



Analyzing the LMS Weight Error Covariance Matrix: An Exact Expectation Approach

Filipe Igreja¹ · Pedro Lara² · Luís Tarrataga² · Laura S. de Assis² ·
Fernanda D. V. R. Oliveira³ · Ana L. F. de Barros¹ · Diego B. Haddad^{1,2}

Received: 3 September 2023 / Revised: 28 February 2024 / Accepted: 29 February 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

One of the most relevant tasks of high-dimensional estimation is the computation of the parameters covariance matrix. This matrix offers valuable insights into the uncertainties associated with the estimation process. In the context of adaptive filtering, the weight error covariance matrix is a key parameter that determines how the filter adapts to input statistics and noise levels. Additional statistical information about the weight error vector allows one to depict a more precise description of the stochastic coupling between the adaptive weights. This paper concentrates on an in-depth study of the least mean squares asymptotic weight error covariance matrix, employing the exact expectation analysis. Through this examination, a key conclusion emerges: The symmetries commonly engendered by traditional analyses are not present in the actual covariance matrix. Instead, these symmetries are artifacts of the widespread assumption of independence. In summary, the advanced analysis performed in this work reveals a significantly more nuanced learning behavior exhibited by the least mean squares algorithm, challenging the conventional understanding put forth by traditional approaches. The theoretical predictions are confirmed by extensive simulations.

Keywords LMS · Stochastic model · Exact expectation analysis · Weight error correlation matrix

1 Introduction

Adaptive filtering, commonly termed as “estimators” or “filters,” plays a pivotal role in extracting relevant information from noisy data across diverse applications such as communications, radar, imaging, biomedical engineering and financial time series

✉ Filipe Igreja
figreja@gmail.com

Extended author information available on the last page of the article

forecasting [4, 40]. These algorithms, striking a balance between parametric and nonparametric approaches [48], prove particularly effective in scenarios where the dynamic nature of the environment or limited prior knowledge poses challenges. Applications span prediction [28, 37, 44], noise cancelation [1, 26, 45], system identification [9, 22, 23], adaptive beamforming [46] and adaptive spectrum analysis [35]. Their significance extends to both military systems and everyday electronic devices. However, the multitude of adaptive filtering algorithms demands scrutiny of their stability, efficiency and tracking abilities, as their learning procedures involve inherent feedback, presenting a non-trivial challenge to ensuring reliable operation.

It is important to distinguish adaptive algorithms from algorithms based on the Wiener method [42]. Such a distinction is rooted in their fundamental principles and applications. Adaptive filtering algorithms are crafted to function within dynamic environments, where the statistical characteristics of input signals may undergo temporal changes. These algorithms continually adjust their parameters, enabling them to dynamically adapt to evolving conditions and enhance their performance. Conversely, Wiener filtering encompasses a specific category of linear filters with the goal of minimizing the mean square error between the desired signal and the output of the filter [27]. Wiener filters achieve optimality by minimizing expected errors based on specific statistical assumptions about input signals and noise. In contrast to adaptive filters, Wiener filters are typically tailored for stationary processes, assuming a priori knowledge of the statistical properties of both the signals and noise. The choice between these approaches hinges on the specific requirements and characteristics of the signal processing task at hand.

In a broad sense, in the realm of adaptive filtering, the least mean squares (LMS) and recursive least squares (RLS) families of algorithms stand out with distinct characteristics [41]. The LMS algorithm, appreciated for its simplicity, demonstrates faster convergence in scenarios of low input correlation, making it suitable for real-time applications with limited computational resources. On the other hand, the recursive least squares (RLS) algorithm, leveraging recursive computations and historical data, excels in environments with high input correlation, exhibiting faster convergence in dynamically changing systems [6]. However, this advantage comes at the cost of increased computational complexity due to recursive matrix inversions and the need to store the entire data history. The selection between LMS and RLS hinges on the specific demands of an application, with LMS favoring simplicity and lower computational requirements, while RLS offers enhanced performance in scenarios where adaptability and rapid convergence in correlated environments are paramount.

Abstracting hardware idiosyncrasies, this paper advances a stochastic model that predicts some operating characteristics of the least mean squares (LMS) algorithm. Therefore, it focuses on stochastic models of adaptive filtering algorithms rather than on physical implementation. Such models usually rely on theoretical constructs and statistical hypotheses that are employed to *explain* and/or *predict* phenomena and are evaluated according to their ability to do so. One of our main objectives consists of determining whether for a certain configuration and parameter choice, the algorithms generate instability. Accordingly, this work intends to offer several predictions about the learning characteristics of the LMS in a more accurate manner than traditional approaches do.

Out of the hypotheses commonly used, the most troublesome one is the widespread *independence assumption* (IA), usually encountered in the stochastic approximations field [47] and widely used in the dynamic convergence analysis of adaptive algorithms [36]. The IA is a strong stochastic hypothesis that states that the adaptive weights are statistically independent of the input samples that currently feed the adaptive structure. This hypothesis is clearly violated in tapped-delay line structures, since these impose a *deterministic coherence* between consecutive input vectors. Due to this fact, some authors claim that the expression “independence *heuristic*” is more adequate than “independence assumption” [43].

The IA assumption renders the resulting stochastic model inaccurate when a non-infinitesimally small step size factor is adopted, especially when non-normalized algorithms (such as the LMS) are utilized. However, some potential advantages of adaptive filtering schemes may become apparent only for *large* step sizes [19]. Furthermore, instability issues are commonly encountered when the step size presents a large magnitude. Consequently, traditional analyses cannot accurately predict an upper bound for the step size that guarantees stability. Such an issue is widely recognized in standard textbooks about adaptive filtering [10, 24, 42]. Since instability is undesirable, analyses capable of providing reliable stability predictions are crucial.

This paper circumvents such limitations by employing the exact expectation analysis (EEA), which was presented in references [11, 14, 17]. In general terms, the EEA is a systematic constructive procedure that utilizes algebraic manipulations for the generation of a system of equations that model in a precise manner the learning behavior of an adaptive filter. The analysis is computationally intensive, requiring a judicious and efficient design of the programming code.

The focus of this paper relies on the covariance matrix of the LMS deviation weight.

$$\tilde{\mathbf{w}}(k) \triangleq \mathbf{w}^* - \mathbf{w}(k), \quad (1)$$

where $\mathbf{w}^* \in \mathbb{R}^N$ denotes the ideal (and unknown) Wiener solution and $\mathbf{w}(k) \in \mathbb{R}^N$ is the adaptive weight vector at the k th iteration.

In the context of adaptive filtering, the weight error covariance matrix (WECM) is a key parameter that determines how the filter adapts to changing input statistics and noise levels. This is because WECM provides an objective estimate of the uncertainties in the estimated filter coefficients. Thus, the WECM is defined as:

$$\mathbf{R}_{\tilde{\mathbf{w}}}(k) \triangleq \mathbb{E} \left[\tilde{\mathbf{w}}(k) \tilde{\mathbf{w}}^T(k) \right] \quad (2)$$

is very important for the characterization of the learning abilities of an adaptive filter. In this context, [3] states:

“In many situations, additional statistical information about the weight vector would be useful. These cases include the detection of a narrowband line component in background noise using the weight vector as a test statistic, using the filtered output as a test statistic and time delay estimation. The algorithm has been used as a canceller as part of a spread spectrum communication system. The output of the canceller acts as the input to a matched filter binary decision

device. Knowledge of the statistics of the canceller output is crucial to predicting error probabilities for the system. Most often, via a central limit argument, it is assumed that the test statistic and/or the weights are Gaussian.”

