# An Exact Expectation Model for the LMS Tracking Abilities

Thiago T. P. Silva, Pedro Lara, Filipe Igreja, Fernanda D. V. R. Oliveira, *member, IEEE*, Luís Tarrataca, and Diego B. Haddad, *member, IEEE*

***Abstract*—Nonstationary environments are ubiquitous in communications and acoustic systems. The ability to track their dynamics is one of the most desirable features of adaptive processing algorithms. The designers of these algorithms employ guidelines derived from stochastic analyses to adjust user-defined parameters to maximize performance or avoid stability issues. It is therefore important that analyzes of adaptive processing algorithms take into account the built-in sophisticated learning capabilities. This work presents a comprehensive model of the performance of the least mean square algorithm, operating under Markovian time-varying channels. Our advanced analysis considers both transient and steady-state regimes. Furthermore, in our analysis, the popular *independence assumption* is not adopted, resulting in a stochastic model which is accurate even when: (i) the step size is not infinitesimally small; or (ii) when the unknown system presents a high nonstationary degree. In addition, our evaluation is also able to provide a deterministic theoretical step-size sequence that optimizes algorithmic performance, as well as an accurate step size upper bound that guarantees algorithm stability. Computer simulations performed are in accordance with our theoretical predictions.**

***Index Terms*—Adaptive Filtering, Tracking, Exact Expectation Analysis.**

## I. Introduction

In applications such as channel equalization, hearing aids and echo cancellation, it is crucial that adaptive algorithms present a low asymptotic mean-squared error (MSE) [1]. It is sometimes argued that the choice of a fixed learning factor (or step size) $\beta$ is influenced by a trade-off between high convergence rate and good steady-state performance [2]. In practice, such a trade-off is more subtle, since time-varying channels introduce a "lag" in the adaptive process when emulating the optimal and unknown vector[1] $\boldsymbol{w}^{\star}(k) \in \mathbb{R}^N$ [3]. As a consequence, reducing $\beta$ does not necessarily imply that better steady-state performance will be attained [4].

The least mean squares (LMS) algorithm is a simple yet effective adaptive scheme that combines robustness with a numerically stable implementation. Furthermore, it demands a computational burden that is only proportional to the filter tap-length, since its update equation may be expressed in terms of an inner product of the current input vector. Additionally, the LMS performance dynamics depend on a feedback control mechanism, in which the error $e(k) \in \mathbb{R}$ drives the update of the adaptive coefficients $\boldsymbol{w}(k) \in \mathbb{R}^N$. This feedback procedure is responsible for the complex behaviour of the LMS (*e.g.*, its disturbance rejection feature [5]), which makes the development of analytical stochastic models challenging. Moreover, the fact that a proper adjustment of its learning factor remains a difficult task (since the optimal step size choice is an intricate function of the input statistics [6], [7]) is a driving force for the significant amount of effort invested in the development of new theoretical approaches. In order to maintain the mathematics tractable, LMS advanced models that can be found in the literature often rely on some simplifying assumptions, that are able to provide insights about factors that enhance algorithmic performance even when the resulting predictions are not accurate [3], [8], [9]. The most critical of them is the *independence assumption* (or *independence heuristic* [10]), which considers that the excitation data is statistically independent from the adaptive weights. Despite its wide usage in the stochastic approximations field [3], the independence assumption performs unreliable predictions when the step size parameter assumes large values. Nevertheless, some potential advantages of adaptive filtering schemes may become apparent only for large step sizes [11].

The exact expectation analysis (EEA) [12]–[18] is a refined technique that recursively generates update equations for a set of joint moments (or *state variables*). The joint moments update equations govern the dynamics of the LMS learning. Since EEA does not adopt the independence assumption, its predictions are a better fit for experimental results than standard approaches (*e.g.*, in the specification of an upper bound for the learning factor that ensures convergence [14], [16], [17]).

The EEA technique was proposed in [19], in a configuration where the excitation data is assumed to be white. An unequal-mode convergence behaviour in the variances of the filter coefficients (which contradicts the independence assumption) was theoretically predicted and confirmed by simulations. Work [12] extended the approach to the sign-data LMS algorithm, without assuming a white input data. A more comprehensive analysis of the LMS algorithm (assuming colored input data) under EEA is the focus of [13], [14], which effectively established EEA as a powerful and alternative analysis method for predicting LMS performance. More recently,

This work was partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES) Finance Code 001 and by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). Thiago T. P. Silva, Pedro Lara, Filipe Igreja, Luís Tarrataca and Diego B. Haddad are with Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Brazil (emails: thiagoteodoro501@gmail.com, pedro.lara@cefet-rj.br, figreja@gmail.com, luis.tarrataca@cefet-rj.br, diego.haddad@cefet-rj.br). Fernanda Oliveira is with Universidade Federal do Rio de Janeiro (email: fernanda.dvro@poli.ufrj.br)

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

this technique has experienced some generalizations, namely: [15] removed the whiteness assumption of the additive noise; [18] has demonstrated that coloring the additive noise does not impact mean-square stability; [20] addressed the identification of nonlinear plants; [16] employed EEA to derive an optimal sequence of step size values that optimize performance; and [17] modeled the deficient case that occurs when the adaptive filter length is surpassed by the length of the unknown transfer function the adaptive filter intends to emulate.

In this paper, EEA is for the first time employed in order to perform an in-depth analysis of asymptotic and non-asymptotic behaviour of the LMS under the time-varying plant setting. Popular assumptions (such as the *independence assumption*) are not used, which yields an accurate and more complex model. One may argue that our approach trades off computational cost for adherence with experiments. This point of view may be misleading, since standard approaches may incorrectly predict that the algorithm is stable in cases where it is not. This problem does not exist when the proposed stochastic model is used, and in these configurations there is no trade-off at all.

This paper is structured as follows. Section II concisely describes the LMS algorithm, whereas Section III depicts a standard stochastic model (*i.e.*, one based on the ubiquitous independence assumption) for its performance abilities. The devised exact expectation method is explained in Section IV, which is later utilized for constructing two versions of optimal and *deterministic* step-size sequences in Section V. Although the first version can be seen as an adaptation of the method devised in [16] to the tracking setting, the second version is an entirely new contribution. Comparisons between theoretical predictions and experimental results are presented and discussed in Section VI. Concluding remarks are presented in Section VII which outlines some possible promising future works that could be derived from it.

Throughout this paper, vector and matrices are represented with lowercase and uppercase bold fonts, respectively, while scalars are denoted by italics. The symbol $\otimes$ stands for the Kronecker product of matrices, and $\mathbb{E}[\cdot]$ is the mean operator. $(\cdot)^T$ denotes transpose and the vec$(\mathbf{A})$ operator generates a column-type vector by successively stacking the columns of the matrix $\mathbf{A}$ (*e.g.*, $\text{vec}\left(\begin{bmatrix} a & c \\ b & d \end{bmatrix}\right) = [a\ b\ c\ d\ ]^T$). Finally, $\mathbf{I}_N$ means the $N \times N$ identity matrix, whereas $\mathbf{0}_N$ denotes a null column vector with dimensions $N \times 1$.

## II. LMS

Suppose that the output of the ideal system (excited by a stationary stochastic input signal $x(k)$) is generated by the following noisy affine-in-the-parameters regression model:

$$d(k) = \left[\boldsymbol{w}^\star(k)\right]^T \boldsymbol{x}(k) + \nu(k), \tag{1}$$

where $d(k) \in \mathbb{R}$ denotes the reference signal, $\nu(k) \in \mathbb{R}$ is a zero-mean additive noise and

$$\boldsymbol{x}(k) \triangleq \begin{bmatrix} x(k) & x(k-1) & \dots & x(k-N+1) \end{bmatrix}^T. \tag{2}$$

Linear observation models similar to (1) are popular in the literature, since they are apt to capture many practical cases of interest [21].

The optimal set of coefficients $\boldsymbol{w}^\star(k)$ is updated according to a first-order stochastic multivariate Markovian random walk model:

$$\boldsymbol{w}^\star(k+1) = \boldsymbol{w}^\star(k) + \boldsymbol{q}(k), \tag{3}$$

where $\boldsymbol{q}(k) \in \mathbb{R}^N$ is a driving white perturbation random vector whose components are statistically independent. It is noteworthy that first-order Markovian variations are encountered in several realistic applications such as audio signal processing, acoustic echo cancellation and transmission systems [22]–[25]. In addition, LMS may outperform the Recursive Least Square (RLS) under slow-fading environments [26].

In a system identification setup, the LMS algorithm intends to emulate $\boldsymbol{w}^\star(k)$ from the observable data $\{\boldsymbol{x}(k), d(k)\}$. In order to do that, the LMS uses the stochastic approximation of the iterative steepest descent algorithm. By adopting the instantaneous quadratic error as the cost function, its update equation is derived from

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \beta \nabla_{\boldsymbol{w}(k)} \left[\frac{1}{2} e^2(k)\right], \tag{4}$$

where

$$e(k) \triangleq d(k) - y(k) = d(k) - \boldsymbol{w}^T(k)\boldsymbol{x}(k), \tag{5}$$

is an instantaneous discrepancy measure that is correlated with current adaptive performance. The LMS modification rule arises from (4):

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \boldsymbol{x}(k) e(k), \tag{6}$$

which uses a stochastic gradient of the cost function $\mathbb{E}\left[\frac{1}{2}e^2(k)\right]$. Alternatively, the LMS update equation (6) can be interpreted as an exact solver of the following *deterministic* problem:

$$\min_{\boldsymbol{w}(k+1)} \|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^2 \text{ s.t. } e_{\text{p}}(k) = \left(1 - \beta \|\boldsymbol{x}(k)\|^2\right) e(k), \tag{7}$$

where $e_{\text{p}}(k) \triangleq d(k) - \boldsymbol{w}^T(k+1)\boldsymbol{x}(k)$ denotes the posterior error. The term $\|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^2$ in (7) incorporates the minimum distortion principle, which adopts a conservative guideline that prevents the adaptive coefficients from suffering abrupt changes. Deterministic analyses may elucidate relevant features of the LMS algorithm, as, for example the presence of amplified components of high-frequency disturbances in the error output [5]. Standard statistical assumptions, which are usually applied on LMS stochastic models are discussed in the next section.

