Signal, Image, and Video Processing manuscript No.
(will be inserted by the editor)

Advances on the Analysis of the LMS Algorithm with a Colored Measurement Noise

Pedro Lara[a,1], **Diego B. Haddad**[b,1], **Luís Tarrataca**[c,1]

[1]Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET-RJ, Petrópolis

Received: date / Accepted: date

Abstract Due to the inherent feedback feature of adaptive filtering algorithms, a comprehensive theoretical understanding of their learning process is still challenging. In order to make the mathematics tractable, most stochastic analyses adopt simplifications. One of these is the independence assumption, which presumes statistical independence between adaptive weights and input signal samples. Furthermore, the additive noise is usually assumed to be white, although it may not be the case in practice. This paper advances a novel theoretical model of the least mean square adaptive filter that, under the independence assumption, does not presume a white noise signal. Additionally, a theoretical result is established which implies that the stability properties of such an adaptive filter are not influenced by noise coloring. Such a result does not employ the above-mentioned assumptions, *i.e.*, it is valid for both colored noise and non-infinitesimally small step size values. An optimal step-size sequence that minimizes the mean-square deviation and takes into account the stochastic coupling of the excitation data and adaptive weights is proposed. It is noteworthy that such a design does not induce divergence in practice and does not assume, for the first time, a white measurement noise. The theoretical contributions are confirmed by simulations.

1 Introduction

Adaptive filtering algorithms are recursive stochastic estimators of a set of parameters (such as the taps of an unknown acoustic transfer function stored in the vector $\boldsymbol{w}^\star \in \mathbb{R}^N$) that usually present numerical stability, ease of implementation, low computational burden and ability to address nonstationary environments [1, 2]. Similarly to the majority of stochastic gradient descent methods, the least-mean-square (LMS) algorithm employs an instantaneous error signal sample $e(k)$ to adapt a linear filtering structure. The objective of such a time-variant structure is to extract from the input signal $x(k)$ the information of interest in an efficient way [3]. Despite the simplicity of the LMS update equation, it indeed implements a complex nonlinear estimator [4, 5].

The development of stochastic models capable of predicting both LMS performance and stability is of primary importance, since it provides guidelines and guarantees for the designer. However, modeling the LMS learning ability is a difficult task, especially when common statistical assumptions are not employed [6]. Two of the most critical assumptions (which may introduce significant deviations between theory and experiments) can be stated as follows:

$\star$ Independence Assumption (`IA`): the excitation vector[1] $\boldsymbol{x}(k) \in \mathbb{R}^N$ is statistically independent from the adaptive coefficient vector $\boldsymbol{w}(k) \in \mathbb{R}^N$. Although such an assumption is common in the field of stochastic approximations [3], it implies accurate predictions only when the step size $\beta \in \mathbb{R}_+$ is small [7].

$\star$ White Noise Assumption (`WNA`): the samples of the noise signal $v(k) \in \mathbb{R}$ are independent from each other. Although it is almost universal, this hypothesis is clearly violated in practice when the reference signal $d(k) \in \mathbb{R}$ is corrupted by an additive colored signal (such as a narrowband or a speech signal) [8].

This paper focuses on stochastic properties of the LMS algorithm without the usage of the ubiquitous `WNA`. The learning process under a colored noise signal is modeled using `IA` (see Section 2). Furthermore, Section 3 presents for the first time a proof (through Theorem 1) which establishes that the stability features of the LMS are not influenced by

[a]e-mail: pedro.lara@cefet-rj.br
[b]e-mail: diego.haddad@cefet-rj.br
[c]e-mail: luis.tarrataca@cefet-rj.br

[1]In this paper, all vectors are assumed to be of column-type.

the coloring of the additive noise signal. It is worth noting that such a theorem is very general, since *none* of the previous assumptions are necessary to establish the result. Section 4 devises a theoretical design procedure of an *deterministic* and optimal step-size sequence that considers the coloring of the noise and simultaneously avoids IA. Such a sequence permits one to address in a convenient way the trade-off between a fast convergence and a good steady-state performance. Since this sequence is constructed without the usage of IA, it is much less prone to induce divergence in the first phase of the learning process [9]. Furthermore, it is important to benchmark a given design against a theoretical optimum performance [10]. Theoretical predictions are confirmed by simulations in Section 5, which precedes the concluding remarks of the paper (Section 6).

2 Proposed Stochastic Analysis

Exact expectation analysis (EEA) [8, 11, 12, 9] is a sophisticated technique for predicting both stability and performance evolution of the LMS. One of the most useful metrics for the evaluation of an adaptive algorithm performance is the mean-squared error (MSE), defined as $\xi(k) \triangleq \mathbb{E}\left[e^2(k)\right]$, where $\mathbb{E}[\cdot]$ denotes the expectation operator. MSE is dependent on several joint statistical moments (or *state variables*) of several random variables. When both IA and WNA are employed, it can be approximated by [1]

$$\xi(k) \approx \sigma_v^2 + \mathrm{Tr}\left\{\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right\}, \tag{1}$$