This excerpt emphasizes the common assumption that the adaptive weights are sampled according to a multivariate normal distribution. Under this scenario, a possible sparseness of the WECM provides valuable information about independence and conditional independence properties. Hence, the WECM encodes complete information about such properties, whereas, for general distributions, they merely reveal correlations and partial correlations between variables. Another attractive feature of multivariate normal distributions is that the second moment matrix and its inverse also contain complete information about independence and conditional independence properties. Specifically, a zero entry in the covariance matrix’s ij th position indicates that variables i and j are independent, while a zero entry in the precision (inverse covariance) matrix’s ij th position indicates that the two are conditionally independent [39].

Furthermore, the WECM provides information about the convergence behavior of the filter and the trade-off between convergence speed and steady-state error. Therefore, a more accurate estimate of the WECM can provide insights into the behavior of the algorithm being modeled. Namely, it can help allocate computational resources more efficiently by providing information about the degree of uncertainty in the estimated filter coefficients. In addition, it can also be used to infer the statistical properties of the adaptive estimator. For example, by examining the estimated covariance matrix one can see whether the deviation term variance is equal or presents some symmetry.

Among the contributions of this article, we can enumerate: (i) the calculation of the WECM in steady state without adopting the restrictive independence assumption, through the technique of exact expectation analysis; (ii) noting that several symmetries present in the WECM, commonly found in the literature, are actually artifacts of the IA; (iii) concluding that the correlations between different deviations in asymptotic regime can be quite distinct in magnitude compared to the theoretical values predicted by classical analyses; (iv) validating the theoretical findings with simulations; and (v) conducting real experiments with non-stationary signals and observing that such symmetries do not even exist, thereby indicating that the asymmetries predicted by the exact analysis are not restricted to signals adhering to the stochastic model used in the analysis.

This paper is structured as follows. Section 2 presents fundamental concepts of the LMS algorithm. Section 3 focuses on traditional theoretical results of the WECM, which adopts the IA. Section 4 explains the employment of the EEA to obtain the asymptotic WECM. Section 5 outlines the results of the paper. The concluding remarks of the paper are presented in Sect. 6.

Throughout this paper, vectors and matrices are represented with lowercase and uppercase bold fonts, respectively, while scalars are denoted by italics. All vectors are of column type. $\mathbb{E}[\cdot]$ is the expectation operator. $(\cdot)^T$ represents matrix transposition. Operator $\text{Tr}(\mathbf{A})$ denotes the trace of matrix \mathbf{A} .

2 The LMS Algorithm

The LMS algorithm is the most traditional adaptive filtering algorithm. The simplicity of the LMS implementation has made it a relevant benchmark for other adaptive algorithms [21]. In this section, we succinctly explain the operations performed by an adaptive filter that uses LMS adaptation, to introduce definitions that are employed in the following sections.

The LMS adopts the stochastic gradient optimization method in order to minimize the following stochastic cost function:

$$\mathcal{F}_{\text{LMS}}[\mathbf{w}(k)] \triangleq \frac{1}{2}e^2(k), \quad (3)$$

where $\mathbf{w}(k) \in \mathbb{R}^N$ denotes the adaptive weight vector and

$$e(k) \triangleq d(k) - y(k), \quad (4)$$

where $d(k) \in \mathbb{R}$ denotes the k th sample of the reference signal and $y(k) \in \mathbb{R}$ denotes the filter output at the k th iteration. Throughout this paper, the reference signal is, as usual, considered to be generated according to the following noisy linear regression model:

$$d(k) = [\mathbf{w}^*]^T \mathbf{x}(k) + v(k), \quad (5)$$

where $v(k) \in \mathbb{R}$ accounts for the measurement noise and $\mathbf{w}^* \in \mathbb{R}^N$ is the ideal (and unknown) vector. This paper also assumes a system identification problem, so that the adaptive filter intends to emulate \mathbf{w}^* .

The stochastic minimization of $\mathcal{F}_{\text{LMS}}[\mathbf{w}(k)]$ is performed by the iterative operation

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \beta \nabla_{\mathbf{w}}(k) \mathcal{F}[\mathbf{w}(k)] = \mathbf{w}(k) + \beta \mathbf{x}(k)e(k), \quad (6)$$

where $\beta \in \mathbb{R}_+$ is the step size and

$$\mathbf{x}(k) \triangleq [x(k) \ x(k-1) \ x(k-2) \ \dots \ x(k-N+1)]^T \quad (7)$$

denotes the input vector. Note that definition (7) assumes a tapped-delay line (also known as transversal structure). The output of the filter at the k th iteration can be described as

$$y(k) = \mathbf{w}^T(k) \mathbf{x}(k), \quad (8)$$

where

$$\mathbf{w}(k) \triangleq [w_0(k) \ w_1(k) \ w_2(k) \ \dots \ w_{N-1}(k)]^T \quad (9)$$

contains the N adaptive weights $w_i(k)$ (for $i \in \{0, 1, \dots, N-1\}$) at the k th iteration.

3 Traditional Estimates of the WECM

The ubiquitous IA provides some theoretical predictions w.r.t. the WECM. For example, most models assume that the off-diagonal elements of the WECM decrease monotonically to 0, so that the deviations are supposedly asymptotically uncorrelated [29]. Convergence studies often concentrate on the asymptotic properties of the diagonal elements of the WECM, as a large MSE results from a linear combination of these elements [20].

The adoption of the IA usually introduces some symmetry in the theoretical asymptotic WECM. Thus, for example, considering the Gaussianity of the input signal and traditional assumptions, [42] presents the following recursion for the WECM:

$$\mathbf{R}_{\tilde{\mathbf{w}}(k+1)} = \mathbf{R}_{\tilde{\mathbf{w}}(k)} - \beta [\mathbf{R}\mathbf{R}_{\tilde{\mathbf{w}}(k)} + \mathbf{R}_{\tilde{\mathbf{w}}(k)}\mathbf{R}] + 2\beta^2 \mathbf{R}\mathbf{R}_{\tilde{\mathbf{w}}(k)}\mathbf{R} + \beta^2 \mathbf{R} \{ \sigma_v^2 + \text{Tr}[\mathbf{R}\mathbf{R}_{\tilde{\mathbf{w}}(k)}] \}, \quad (10)$$

where the input autocorrelation matrix is defined by

$$\mathbf{R} \triangleq \mathbb{E} [\mathbf{x}(k)\mathbf{x}^T(k)]. \quad (11)$$

Another standard result states that the WECM can be recursively computed according to [42]:

$$\mathbf{R}_{\tilde{\mathbf{w}}(k+1)} = (\mathbf{I} - \beta \mathbf{R})\mathbf{R}_{\tilde{\mathbf{w}}(k)}(\mathbf{I} - \beta \mathbf{R}) + \beta^2 J_{\min} \mathbf{R}, \quad (12)$$

where J_{\min} denotes the minimum obtainable MSE. The symmetries presented in (10) and (12) give rise to structured asymptotic solutions for the WECM. Such a fact can be seen more clearly in the asymptotic identity [5]:

$$\lim_{k \rightarrow \infty} \mathbf{R}\mathbf{R}_{\tilde{\mathbf{w}}(k)} + \mathbf{R}_{\tilde{\mathbf{w}}(k)}\mathbf{R} = \mathbf{F}, \quad (13)$$

where \mathbf{F} is the following “excitation matrix”:

$$\mathbf{F} = \sum_{l=-\infty}^{\infty} \mathbb{E} [v(k)v(k-l)] \mathbb{E} [\mathbf{x}(k)\mathbf{x}^T(k-l)]. \quad (14)$$

In the case of white process $v(k)$, Eq. (14) leads to the following closed-form solution for the WECM [5]:

$$\lim_{k \rightarrow \infty} \mathbf{R}_{\tilde{\mathbf{w}}(k)} = \frac{1}{2} \sigma_v^2 \mathbf{I}. \quad (15)$$

Result (15) helps to explain, in an intuitive manner, some traditional theoretical findings. For example, frequently, the WECM is assumed to be approximately uncorrelated from tap to tap [12]. Further, under a white input signal, the WECM is expected to be a diagonal matrix, with equal diagonal elements [38]. These predictions are obtained

from usual stochastic approximations (i.e., the IA) and can be refined with the more sophisticated tools brought about by the EEA, which is the focus of the next section.