## III. Standard LMS Tracking Analysis

This section presents a concise theoretical tracking analysis of the LMS, which helps to emphasize by contrast some import features of the EEA. Among the reasons that motivate EEA, one may highlight that adaptive filters exhibiting good learning behavior in stationary environments do not necessarily present good tracking ability when identifying time-variant plants [27]. Considering the deviation vector $\tilde{\boldsymbol{w}}(k)$ as

$$\tilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}^\star(k) - \boldsymbol{w}(k), \tag{8}$$

and using (3) and (6), the following nonhomogeneous stochastic difference equation can be derived:

$$\tilde{\boldsymbol{w}}(k+1) = \left[\boldsymbol{I}_N - \beta \boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\tilde{\boldsymbol{w}}(k) - \beta \boldsymbol{x}(k)\nu(k) + \boldsymbol{q}(k), \tag{9}$$

where $\boldsymbol{I}_N$ is the identity matrix of size $N$. Rigorously speaking, the driving forces $-\beta\boldsymbol{x}(k)\nu(k)$ and $\boldsymbol{q}(k)$ in (9) precludes asymptotic convergence of the deviation vector energy to zero, even if the algorithm operates in the stability region. The average behavior of the deviation vector can be studied by applying the expectation operator in (9). Some complicated joint moments usually are not computed due to the use of several assumptions, namely:

$\star$ *Independence Assumption* (IA). The components of the instantaneous input vector $\boldsymbol{x}(k)$ are statistically independent from the adaptive weights $w_i(k)$, for $i \in \{0, \ldots, N-1\}$.

$\star$ *Noise Independence Assumption* (NIA). The zero-mean additive noise $\nu(k)$ is statistically independent from the excitation data.

*Remarks*: Although almost ubiquitous, IA-based analyses[2] are accurate only when small step sizes are utilized [14]. Although even more common, NIA is not valid in the presence of physical nonlinear phenomena, such as operation near the saturation region, diffusion capacitance, and intermodulation distortion [28]. In these cases the additive noise contains nonlinear components of the excitation data.

Using IA and NIA, the mean weight tracking behavior of the LMS can be written as

$$\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] = \left[\boldsymbol{I}_N - \beta\boldsymbol{R}_x\right]\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right], \tag{10}$$

where $\mathbb{E}[\cdot]$ represents the expectation operator and the term $\boldsymbol{R}_x \triangleq \mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]$ is the input autocorrelation matrix. Eq. (10) implies that the LMS performs an unbiased asymptotic estimation and that the adaptive coefficients, despite their Brownian motion, behave on average like the weights of the steepest descent algorithm. Moreover, under mild conditions, it may be stated that the expected value of $\boldsymbol{w}(k)$ converges to $\boldsymbol{w}^\star$ (in a stationary setting) if the step size satisfies

$$\beta < \frac{2}{\rho\left[\boldsymbol{R}_x\right]}, \tag{11}$$

with $\rho\left[\boldsymbol{R}_x\right]$ standing for the *spectral radius* of $\boldsymbol{R}_x$, *i.e.*, $\rho\left[\boldsymbol{R}_x\right] \triangleq \max_i \left|\lambda_i\left(\boldsymbol{R}_x\right)\right|$, where $\lambda_i$ is the $i$-th eigenvalue of $\boldsymbol{R}_x$.

Since stable-in-the-mean adaptive filters can diverge in practice due to an unbounded variance of the adaptive weight vector, the first-order analysis (10) is not very useful. Alternatively, a second-order (or mean square) analysis may be employed to estimate either the MSD or the MSE, defined as

$$\begin{aligned} \mathrm{MSD}(k) &= \varsigma(k) = \mathbb{E}\left[\left\|\tilde{\boldsymbol{w}}(k)\right\|^2\right], &(12)\\ \mathrm{MSE}(k) &= \xi(k) \approx \sigma_\nu^2 + \mathbb{E}\left[\boldsymbol{R}_x\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right], &(13)\end{aligned}$$

where $\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) \triangleq \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right]$ is the system mismatch covariance matrix and $\sigma_\nu^2$ is the additive noise variance. Approximation (13) employs the following additional assumption w.r.t. the stochastic characteristic of the additive noise:

$\star$ *White Noise Assumption* (WNA). The additive noise $\nu(k)$ is white.

*Remarks*: WNA remains as an almost universal hypothesis. Although it often is approximately true when $\nu(k)$ derives directly from quantization issues, one cannot assure that it will be always the case, since a colored interference (*e.g.*, a speech signal) may be captured by the transducers. Papers [15], [29] are important references that do not assume whiteness of the additive noise. Furthermore, approximation (13) helps to elucidate the major role that $\boldsymbol{R}_x$ plays in the standard statistical analyses of adaptive filtering algorithms. As will be clearer in the next section, such a role is less pronounced under the EEA paradigm.

Multiplying (9) by its transpose, one obtains

$$\begin{aligned} \boldsymbol{\Theta}(k+1) &= \boldsymbol{\Theta}(k) - \beta\boldsymbol{\Theta}(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k) + \mathcal{O}\left[\nu(k), \boldsymbol{q}(k)\right]\\ &\quad -\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\boldsymbol{\Theta}(k) + \beta^2\nu^2(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k) \quad (14)\\ &\quad +\beta^2\boldsymbol{x}(k)\boldsymbol{x}^T(k)\boldsymbol{\Theta}(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k) + \boldsymbol{q}(k)\boldsymbol{q}^T(k),\end{aligned}$$

where $\boldsymbol{\Theta}(k) \triangleq \tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)$ and $\mathcal{O}\left[\nu(k), \boldsymbol{q}(k)\right]$ contains first-order components of $\nu(k)$ and $\boldsymbol{q}(k)$ that do not impact the following derivations.

Obtaining a recursion for $\mathbb{E}\left[\boldsymbol{\Theta}(k)\right]$ requires us to consider operator vec$(\boldsymbol{A})$, using the identity

$$\mathrm{vec}(\boldsymbol{XYZ}) = \left[\boldsymbol{Z}^T \otimes \boldsymbol{X}\right]\mathrm{vec}(\boldsymbol{Y}), \tag{15}$$

and the linearity of the expectation operator, Eq. (14) leads to the following dynamic *time-invariant* state space model of the deviation error covariance matrix:

$$\boldsymbol{y}^{(\mathrm{IA},2)}(k+1) = \boldsymbol{A}^{(\mathrm{IA},2)}\boldsymbol{y}^{(\mathrm{IA},2)}(k) + \boldsymbol{d}^{(\mathrm{IA},2)}, \tag{16}$$

where the superscript $(\mathrm{IA}, 2)$ means that the state space linear model is obtained by adopting IA in a mean square (*i.e.*, second-order) stochastic analysis and

$$\begin{aligned} \boldsymbol{y}^{(\mathrm{IA},2)}(k) &\triangleq \mathbb{E}\left[\boldsymbol{\Theta}(k)\right], &(17)\\ \boldsymbol{A}^{(\mathrm{IA},2)} &\triangleq \boldsymbol{I}_{N^2} - \beta\mathbb{E}\left[\boldsymbol{I}_N \otimes \boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\\ &\quad -\beta\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k) \otimes \boldsymbol{I}_N\right] &(18)\\ &\quad +\beta^2\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k) \otimes \boldsymbol{x}(k)\boldsymbol{x}^T(k)\right],\\ \boldsymbol{d}^{(\mathrm{IA},2)} &\triangleq \beta^2\sigma_\nu^2\boldsymbol{R}_x + \mathbb{E}\left[\boldsymbol{q}(k)\boldsymbol{q}^T(k)\right]. &(19)\end{aligned}$$

Note that the *transition matrix* $\boldsymbol{A}^{(\mathrm{IA},2)}$ depends on $\beta$ and the analysis predicts stable operation if the magnitude of its maximum eigenvalue is less than unity [14]. Also important is that such an eigenvalue can be efficiently computed, for a fixed step size, using the power method [30]. Moreover, if the algorithm does not diverge, the relevant steady-state statistical quantities may be estimated by [16]

$$\lim_{k\to\infty} \boldsymbol{y}^{(\mathrm{IA},2)}(k) = \left(\boldsymbol{I}_N - \boldsymbol{A}^{(\mathrm{IA},2)}\right)^{-1}\boldsymbol{d}^{(\mathrm{IA},2)}. \tag{20}$$

The advanced analysis of this paper also aims to construct a linear state space equation system. However, the entire procedure, as will be described in the next section, is much more complex, because IA is no longer employed.

[2]In this paper, the acronym IA refers to independence assumption, where IA denotes the analysis that relies on the independence assumption.

## IV. Exact Expectation Tracking Analysis

Model (3) is the most commonly employed in the study of the effects of nonstationarity in adaptive filtering systems [31]. In this section, an analysis of the LMS tracking abilities is performed without IA for the first time. Several theoretical analyses also assume independence between samples of the input sequence, which can be stated as follows:

⋆ *White Input Assumption* (WIA). The excitation data $x(k)$ is a stationary sequence of independent random variables.

*Remark*: WIA is a very popular assumption that renders the analysis more tractable, but can be incoherent with practical settings [32], [33]. In this work, WIA is circumvented by modeling $x(k)$ as an output sequence of an $M$-length moving average system $B(z)$ applied to an unitary-variance white signal with an even probability density function (pdf) $u(k)$:

$$x(k) = \sum_{i=0}^{M-1} b_i u(k-i). \tag{21}$$

*Remark*: Note that model (21) does not assume a Gaussian $x(k)$, which is another popular hypothesis [8], [13], [34]–[36]. Its major restriction applies to the correlation between $x(k_1)$ and $x(k_2)$, which is supposed to be zero whenever $|k_1 - k_2| > M$.

Loosely speaking, the EEA is a systematic procedure that recursively generates update equations for the statistical quantities of interest [14]. The method constructs a linear state space model similar to (16), which contains the necessary information of the algorithm performance dynamics.

Our proposal does not employ WNA, because the additive noise can present a finite memory character, since it is assumed to be generated by an $L$-length moving average model:

$$\nu(k) \triangleq \sum_{i=0}^{L-1} a_i v(k-i), \tag{22}$$

where $v(k)$ is an unitary-variance i.i.d. process with a symmetric distribution. Also, (22) does not assume that $\nu(k)$ is Gaussian.