where

$$\boldsymbol{x}(k) \triangleq \left[x(k)\; x(k-1)\; \ldots x(k-N+1)\right]^T,$$

$\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) \triangleq \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right]$ is the autocorrelation matrix of the deviation vector $\tilde{\boldsymbol{w}}(k) \in \mathbb{R}^N$ (defined by $\tilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}^\star - \boldsymbol{w}(k)$), σ_v^2 is the variance of the noise $v(k)$ and $\mathrm{Tr}[\boldsymbol{A}]$ denotes the trace of matrix $\boldsymbol{A}$. Note that approximation (1) does not depend on joint moments between input signal and deviation coefficient random variables. When standard EEA is adopted, such a fact is no longer valid. This analysis implements a systematic procedure that generates a time-invariant state equation system that permits an affine-in-the-parameters update of the required state variables. Originally, in order to avoid the IA it was proposed to incorporate the stochastic coupling between deviation weights $\tilde{w}_i(k) \in \mathbb{R}$ (for $i \in \{0, 1, \ldots, N-1\}$) and the input signal $x(k)$, assumed to be white [11]. The analysis performed in this paper (such as those presented in [12, 9, 8]) does not have this last restriction, since the excitation signal is assumed to be generated through an M-th order moving average (MA) process

$$x(k) = \sum_{m=0}^{M-1} b_m u(k-m), \tag{2}$$

where $u(k)$ is an i.i.d. signal distributed according to an even-symmetric pdf (which permits one to take into account zero-mean uniform, Gaussian and Laplacian distributions, for example).

The stochastic-gradient based update equation of the LMS adaptive filter may be written as the following

$$\begin{aligned}\boldsymbol{w}(k+1) &= \boldsymbol{w}(k) - \beta(k)\nabla_{\boldsymbol{w}(k)}\left[\frac{1}{2}e^2(k)\right]\\ &= \boldsymbol{w}(k) + \beta(k)\boldsymbol{x}(k)e(k),\end{aligned} \tag{3}$$

where $\beta(k)$ is the step size, learning factor or convergence factor in the k-th iteration [13], the error $e(k)$ is given by $e(k) \triangleq d(k) - \boldsymbol{w}^T(k)\boldsymbol{x}(k)$ and

$$\boldsymbol{w}(k) \triangleq \left[w_0(k)\; w_1(k)\; \ldots\; w_{N-1}(k)\right]^T.$$

WNA can be overcome by assuming that the noise signal is obtained from an L-th order MA model [8]:

$$v(k) = \sum_{l=0}^{L-1} a_l \eta(k-l), \tag{4}$$

where $\eta(k)$ is a white signal whose pdf, for simplicity, has the same restrictions than of the $u(k)$. The EEA procedure recursively incorporates into a state vector $\boldsymbol{y}(k) \in \mathbb{R}^P$ the state variables required for the analysis, which eventually provides a state space equation model

$$\boldsymbol{y}(k+1) = \boldsymbol{A}\boldsymbol{y}(k) + \boldsymbol{b}, \tag{5}$$

where transition matrix $\boldsymbol{A} \in \mathbb{R}^{P\times P}$ incorporates[2] all information necessary to predict the stability of the algorithm [12]. In a general setting, it is necessary to rewrite (3) in order to derive a recursion of the deviation vector, which provides the following non-homogeneous stochastic difference equation:

$$\tilde{\boldsymbol{w}}(k+1) = \left[\boldsymbol{I} - \beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\tilde{\boldsymbol{w}}(k) - \beta\boldsymbol{x}(k)v(k). \tag{6}$$

Note that (6) can be used to derive the following second-order recursion:

$$\begin{aligned}\tilde{\boldsymbol{w}}(k+1)\tilde{\boldsymbol{w}}^T(k+1) = {}& \tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k) - \beta\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\\ &+ \beta^2\boldsymbol{x}(k)\boldsymbol{x}^T(k)v^2(k) + \mathscr{O}[v(k)]\\ &- \beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\\ &+ \beta^2\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k),\end{aligned} \tag{7}$$

where $\mathscr{O}[v(k)]$ contains the first-order noise terms. In order to illustrate the proposed analysis procedure, please consider the case $(N, M, L) = (1, 2, 2)$, Describing the MSE of such a configuration requires an additional simplification:

[2] Note that P depends on N and M, as will be made clear further along.

⋆ Noise Independence Assumption (NIA): the zero-mean noise signal $\nu(k)$ is statistically independent from the input signal $x(k)$.

Remark: note that NIA is an almost universal hypothesis, which is plausible to assume in practice [14].

Using NIA, the MSE for the considered configuration can be expressed as

$$\xi^{(\text{EEA})}(k) \triangleq b_0^2\gamma_2\mathbb{E}[\tilde{w}_0^2(k)] + b_1^2\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] - 2b_1a_1b_0a_0\sigma_\eta^2\beta\gamma_2 + (a_0^2+a_1^2)\sigma_\eta^2, \quad (8)$$

where $\gamma_n \triangleq \mathbb{E}[u^n(k)]$ and σ_η^2 is the variance of $\eta(k)$. Note that (8) presents two state variables: $y_0(k) = \mathbb{E}[\tilde{w}_0^2(k)]$ and $y_1(k) = \mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)]$, with the last of these being approximated by $\gamma_2\mathbb{E}[\tilde{w}_0^2(k)]$ in the common scenario where IA is assumed to be valid. Since the MSE at the k-th iteration is dependent on these state variables, it is necessary to construct recursions for them. This permits one to predict the performance evolution along the iterations. The recursion for $y_0(k)$ can be found using (7), which degenerates in the considered setting into a scalar identity. Using NIA and the linearity property of the expectation operator, one obtains:

$$\begin{aligned}\mathbb{E}[\tilde{w}_0^2(k+1)] = {}& (b_0^4\beta^2\gamma_4 + 1 - 2b_0^2\gamma_2\beta)\mathbb{E}[\tilde{w}_0^2(k)] \\ &+ (6b_0^2\beta^2b_1^2\gamma_2 - 2b_1^2\beta)\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\ &+ b_1^4\beta^2\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)] \\ & 2b_1a_1\beta^2b_0a_0\sigma_\eta^2\gamma_2 - 6b_0^3\beta^3b_1a_1\gamma_2^2a_0\sigma_\eta^2 \\ &- 2b_1^3\beta^3a_1b_0a_0\sigma_\eta^2\gamma_4 + b_0^2a_0^2\sigma_\eta^2\beta^2\gamma_2 \\ &+ b_0^2a_1^2\sigma_\eta^2\beta^2\gamma_2 + b_1^2a_0^2\sigma_\eta^2\beta^2\gamma_2 \\ &+ b_1^2a_1^2\sigma_\eta^2\beta^2\gamma_2, \end{aligned} \quad (9)$$

where some *time-invariant* state variables are replaced by their equivalent expressions (e.g., the term $\mathbb{E}[u(k-1)v(k-1)\tilde{w}_0(k)]$ can be written as $-b_0a_0\sigma_2\beta\gamma_\eta^2$, which does not depend on k). Note that (9) depends on state variables $y_0(k)$ and $y_1(k)$ and also on a the nuisance term

$$y_2(k) = \mathbb{E}\left[u^4(k-1)\tilde{w}_0^2(k)\right].$$

Joint moment $y_2(k)$ is denominated as a *nuisance term* because one is not primarily interested in (see Eq. (8)), even though its estimation is necessary to establish the recursion for the statistical quantities of interest [15]. Multiplying both sides of (7) by convenient terms before the application of the expectation operator and using NIA, it is possible to obtain recursions for both $y_1(k)$ and $y_2(k)$. These recursions provide the foundation for the model (5), whose transition matrix $\boldsymbol{A}$ and state vector $\boldsymbol{y}(k)$ (for the considered configuration) can be described as

$$\mathbb{E}[u(k)v(k)\tilde{w}_0(k+1)] = -b_0a_0\sigma_\eta^2\beta\gamma_2, \quad (10)$$

$$\mathbb{E}[u^3(k)v(k)\tilde{w}_0(k+1)] = -b_0a_0\sigma_\eta^2\beta\gamma_4, \quad (11)$$

$$\begin{aligned}\mathbb{E}[u^2(k)\tilde{w}_0^2(k+1)] = {}& (b_0^4\beta^2\gamma_6 + \gamma_2 - 2b_0^2\gamma_4\beta)\mathbb{E}[\tilde{w}_0^2(k)] \\ &+ (6b_0^2\beta^2b_1^2\gamma_4 - 2b_1^2\beta\gamma_2)\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\ &+ b_1^4\beta^2\gamma_2\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)]2b_1a_1\beta^2\gamma_2^2b_0a_0\sigma_\eta^2 \\ &- 6b_0^3\beta^3b_1a_1\gamma_4a_0\sigma_\eta^2\gamma_2 - 2b_1^3\beta^3a_1\gamma_2b_0a_0\sigma_\eta^2\gamma_4 \\ &+ b_0^2a_0^2\sigma_\eta^2\beta^2\gamma_4 + b_0^2a_1^2\sigma_\eta^2\beta^2\gamma_4 + b_1^2a_0^2\sigma_\eta^2\beta^2\gamma_2^2 \\ &+ b_1^2a_1^2\sigma_\eta^2\beta^2\gamma_2^2. \end{aligned} \quad (12)$$

Recursions (9) and (12) are not self-contained, since they depend on the nuisance term $y_2(k)$. Using a similar procedure (*i.e.*, multiplying (7) by a convenient term before the application of the expectation operator), such a recursion can be derived as

$$\begin{aligned}\mathbb{E}[u^4(k)\tilde{w}_0^2(k+1)] = {}& (b_0^4\beta^2\gamma_8 + \gamma_4 - 2b_0^2\gamma_6\beta)\mathbb{E}[\tilde{w}_0^2(k)] \\ &+ (6b_0^2\beta^2b_1^2\gamma_6 - 2b_1^2\beta\gamma_4)\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\ &+ b_1^4\beta^2\gamma_4\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)]2b_1a_1\beta^2\gamma_4b_0a_0\sigma_\eta^2\gamma_2 \\ &- 6b_0^3\beta^3b_1a_1\gamma_6a_0\sigma_\eta^2\gamma_2 - 2b_1^3\beta^3a_1\gamma_4^2b_0a_0\sigma_\eta^2 \\ &+ b_0^2a_0^2\sigma_\eta^2\beta^2\gamma_6 + b_0^2a_1^2\sigma_\eta^2\beta^2\gamma_6 + b_1^2a_0^2\sigma_\eta^2\beta^2\gamma_4\gamma_2 \\ &+ b_1^2a_1^2\sigma_\eta^2\beta^2\gamma_4\gamma_2, \end{aligned} \quad (13)$$

so that, for the considered configuration, model (6) can be defined by (14)-(16).

$$\boldsymbol{A} = \begin{bmatrix} 1-2b_0^2\gamma_2\beta + b_0^4\beta^2\gamma_4 & -2b_1^2\beta + 6b_0^2\beta^2b_1^2\gamma_2 & b_1^4\beta^2 \\ \gamma_2 - 2b_0^2\gamma_4\beta + b_0^4\beta^2\gamma_6 & -2b_1^2\beta\gamma_2 + 6b_0^2\beta^2b_1^2\gamma_4 & b_1^4\beta^2\gamma_2 \\ \gamma_4 - 2b_0^2\gamma_6\beta + b_0^4\beta^2\gamma_8 & -2b_1^2\beta\gamma_4 + 6b_0^2\beta^2b_1^2\gamma_6 & b_1^4\beta^2\gamma_4 \end{bmatrix}, \quad (14)$$

$$\boldsymbol{y}(k) = \left[\mathbb{E}[\tilde{w}_0^2(k)]\ \mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)]\ \mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)]\right]^T. \quad (15)$$