4 Exact Expectation Prediction of the WECM

Given the importance of the WECM, in this paper, EEA is employed to obtain a more accurate estimate of the LMS algorithm asymptotic WECM. The EEA [13, 15, 18, 30–33] 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 IA, 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 [18, 30, 33]).

The EEA technique was proposed in [16], in a configuration where the excitation data is assumed to be white. An unequal-mode convergence behavior in the variances of the filter coefficients (which contradicts the IA) was theoretically predicted and confirmed by simulations. Work [13] extends 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 [15, 18], which effectively established EEA as a powerful and alternative analysis method for predicting LMS performance. More recently, this technique has experienced some generalizations; namely, [31] removed the whiteness assumption of the additive noise; [32] has demonstrated that coloring the additive noise does not impact mean square stability; [34] addressed the identification of nonlinear plants; [30] employed EEA to derive an optimal sequence of step size values that optimize performance; and [33] 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 order to explain the EEA, let us consider a simple configuration, for didactic purposes. This setup is the same as the one presented in [17]. Even in this simple configuration, the EEA results in lengthy algebraic manipulations. We will begin by describing the statistical assumptions.

Moving Average Assumption Suppose that the input signal is generated by the following moving average process

$$x(k) = b_0u(k) + b_1u(k - 1), \quad (16)$$

so that the analysis is not restricted to white input signals, although the signal $u(k)$ is modeled as a zero-mean i.i.d. random process, whose probability density function is even, so that

$$\mathbb{E}[u^n(k)] = \begin{cases} 0, & \text{for odd } n \\ \gamma_n, & \text{for even } n \end{cases}, \quad (17)$$

where γ_n depends on the specific considered distribution.

Note that the correlation between $x(k_1)$ and $x(k_2)$ should be zero whenever $|k_1 - k_2| > M$, where M is finite. This is a restriction imposed by the EEA analysis (otherwise, its recursion procedure never halts), which is guaranteed by model (16).

Noise Assumption The zero-mean noise $v(k)$ is statistically independent of the remaining random variables. so that terms like $\mathbb{E} [v(k)u(k)\tilde{w}_0(k)]$ and $\mathbb{E} [v(k)u(k-1)\tilde{w}_0(k)]$ can be canceled.

Independence Assumption (IA) Most of the time, stochastic models of adaptive filtering algorithms employ the IA, which assumes that the adaptive weights in $\mathbf{w}(k)$ are statistically independent of the elements of $\mathbf{x}(k)$. IA is clearly violated in tapped-delay structures since these structures impose a deterministic coherence between consecutive input vectors. Since previous input vectors were utilized for adaptation purposes, the adaptive vector $\mathbf{w}(k)$ is indeed statistically coupled with $\mathbf{x}(k)$. Nevertheless, IA is a widespread assumption, since it simplifies the resulting model equations. In practice, it leads to inaccurate results when β is large [5].

Using the assumptions, the following theorem will be demonstrated:

Theorem. Suppose that the adaptive filter has only one adaptive tap (i.e., $N = 1$). Using the moving average, noise and independence assumptions, one may establish the following *linear* time-invariant state space description:

$$\mathbf{y}(k+1) = \mathbf{A}\mathbf{y}(k) + \mathbf{b}, \quad (18)$$

where the state vector is

$$\mathbf{y}(k) = \begin{bmatrix} \mathbb{E} [\tilde{w}_0^2(k)] \\ \mathbb{E} [\tilde{w}_0^2(k)u^2(k-1)] \\ \mathbb{E} [\tilde{w}_0^2(k)u^4(k-1)] \end{bmatrix}, \quad (19)$$

the *transition matrix* is

$$\mathbf{A} = \begin{bmatrix} 1 - 2b_0^2\beta\gamma_2 + b_0^4\beta^2\gamma_4 & 6b_0^2b_1^2\beta^2\gamma_2 - 2b_1^2\beta & b_1^4\beta^2 \\ \gamma_2 - 2b_0^2\beta\gamma_4 + b_0^4\beta^2\gamma_6 & 6b_0^2b_1^2\beta^2\gamma_4 - 2b_1^2\beta\gamma_2 & b_1^4\beta^2\gamma_2 \\ \gamma_4 - 2b_0^2\beta\gamma_6 + b_0^4\beta^2\gamma_8 & 6b_0^2b_1^2\beta^2\gamma_6 - 2b_1^2\beta\gamma_4 & b_1^4\beta^2\gamma_4 \end{bmatrix} \quad (20)$$

and

$$\mathbf{b} = \begin{bmatrix} b_0^2\beta^2\sigma_v^2\gamma_2 + b_1^2\beta^2\sigma_v^2\gamma_2 \\ b_0^2\beta^2\sigma_v^2\gamma_4 + b_1^2\beta^2\sigma_v^2\gamma_2^2 \\ b_0^2\beta^2\sigma_v^2\gamma_6 + b_1^2\beta^2\sigma_v^2\gamma_2\gamma_4 \end{bmatrix}. \quad (21)$$

Proof. When $N = 1$, recursion (6) degenerates into

$$\tilde{w}_0(k+1) = (1 - \beta x^2(k))\tilde{w}_0(k) + \beta v(k)x(k), \quad (22)$$

where $\tilde{w}_i(k) \in \mathbb{R}$ denotes the i th element of vector $\tilde{\mathbf{w}}(k)$ (see (9)). From (4), the error can be written as

$$e(k) = \tilde{w}_0(k)x(k) + v(k) = b_0\tilde{w}_0(k)u(k) + b_1\tilde{w}_0(k)u(k-1) + v(k), \quad (23)$$

so that the square of the error, in terms of signal $u(k)$, is

$$e^2(k) = b_0^2 u^2(k) \tilde{w}_0^2(k) + 2b_0 b_1 u(k) u(k-1) \tilde{w}_0^2(k) + 2b_0 v(k) u(k) \tilde{w}_0(k) + b_1^2 u^2(k-1) \tilde{w}_0^2(k) + 2b_1 v(k) u(k-1) \tilde{w}_0(k) + v^2(k). \quad (24)$$

The MSE can be obtained by the application of the expectation operator in (24), which leads to

$$\begin{aligned} \xi(k) \triangleq \mathbb{E} [e^2(k)] &= b_0^2 \mathbb{E} [u^2(k) \tilde{w}_0^2(k)] + 2b_0 b_1 \mathbb{E} [u(k) u(k-1) \tilde{w}_0^2(k)] \\ &\quad + 2b_0 \mathbb{E} [v(k) u(k) \tilde{w}_0(k)] + b_1^2 \mathbb{E} [u^2(k-1) \tilde{w}_0^2(k)] \\ &\quad + 2b_1 \mathbb{E} [v(k) u(k-1) \tilde{w}_0(k)] + \mathbb{E} [v^2(k)], \end{aligned} \quad (25)$$

which simplifies (25):