Consider the configuration $(N, M, L) = (1, 2, 2)$, which maintains equation lengths tractable, in the following derivations, where $N$ is the filter length and parameters $M$ and $L$ are related to the length of the filters responsible for the coloring of $x(k)$ and $\nu(k)$ (see (21)-(22)). Although the validity of the equations presented next is restricted to the considered configuration, the construction procedure is the same for more complex settings.

### A. Mean Weight Analysis

Recursion (9) allows for the writing of the following recursion:

$$\tilde{w}_0(k+1) = (1 - \beta x^2(k))\tilde{w}_0(k) - \beta x(k)\nu(k) + q_0(k), \tag{23}$$

where $\tilde{w}_i(k)$ (resp. $q_i(k)$) is the $i$-th element of vector $\tilde{\boldsymbol{w}}(k)$ (resp. $\boldsymbol{q}(k)$), for $i \in \{0, 1, \ldots, N-1\}$. In a standard stochastic approach (*i.e.*, an IA-based one), the application of the expectation operator in (23) together with NIA and (21) leads to

$$\mathbb{E}\left[\tilde{w}_0(k+1)\right] = \left[1 - \beta(b_0^2 + b_1^2)\gamma_2\right]\mathbb{E}\left[\tilde{w}_0(k)\right], \tag{24}$$

where $\gamma_n \triangleq \mathbb{E}\left[u^n(k)\right]$ and $b_i$ (for $i \in \{0, 1\}$) derives from model (21). Eq. (24) is a simpler recursion than the one obtained with EEA, since in the EEA case one cannot make use of the IA-based approximation[3] $\mathbb{E}\left[u^n(k-j)\tilde{w}_i^m(k)\right] \approx \gamma_n \mathbb{E}\left[\tilde{w}_i^m(k)\right]$, for $j > 0$. Therefore, EEA generates the following update equation for the state variable $\mathbb{E}\left[\tilde{w}_0(k)\right]$:

$$\mathbb{E}[\tilde{w}_0(k+1)] = (1 - b_0^2\beta\gamma_2)\mathbb{E}[\tilde{w}_0(k)] - b_1^2\beta\mathbb{E}[u^2(k-1)\tilde{w}_0(k)], \tag{25}$$

where one may note the emergency of a *nuisance* state variable $\mathbb{E}[u^2(k-1)\tilde{w}_0(k)]$. This joint moment is termed as a nuisance due to fact that it is necessary to compute it in order to update another state variable of interest (*i.e.*, $\mathbb{E}[\tilde{w}_0(k)]$) [37]. Since (25) is no longer a self-contained recursion (like (24)), an additional recursion for the nuisance term $\mathbb{E}[u^2(k-1)\tilde{w}_0(k)]$ is necessary. This recursion may be obtained by multiplying both sides of (23) by $u^2(k)$, before the application of the expectation operator, which leads to:

$$\begin{aligned}\mathbb{E}[u^2(k)\tilde{w}_0(k+1)] &= (\gamma_2 - b_0^2\beta\gamma_4)\mathbb{E}[\tilde{w}_0(k)] \\ &\quad - b_1^2\beta\gamma_2\mathbb{E}[u^2(k-1)\tilde{w}_0(k)].\end{aligned} \tag{26}$$

Equations (25) and (26) may be concisely rewritten as a self-contained linear state space equation system:

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

where

$$\boldsymbol{y}^{(1)}(k) = \begin{bmatrix}\mathbb{E}[\tilde{w}_0(k)] \\ \mathbb{E}[u^2(k-1)\tilde{w}_0(k)]\end{bmatrix} \tag{28}$$

contains the *state* of the statistical quantities the theoretical analysis is interested in at the $k$-th iteration and the time-invariant transition matrix is

$$\boldsymbol{A}^{(1)} = \begin{bmatrix}1 - b_0^2\beta\gamma_2 & -b_1^2\beta \\ \gamma_2 - b_0^2\beta\gamma_4 & -b_1^2\beta\gamma_2\end{bmatrix}. \tag{29}$$

The two eigenvalues of matrix $\boldsymbol{A}^{(1)}$ can be written as

$$\lambda_1 = \frac{1 + \sqrt{\Delta} + \left(-b_0^2 - b_1^2\right)\beta\gamma_2}{2}, \tag{30}$$

$$\lambda_2 = \frac{1 - \sqrt{\Delta} + \left(-b_0^2 - b_1^2\right)\beta\gamma_2}{2}, \tag{31}$$

where

$$\Delta = \beta^2 b_1^4\gamma_2^2 - 2\beta\left(\beta\left(\gamma_2^2 - 2\gamma_4\right)b_0^2 + \gamma_2\right)b_1^2 + \left(b_0^2\beta\gamma_2 - 1\right)^2. \tag{32}$$

The eigenvalues of the transition matrices depend on $\beta$ in a complex way, which indicates the higher sophistication of taking into account additional stochastic couplings. One may state that in the considered configuration the LMS is first-order stable in the exact expectation sense if $\beta$ is chosen so that $\max\{|\lambda_1|, |\lambda_2|\} \leq 1$. Note that neither the additive noise variance $\sigma_\nu^2$ nor $\sigma_q^2$ (the variance of $q_i(k)$, for $i \in \{0, 1, \ldots, N-1\}$) influence the stability in the mean sense.

[3]Random variable $\tilde{w}_i(k)$ is indeed statistically dependent from $u(k)$, although it is not the case with $x(k)$, unless when $x(k)$ is a white input signal (*i.e.*, $M = 1$ in (21)).

### B. Mean Square Analysis

Under the considered setting and circumventing the employment of the ubiquitous IA, the MSE at the $k$-th iteration may be expressed as

$$\begin{aligned}\text{MSE}(k) &= \mathbb{E}\left[e^2(k)\right] = \mathbb{E}\left[(\tilde{w}_0(k)x(k)+\nu(k))^2\right] \\ &= (a_0^2+a_1^2)\sigma_v^2 + b_0^2\gamma_2\mathbb{E}[\tilde{w}_0^2(k)] + b_1^2\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)].\end{aligned} \tag{33}$$

Thus, a theoretical EEA-based predicion of the MSE requires two statistical quantities:

$$\mathbb{E}[\tilde{w}_0^2(k)] \text{ and } \mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)]. \tag{34}$$

Finding a recursion for the former quantity may be done by squaring both sides of (23) and applying the expectation operator, resulting in

$$\begin{aligned}\mathbb{E}[\tilde{w}_0^2(k+1)] &= (1-2b_0^2\gamma_2\beta+b_0^4\beta^2\gamma_4)\mathbb{E}[\tilde{w}_0^2(k)]+\sigma_q^2 \\ &\quad +(6b_0^2\beta^2b_1^2\gamma_2-2b_1^2\beta)\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\ &\quad +(a_0^2+a_1^2)b_0^2\sigma_\nu^2\beta^2\gamma_2+(a_0^2+a_1^2)b_1^2\sigma_v^2\beta^2\gamma_2 \\ &\quad +b_1^4\beta^2\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)],\end{aligned} \tag{35}$$

where $\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)]$ is a nuisance state variable, whereas $\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)]$ is not, because it is required for computing the MSE (see (33)). Eq. (35) is not self-contained, since it requires the evaluation of two state variables, namely: $\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)]$ and $\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)]$. By carrying out similar steps to those that lead to (26), one may conclude that

$$\begin{aligned}\mathbb{E}[u^2(k)\tilde{w}_0^2(k+1)] &= (\gamma_2-2b_0^2\gamma_4\beta+b_0^4\beta^2\gamma_6)\mathbb{E}[\tilde{w}_0^2(k)]+\gamma_2\sigma_q^2 \\ &\quad +(6b_0^2\beta^2b_1^2\gamma_4-2b_1^2\beta\gamma_2)\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\ &\quad +b_0^2\left(a_0^2+a_1^2\right)\sigma_v^2\beta^2\gamma_4 \\ &\quad +b_1^2\left(a_0^2+a_1^2\right)\sigma_v^2\beta^2\gamma_2^2 \\ &\quad +b_1^4\beta^2\gamma_2\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)],\end{aligned} \tag{36}$$

and

$$\begin{aligned}\mathbb{E}[u^4(k)\tilde{w}_0^2(k+1)] &= (\gamma_4-2b_0^2\gamma_6\beta+b_0^4\beta^2\gamma_8)\mathbb{E}[\tilde{w}_0^2(k)]+\gamma_4\sigma_q^2 \\ &\quad +(6b_0^2\beta^2b_1^2\gamma_6-2b_1^2\beta\gamma_4)\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\ &\quad +b_0^2\left(a_0^2+a_1^2\right)\sigma_v^2\beta^2\gamma_6 \\ &\quad +b_1^2\left(a_0^2+a_1^2\right)\sigma_v^2\beta^2\gamma_4\gamma_2 \\ &\quad +b_1^4\beta^2\gamma_4\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)].\end{aligned} \tag{37}$$

A linear time-domain state space equation may concisely represent (35)-(37):

$$\boldsymbol{y}^{(2)}(k+1) = \boldsymbol{A}^{(2)}\boldsymbol{y}^{(2)}(k)+\boldsymbol{d}^{(2)}, \tag{38}$$

where

$$\boldsymbol{A}^{(2)} = \begin{bmatrix} 1-2b_0^2\gamma_2\beta+b_0^4\beta^2\gamma_4 & -2b_1^2\beta+6b_0^2\beta^2b_1^2\gamma_2 & b_1^4\beta^2 \\ \gamma_2-2b_0^2\gamma_4\beta+b_0^4\beta^2\gamma_6 & -2b_1^2\beta\gamma_2+6b_0^2\beta^2b_1^2\gamma_4 & b_1^4\beta^2\gamma_2 \\ \gamma_4-2b_0^2\gamma_6\beta+b_0^4\beta^2\gamma_8 & -2b_1^2\beta\gamma_4+6b_0^2\beta^2b_1^2\gamma_6 & b_1^4\beta^2\gamma_4 \end{bmatrix}, \tag{39}$$

$$\boldsymbol{y}^{(2)}(k) = \begin{bmatrix} \mathbb{E}[\tilde{w}_0^2(k)] \\ \mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\ \mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)] \end{bmatrix}, \tag{40}$$

and

$$\boldsymbol{d}^{(2)} = \begin{bmatrix} \sigma_q^2+b_0^2(a_0^2+a_1^2)\sigma_v^2\beta^2\gamma_2+b_1^2(a_0^2+a_1^2)\sigma_v^2\beta^2\gamma_2 \\ \gamma_2\sigma_q^2+b_0^2(a_0^2+a_1^2)\sigma_v^2\beta^2\gamma_4+b_1^2(a_0^2+a_1^2)\sigma_v^2\beta^2\gamma_2^2 \\ \gamma_4\sigma_q^2+b_0^2(a_0^2+a_1^2)\sigma_v^2\beta^2\gamma_6+b_1^2(a_0^2+a_1^2)\sigma_v^2\beta^2\gamma_4\gamma_2 \end{bmatrix}. \tag{41}$$

The model under consideration requires $R=3$ internal state variables, of which only one is a nuisance parameter. The increase of $N$ and $M$ implies a substantial increment of $R$, and for most configurations the majority state variables are nuisance joint moments (*e.g.*, for $(N,M,L)=(8,1,1)$, $R=2,438,009$, whereas the number of non-nuisance terms is only 29). The large amount of equations necessary for performing EEA means that commercially available algebraic symbolic packages are not useful. As a result, we decided to develop our own efficient C++-based code responsible for tackling the issue.

