mathrm{EA})}(k)$ are functions of $\beta(k)$. At iteration k, the value of the step size that minimizes $\xi(k+1)$ can be obtained by differentiating Eq. (15) with respect to $\beta(k)$ and equating it to zero, which leads to

$$\sum_{j=0}^{N-1}\left[\sum_{l=0}^{L-1}\frac{\partial a_{i_j,l}^{(\mathrm{EA})}(k)}{\partial\beta(k)}y_l^{(\mathrm{EA})}(k)+\frac{\partial b_{i_j}^{(\mathrm{EA})}(k)}{\partial\beta(k)}\right]=0. \qquad (16)$$

Since $a_{i_j,l}^{(\mathrm{EA})}(k)$ and $b_{i_j}^{(\mathrm{EA})}(k)$ depend quadratically on $\beta(k)$, solving Eq. (16) is a simple task. Note that $\beta_{\mathrm{IA}}(k)$ can be computed in a similar manner using $a_{i_j,l}^{(\mathrm{IA})}(k)$ and $b_{i_j}^{(\mathrm{IA})}(k)$ in the above equation.

Fig. 3 shows the evolutions of the sequences $\beta_{\mathrm{EA}}(k)$ and $\beta_{\mathrm{IA}}(k)$ for the configurations of Fig. 1. The exact analysis suggests the use of more conservative step size values in the first training phase when compared to the IA sequence. As the iterations evolve, both sequences converge to the same values, which can be explained by the fact that the IA is accurate for small step sizes.

Fig. 4 shows the evolution of $|\lambda_{\max}^{(\mathrm{EA})}(k)|$ when the exact analysis is employed to derive the optimal step-size sequence. The stability upper bound is not violated in general, except in 3 only iterations (in Fig. 4(a)) and in 1 iteration in Fig. 4(b). Such violations are due to the fact that $\beta_{\mathrm{EA}}(k)$ was designed to minimize the sum of a subset of $\boldsymbol{y}^{(\mathrm{EA})}(k)$ elements, without considering the global stability of the system of Eq. (13). Contrary to what occurs in the IA analysis, it is easy to correct such atypical cases, since $\boldsymbol{A}^{(\mathrm{EA})}(k)$ is known and its largest magnitude eigenvalue (in terms of $\beta(k)$) can be computed, thereby allowing the reduction of the values of $\beta_{\mathrm{EA}}(k)$ if a stability issue is detected. This case is not addressed in this paper, since even with very large values of K (e.g., $K=10^7$) no divergence was experimentally observed, which is in agreement with the instantaneous MSD minimization undertaken by the solution of Eq. (16). Whereas the oscillatory behavior of $\beta_{\mathrm{EA}}(k)$ at the beginning of Fig. 4(a) cannot be emulated by a practical data-dependent VSS algorithm, a proper design may reduce the differences between the theoretical curve and the actual one, which can be accomplished by establishing an actual step-size sequence with values at approximately the center of the oscillations.

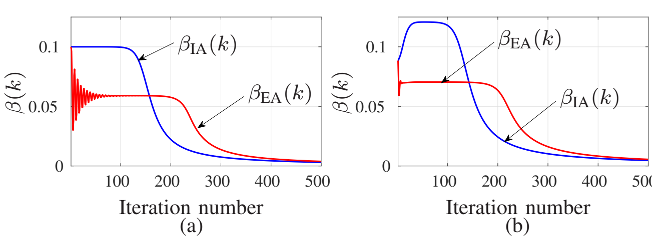

Fig. 3. Evolutions of the step-size sequences $\beta_{\mathrm{EA}}(k)$ and $\beta_{\mathrm{IA}}(k)$: (a) configuration of Fig. 1(a); (b) configuration of Fig. 1(b).

V. Evaluation of the Proposed Step-size Sequence

The final test for a candidate sequence consists in applying it to an actual learning configuration. Using the same features of the theoretical scenarios depicted in Fig. 1, Fig. 5 shows the MSD evolution of the VSS LMS algorithm using $\beta_{\mathrm{IA}}(k)$

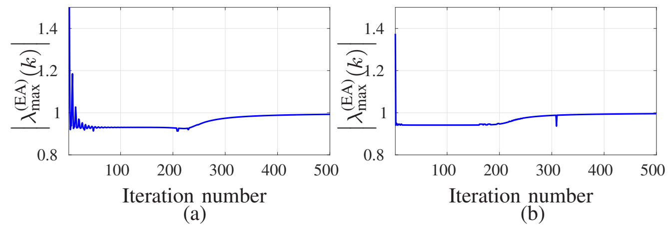

Fig. 4. Evolution of the largest magnitude eigenvalue of the transition matrix $\boldsymbol{A}^{(\mathrm{EA})}$ for step-size sequence $\beta_{\mathrm{EA}}(k)$: (a) configuration of Fig. 1(a); (b) configuration of Fig. 1(b).

and $\beta_{\mathrm{EA}}(k)$ as step-size sequences for $K=10^8$ experiments. Note that the exact analysis-based theoretic sequence presents better stability and satisfactory convergence properties.