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

Note that $\tilde{w}_0(k)$ depends directly on $x(k-1)$, since (see (22)):

$$\tilde{w}_0(k) = (1 - \beta x^2(k-1)) \tilde{w}_0(k-1) + \beta v(k-1) x(k-1), \quad (27)$$

which implies that $\tilde{w}_0(k)$ depends statistically on sample $x(k)$, due to the fact that $x(k)$ is colored (see (16)) and $x(k)$ is correlated with $x(k-1)$. Due to this fact, (26) was written in terms of $u(k)$. Being $u(k)$ an i.i.d. process, $\tilde{w}_0(k)$ depends on $u(k-1)$ and is indeed statistically independent from $u(k)$, since $u(k)$ is statistically independent from $u(k-1)$. This implies that

$$\mathbb{E} [u^2(k) \tilde{w}_0^2(k)] = \mathbb{E} [u^2(k)] \mathbb{E} [\tilde{w}_0^2(k)] = \gamma_2 \mathbb{E} [\tilde{w}_0^2(k)], \quad (28)$$

$$\mathbb{E} [u(k) u(k-1) \tilde{w}_0^2(k)] = \overbrace{\mathbb{E} [u(k)]}^{=0} \mathbb{E} [u(k-1) \tilde{w}_0^2(k)] = 0. \quad (29)$$

The combination of (28) and (29) with (26) leads to

$$\xi(k) = \sigma_v^2 + \mathbb{E} [\tilde{w}_0(k) x^2(k)] = \sigma_v^2 + b_0^2 \gamma_2 \mathbb{E} [\tilde{w}_0^2(k)] + b_1^2 \mathbb{E} [\tilde{w}_0^2(k) u^2(k-1)], \quad (30)$$

which means that the estimation of the MSE can be equivalently converted in the estimation of the evolution of the statistical quantities $\mathbb{E} [\tilde{w}_0^2(k)]$ and $\mathbb{E} [\tilde{w}_0^2(k) u^2(k-1)]$. Such quantities are also termed as *state variables*.

The EEA is a systematic procedure that intends to obtain recursive equations that describe the evolution of the state variables. At the moment, there are two state variables: $\mathbb{E}[\tilde{w}_0^2(k)]$ and $\mathbb{E}[\tilde{w}_0^2(k)u^2(k-1)]$. The first step to obtaining a recursion for the first state variable requires squaring of both sides of (22):

$$\begin{aligned}\tilde{w}_0^2(k+1) = & b_0^4\beta^2u^4(k)\tilde{w}_0^2(k) + 4b_0^3b_1\beta^2u^3(k)u(k-1)\tilde{w}_0^2(k) - 2b_0^3\beta^2v(k)u^3(k)\tilde{w}_0(k) \\ & + 6\beta^2b_0^2b_1^2u^2(k)u^2(k-1)\tilde{w}_0^2(k) - 6b_0^2b_1\beta^2v(k)u^2(k)u(k-1)\tilde{w}_0(k) \\ & + b_0^2\beta^2v^2(k)u^2(k) - 2b_0^2\beta u^2(k)\tilde{w}_0^2(k) + 4b_0b_1^3\beta^2u(k)u^3(k-1)\tilde{w}_0^2(k) \\ & - 6b_0b_1^2\beta^2v(k)u(k)u^2(k-1)\tilde{w}_0(k) + 2b_0b_1\beta^2v^2(k)u(k)u(k-1) \\ & - 4b_0b_1\beta u(k)u(k-1)\tilde{w}_0^2(k) + 2b_0\beta v(k)u(k)\tilde{w}_0(k) + b_1^4\beta^2u^4(k-1)\tilde{w}_0^2(k) \\ & - 2b_1^3\beta^2v(k)u^3(k-1)\tilde{w}_0 + b_1^2\beta^2v^2(k)u^2(k-1) - 2b_1^2\beta u^2(k-1)\tilde{w}_0^2(k) \\ & + 2b_1\beta v(k)u(k-1)\tilde{w}_0(k) + \tilde{w}_0^2(k).\end{aligned}\quad (31)$$

After applying the expectation operator in (31) and using the simplifications that arrive from the combination of both the statistical independence of $v(k)$ w.r.t. the remaining random variables and the whiteness of $u(k)$, one has

$$\begin{aligned}\mathbb{E}[\tilde{w}_0^2(k+1)] = & (1 - 2b_0^2\beta\gamma_2 + b_0^4\beta^2\gamma_4)\mathbb{E}[\tilde{w}_0^2(k)] + b_1^4\beta^2\mathbb{E}[\tilde{w}_0^2(k)u^4(k-1)] \\ & + (6b_0^2b_1^2\beta^2\gamma_2 - 2b_1^2\beta)\mathbb{E}[\tilde{w}_0^2(k)u^2(k-1)] \\ & + (b_0^2\beta^2\sigma_v^2\gamma_2 + b_1^2\beta^2\sigma_v^2\gamma_2),\end{aligned}\quad (32)$$

where we observe the emergence of a novel state variable: $\mathbb{E}[\tilde{w}_0^2(k)u^4(k-1)]$. Note that we were not initially interested in estimating such a quantity in order to obtain the MSE (see Eq. (30)). For this reason, the new term is named as a *nuisance state variable*, because its estimation is necessary for the estimation of one statistical quantity of interest [7]. Unfortunately, in more complex configurations, the nuisance parameters may be the large majority of the state variables.

After obtaining the recursion for $\mathbb{E}[\tilde{w}_0^2(k)]$ in (32), two state variables remain to be addressed: The term $\mathbb{E}[\tilde{w}_0^2(k)u^2(k-1)]$, which is originally encountered in the MSE expression (25) and the *nuisance* state variable $\mathbb{E}[\tilde{w}_0^2(k)u^4(k-1)]$. The recursions for these state variables can be derived by multiplying both sides of (31) by convenient terms before the application of the expectation operator and the subsequent simplification by both the statistical independence of the measurement noise and the whiteness of $u(k)$. Due to the length of the necessary manipulations, the derivation is omitted and the final recursions are:

$$\begin{aligned}\mathbb{E}[\tilde{w}_0^2(k+1)u^2(k)] = & (\gamma_2 - 2b_0^2\beta\gamma_4 + b_0^4\beta^2\gamma_6)\mathbb{E}[\tilde{w}_0^2(k)] + b_1^4\beta^2\gamma_2\mathbb{E}[\tilde{w}_0^2(k)u^4(k-1)] \\ & + (6b_0^2b_1^2\beta^2\gamma_4 - 2b_1^2\beta\gamma_2)\mathbb{E}[\tilde{w}_0^2(k)u^2(k-1)] \\ & + (b_0^2\beta^2\sigma_v^2\gamma_4 + b_1^2\beta^2\sigma_v^2\gamma_2^2),\end{aligned}\quad (33)$$

$$\begin{aligned}\mathbb{E}[\tilde{w}_0^2(k+1)u^4(k)] = & (\gamma_4 - 2b_0^2\beta\gamma_6 + b_0^4\beta^2\gamma_8)\mathbb{E}[\tilde{w}_0^2(k)] + b_1^4\beta^2\gamma_4\mathbb{E}[\tilde{w}_0^2(k)u^4(k-1)] \\ & + (6b_0^2b_1^2\beta^2\gamma_6 - 2b_1^2\beta\gamma_4)\mathbb{E}[\tilde{w}_0^2(k)u^2(k-1)] \\ & + (b_0^2\beta^2\sigma_v^2\gamma_6 + b_1^2\beta^2\sigma_v^2\gamma_2\gamma_4).\end{aligned}\quad (34)$$

Through straightforward manipulations, Eqs. (32)–(34) result in Eqs. (18)–(21). \square
Remarks When IA is adopted, Equation (30) can be simplified, since

$$\mathbb{E} \left[\tilde{w}_0^2(k) u^2(k-1) \right] \approx \mathbb{E} \left[\tilde{w}_0^2(k) \right] \mathbb{E} \left[u^2(k-1) \right] = \gamma_2 \mathbb{E} \left[\tilde{w}_0^2(k) \right]. \quad (35)$$

Nevertheless, this approximation is not used in the EEA.