It is noteworthy that most state variables in (15) are not nuisance terms. For more complex settings, such a fact is no longer valid: typically the nuisance terms are much more numerous than the joint moments we are primarily interested in. Note that model (14)-(16) (recently derived in [8]) is somewhat complex, although the underlying configuration is very simple. This is due to the avoidance of IA. The first contribution of this paper consists of combining IA without using WNA, which allows one to simplify the final model. This simplification is important, since the number of equations required in the standard EEA technique grows rapidly with N and M. For instance, for the configuration $(N,M) = (8,1)$ the EEA needs P = 2,438,009 state equations (which coincides with the number of rows P of matrix $\boldsymbol{A}$), whereas IA requires only 8 equations. Under the considered scenario, the resulting model can be simplified to the following relation

$$\begin{aligned}\mathbb{E}[\tilde{w}_0^2(k+1)] = {}& (1 - 2b_0^2\gamma_2\beta - 2b_1^2\gamma_2\beta + b_0^4\beta^2\gamma_4 + 6b_0^2\beta^2b_1^2\gamma_2^2 \\ &+ b_1^4\beta^2\gamma_4)\mathbb{E}[\tilde{w}_0^2(k)] + b_0^2a_0^2\sigma_\eta^2\beta^2\gamma_2 + b_0^2a_1^2\sigma_\eta^2\beta^2\gamma_2 \\ &+ b_1^2a_0^2\sigma_\eta^2\beta^2\gamma_2 + b_1^2a_1^2\sigma_\eta^2\beta^2\gamma_2, \end{aligned} \quad (17)$$

$$\boldsymbol{b} = \begin{bmatrix} 2b_1a_1\beta^2b_0a_0\sigma_2\gamma_2 - 6b_0^3\beta^3b_1a_1\gamma_2^2a_0\sigma_2 - 2b_1^3\beta^3a_1b_0a_0\sigma_2\gamma_4 + b_0^2a_0^2\sigma_2\beta^2\gamma_2 + b_0^2a_1^2\sigma_2\beta^2\gamma_2 + b_1^2a_0^2\sigma_2\beta^2\gamma_2 + b_1^2a_1^2\sigma_2\beta^2\gamma_2 \\ 2b_1a_1\beta^2\gamma_2^2b_0a_0\sigma_2 - 6b_0^3\beta^3b_1a_1\gamma_4a_0\sigma_2\gamma_2 - 2b_1^3\beta^3a_1\gamma_2b_0a_0\sigma_2\gamma_4 + b_0^2a_0^2\sigma_2\beta^2\gamma_4 + b_0^2a_1^2\sigma_2\beta^2\gamma_4 + b_1^2a_0^2\sigma_2\beta^2\gamma_2^2 + b_1^2a_1^2\sigma_2\beta^2\gamma_2^2 \\ 2b_1a_1\beta^2\gamma_4b_0a_0\sigma_2\gamma_2 - 6b_0^3\beta^3b_1a_1\gamma_6a_0\sigma_2\gamma_2 - 2b_1^3\beta^3a_1\gamma_4^2b_0a_0\sigma_2 + b_0^2a_0^2\sigma_2\beta^2\gamma_6 + b_0^2a_1^2\sigma_2\beta^2\gamma_6 + b_1^2a_0^2\sigma_2\beta^2\gamma_4\gamma_2 + b_1^2a_1^2\sigma_2\beta^2\gamma_4\gamma_2 \end{bmatrix}. \quad (16)$$

which is much simpler than the previous one. It is noteworthy that the proposed simplification implies less accurate predictions (compared to the EEA approach), but still is able to take into account the main factors driving algorithm learning performance. In this paper, all required symbolic manipulations are performed by a computationally-efficient C++ code.

3 Stability Analysis

The main contribution of this paper (see Theorem 1 below) is a demonstration that matrix $\boldsymbol{A}$ does *not* depend on the coloring of the additive measurement noise. Such a demonstration requires the following lemma.

Lemma 1 Expected value $\mathbb{E}\left[\eta(k-l)\prod_j u(k-m_j)^{e_j}\tilde{w}_i(k)\right]$ for $l \in [0, L-1]$, $m_j \in [0, M-1]$, $i \in [0, N-1]$, $e_j \in \mathbb{N}$, $k \in \mathbb{N}$ is time-invariant, *i.e.*, it does not depend on k.

Proof. We can write a scalar form of (6) as

$$\tilde{w}_i(k+1) = \tilde{w}_i(k) - \beta x(k-i)\left[\sum_{j=0}^{N-1} x(k-j)\tilde{w}_j(k)\right] - \beta v(k)x(k-i). \quad (18)$$