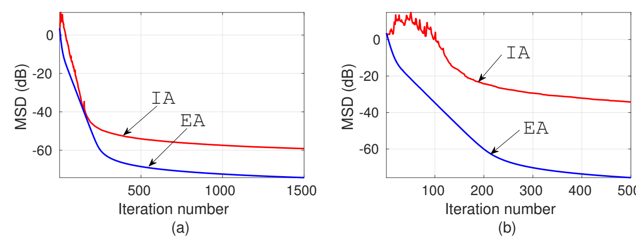

Fig. 5. Resulting MSD (dB) for the step-size sequences $\beta_{\mathrm{IA}}(k)$ and $\beta_{\mathrm{EA}}(k)$: (a) configuration of Fig. 1(a); (b) configuration of Fig. 1(b).

The *learning plane* is a simultaneous *theoretical* representation of the cost function and step size evolutions, and is useful for the design of VSS algorithms [8]. Fig. 6 presents the learning plane of both configurations considered in this paper (see Fig. 1). Note that there is a large difference between the two theoretical curves, which means that the proposed step-size sequence may influence both evaluation and design of new VSS algorithms [8].

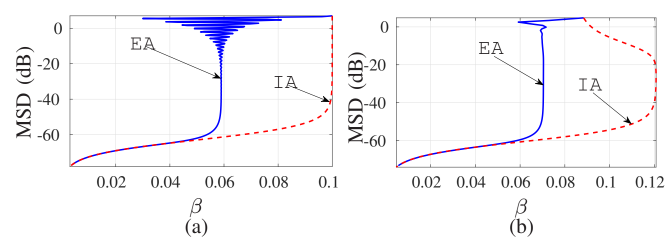

Fig. 6. Resulting MSD (dB) for the step-size sequences $\beta_{\mathrm{IA}}(k)$ and $\beta_{\mathrm{EA}}(k)$: (a) configuration of Fig. 1(a); (b) configuration of Fig. 1(b).

VI. Conclusion

In this work, it is shown that the use of the optimal deterministic step-size sequence derived using the independence assumption is prone to cause divergence, especially for colored and/or heavy-tailed input signals, as verified experimentally and theoretically. An alternative step-size sequence, based on the exact expectation analysis technique, is advanced, which can lead to new VSS adaptive filters with improved learning performance.

References

[1] A. H. Sayed, *Adaptive filters*. John Wiley & Sons, 2011.
[2] R. Harris, D. Chabries, and F. Bishop, "A variable step (vs) adaptive filter algorithm," *IEEE Transactions on Acoustics, Speech, and 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, Jul 1992.
[4] J. Benesty, H. Rey, L. R. Vega, and S. Tressens, "A nonparametric vss nlms algorithm," *IEEE Signal Processing Letters*, vol. 13, pp. 581–584, Oct 2006.
[5] S. B. Gelfand, Y. Wei, and J. V. Krogmeier, "Stability of variable and random stepsize lms," in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1961–1963 vol.3, Apr 1997.
[6] T. Aboulnasr and K. Mayyas, "A robust variable step-size lms-type algorithm: analysis and simulations," *IEEE Transactions on Signal Processing*, vol. 45, pp. 631–639, Mar 1997.
[7] L. C. Resende, D. B. Haddad, and M. R. Petraglia, "A variable step-size nlms algorithm with adaptive coefficient vector reusing," in *2018 IEEE International Conference on Electro/Information Technology*, pp. 1–6, May 2018.
[8] C. G. Lopes and J. C. M. Bermudez, "Evaluation and design of variable step size adaptive algorithms," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, vol. 6, pp. 3845–3848 vol.6, May 2001.
[9] S. O. Haykin, *Adaptive filter theory*. Pearson Higher Ed, 2013.
[10] S. Florian and A. Feuer, "Performance analysis of the lms algorithm with a tapped delay line (two-dimensional case)," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, pp. 1542–1549, Dec 1986.
[11] S. C. Douglas and T. H. Y. Meng, "Exact expectation analysis of the lms adaptive filter without the independence assumption," in *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 61–64 vol.4, Mar 1992.
[12] S. C. Douglas and W. Pan, "Exact expectation analysis of the lms adaptive filter," *IEEE Transactions on Signal Processing*, vol. 43, pp. 2863–2871, Dec 1995.
[13] S. Zhang and J. Zhang, "New steady-state analysis results of variable step-size lms algorithm with different noise distributions," *IEEE Signal Processing Letters*, vol. 21, pp. 653–657, June 2014.
[14] J. E. Mazo, "On the independence theory of equalizer convergence," *The Bell System Technical Journal*, vol. 58, pp. 963–993, May 1979.
[15] S. Jones, R. Cavin, and W. Reed, "Analysis of error-gradient adaptive linear estimators for a class of stationary dependent processes," *IEEE Transactions on Information Theory*, vol. 28, pp. 318–329, Mar 1982.
[16] V. H. Nascimento and A. H. Sayed, "Are ensemble-average learning curves reliable in evaluating the performance of adaptive filters?," in *Conference Record of Thirty-Second Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1171–1175, Nov 1998.
[17] H.-C. Shin, W.-J. Song, and A. H. Sayed, "Mean-square performance of data-reusing adaptive algorithms," *IEEE Signal Processing Letters*, vol. 12, pp. 851–854, Dec 2005.
[18] S. Zhang and J. Zhang, "Transient analysis of zero attracting nlms algorithm without gaussian inputs assumption," *Signal Processing*, vol. 97, pp. 100 – 109, 2014.
[19] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd Ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996.
[20] V. H. Nascimento and A. H. Sayed, "On the learning mechanism of adaptive filters," *IEEE Transactions on Signal Processing*, vol. 48, pp. 1609–1625, Jun 2000.