Since recursions (32)–(34) are self-contained, for the considered setup, the resulting state space model contains $R = 3$ equations. Unfortunately, R can be very large, even for simple configurations. Note that the WECM can be recovered from the state vector $\mathbf{y}(k)$ and that the steady-state value of $\mathbf{y}(k)$ can be obtained in a closed-form manner:

$$\lim_{k \rightarrow \infty} \mathbf{y}(k) = \mathbf{y}_{ss} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{b}. \quad (36)$$

In order to perform the demanded algebraic operations, efficient C++-based codes were implemented. Thus, commercial general-purpose software for symbolic manipulations was avoided. This allowed us to model configurations that have never been considered using EEA. The implemented code also allows the imposition of the IA, in order to obtain results derived by traditional analyses. In the next section, the predictions obtained by EEA are compared with the ones of the traditional analysis (i.e., one that adopts IA).

5 Results

In this section, we will conduct some experiments with simulated data and one experiment with real data. In the following, the steady-state WECM (SS-WECM) of several simulated scenarios is considered. Three distinct SS-WECMs are computed: (i) the empirical one (EMP-SS-WECM, obtained through simulations); (ii) the classical one (IA-SS-WECM, computed with the IA); and (iii) the WECM obtained with the EEA (EEA-SS-WECM).

The settings of the distinct scenarios are summarized in Table 1. In all scenarios, the optimum filter is $w_i^* = 1$, for $i \in \{0, 1, \dots, N-1\}$, except for Scenario 6 where $w_i^* = 10$. Scenarios 6 and 5 have similar settings except for $w_i^* = 10$, for $i \in \{0, 1, \dots, N-1\}$. Since we are interested in the asymptotic operation, a significant influence of the optimal solution in the WECM is not expected. The number of equations demanded for each scenario is presented in Table 2.

5.1 Scenario 1

In Scenario 1, the IA-SS-WECM is:

$$\begin{bmatrix} 5.826 \times 10^{-8} & -4.5 \times 10^{-9} & 4.091 \times 10^{-10} \\ -4.5 \times 10^{-9} & 5.867 \times 10^{-8} & -4.5 \times 10^{-9} \\ 4.091 \times 10^{-10} & -4.5 \times 10^{-9} & 5.826 \times 10^{-8} \end{bmatrix} \quad (37)$$

Table 1 Properties of the scenarios

# Scenario	N	M	L	β	σ_v^2	N_{samples}	$H(z)$	w_i^*
1	3	2	1	0.075	10^{-6}	10^7	$1 - 0.9z^{-1}$	1
2	3	2	2	0.075	10^{-6}	10^7	$1 - 0.9z^{-1}$	1
3	4	2	1	0.05	10^{-6}	10^6	1	1
4	5	1	1	0.075	10^{-6}	10^7	1	1
5	6	1	1	0.055	10^{-6}	10^6	1	1
6	6	1	1	0.055	10^{-6}	10^7	1	10

N_{samples} denotes the number of independent Monte Carlo trials. $H(z)$ is the coloring filter of the input signal

Table 2 Number of equations required for the linear state space equation for the considered scenarios and the number of the iteration considered for the asymptotic operation

# Scenario	# Equations (IA)	# Equations (EEA)	Steady-state iteration
1	6	698	299
2	6	830	299
3	10	9578	649
4	15	5313	199
5	21	49,695	399
6	21	49,695	399

Note that $\mathbb{E}[\tilde{w}_0^2(k)]$ and $\mathbb{E}[\tilde{w}_2^2(k)]$ (the initial and the final element of the main diagonal) are equal. This symmetry, as discussed in Sect. 3, is frequent when the IA is adopted.

In the same scenario, the EEA-SS-WECM is:

$$\begin{bmatrix} 6.556 \times 10^{-8} & 1.051 \times 10^{-9} & 1.86 \times 10^{-9} \\ 1.051 \times 10^{-9} & 6.97 \times 10^{-8} & -1.786 \times 10^{-9} \\ 1.86 \times 10^{-9} & -1.786 \times 10^{-9} & 7.873 \times 10^{-8} \end{bmatrix}. \quad (38)$$

EEA-SS-WECM predicts that $\mathbb{E}[\tilde{w}_i^2(k)] \neq \mathbb{E}[\tilde{w}_j^2(k)]$, when $i \neq j$, which implies that the supposed symmetry is indeed a by-product engendered by the IA. Further, the EEA predicts that $\mathbb{E}[\tilde{w}_2^2(k)] > \mathbb{E}[\tilde{w}_1^2(k)] > \mathbb{E}[\tilde{w}_0^2(k)]$, and that $\mathbb{E}[\tilde{w}_0(k)\tilde{w}_2(k)]$ (i.e., 1.86×10^{-9}) is much larger than the value predicted by the traditional method (4.091×10^{-10}), indicating that distinct deviations are much more correlated than presumed by traditional analyses. This fact means that the stochastic coupling between the adaptive weights is actually much more intricate than standard analyses portray.

The EMP-SS-WECM (i.e., the WECM obtained through actual executions of the LMS) in this scenario is:

Table 3 Results for Scenario 1

$\mathbb{E}[\tilde{w}_i(k)\tilde{w}_j(k)]$	IA	EEA	EMP	$\% \Delta_{IA,EMP}$	$\% \Delta_{EEA,EMP}$
$\mathbb{E}[\tilde{w}_0^2(k)]$	5.826×10^{-8}	6.556×10^{-8}	6.358×10^{-8}	- 8.36%	+ 3.11%
$\mathbb{E}[\tilde{w}_1^2(k)]$	5.867×10^{-8}	6.97×10^{-8}	6.76×10^{-8}	- 13.21%	+ 3.1%
$\mathbb{E}[\tilde{w}_2^2(k)]$	5.826×10^{-8}	7.873×10^{-8}	7.342×10^{-8}	- 20.64%	+ 7.23%
$\mathbb{E}[\tilde{w}_0(k)\tilde{w}_2(k)]$	4.091×10^{-10}	1.86×10^{-9}	1.296×10^{-9}	- 68.43%	+ 43.51%

$$\begin{bmatrix} 6.358 \times 10^{-8} & 2.066 \times 10^{-9} & 1.296 \times 10^{-9} \\ 2.066 \times 10^{-9} & 6.76 \times 10^{-8} & -8.467 \times 10^{-10} \\ 1.296 \times 10^{-9} & -8.467 \times 10^{-10} & 7.342 \times 10^{-8} \end{bmatrix}. \tag{39}$$

To help the comparisons, Table 3 summarizes some important results for Scenario 1. Note that the simulations confirm the EEA prediction that $\mathbb{E}[\tilde{w}_2^2(k)] > \mathbb{E}[\tilde{w}_1^2(k)] > \mathbb{E}[\tilde{w}_0^2(k)]$. The empirical moments $\mathbb{E}[\tilde{w}_0^2(k)]$, $\mathbb{E}[\tilde{w}_1^2(k)]$ and $\mathbb{E}[\tilde{w}_2^2(k)]$ are 9,1%, 15% and 26% larger (respectively) than the traditional estimates, whereas the EEA estimates exceeds the empirical correlations by 3,1%, 3,1% and 7,2% (respectively). Thus, it is possible to conclude that the EEA displays a much higher level of adherence to empirical outcomes. Please observe that the total absence of symmetry in the theoretical estimate is a feature of the proposed analysis, which has no precedent in the literature regarding LMS.