Vector $\boldsymbol{d}^{(2)}$ can be decomposed into two parts:

$$\boldsymbol{d}^{(2)} = \boldsymbol{d}_{\text{std}}^{(2)} + \boldsymbol{d}_q^{(2)}, \tag{42}$$

where $\boldsymbol{d}_q^{(2)} \in \mathbb{R}^R$ only contains terms that depends on statistic features of vector $\boldsymbol{q}(k)$. For example, by applying the decomposition on $\boldsymbol{d}^{(2)}$ from (41), $\boldsymbol{d}_q^{(2)}$ can be written as:

$$\boldsymbol{d}_q^{(2)} = \begin{bmatrix}\sigma_q^2 & \gamma_2\sigma_q^2 & \gamma_4\sigma_q^2\end{bmatrix}^T. \tag{43}$$

Decomposition (42) is important because it can describe asymptotic operation (assuming algorithm stability) as a sum of two terms:

$$\lim_{k\to\infty}\boldsymbol{y}^{(2)}(k) = \underbrace{\left(\boldsymbol{I}_R-\boldsymbol{A}^{(2)}\right)^{-1}\boldsymbol{d}_{\text{std}}^{(2)}}_{\triangleq\boldsymbol{y}_{\text{ss,std}^{(2)}}} + \underbrace{\left(\boldsymbol{I}_R-\boldsymbol{A}^{(2)}\right)^{-1}\boldsymbol{d}_q^{(2)}}_{\triangleq\boldsymbol{y}_q^{(2)}}, \tag{44}$$

where $\boldsymbol{y}_q^{(2)}$ isolates the contribution of the tracking task to steady-state statistical quantities. Obviously, such a decoupling is also possible with IA-based analyses.

Employing EEA, the upper bound $\beta_{\max}$ on the step size that guarantees that the state space realization is stable (*i.e.*, the algorithm converges) can be written as

$$\beta_{\max} \triangleq \sup\left\{\beta \text{ such that } \rho\left[\boldsymbol{A}^{(2)}\right]<1\right\}. \tag{45}$$

*Remarks*: Note that (45) is able to provide a more accurate upper bound, guaranteeing stability of the LMS, than the one provided by IA-based standard approaches, because instability issues occur for large step sizes, when such approaches are not precise [17]. The number of equations of even a slightly complex configuration (*i.e.* $(N,M,L)=(2,2,1)$) is too large to present in this paper because it requires 48 recursions, which can be obtained by following the same steps described in this section. In this case, the length of the recursions (which are not the larger ones of the considered configuration) of the state variables $\mathbb{E}[\tilde{w}_0^2(k)]$ and $\mathbb{E}[\tilde{w}_1^2(k)]$ are somewhat cumbersome:

$$\begin{aligned}\mathbb{E}[\tilde{w}_0^2(k+1)] = (1 &- 2b_0^2\gamma_2\beta + b_0^4\beta^2\gamma_4)\mathbb{E}[\tilde{w}_0^2(k)] \\ &+ (-2b_1^2\beta + 6b_0^2\beta^2 b_1^2\gamma_2)\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\ &+ (-2b_1 b_0\beta + 6b_0^3\beta^2 b_1\gamma_2)\mathbb{E}[u^2(k-1)\tilde{w}_1(k)\tilde{w}_0(k)] \\ &+ \vartheta\mathbb{E}[u(k-2)u(k-1)\tilde{w}_1(k)\tilde{w}_0(k)] + \sigma_q^2 \\ &+ b_0^2 a_0^2\sigma_\nu^2\beta^2\gamma_2 + b_0^4\beta^2\gamma_2\mathbb{E}[u^2(k-1)\tilde{w}_1^2(k)] \\ &+ 2b_0^3\beta^2 b_1\gamma_2\mathbb{E}[u(k-2)u(k-1)\tilde{w}_1^2(k)] \\ &+ b_0^2 b_1^2\beta^2\gamma_2\mathbb{E}[u^2(k-2)\tilde{w}_1^2(k)] + b_1^2 a_0^2\sigma_\nu^2\beta^2\gamma_2 \\ &+ b_1^4\beta^2\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)] \\ &+ 2b_1^3\beta^2 b_0\mathbb{E}[u^4(k-1)\tilde{w}_0(k)\tilde{w}_1(k)] \\ &+ 2b_1^4\beta^2\mathbb{E}[u^3(k-1)u(k-2)\tilde{w}_0(k)\tilde{w}_1(k)] \\ &+ b_1^2 b_0^2\beta^2\mathbb{E}[u^4(k-1)\tilde{w}_1^2(k)] \\ &+ 2b_1^3 b_0\beta^2\mathbb{E}[u^3(k-1)u(k-2)\tilde{w}_1^2(k)] \\ &+ b_1^4\beta^2\mathbb{E}[u^2(k-1)u^2(k-2)\tilde{w}_1^2(k)], \end{aligned} \tag{46}$$

where $\vartheta \triangleq -2b_1^2\beta + 6b_0^2\beta^2 b_1^2\gamma_2$, and

$$\begin{aligned}\mathbb{E}[\tilde{w}_1^2(k+1)] = \mathbb{E}[\tilde{w}_1^2(k)] &- 2b_0 b_1\beta\mathbb{E}[u^2(k-1)\tilde{w}_0(k)\tilde{w}_1(k)] \\ &- 2b_0^2\beta\mathbb{E}[u^2(k-1)\tilde{w}_1^2(k)] \\ &- 4b_0 b_1\beta\mathbb{E}[u(k-2)u(k-1)\tilde{w}_1^2(k)] \\ &- 2b_1^2\beta\mathbb{E}[u(k-1)u(k-2)\tilde{w}_0(k)\tilde{w}_1(k)] \\ &- 2b_1^2\beta\mathbb{E}[u^2(k-2)\tilde{w}_1^2(k)] + \sigma_q^2 + b_0^2 a_0^2\sigma_\nu^2\beta^2\gamma_2 \\ &+ b_0^4\gamma_2\beta^2\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\ &+ 2b_0^3\gamma_2\beta^2 b_1\mathbb{E}[u(k-1)u(k-2)\tilde{w}_0^2(k)] \\ &+ b_0^2 b_1^2\beta^2\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)] \\ &+ 2b_0^3 b_1\beta^2\mathbb{E}[u^4(k-1)\tilde{w}_0(k)\tilde{w}_1(k)] \\ &+ 6b_0^2 b_1^2\beta^2\mathbb{E}[u^3(k-1)u(k-2)\tilde{w}_0(k)\tilde{w}_1(k)] \\ &+ 2b_0 b_1^3\beta^2\mathbb{E}[u^3(k-1)u(k-2)\tilde{w}_0^2(k)] \\ &+ 6b_0 b_1^3\beta^2\mathbb{E}[u^2(k-1)u^2(k-2)\tilde{w}_0(k)\tilde{w}_1(k)] \\ &+ b_0^4\beta^2\mathbb{E}[u^4(k-1)\tilde{w}_1^2(k)] \\ &+ 4b_0^3\beta^2 b_1\mathbb{E}[u^3(k-1)u(k-2)\tilde{w}_1^2(k)] \\ &+ 6b_0^2\beta^2 b_1^2\mathbb{E}[u^2(k-1)u^2(k-2)\tilde{w}_1^2(k)] \\ &+ 4b_0 b_1^3\beta^2\mathbb{E}[u(k-1)u^3(k-2)\tilde{w}_1^2(k)] + b_1^2 a_0^2\sigma_\nu^2\beta^2\gamma_2 \\ &+ b_1^2 b_0^2\gamma_2\beta^2\mathbb{E}[u^2(k-2)\tilde{w}_0^2(k)] \\ &+ b_1^4\beta^2\mathbb{E}[u^2(k-2)u^2(k-1)\tilde{w}_0^2(k)] \\ &+ 2b_1^4\beta^2\mathbb{E}[u^3(k-2)u(k-1)\tilde{w}_0(k)\tilde{w}_1(k)] \\ &+ b_1^4\beta^2\mathbb{E}[u^4(k-2)\tilde{w}_1^2(k)]. \end{aligned} \tag{47}$$

In order to gain a better grasp of the computational burden of EEA, Table I presents a comparison of some key computational metrics between the EEA and IA methods. The `C++` program was compiled using the `-Ofast` optimization directive and we employed the `perf` performance analyzing tool in Linux to get the results. The table describes the number of instructions, cycles, amount of memory used (peak), equations generated and time elapsed for two different configurations. Namely, the number of CPU cycles and instructions is, roughly, between an order and two orders of magnitude worse for the EEA. Memory usage is increased by a constant factor. The number of equations required by the IA is in the single digits, and the execution time is reduced to less than one second. Overall, in all observed measures, the EEA method presents a significantly worse computational performance than the IA method.