By multiplying both sides of (18) by appropriate terms and taking the expected value in order to obtain $\mathbb{E}\left[\eta(k-l)\prod_j u(k-m_j)^{e_j}\tilde{w}_i(k)\right]$, it is possible to derive the following recursion:

$$\mathbb{E}\left[\eta(k-l+1)\prod_j u(k-m_j+1)^{e_j}\tilde{w}_i(k+1)\right] = \mathbb{E}\left[\eta(k-l+1)\prod_j u(k-m_j+1)^{e_j}\tilde{w}_i(k)\right] + C_1 - \beta\sum_{m=0}^{M-1}\sum_{j=0}^{N-1}\sum_{m'=0}^{M-1} b_m b_{m'}\mathbb{E}\left[\eta(k-l+1)\prod_j u(k-\overline{m_j})^{\overline{e_j}}\tilde{w}_j(k)\right], \quad (19)$$

where C_1 is time-invariant. Note that in (19) new terms appear which require new recursions. This recursive behaviour eventually halts after $l+1$ steps (with the l-th step associated with a time-invariant term C_l). This happens because of the whiteness of $\eta(k)$, which guarantees the emergence of a null term in the expectation value of the base case. One may obtain new recursive state equations by increasing the temporal indices of η and u and multiplying the result by (18). After applying the expectation operation and performing this process $l+1$ times one has

$$\mathbb{E}\left[\eta(k+1)\prod_j u(k-m_j+l+1)^{e_j}\tilde{w}_i(k)\right] = \mathbb{E}\left[\eta(k+1)\right]\mathbb{E}\left[\prod_j u(k-m_j+l+1)^{e_j}\tilde{w}_i(k)\right] = 0. \quad (20)$$

Finally, one obtains $\mathbb{E}\left[\eta(k-l)\prod_j u(k-m_j)^{e_j}\tilde{w}_i(k)\right] = \sum_{l=1}^{L} C_l$ by recursively replacing previous state equations. □

Theorem 1: Matrix $\boldsymbol{A}$ does not depend neither on the value of a_i nor σ_v^2.

Proof. Squaring (18), one obtains:

$$\tilde{w}_i^2(k+1) = \tilde{w}_i^2(k) - 2\beta x(k-i)\left[\sum_{j=0}^{N-1} x(k-j)\tilde{w}_j(k)\right]\tilde{w}_i(k) \underbrace{-2\beta v(k)x(k-i)\tilde{w}_i(k)}_{*} + \beta^2x^2(k-i)\left[\sum_{j=0}^{N-1} x(k-j)\tilde{w}_j(k)\right]^2 \underbrace{-2\beta^2x^2(k-i)v(k)\left[\sum_{j=0}^{N-1} x(k-j)\tilde{w}_j(k)\right]}_{\Upsilon} + \underbrace{\beta^2v^2(k)x^2(k-i)}_{\dagger}. \quad (21)$$

Applying (4) and the expectation operator in terms containing $v(k)$ (see (†)):

$$\mathbb{E}[\beta^2v^2(k)x^2(k-i)] = \beta^2\mathbb{E}[v^2(k)x^2(k-i)] = \beta^2\sigma_\eta^2\sigma_u^2\sum_{l=0}^{L-1}\sum_{m=0}^{M-1} a_l^2 b_m^2. \quad (22)$$

Looking at $(*)$ we have

$$\mathbb{E}[-2\beta v(k)x(k-i)\tilde{w}_i(k)] = -2\beta\sum_{l=0}^{L-1}\sum_{m=0}^{M-1} a_l b_m\mathbb{E}[\eta(k-l)u(k-i-m)\tilde{w}_i(k)]. \quad (23)$$

By Lemma 1, this is time-invariant. For (Υ) we have (24) which, by Lemma 1, does not depend on k, with such terms located in vector $\boldsymbol{b}$ (see Eq. (5)). □

4 Variable Step-Size Design

As stated before, the employment of a fixed step size implies a trade-off between small steady-state error and fast convergence rate [16, 17]. Such an issue can be overcome by the usage of a variable step size (VSS) scheme, which imposes a fast convergence in the first phase of the learning process and, after some iterations, reduces its magnitude in order to enhance steady-state performance [18]. It is also desirable that such a sequence guarantees good tracking ability and is coherent with the state of the algorithm [13, 19]. The theoretical question concerning the optimal step size sequence for a given application is crucial for the adjustment of the algorithm's parameters in order to enforce such a sequence in practice [10].

The derivation of the optimal sequence is driven by the underlying hypothesis and the performance metric one desires to optimize. For example, assuming a zero-mean Gaussian input, `IA`, `NIA` and `WNA`, the step-size sequence that minimizes the mean square deviation (MSD) $\xi(k)$ defined by

$$\xi(k) = \mathrm{MSD}(k) \triangleq \mathbb{E}\left[\|\boldsymbol{w}^{\star} - \boldsymbol{w}(k)\|^2\right] = \sum_{i=0}^{N-1} \mathbb{E}\left[w_i^{\star} - w_i(k)\right]^2 \tag{25}$$

is described by [10, 20]

$$\beta(k) = \frac{\mathrm{Tr}\left[\boldsymbol{R}_{\boldsymbol{x}}^2 \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right]}{\mathrm{Tr}\left[\boldsymbol{R}_{\boldsymbol{x}}^2\right] \mathrm{Tr}\left[\boldsymbol{R}_{\boldsymbol{x}} \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right] + 2\mathrm{Tr}\left[\boldsymbol{R}_{\boldsymbol{x}}^3 \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right] + \sigma_{\mathrm{v}}^2 \mathrm{Tr}\left[\boldsymbol{R}_{\boldsymbol{x}}^2\right]}, \tag{26}$$

where σ_{v}^2 denotes the variance of the white noise, and $\boldsymbol{R}_{\boldsymbol{x}} \triangleq \mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]$ and $\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) \triangleq \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right]$ are, respectively, the autocorrelation matrices of the input data and of the deviation coefficients.