It is much more challenging to obtain strong empirical evidence in the case of cross-correlations between distinct deviations since such correlations are much smaller than the ones located in the main diagonal of the WECM. But, even with a finite number of Monte Carlo trials, the experimental correlation $\mathbb{E}[\tilde{w}_0(k)\tilde{w}_2(k)]$ (1.296×10^{-9}) is more than three times that predicted by the traditional method, an effect that is captured by the EEA analysis (whose estimate, $1, 86 \times 10^{-9}$, exceeds the empirical estimate by approximately 40%).

5.2 Scenario 2

The major contrast between Scenario 2 and Scenario 1 is that the former, as opposed to the latter, includes colored measurement noise. In Scenario 2, the IA-SS-WECM is given by:

$$\begin{bmatrix} 1.054 \times 10^{-7} & -8.204 \times 10^{-9} & 6.984 \times 10^{-10} \\ -8.204 \times 10^{-9} & 1.061 \times 10^{-7} & -8.204 \times 10^{-9} \\ 6.984 \times 10^{-10} & -8.204 \times 10^{-9} & 1.054 \times 10^{-7} \end{bmatrix}, \tag{40}$$

where once again a symmetry is observed, given that $\mathbb{E}[\tilde{w}_0^2(k)] = \mathbb{E}[\tilde{w}_2^2(k)] = 1.054 \times 10^{-7}$.

In the same scenario, the EEA-SS-WECM is:

Table 4 Results for Scenario 2

$\mathbb{E}[\tilde{w}_i(k)\tilde{w}_j(k)]$	IA	EEA	EMP	% $\Delta_{\text{IA,EMP}}$	% $\Delta_{\text{EEA,EMP}}$
$\mathbb{E}[\tilde{w}_0^2(k)]$	1.054×10^{-7}	1.234×10^{-7}	1.23×10^{-7}	-14.37%	+0.02%
$\mathbb{E}[\tilde{w}_1^2(k)]$	1.061×10^{-7}	1.11×10^{-7}	1.085×10^{-7}	-2.21%	2.3%
$\mathbb{E}[\tilde{w}_2^2(k)]$	1.054×10^{-7}	1.257×10^{-7}	1.214×10^{-7}	-13.17%	+3.54%
$\mathbb{E}[\tilde{w}_0(k)\tilde{w}_1(k)]$	-8.204×10^{-9}	-2.98×10^{-8}	-3.036×10^{-8}	-72.97%	-1.84%
$\mathbb{E}[\tilde{w}_0(k)\tilde{w}_2(k)]$	6.984×10^{-10}	1.803×10^{-9}	1.061×10^{-9}	-34.17%	+69.93%
$\mathbb{E}[\tilde{w}_1(k)\tilde{w}_2(k)]$	-8.204×10^{-9}	-3.003×10^{-8}	1.214×10^{-7}	106.75%	124.73%

$$\begin{bmatrix} 1.234 \times 10^{-7} & -2.98 \times 10^{-8} & 1.803 \times 10^{-9} \\ -2.98 \times 10^{-8} & 1.11 \times 10^{-7} & -3.003 \times 10^{-8} \\ 1.803 \times 10^{-9} & -3.003 \times 10^{-8} & 1.257 \times 10^{-7} \end{bmatrix}, \quad (41)$$

which implies that the statistical correlations between distinct deviations are much more intense than the ones predicted by IA. Such a finding can be confirmed by the respective EMP-SS-WECM, which is:

$$\begin{bmatrix} 1.23 \times 10^{-7} & -3.036 \times 10^{-8} & 1.061 \times 10^{-9} \\ -3.036 \times 10^{-8} & 1.085 \times 10^{-7} & -2.848 \times 10^{-8} \\ 1.061 \times 10^{-9} & -2.848 \times 10^{-8} & 1.214 \times 10^{-7} \end{bmatrix}. \quad (42)$$

Table 4 suggests that the EEA-SS-WECM fits the data better than IA-SS-WECM. For instance, EEA predicts that $\mathbb{E}[\tilde{w}_0(k)\tilde{w}_1(k)]$ is -2.98×10^{-8} , which is 3.63 times larger than the value predicted by traditional analysis (-8.204×10^{-9}). The simulated value of $E[\tilde{w}_0(k)\tilde{w}_1(k)]$ is -3.036×10^{-8} , just 1.88% above the value predicted by the advanced analysis. In addition, EEA predicts that $\mathbb{E}[\tilde{w}_0^2(k)]$ is 17.08% above the value estimated by IA. The simulations uphold this finding, revealing a value for $\mathbb{E}[\tilde{w}_0^2(k)]$ that is 16.70% higher than the prediction made by traditional analysis. Even in the case of $\mathbb{E}[\tilde{w}_1(k)\tilde{w}_2(k)]$, where the exact analysis is farther from the actual value than the traditional method (both including a change of sign), it correctly assumes that this combined moment (in absolute terms) has a much greater intensity than that predicted by the classical analysis.

5.3 Scenario 3

Scenario 3 uses a filter of size $N = 4$, larger than the size adopted in Scenarios 1 and 2. In this scenario, the IA-SS-WECM is given by:

$$\begin{bmatrix} 3.483 \times 10^{-8} & -1.605 \times 10^{-9} & 2.033 \times 10^{-10} & 6.838 \times 10^{-11} \\ -1.605 \times 10^{-9} & 3.503 \times 10^{-8} & -1.537 \times 10^{-9} & 2.033 \times 10^{-10} \\ 2.033 \times 10^{-10} & -1.537 \times 10^{-9} & 3.503 \times 10^{-8} & -1.605 \times 10^{-9} \\ 6.838 \times 10^{-11} & 2.033 \times 10^{-10} & -1.605 \times 10^{-9} & 3.483 \times 10^{-8} \end{bmatrix}, \quad (43)$$

Table 5 Some results for Scenario 3

$\mathbb{E} [\tilde{w}_i(k)\tilde{w}_j(k)]$	IA	EEA	EMP	$\% \Delta_{IA,EMP}$	$\% \Delta_{EEA,EMP}$
$\mathbb{E} [\tilde{w}_0(k)\tilde{w}_1(k)]$	$-1.605e-9$	$-2.159e-10$	$-3.137e-10$	+ 511.63%	- 31.17%
$\mathbb{E} [\tilde{w}_2(k)\tilde{w}_3(k)]$	$-1.605e-9$	$-4.636e-10$	$-2.054e-10$	+ 781.4%	+ 225.7%

where the expect symmetry also happens, since $\mathbb{E} [\tilde{w}_0^2(k)] = \mathbb{E} [\tilde{w}_3^2(k)]$, $\mathbb{E} [\tilde{w}_1^2(k)] = \mathbb{E} [\tilde{w}_2^2(k)]$ and $\mathbb{E} [\tilde{w}_0(k)\tilde{w}_1(k)] = \mathbb{E} [\tilde{w}_2(k)\tilde{w}_3(k)]$, for example. In other words, classical analysis constrains the freedom of the WECM to a greater extent than the constraint imposed by the symmetry of the WECM.