TABLE I
COMPUTATIONAL PERFORMANCE COMPARISON FOR $(N, M, L) = (6, 1, 2)$ AND $(N, M, L) = (2, 4, 6)$.

| | $(N,M,L)=(6,1,2)$ | | $(N,M,L)=(2,4,6)$ | |
|---|---|---|---|---|
| | EEA | IA | EEA | IA |
| CPU Cycles | 106G | 23M | 961G | 1G |
| Instructions | 163G | 53M | 1.2T | 894M |
| Memory | 583M | 5M | 896M | 10M |
| Equations | 28181 | 6 | 3517 | 3 |
| Time | 33s | 0.02s | 291s | 0.33s |

## V. THEORETICAL DERIVATION OF AN OPTIMUM VARIABLE STEP-SIZE SEQUENCE

Once a constant step size has been chosen, a trade-off between convergence rate, steady-state performance and tracking ability is established. This trade-off may be circumvented by the adoption of a time-variant step size $\beta(k)$ [2], [38], whose adjustment relies on a data-dependent control policy [39]. This mechanism typically employs observable quantities, *e.g.:* the instantaneous quadratic error, estimated correlation between adjacent error samples or even the estimated correlation between the error and the input vector [40]. Paper [41] argues that the design (*i.e.*, the parameters tuning for better performance) of a variable step size (VSS) scheme depends on the evaluation of its efficiency, which may be performed by comparing it with an optimal theoretical sequence. Unfortunately, the use of `IA`-based derivation of such a *deterministic optimal* sequence may induce divergence in practice, because standard analyses overestimate the $\beta$ upper bound that guarantees stability [16]. Our work generalizes the EEA-based derivation of an optimal step-size sequence (proposed in [16]) in order to address the MSD minimization in the tracking setting. In this context, the involved optimum step sizes do not asymptotically converge to zero. Two versions of such an extension are described:

⋆ *Version I*: at each iteration, the step size $\beta(k)$ is the same for all adaptive weights;

⋆ *Version II*: each adaptive coefficient $w_i(k)$ is updated by the utilization of a particular step size $\beta_i(k)$.

*Remark*: The second version has the potential to enhance the design of proportionate adaptive algorithms, which distribute the "adaptation energy" in order to provide faster convergence of high-magnitude coefficients [8], [42], [43]. Since the updating process of each adaptive coefficient is independent from the other coefficients, the adaptation gains are distributed in order to speed up their convergence [44]. Proportionate algorithms are practical mainly because they determine the direction update based on current filter estimates without demanding any prior information [45].

### A. Theoretical VSS Sequence - Version I

From (12), the MSD at the $k$-th iteration can be written as:

$$\varsigma(k+1) = \sum_{i=0}^{N-1}\mathbb{E}\left[\tilde{w}_i^2(k)\right] = \sum_{i=0}^{N-1} y_{\mathcal{I}_i}^{(2)}(k+1), \tag{48}$$

where $\mathcal{I}_i$ is the position of vector $\boldsymbol{y}^{(2)}(k+1)$ containing state variable $\mathbb{E}\left[\tilde{w}_i^2(k)\right]$. Using (38), Eq. (48) can be explicitly rewritten as a function of elements of vector $\boldsymbol{y}^{(2)}(k)$:

$$\varsigma(k+1) = \sum_{i=0}^{N-1}\sum_{j=0}^{R-1} a_{\mathcal{I}_i,j}^{(2)}(k) y_j^{(2)}(k) + \sum_{i=0}^{N-1} d_{\mathcal{I}_i}^{(2)}(k), \tag{49}$$

where $a_{\mathcal{I}_i,j}^{(2)}(k)$ is the $(I_i, j)$ element of matrix $\boldsymbol{A}^{(2)}(k)$ and $d_{\mathcal{I}_i}^{(2)}(k)$ is the $\mathcal{I}_i$-th element of vector $\boldsymbol{d}^{(2)}(k)$. Note that these quantities are no longer time-invariant, since they now depend on a time-variant step size $\beta(k)$. Since $\beta(k)$ is an user-defined parameter, it can be adjusted in order to minimize the posterior MSD $\varsigma(k+1)$ by zeroing the following derivative:

$$\frac{\partial\varsigma(k+1)}{\partial\beta(k)} = \sum_{i=0}^{N-1}\sum_{j=0}^{R-1}\frac{\partial a_{\mathcal{I}_i,j}^{(2)}(k)}{\partial\beta(k)} y_j^{(2)}(k) + \sum_{i=0}^{N-1}\frac{\partial d_{\mathcal{I}_i}^{(2)}(k)}{\beta(k)} = 0, \tag{50}$$

which results in a linear equation w.r.t. $\beta(k)$, because both $a_{\mathcal{I}_i,j}^{(2)}(k)$ and $d_{\mathcal{I}_i}^{(2)}(k)$ depend quadratically on $\beta(k)$, see (39) and (41).

*Remarks*: The sequence $\{\beta(k)\}$ that solves (50) is *deterministic*. Furthermore, unlike traditional approaches, the optimization that provides the sequence construction procedure takes into account the stochastic coupling between the input data and the adaptive coefficients. This reduces the probability of divergence, especially in the initial phase of the learning curve [16]. The optimal sequence is necessary to plot the optimal trajectory in the *learning plane*, which provides useful guidelines for the designer of practical VSS strategies that rely on observable data [41]. It should be emphasized that in experiments, the deterministic sequence $\{\beta(k)\}$, chosen through (50), tends to assume larger values in the initial learning phase, and provides small steps

when the error magnitude is small, as expected. This type of behaviour avoids the trade-off between steady-state misadjusment and convergence rate that exists when a fixed learning factor is employed [46].

### B. Theoretical VSS Sequence - Version II

It is widely recognized that the use of different step sizes for each adaptive weight may enhance the convergence rate of the adaptive filter [8], [42], [43]. In this case, the LMS update equation is

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \boldsymbol{\Gamma}(k)\boldsymbol{x}(k)e(k), \tag{51}$$

where $\boldsymbol{\Gamma}(k) \in \mathbb{R}^{N\times N}$ is a diagonal matrix, with the $i$-th element of its main diagonal denoted as $\beta_i(k) \in \mathbb{R}_+$ (*i.e.*, the step size associated with the adaptive coefficient $w_i(k)$). Using similar steps to those presented in Section IV, it is possible to write a state space linear *time-variant* equation system:

$$\boldsymbol{y}^{(2)}(k+1) = \boldsymbol{A}^{(2,\mathrm{II})}(k)\boldsymbol{y}^{(2)}(k) + \boldsymbol{d}^{(2,\mathrm{II})}(k), \tag{52}$$

where $\boldsymbol{A}^{(2,\mathrm{II})}(k)$ and $\boldsymbol{d}^{(2,\mathrm{II})}(k)$ depend on $\{\beta_i(k)\}$ (for $i \in \{0, 1, \ldots, N-1\}$), and no longer on $\beta(k)$. This fact is indicated by the superscript "II". The optimal step sizes $\{\beta_i(k)\}$ for the $k$-th iteration can be evaluated by zeroing the following gradient

$$\nabla_{\boldsymbol{\beta}(k)}\left[\sum_{i=0}^{N-1}\sum_{j=0}^{R-1} a^{(2,\mathrm{II})}_{\mathcal{I}_i,j}(k)y^{(2)}_j(k) + \sum_{i=0}^{N-1} d^{(2,\mathrm{II})}_{\mathcal{I}_i}(k)\right] = \mathbf{0}_N, \tag{53}$$

where $\boldsymbol{\beta}(k) \triangleq \begin{bmatrix}\beta_0(k) & \beta_1(k) & \ldots & \beta_{N-1}(k)\end{bmatrix}^T$.

*Remark*: Usually, it is assumed that the step size sequence should asymptotically approach zero, in order to enhance steady-state performance by reducing the variance of the adaptive estimator. This is not true for the tracking scenario, which is the one considered in this paper.

## VI. Results

This section presents simulation results to confirm the theoretical analysis presented earlier. In these simulations we compare the results of the proposed analysis (EEA) and of the standard analysis (IA-based) with empirical results derived from computer-based LMS simulations, that mirror the actual behavior of the LMS. The analysis that better matches the empirical result presents a more accurate prediction of the LMS performance. The empirical curves were obtained by computing the LMS behavior for $K$ independent Monte Carlo trials.

In the following experiments, the elements of the initial vector $\boldsymbol{w}^\star(0)$ are equal to unity, and the adaptive coefficient vector are initialized with zeros. The zero-mean perturbation vector $\boldsymbol{q}(k)$ is white and Gaussian. Furthermore, their elements are statistically independent from the remaining random variables, which is coherent with the adopted Markovian model. Unless stated otherwise, $u(k)$ and the noise $\nu(k)$ are zero-mean, white and Gaussian. The input signal is generated by filtering a zero-mean unitary Gaussian signal using the filter $B(z) = 1-0.8z^{-1}$, in order to demonstrate that the advanced analysis does not require a white input signal.

Fig. 1 depicts the evolution of the expected value $\mathbb{E}[w_0(k)]$ along the iterations, for a scenario where $N = 5$, $M = 2$, $\beta = 0.075$, and $\sigma^2_\nu = \sigma^2_q = 10^{-2}$ . Since it is related to first-order statistics, one usually does not expect significant divergence between simulated and theoretical curves, even under IA. Fig. 1 demonstrates that this is not always the case, because the large value of $\beta$ in this configuration (relative to the stability upper bound) emphasizes the discrepancy of empirical results w.r.t. the standard model (*i.e.*, one based on IA). From Fig. 1, we can see that larger discrepancies happen in the beginning of the simulation. At iteration 25, the IA-based simulation expected value is around 15% higher than the EEA and empirical expected values.

Under stable operation, the MSE curve usually monotonically decreases up to some point, after which the algorithm reaches the steady-state operation. For some scenarios, though, as the one presented in Fig. 2, this description is inaccurate, but EEA can predict the correct behaviour.

As an example, consider the configuration $N = 3$, $\sigma^2_\nu = 10^{-3}$ and $\sigma^2_q = 10^{-2}$. Fig. 2 presents the MSE variation through the iterations for this configuration. In this scenario, EEA predicts an initial stage in which the algorithms experiment a performance worse than the one obtained during initialization. Fig. 2 shows that this theoretical EEA forecast matches with the experimental curve, since the empirical curve shows an MSE increase in the first iterations. The standard analysis, on the other hand, is not capable of predicting the MSE increase and instead shows a more conventional behaviour where MSE always decreases.