Assuming a time-variant step size, matrix $\boldsymbol{A}$ (that depends on $\beta(k)$) is no longer constant, so that model (5) should be generalized to

$$\boldsymbol{y}(k+1) = \boldsymbol{A}(k)\boldsymbol{y}(k) + \boldsymbol{b}, \tag{27}$$

where the MSD at the iteration $k+1$ can be written as

$$\xi(k+1) = \sum_{j=0}^{N-1} y_{g_j}(k+1), \tag{28}$$

where g_j denotes the index of the term $\mathbb{E}\left[\tilde{w}_j^2(k+1)\right]$ in the vector $\boldsymbol{y}(k)$. Note that the sum $\sum_{i=0}^{N-1} \mathbb{E}\left[\tilde{w}_i^2(k+1)\right]$ is required to predict $\xi(k+1)$ (see Eq. (25)). Each term $y_{g_j}(k+1)$ in Eq. (28) can be described as (see Eq. (27))

$$y_{g_j}(k+1) = \sum_{i=0}^{P-1} a_{g_j,i}(k) y_{g_j}(k) + b_{g_j}(k), \tag{29}$$

where P is the number of rows of the square matrix $\boldsymbol{A}(k)$, whose element at the (i,j) location is denoted by $a_{i,j}(k)$. Using (28)-(29), one concludes that minimizing $\xi(k+1)$ implies finding the $\beta(k)$ that enforces

$$\frac{\partial}{\partial \beta(k)} \left\{ \sum_{j=0}^{N-1} \left[\sum_{i=0}^{P-1} a_{g_j,i}(k) y_{g_j}(k) + b_{g_j}(k) \right] \right\} = 0. \tag{30}$$

Note that the solution of (30) for the first time takes into account *both* the noise coloring as well as the statistical dependence between the excitation data and the adaptive weights. Due to this last feature, the derivation of the theoretical sequence $\beta(k)$ is suitable even for non-infinitesimal step sizes. In variable step-size strategies, it is known that the usage of `IA` for obtaining an optimal deterministic step-size sequence can engineer divergence in practice [9]. It is noteworthy that such a divergence can occur with low probability, which implies that employing a small number of independent Monte Carlo trials may not reveal such an issue [9]. Since the derivation of the theoretical sequence $\beta(k)$ (obtained by solving (30)) does not employ `IA`, the employment of such a sequence does not engineer divergence in practice. In the next section, such a claim is confirmed by experiments.

5 Simulations

This section illustrates the theoretical contributions through three simulations. In these experiments, signals $u(k)$ and $\eta(k)$ are zero-mean Gaussian, $\sigma_u^2 = 1$, $\sigma_\eta^2 = 0.01$, and all coefficients of the ideal transfer function are equal to one.

5.1 Simulation I

In this experiment, analytic expressions for steady-state performance are derived under distinct assumptions. Such an expression can be very lengthy. Accordingly, the simple scenario $(N,M,L) = (1,2,2)$ is assumed, with $a_0 = b_0 = 1$, and $a_1 = b_1 = -0.9$. Due to the fact that the MSE depends on a linear combination of some state variables (see (8)), a theoretical prediction of the steady-state performance can be derived if one computes vector $\boldsymbol{y}_\infty \triangleq \lim_{k\to\infty} \boldsymbol{y}(k)$. If the algorithm indeed converges, such a vector can be evaluated using the closed-form expression $\boldsymbol{y}_\infty = (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{b}$ [12]. By incorporating in the stochastic model the noise coloring without the usage of `IA`, the steady-state MSE can be described as in (31), whereas the adoption of `IA` results in a simpler expression.

$$-2\beta^2\mathbb{E}\left[x^2(k-i)v(k)\left[\sum_{j=0}^{N-1}x(k-j)\tilde{w}_j(k)\right]\right] = -2\beta^2\sum_{m=0}^{M-1}\sum_{m'=0}^{M-1}\sum_{m''=0}^{M-1}\sum_{l=0}^{L-1}\sum_{j=0}^{N-1}\mathbb{E}[v(k-l)u(k-j-m)u(k-i-m')u(k-i-m'')\tilde{w}_i(k)]. \quad (24)$$

$$\lim_{k\to\infty}\text{MSE}^{(\texttt{EEA})}(k) \approx \frac{4.3\cdot10^{-4}-8.1\cdot10^{-3}\beta^6+1.5\cdot10^{-2}\beta^5+6.7\cdot10^{-3}\beta^4-1.3\cdot10^{-2}\beta^3+3.8\cdot10^{-3}\beta^2-1.7\cdot10^{-3}\beta}{2.4\cdot10^{-2}+\beta^5-2.6\cdot10^{-16}\beta^4-7.2\cdot10^{-1}\beta^3+2.3\cdot10^{-1}\beta^2-10^{-1}\beta}. \quad (31)$$

$$\lim_{k\to\infty}\text{MSE}^{(\texttt{IA})}(k) \approx \frac{5.93\cdot10^{-2}\beta^2}{-9.83\beta^2+3.62\beta}+1.81\cdot10^{-2}. \quad (32)$$

Notice that both expressions yield the same value (*i.e.*, the noise variance) for the steady-state MSE when $\beta \to 0$. More complex configurations than the chosen setting present lengthy steady-state equations, which cannot be shown in this paper, due to lack of space. Fig. 1 compares theoretical and empirical results, the last of these was obtained by using 10^6 Monte Carlo independent trials. Note that the IA-based theoretical curve underestimates the steady-state performance of the adaptive filter, especially when β presents high values.