In the same scenario, the computation of EEA-SS-WECM leads to

$$\begin{bmatrix} 3.499 \times 10^{-8} & -2.159 \times 10^{-10} & 2.002 \times 10^{-10} & 8.654 \times 10^{-11} \\ -2.159 \times 10^{-10} & 3.534 \times 10^{-8} & -6.934 \times 10^{-11} & 4.258 \times 10^{-10} \\ 2.002 \times 10^{-10} & -6.934 \times 10^{-11} & 3.539 \times 10^{-8} & -4.636 \times 10^{-10} \\ 8.654 \times 10^{-11} & 4.258 \times 10^{-10} & -4.636 \times 10^{-10} & 3.537 \times 10^{-8} \end{bmatrix}, \quad (44)$$

whereas the EMP-SS-WECM is

$$\begin{bmatrix} 3.472 \times 10^{-8} & -3.137 \times 10^{-10} & 1.299 \times 10^{-10} & 8.762 \times 10^{-11} \\ -3.137 \times 10^{-10} & 3.491 \times 10^{-8} & -1.35 \times 10^{-10} & 3.428 \times 10^{-10} \\ 1.299 \times 10^{-10} & -1.35 \times 10^{-10} & 3.496 \times 10^{-8} & -2.054 \times 10^{-10} \\ 8.762 \times 10^{-11} & 3.428 \times 10^{-10} & -2.054 \times 10^{-10} & 3.44 \times 10^{-8} \end{bmatrix}. \quad (45)$$

Table 5 presents two cases in which the IA-SS-WECM’s estimation of the joint moment differs from the empirical value by an order of magnitude. Conversely, EEA yields estimates that are much closer to the simulated values, always within the same order of magnitude.

5.4 Scenario 4

Scenario 4 features a larger filter size ($N = 5$) compared to the previous scenarios. This is the first scenario in which the input signal is white. In this case, as observed, classical analysis tends to generate estimates of very low magnitude for the deviations cross-correlation (i.e., $\mathbb{E} [\tilde{w}_i(k)\tilde{w}_j(k)]$ for $i \neq j$). In this scenario, the IA-SS-WECM is given by:

$$\begin{bmatrix} 5.092 \times 10^{-8} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} \\ 1.232 \times 10^{-13} & 5.092 \times 10^{-8} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} \\ 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 5.092 \times 10^{-8} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} \\ 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 5.092 \times 10^{-8} & 1.232 \times 10^{-13} \\ 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 5.092 \times 10^{-8} \end{bmatrix}, \quad (46)$$

which makes a strong prediction: that all cross-correlations are equal to 1.232×10^{-13} .

On the other hand, EEA-SS-WECM assumes that such cross-correlations have significantly higher (and heterogeneous) magnitudes, depicting a much richer learning behavior:

$$\begin{bmatrix} 5.102 \times 10^{-8} & 2.96 \times 10^{-11} & -4.416 \times 10^{-11} & 2.254 \times 10^{-11} & -6.17 \times 10^{-11} \\ 2.96 \times 10^{-11} & 5.064 \times 10^{-8} & 3.466 \times 10^{-11} & 2.08 \times 10^{-11} & 2.147 \times 10^{-11} \\ -4.416 \times 10^{-11} & 3.466 \times 10^{-11} & 5.034 \times 10^{-8} & 3.149 \times 10^{-11} & 4.696 \times 10^{-11} \\ 2.254 \times 10^{-11} & 2.08 \times 10^{-11} & 3.149 \times 10^{-11} & 5.005 \times 10^{-8} & 2.244 \times 10^{-11} \\ -6.17 \times 10^{-11} & 2.147 \times 10^{-11} & 4.696 \times 10^{-11} & 2.244 \times 10^{-11} & 4.968 \times 10^{-8} \end{bmatrix} \quad (47)$$

The EMP-SS-WECM computed in this scenario is as follows:

$$\begin{bmatrix} 5.101 \times 10^{-8} & 4.916 \times 10^{-12} & -3.739 \times 10^{-11} & 7.144 \times 10^{-12} & -3.448 \times 10^{-11} \\ 4.916 \times 10^{-12} & 5.063 \times 10^{-8} & 3.611 \times 10^{-11} & 3.229 \times 10^{-11} & 3.522 \times 10^{-11} \\ -3.739 \times 10^{-11} & 3.611 \times 10^{-11} & 5.035 \times 10^{-8} & 4.145 \times 10^{-11} & 5.355 \times 10^{-11} \\ 7.144 \times 10^{-12} & 3.229 \times 10^{-11} & 4.145 \times 10^{-11} & 5.005 \times 10^{-8} & 2.622 \times 10^{-11} \\ -3.448 \times 10^{-11} & 3.522 \times 10^{-11} & 5.355 \times 10^{-11} & 2.622 \times 10^{-11} & 4.970 \times 10^{-8} \end{bmatrix}. \quad (48)$$

The EMP-SS-WECM corroborates the findings of the exact analysis, validating that the stochastic coupling among the various deviations is far more intricate than implied by conventional analysis. It is worth noting that in some cases, classical analysis (and only classical analysis) even gets the sign of the joint moment wrong, such as in the estimates of $\mathbb{E}[\tilde{w}_0(k)\tilde{w}_2(k)]$ and $\mathbb{E}[\tilde{w}_0(k)\tilde{w}_4(k)]$.

5.5 Scenario 5

In Scenario 5, an even larger filter size is observed (i.e., $N = 6$). In this scenario, the IA-SS-WECM is given by:

$$\begin{bmatrix} 3.525 \times 10^{-8} & 9.562 \times 10^{-20} \\ 9.562 \times 10^{-20} & 3.525 \times 10^{-8} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} \\ 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 3.525 \times 10^{-8} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} \\ 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 3.525 \times 10^{-8} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} \\ 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 3.525 \times 10^{-8} & 9.562 \times 10^{-20} \\ 9.562 \times 10^{-20} & 3.525 \times 10^{-8} \end{bmatrix}, \quad (49)$$

whereas the EEA-SS-WECM is given by

$$\begin{bmatrix} 3.528 \times 10^{-8} & 1.659 \times 10^{-16} & -3.721 \times 10^{-11} & 1.255 \times 10^{-16} & -4.317 \times 10^{-11} & 1.48 \times 10^{-16} \\ 1.659 \times 10^{-16} & 3.514 \times 10^{-8} & 1.889 \times 10^{-16} & -1.981 \times 10^{-11} & 1.806 \times 10^{-16} & -2.631 \times 10^{-11} \\ -3.721 \times 10^{-11} & 1.889 \times 10^{-16} & 3.501 \times 10^{-8} & 2.341 \times 10^{-16} & -6.932 \times 10^{-12} & 1.179 \times 10^{-16} \\ 1.255 \times 10^{-16} & -1.981 \times 10^{-11} & 2.341 \times 10^{-16} & 3.49 \times 10^{-8} & 1.719 \times 10^{-16} & 4.568 \times 10^{-12} \\ -4.317 \times 10^{-11} & 1.806 \times 10^{-16} & -6.932 \times 10^{-12} & 1.719 \times 10^{-16} & 3.478 \times 10^{-8} & 1.382 \times 10^{-16} \\ 1.48 \times 10^{-16} & -2.631 \times 10^{-11} & 1.179 \times 10^{-16} & 4.568 \times 10^{-12} & 1.382 \times 10^{-16} & 3.463 \times 10^{-8} \end{bmatrix}. \quad (50)$$