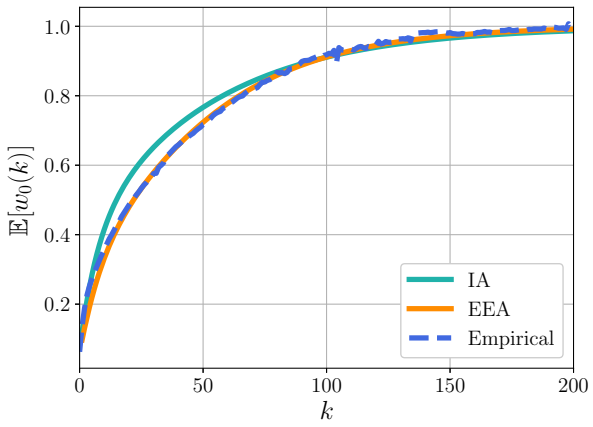


Fig. 1. Evolution of $\mathbb{E}[w_0(k)]$, with $K = 10^6$ independent Monte Carlo trials.

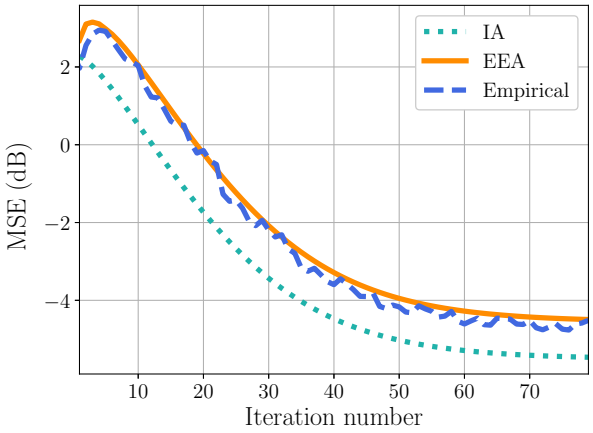


Fig. 2. Evolution of the MSE (in dB). The empirical results were obtained by the usage of $K = 10^{10}$ independent Monte Carlo trials.

Consider a scenario with $N = 3$, $M = 2$, $\sigma^2_q = 10^{-2}$, $\sigma^2_\nu = 10^{-3}$, which illustrates the advantage of our method in the determination of an accurate step size upper bound that avoids divergence. For each step size value, the LMS runs for 1000 iterations. The realization is assumed to experience instability if, in a specific realization, at least one adaptive parameter exceeds the threshold of 10 (*i.e.*, $|w_i(k)| > 10$). Using $K = 10^4$ independent trials, it is possible to estimate the divergence probability for each step size in a experimental manner. Fig. 3 shows the result of this estimation, that is, the probability of divergence as a function of $\beta$ for this scenario, in solid line, along with the theoretical upper bound on $\beta$ advocated by both IA and EEA analyses. The upper bound is obtained by the larger $\beta$ value for which the magnitude of the maximum eigenvalue of the transition matrix is less than unity [14]. Note that the EEA upper bound actually avoids divergence, when it is not violated, whereas the IA-based upper bound on $\beta$ is not trustworthy, since it results in approximately 70% of of divergence probability.

The asymptotic MSE for different values of the step size $\beta$ is shown in Fig. 4, for a case where $N = 6$, $M = 1$, $\sigma^2_\nu = 10^{-8}$ and $L = 5$. The colored additive noise is generated by filtering an white Gaussian noise by the following filter

$$A(z) = 1 - 0.8z^{-1} + 0.8z^{-2} - 0.7z^{-3} + 0.6z^{-4}. \tag{54}$$

As expected, Fig. 4 shows that larger step sizes lead to lower accuracy of IA-based analysis. That is, with the IA-based analysis we are not able to precisely predict the asymptotic MSE for $\beta$ higher than 0.05.

The misadjustment, given by

$$\mathcal{M}(k) \triangleq \frac{\mathbb{E}[e^2(k)] - \sigma^2_\nu}{\sigma^2_\nu}, \tag{55}$$

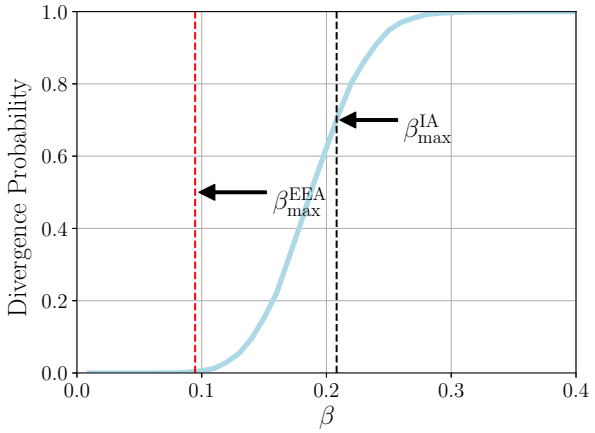


Fig. 3. Divergence probability as a function of $\beta$. The empirical curve (in blue) was obtained by the use of $K = 10^4$ independent Monte Carlo trials.

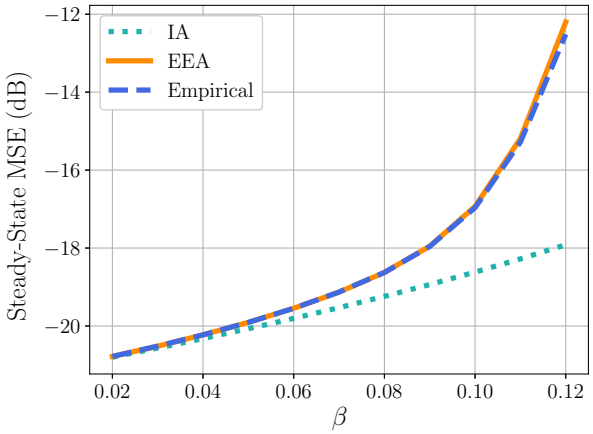


Fig. 4. Steady-state MSE as a function of the step size. The empirical curve was generated with $K = 10^5$ independent Monte Carlo trials.

is another important criterion used to asses the filter performance. The misadjustment is directly related to the excess MSE. In Fig. 5 we plot the evolution of the misadjustment considering $N = 3$, $M = 2$, $L = 1$, $\beta = 0.082$, $\sigma_\nu^2 = 10^{-2}$, and $\sigma_q^2 = 10^{-8}$. In this scenario, IA does not predict the misadjustment increase, in the first iterations which is a behavior similar to the one seen in Fig. 2. Using EEA, on the other hand, we are capable of correctly computing the misadjustment. Even in steady-state, the IA-based analysis predicts that the misadjustment is 0.5 dB smaller than actually is.

Fig. 6 presents steady state MSE, but for a scenario with $N = 2$, $\sigma_\nu^2 = 10^{-3}$, and $\sigma_q^2 = 10^{-2}$. The curves of both Figs. 4 and 6 are important because they convey information that can be used as a guideline for the choice of the best step size (*i.e.*, one that yields the lower asymptotic MSE). Note that in Fig. 6 the adoption of the IA-based analysis leads to the misconception that the best step size is close to 0.14, but from the empirical results one can see that this step size value is actually located near the instability region. The EEA, on the other hand, predicts precisely the steady-state MSE for each step size, allowing for the designer to choose a $\beta$ that results in the optimal MSE in practice. Additionally, the impact of the Markovian noise $\boldsymbol{q}(k)$ can be assessed through decomposition (44), because the asymptotic MSE can be decoupled by a sum of two easily distinguishable contributions:

$$\lim_{k\to\infty} \text{MSE}(k) = \text{MSE}_{\text{ss,std}} + \text{MSE}_{\text{ss},q}, \tag{56}$$

where $\text{MSE}_{\text{ss},q}$ (*i.e.*, the tracking delay [47]) isolates the influence of the perturbation $\boldsymbol{q}(k)$, whereas $\text{MSE}_{\text{ss,std}}$ coincides with the steady-state mean squared error of the stationary case. The decomposition results are presented in Figs. 7 and 8. Fig. 7 depicts $\text{MSE}_{\text{ss},q}$ with respect to different $\beta$ values. From Fig. 7 we can see that the standard analysis is not able to accurately model the impact of large step sizes, which is also valid for $\text{MSE}_{\text{ss,std}}$, as depicted in Fig. 8. It is interesting to note that such a lag contribution is not $\beta$-decreasing, as argued by some papers (*e.g.*, [47]). Furthermore, notice, by comparing Figs. 6, 7 and 8, that the contribution of $\text{MSE}_{\text{ss},q}$ is much more significant to the overall MSE (Fig. 6), under the considered setting.

The ability of the EEA to determine the optimal step size even in a high nonstationary scenario (*i.e.*, one with a large $\sigma_q^2$) can be seen in Fig. 9, in which the distribution of the signal $u(k)$ is Laplacian[4]. This experiment was performed with $N = M = 2$ and $\sigma_\nu^2 = 10^{-3}$. In view of the unimodality of the steady-state MSE as a function of $\beta$ (see Fig. 6), for each value of $\sigma_q^2$ a ternary search was used to obtain the $\beta$ that minimizes the empirical steady-state value of the MSE. In addition, EEA indicates a proper step size value that optimizes asymptotic performance.

As stated in Section V, the employment of a fixed step size value is often not an optimal choice, due to the inherent trade-off between convergence rate, asymptotic performance and instability issues. The impact of using the deterministic variable step size (VSS) derived in Section V can be seen in Fig. 10, which presents the evolution of a theoretical sequence $\beta(k)$, obtained using Version I of the VSS method described in Section V. The considered setup employs $N = 3$, $M = 2$, $\sigma_\nu^2 = 10^{-4}$ and $\sigma_q^2 = 10^{-4}$. It should be emphasized that the standard model suggests a larger step size for

[4]The Laplacian distribution was chosen to demonstrate that the advanced model does not require a Gaussian input signal

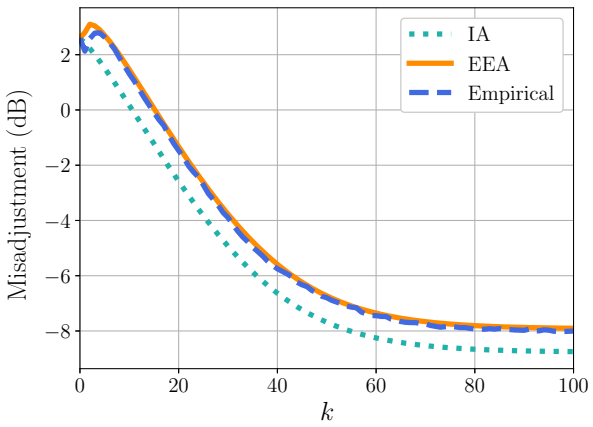


Fig. 5. Misadjustment evolution (in dB). The empirical results were obtained by the usage of $K = 10^7$ independent Monte Carlo Trials.