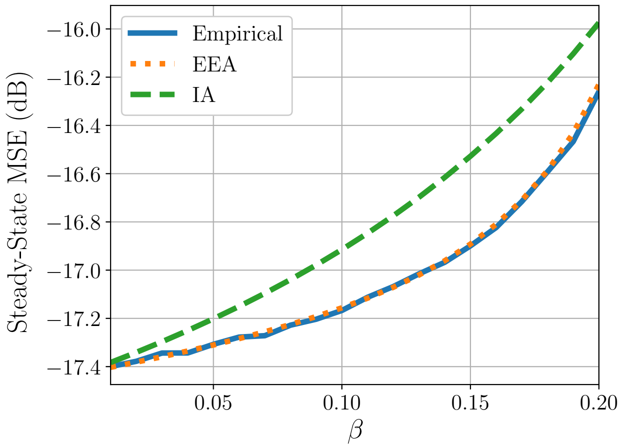

Fig. 1: Steady-state MSE (in dB) for the configuration $(N,M,L)=(1,2,2)$.

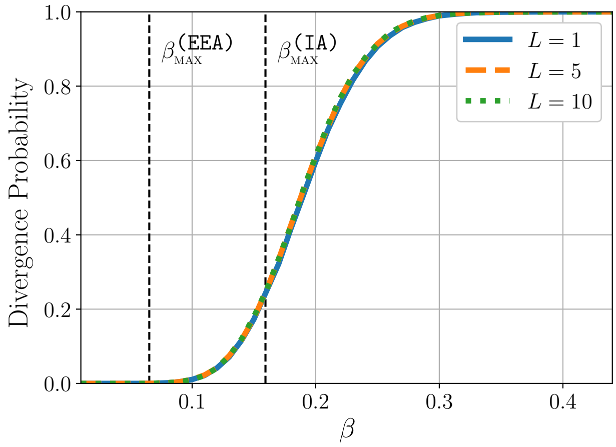

Fig. 2: Divergence probability of β for the configuration $(N,M)=(2,3)$.

are tested, for $L=1$ (white noise), and $L\in\{5,10\}$ (colored noise). Consider $\beta_{\text{MAX}}^{(\texttt{EEA})}$ (resp. $\beta_{\text{MAX}}^{(\texttt{IA})}$) as the maximum value of β that guarantees that the respective transition matrices (from analysis without IA and with IA, respectively) have absolute eigenvalues lower than the unity. Such an uppper bound can be evaluated through the Power Method [21] and theoretically specifies a stability region. Note in Fig. 2 that the empirical probability of divergence is zero when $\beta=\beta_{\text{MAX}}^{(\text{EEA})}$ and is approximately 25% when $\beta=\beta_{\text{MAX}}^{(\text{IA})}$. One may so conclude that the employment of the EEA provides a more reliable upper bound for the step size that guarantess stability. Furthermore, the divergence probability, as inferred from Theorem 1, is independent from the value of L.

5.2 Simulation II

In this experiment, the LMS is executed for different β values, with $b_n=(-1)^n\left[\frac{10-n}{10}\right]$ (for $n\in\{0,1,\ldots,L-1\}$). For each step size value, 2×10^5 realizations of the LMS learning process are performed, each of them along 340 iterations. If the absolute value of any adaptive coefficient surpasses 10 (*i.e.*, if there exists at least a single k for which $|w_i(k)|>10$, for $i\in\{0,1,\ldots,N\}$), we consider that the realization being evaluated has diverged. Different noise signals

5.3 Simulation III

Consider $\beta_{\texttt{EEA}}(k)$ (resp. $\beta_{\texttt{IA}}(k)$) the optimal step-size derived under EEA (resp. IA). Figure 3 depicts the theoretical step-size sequences obtained for both standard (i.e., IA-based) and exact analyses. Figure 4 presents the theoretical MSD evolution (obtained by the usage of IA) using $\beta_{\texttt{IA}}(k)$. Note that the empirical MSD curve is very different from the predicted one. This fact is due to the occurrence of divergence

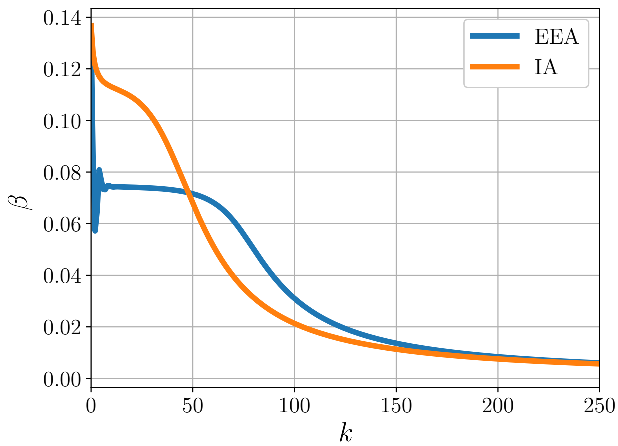

Fig. 3: Optimal theoretical step-size sequence for configuration $(N,M,L) = (3,2,3)$ employing both `IA`- and `EA`-based techniques.