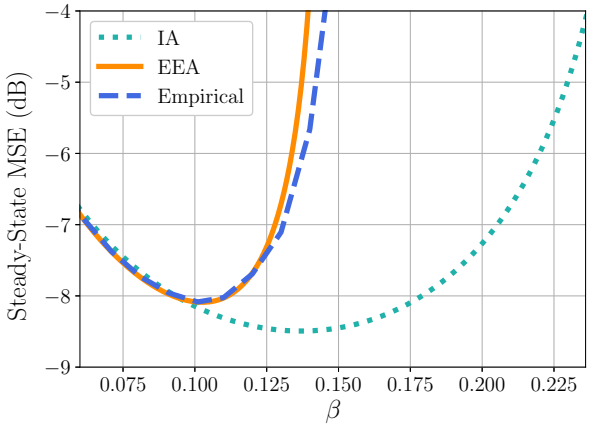


Fig. 6. Steady-state MSE as a function of the step size. The empirical curve was generated with $K = 10^6$ independent Monte Carlo trials.

enhanced performance, whereas our proposal indicates a more conservative step size magnitude, which is coherent with the discussion presented in[5] [16].

[5]However [16] does not address the LMS tracking abilities.

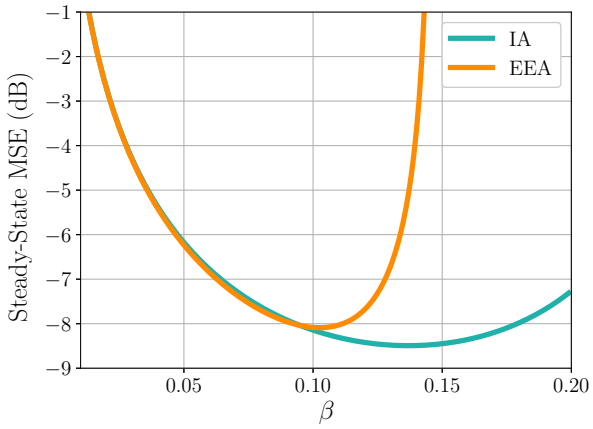


Fig. 7. Theoretical steady-state $\text{MSE}_{\text{ss},q}$ w.r.t. $\beta$ using the configuration of Fig. 6.

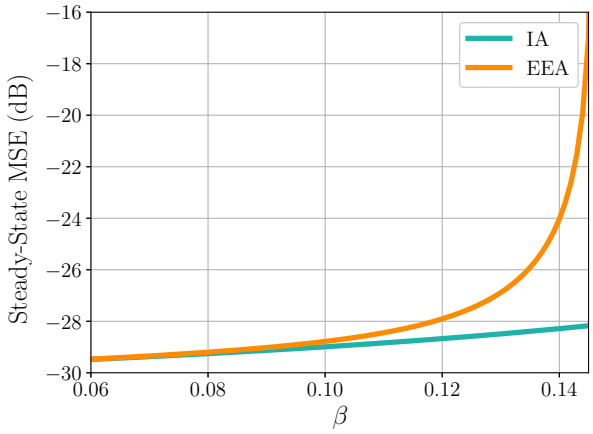


Fig. 8. Theoretical steady-state $\text{MSE}_{\text{ss,std}}$ w.r.t. $\beta$ using the same configuration of Fig. 6.

Furthermore, the different sequences do not converge asymptotically to zero, due to the fact that the use of infinitesimally small step sizes hampers the tracking capabilities of the adaptive algorithm.

The effectiveness of the optimal deterministic step size sequence obtained using both theoretical models is assessed in Fig. 11. This figure depicts the resulting empirical MSE evolution obtained by the utilization of the referred sequences. The employment of the sequence suggested by the standard model tends to induce divergence in practice, especially in the first phase of learning, which is again coherent with the results of [16]. Although the divergence occurs with low probability (which explains the elevated number of independent Monte Carlo trials used in the experiment in order to capture its effects), this is due to a catastrophic failure whose avoidance is crucial in practice. We can also see from Fig. 11 that by using the EEA sequence, MSE reaches steady-state approximately 50 iterations before the solution that uses the IA sequence.

Fig. 12 shows deterministic step size sequences (for both standard and advanced models) for Version II of the VSS technique derived in Section V, which allows for distinct step size values for each adaptive coefficient. The figure setup has the following specifications: $N = 3$, $M = 2$, $\sigma_\nu^2 = 10^{-4}$, and $\sigma_q^2 = 10^{-4}$. Once again, the standard model suggests a more aggressive step size policy, in order to maximize the convergence rate. In theory, this type of step sizes may induce divergence, since an IA-based technique does not infer correctly the upper bound of $\beta$ that avoids instability, as demonstrated in Fig. 3. The sequences suggested by both analyses can be compared by the empirical evaluation of the different step size sequences. This scenario, the same of Fig. 12, is shown in Fig. 13 which reveals that the theoretical step size sequence engineered by an IA-based stochastic model behaves in a far-from-optimal manner. On the other hand, with the EEA sequence the

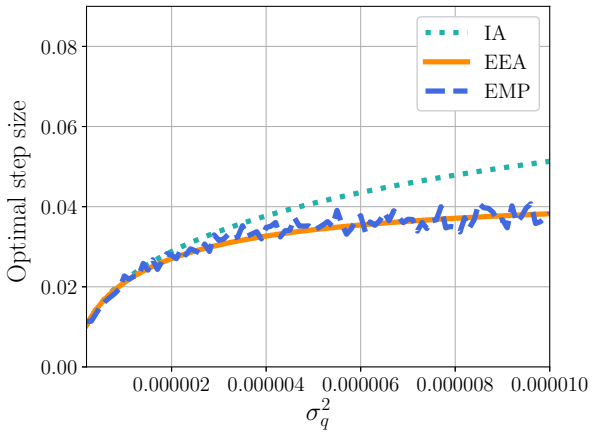


Fig. 9. This plot shows the optimal step size choice in function of $\sigma_q^2$. A total of $10^5$ independent Monte Carlo trials were used to achieve the empirical results.

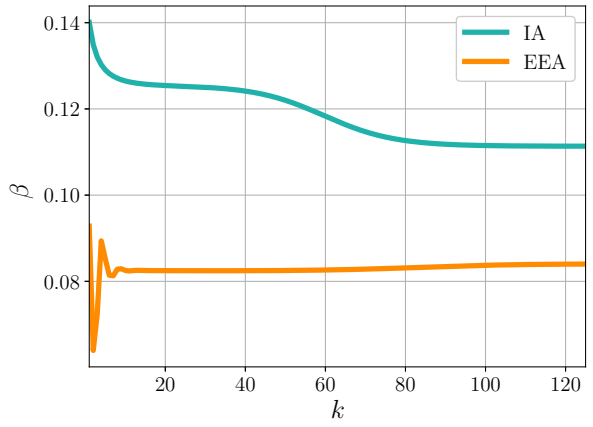


Fig. 10. Theoretical sequences $\beta(k)$ obtained by EEA and IA analyses.

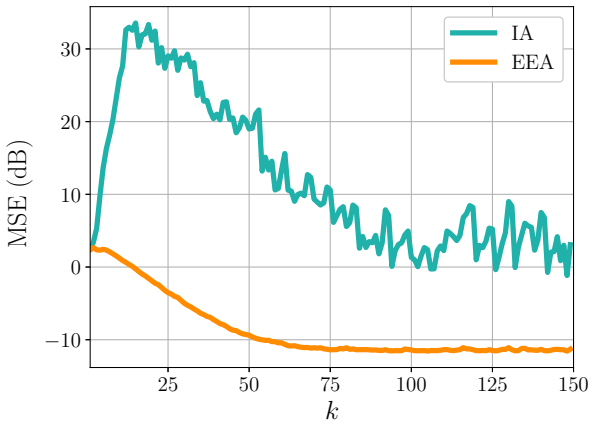


Fig. 11. Evolution of the empirical MSE using theoretical step size sequences from the standard (IA) and from the proposed (EEA) analyses, for $K = 10^7$ independent Monte Carlo trials.

MSD, shown in Fig. 13, decreases throughout the iterations, as desired. By using the EEA sequence, MSD reaches steady-state performance around the $80^{th}$ iteration mark, by contrast, when utilizing IA generated sequence, 140

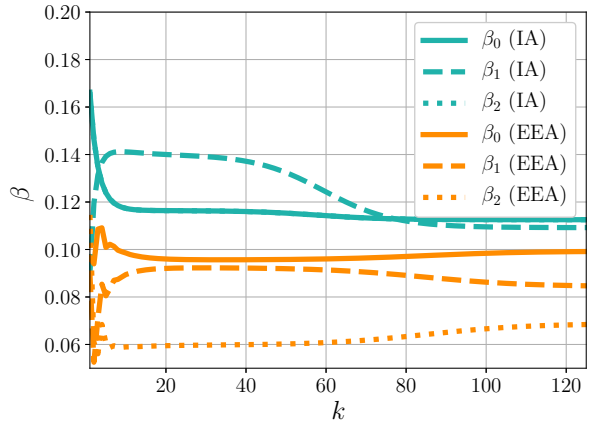


Fig. 12. Theoretical sequences $\beta_i(k)$ (for $i \in \{0, 1, \ldots, N-1\}$ prescribed by standard (IA) versus advanced (EEA) analyses.

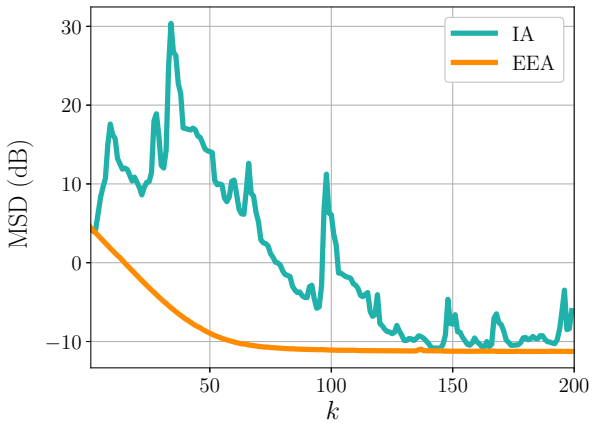


Fig. 13. Evolution of the empirical MSD (in dB) in the same scenario of Fig. 12, using the step size sequences suggested by IA- and EEA-based analyses. The empirical results were obtained by the usage of $K = 10^7$ independent Monte Carlo trials.

iterations are necessary to reach a similar MSD value.