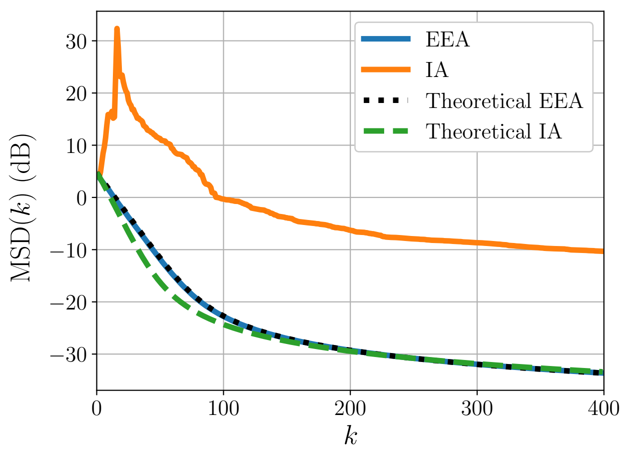

Fig. 4: Theoretical and empirical MSD (in dB) for both `EEA` and `IA`. Configuration $(N,M,L) = (3,2,3)$.

in the first iterations, since one cannot guarantee that the adaptive filter is stable under `IA` (see Fig. 2). However, employing `EEA` such a divergence is no longer observed (see [9] for a related discussion about this phenomenon). Furthermore, there is a good adherence between simulated and theoretical performance evolution. Notice that combining the optimal sequence $\beta(k)$ (see Fig. 4) with the theoretical performance evolution in Fig. 4 permits one to construct a *learning plane*, which provides to the designer the ability of judiciously selecting parameters that maximize both transient and steady-state performance of VSS-based schemes [10, 9].

6 Conclusions

This paper advances for the first time a simplified stochastic model of the LMS adaptive filter that incorporates the non-whiteness of the additive noise signal, assuming that `IA` is strictly valid. Furthermore, a theorem was presented that solves a long standing problem. Such a result forecasts that the stability property of the LMS remains, independently from noise coloring. This result is also remarkable because it was proven under very weak assumptions (e.g., without employing neither `IA` nor `WNA`). Furthermore, an optimal and deterministic step-size sequence is derived for both `IA` and `EEA` approches, *without* assuming a white noise signal. Such a sequence is important for adjusting the parameters of variable step-size algorithms. The theoretical results are confirmed by simulations.

Acknowledgements

This work has been supported by CNPq, FAPERJ and CAPES.

References

1. Simon Haykin. *Adaptive filter theory*. Prentice Hall, Upper Saddle River, NJ, 4th edition, 2002.
2. M. T. Khan and R. A. Shaik. Optimal complexity architectures for pipelined distributed arithmetic-based lms adaptive filter. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 66(2):630–642, Feb 2019.
3. B Widrow, John M. McCool, Michael Larimore, and C.Richard Johnson Jr. Stationary and nonstationary learning characteristics of the lms adaptive filter. *Proceedings of the IEEE*, 64:1151 – 1162, 09 1976.
4. K. J. Quirk, L. B. Milstein, and J. R. Zeidler. A performance bound for the lms estimator. *IEEE Transactions on Information Theory*, 46(3):1150–1158, May 2000.
5. B. S. Bourgeois and A. B. Martinez. Nonlinear behavior of the adaptive lms algorithm. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 42(2):122–124, Feb 1995.
6. V. H. Nascimento and A. H. Sayed. On the learning mechanism of adaptive filters. *IEEE Transactions on Signal Processing*, 48(6):1609–1625, June 2000.
7. L. Dogariu, S. Ciochina, and C. Paleologu. On the properties of the system mismatch covariance matrix in the LMS adaptive algorithm. In *2018 International Conference on Communications*, pages 39–44, June 2018.
8. Pedro Lara, Karen da S. Olinto, Felipe R. Petraglia, and Diego B. Haddad. Exact analysis of the least-mean-square algorithm with coloured measurement noise. *Electronics Letters*, 54:1401–1403(2), November 2018.

9. P. Lara, F. Igreja, L. D. T. J. Tarrataca, D. B. Haddad, and M. R. Petraglia. Exact expectation evaluation and design of variable step-size adaptive algorithms. *IEEE Signal Processing Letters*, 26(1):74–78, Jan 2019.
10. C. Guimaraes Lopes and J. C. Moreira Bermudez. Evaluation and design of variable step size adaptive algorithms. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*, volume 6, pages 3845–3848 vol.6, May 2001.
11. S. C. Douglas and T. H. . Meng. Exact expectation analysis of the lms adaptive filter without the independence assumption. In *ICASSP*, volume 4, pages 61–64, March 1992.
12. S. C. Douglas and Weimin Pan. Exact expectation analysis of the LMS adaptive filter. *IEEE Transactions on Signal Processing*, 43(12):2863–2871, Dec 1995.
13. D. I. Pazaitis and A. G. Constantinides. A novel kurtosis driven variable step-size adaptive algorithm. *IEEE Transactions on Signal Processing*, 47(3):864–872, March 1999.
14. M. Moinuddin and A. Zerguine. Steady-state analysis of the normalized least mean fourth algorithm without the independence and small step size assumptions. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3097–3100, April 2009.
15. J.-François Cardoso. Blind signal separation: statistical principles. *Proceedings of the IEEE*, 86(10):2009–2025, Oct 1998.
16. 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 (EIT)*, pages 0181–0186, May 2018.
17. W. Huang, L. Li, Q. Li, and X. Yao. Diffusion robust variable step-size lms algorithm over distributed networks. *IEEE Access*, 6:47511–47520, Sept. 2018.
18. T. Aboulnasr and K. Mayyas. A robust variable step-size lms-type algorithm: analysis and simulations. *IEEE Transactions on Signal Processing*, 45(3):631–639, Mar 1997.
19. F. Yang and J. Yang. Optimal step-size control of the partitioned block frequency-domain adaptive filter. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 65(6):814–818, June 2018.
20. Simon Haykin. *Adaptive filter theory*. Pearson Higher Ed, 2013.
21. Gene H. Golub and Charles F. Van Loan. *Matrix Computations (3rd Ed.)*. Johns Hopkins University Press, Baltimore, MD, USA, 1996.