## VII. Conclusions

This work presents a comprehensive model of the tracking capabilities of the LMS algorithm, using an EEA-based approach that does not assume neither a white nor a Gaussian input signal. The model is able to accurately predict first-order and second-order learning behaviours of the LMS, as well as to provide proper step size upper bounds that effectively avoid stability issues. Furthermore, it provides proper guidelines for a step size value choice that optimizes performance, even when a "condition of slow variations" [47] of the ideal plant is not satisfied.

Additionally, the advanced stochastic model is capable of engineering a theoretical step size sequence that optimizes the overall LMS performance, and even the restriction of an equal step size for each adaptive coefficient can be removed, in a similar way to what proportionate algorithms do. The evaluation of such sequences in order to avoid divergence issues in practice is an important feature, since it may help the design of data-dependent variable step size algorithms.

The avoidance of the IA is what lends EEA both its significance and limitations. The former derives from the ability to accurately predict algorithm performance when the step size attains large values or when the unknown plant has a significant non-stationary characteristic. EEA main constraint derives from its high algebraic complexity, which limits the range of scenarios the advanced analysis can model. The authors are also investigating whether it is possible to construct an intermediary analysis method, which assumes a weak formulation of IA in order to require a smaller computational burden than the one demanded by the EEA.

## References

[1] S. Marcos and O. Macchi, "Tracking capability of the least mean square algorithm: Application to an asynchronous echo canceller," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, pp. 1570–1578, Nov. 1987.

[2] 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)*, pp. 0181–0186, May. 2018.

[3] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proceedings of the IEEE*, vol. 64, pp. 1151–1162, Aug. 1976.

[4] C. Hsia, Y.-P. Yeh, T.-C. Wu, J. Chiang, and Y.-J. Liou, "Low resolution method using adaptive LMS scheme for moving objects detection and tracking," in *2010 International Symposium on Intelligent Signal Processing and Communication Systems*, pp. 1–4, Dec. 2010.

[5] W. Hernandez, M. E. Dominguez, and G. Sansigre, "Analysis of the error signal of the LMS algorithm," *IEEE Signal Processing Letters*, vol. 17, pp. 229–232, Mar. 2010.

[6] S. C. Douglas, Q. Zhu, and K. F. Smith, "A pipelined LMS adaptive FIR filter architecture without adaptation delay," *IEEE Transactions on Signal Processing*, vol. 46, pp. 775–779, Mar. 1998.

[7] A. Rusu, S. Ciochină, and C. Paleologu, "On the step-size optimization of the LMS algorithm," in *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, pp. 168–173, Jul. 2019.

[8] D. B. Haddad and M. R. Petraglia, "Transient and steady-state MSE analysis of the IMPNLMS algorithm," *Digital Signal Processing*, vol. 33, pp. 50 – 59, Oct. 2014.

[9] K. da S. Olinto, D. B. Haddad, and M. R. Petraglia, "Transient analysis of $\ell_0$-LMS and $\ell_0$-NLMS algorithms," *Signal Processing*, vol. 127, pp. 217 – 226, Oct. 2016.

[10] V. Solo, "LMS: Past, present and future," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7740–7744, May. 2019.

[11] E. Eweda, "A new approach for analyzing the limiting behavior of the normalized LMS algorithm under weak assumptions," *Signal Processing*, vol. 89, no. 11, pp. 2143 – 2151, 2009.

[12] S. C. Douglas, "Exact expectation analysis of the sign-data LMS algorithm for i.i.d. input data," in *[1992] Conference Record of the Twenty-Sixth Asilomar Conference on Signals, Systems Computers*, pp. 566–570 vol.1, Oct. 1992.

[13] S. C. Douglas, "Exact expectation analysis of the LMS adaptive filter for correlated Gaussian input data," in *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 519–522 vol.3, Apr. 1993.

[14] 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.

[15] P. Lara, K. d. S. Olinto, F. R. Petraglia, and D. B. Haddad, "Exact analysis of the least-mean-square algorithm with coloured measurement noise," *Electronics Letters*, vol. 54, pp. 1401–1403, Nov. 2018.

[16] 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*, vol. 26, pp. 74–78, Jan. 2019.

[17] P. Lara, L. D. Tarrataca, and D. B. Haddad, "Exact expectation analysis of the deficient-length LMS algorithm," *Signal Processing*, vol. 162, pp. 54 – 64, 2019.

[18] P. Lara, D. B. Haddad, and L. Tarrataca, "Advances on the analysis of the LMS algorithm with a colored measurement noise," *Signal, Image and Video Processing*, Oct. 2019.

[19] 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, 1992.

[20] P. Lara, F. Igreja, T. T. P. Silva, L. D. T. J. Tarrataca, and D. B. Haddad, "Exact expectation analysis of the LMS adaptive identification of non-linear systems," *Electronics Letters*, vol. 56, pp. 45–48, Jan. 2020.

[21] F. S. Cattivelli and A. H. Sayed, "Analysis of spatial and incremental LMS processing for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 59, pp. 1465–1480, Apr. 2011.

[22] M. T.-H. Alouane and M. Jaidane-Saidane, "A new nonstationary LMS algorithm for tracking markovian time varying systems," *Signal Processing*, vol. 86, pp. 50 – 70, Jan. 2006.
[23] N. R. Yousef and A. H. Sayed, "Ability of adaptive filters to track carrier offsets and channel nonstationarities," *IEEE Transactions on Signal Processing*, vol. 50, pp. 1533–1544, Jul. 2002.
[24] L. M. van de Kerkhof and W. J. W. Kitzen, "Tracking of a time-varying acoustic impulse response by an adaptive filter," *IEEE Transactions on Signal Processing*, vol. 40, pp. 1285–1294, Jun. 1992.
[25] W. Fong, S. J. Godsill, A. Doucet, and M. West, "Monte Carlo smoothing with application to audio signal enhancement," *IEEE Transactions on Signal Processing*, vol. 50, pp. 438–449, Feb. 2002.
[26] P. Lin, P. Rapajic, and Z. Krusevac, "On the tracking performance of LMS and RLS algorithms in an adaptive MMSE CDMA receiver," in *2005 Australian Communications Theory Workshop*, pp. 175–178, Feb. 2005.
[27] M. T. M. Silva and V. H. Nascimento, "Convex combination of adaptive filters with different tracking capabilities," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 3, pp. III–925–III–928, Apr. 2007.
[28] L. S. de Assis, J. R. de P. Junior, L. Tarrataca, A. R. Fontoura, and D. B. Haddad, "Efficient Volterra systems identification using hierarchical genetic algorithms," *Applied Soft Computing*, p. 105745, Dec. 2019.
[29] H. J. Butterweck, "A wave theory of long adaptive filters," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, pp. 739–747, Jun. 2001.
[30] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd Ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996.
[31] J. C. Bermudez, N. J. Bershad, and E. Eweda, "Stochastic analysis of the LMS algorithm for cyclostationary colored Gaussian inputs," *Signal Processing*, vol. 160, pp. 127 – 136, 2019.
[32] E. Eweda and N. J. Bershad, "Stochastic analysis of a stable normalized least mean fourth algorithm for adaptive noise canceling with a white Gaussian reference," *IEEE Transactions on Signal Processing*, vol. 60, pp. 6235–6244, Dec. 2012.
[33] S. Zhang, J. Zhang, and H. C. So, "Mean square deviation analysis of LMS and NLMS algorithms with white reference inputs," *Signal Processing*, vol. 131, pp. 20 – 26, Feb. 2017.
[34] E. Eweda, N. J. Bershad, and J. C. M. Bermudez, "Stochastic analysis of the LMS and NLMS algorithms for cyclostationary white Gaussian and non-Gaussian inputs," *IEEE Transactions on Signal Processing*, vol. 66, pp. 4753–4765, Sep. 2018.
[35] P. Hriljac, "Learning curves for LMS and regular Gaussian processes," *IEEE Transactions on Automatic Control*, vol. 47, pp. 284–289, Feb. 2002.
[36] K. Mayyas and T. Aboulnasr, "Leaky LMS algorithm: MSE analysis for gaussian data," *IEEE Transactions on Signal Processing*, vol. 45, pp. 927–934, Apr. 1997.
[37] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 86, pp. 2009–2025, Oct. 1998.
[38] 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.
[39] 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.
[40] J. Hwang and Y. Li, "Variable step-size LMS algorithm with a gradient-based weighted average," *IEEE Signal Processing Letters*, vol. 16, pp. 1043–1046, Dec. 2009.
[41] C. G. aes Lopes and J. C. M. Bermudez, "Evaluation and design of variable step size adaptive algorithms," in *ICASSP 2001*, vol. 6, pp. 3845–3848 vol.6, May. 2001.
[42] M. R. Petraglia and D. B. Haddad, "New adaptive algorithms for identification of sparse impulse responses - analysis and comparisons," in *2010 7th International Symposium on Wireless Communication Systems*, pp. 384–388, Sep. 2010.
[43] T. N. Ferreira, M. V. S. Lima, P. S. R. Diniz, and W. A. Martins, "Low-complexity proportionate algorithms with sparsity-promoting penalties," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 253–256, May. 2016.
[44] J. Benesty, C. Paleologu, and S. Ciochină, "Proportionate adaptive filters from a basis pursuit perspective," *IEEE Signal Processing Letters*, vol. 17, pp. 985–988, Dec. 2010.
[45] M. Yukawa, "Krylov-proportionate adaptive filtering techniques not limited to sparse systems," *IEEE Transactions on Signal Processing*, vol. 57, pp. 927–943, Mar. 2009.
[46] S. B. Gelfand, Y. Wei, and J. V. Krogmeier, "The stability of variable step-size LMS algorithms," *IEEE Transactions on Signal Processing*, vol. 47, pp. 3277–3288, Dec. 1999.
[47] O. Macchi, "Optimization of adaptive identification for time-varying filters," *IEEE Transactions on Automatic Control*, vol. 31, pp. 283–287, Mar. 1986.