Comparing the EEA-SS-WECM with IA-SS-WECM, it can be observed that the exact analysis predicts that the off-diagonal elements of the WECM have a much higher magnitude than predicted by the classical analysis. This finding is confirmed by the computed EMP-SS-WECM:

$$\begin{bmatrix} 3.527 \times 10^{-8} & 9.484 \times 10^{-13} & -5.0262 \times 10^{-11} & 2.528 \times 10^{-13} & -5.334 \times 10^{-11} & 8.317 \times 10^{-12} \\ 9.484 \times 10^{-13} & 3.514 \times 10^{-8} & -1.258 \times 10^{-11} & -2.08 \times 10^{-11} & 1.51 \times 10^{-12} & -2.954 \times 10^{-11} \\ -5.02 \times 10^{-11} & -1.258 \times 10^{-11} & 3.503 \times 10^{-8} & 1.853 \times 10^{-12} & -2.397 \times 10^{-11} & 1.212 \times 10^{-11} \\ 2.528 \times 10^{-13} & -2.08 \times 10^{-11} & 1.853 \times 10^{-12} & 3.493 \times 10^{-8} & -2.811 \times 10^{-11} & 3.357 \times 10^{-11} \\ -5.33 \times 10^{-11} & 1.51 \times 10^{-12} & -2.397 \times 10^{-11} & -2.811 \times 10^{-11} & 3.478 \times 10^{-8} & -1.688 \times 10^{-12} \\ 8.317 \times 10^{-12} & -2.954 \times 10^{-11} & 1.212 \times 10^{-11} & 3.357 \times 10^{-11} & -1.688 \times 10^{-12} & 3.463 \times 10^{-8} \end{bmatrix}, \tag{51}$$

However, it is worth mentioning that even the exact analysis occasionally misjudges the order of magnitude of the cross-correlations of the deviations. Nonetheless, in all cases, the error of the exact analysis is lower than the error of the classical analysis. By comparing the scenarios, it is possible to notice that with the increase in N , there appears to be a tendency for a reduction in the magnitude of correlations between different deviations. Thus, even with the adoption of a large number of independent Monte Carlo trials (i.e., 10^6 trials) for Scenario 5, it is challenging to compute a reasonable approximation of such small numbers.

5.6 Scenario 6

The fundamental difference in Scenario 6 compared to the previous scenarios lies in the optimal values of the adaptive coefficients (they all are now set to 10). The purpose of this scenario is to determine whether there is a notable difference in the asymptotic behavior when the final values of the filters change. Intuitively, after convergence, the variance and deviation correlations are not highly sensitive to the values around which the adaptive estimator oscillates (as LMS is an unbiased algorithm). The objective of the simulations in this scenario is to validate or refine such intuition.

In this scenario, the IA-SS-WECM is given by:

$$\begin{bmatrix} 3.525 \times 10^{-8} & 9.562 \times 10^{-18} \\ 9.562 \times 10^{-18} & 3.525 \times 10^{-8} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} \\ 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 3.525 \times 10^{-8} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} \\ 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 3.525 \times 10^{-8} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} \\ 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 3.525 \times 10^{-8} & 9.562 \times 10^{-18} \\ 9.562 \times 10^{-18} & 3.525 \times 10^{-8} \end{bmatrix}, \tag{52}$$

whereas the EEA-SS-WECM gives us:

$$\begin{bmatrix} 3.528 \times 10^{-8} & 1.659 \times 10^{-14} & -3.719 \times 10^{-11} & 1.255 \times 10^{-14} & -4.316 \times 10^{-11} & 1.48 \times 10^{-14} \\ 1.659 \times 10^{-14} & 3.514 \times 10^{-8} & 1.889 \times 10^{-14} & -1.979 \times 10^{-11} & 1.806 \times 10^{-14} & 2.63 \times 10^{-11} \\ -3.719 \times 10^{-11} & 1.889 \times 10^{-14} & 3.501 \times 10^{-8} & 2.341 \times 10^{-14} & -6.915 \times 10^{-12} & 1.179 \times 10^{-14} \\ 1.255 \times 10^{-14} & -1.979 \times 10^{-11} & 2.341 \times 10^{-14} & 3.49 \times 10^{-8} & 1.719 \times 10^{-14} & 4.579 \times 10^{-12} \\ -4.316 \times 10^{-11} & 1.806 \times 10^{-14} & -6.915 \times 10^{-12} & 1.719 \times 10^{-14} & 3.478 \times 10^{-8} & 1.382 \times 10^{-14} \\ 1.48 \times 10^{-14} & -2.63 \times 10^{-11} & 1.179 \times 10^{-14} & 4.579 \times 10^{-12} & 1.382 \times 10^{-14} & 3.463 \times 10^{-8} \end{bmatrix} \tag{53}$$

The EMP-SS-WECM in this scenario is:

$$\begin{bmatrix} 3.531 \times 10^{-8} & -1.489 \times 10^{-11} & -3.799 \times 10^{-11} & -1.79 \times 10^{-11} & -4.126 \times 10^{-11} & 3.896 \times 10^{-13} \\ -1.489 \times 10^{-11} & 3.514 \times 10^{-8} & -1.876 \times 10^{-13} & -1.877 \times 10^{-11} & 1.069 \times 10^{-12} & -4.442 \times 10^{-11} \\ -3.799 \times 10^{-11} & -1.876 \times 10^{-13} & 3.502 \times 10^{-8} & 3.381 \times 10^{-12} & -1.722 \times 10^{-11} & -2.047 \times 10^{-11} \\ -1.79 \times 10^{-11} & -1.877 \times 10^{-11} & 3.381 \times 10^{-12} & 3.489 \times 10^{-8} & 1.175 \times 10^{-12} & 2.373 \times 10^{-11} \\ -4.126 \times 10^{-11} & 1.069 \times 10^{-12} & -1.722 \times 10^{-11} & 1.175 \times 10^{-12} & 3.476 \times 10^{-8} & -1.029 \times 10^{-12} \\ 3.896 \times 10^{-13} & -4.442 \times 10^{-11} & -2.047 \times 10^{-11} & 2.373 \times 10^{-11} & -1.029 \times 10^{-12} & 3.464 \times 10^{-8} \end{bmatrix}, \quad (54)$$

from which we can extract very similar conclusions to those of Scenario 5. The results of this scenario support the intuition that the statistical properties of the adaptive estimator in steady state are reasonably unaffected by the optimal value of the coefficients.

5.7 Experiment with Real Data

Real data commonly exhibit non-analytical probability density functions and are often non-stationary. Therefore, it cannot be expected that the exact expectation analysis technique, which assumes a highly idealized distribution of input signals, can accurately infer phenomena occurring with adaptive filters fed by real data. Nevertheless, in this section, we seek to determine whether the primary conclusion drawn through exact analysis in this paper (namely, that the symmetries present in the WECM calculated through classical analysis actually do not exist) would hold true with real data.

Therefore, consider the AudioMNIST dataset of voice signals [2]. The dataset comprises 30,000 audio samples featuring spoken digits (0–9) from 60 distinct speakers. To construct each input signal, we concatenate the first 10 utterances from each speaker for each of the digits (0–9). Thus, there is a total of $600 = 60$ (speakers) \times 10 (digits) distinct input signals. Each input signal has been normalized to have unit variance. The speech is filtered by the optimal plant defined by

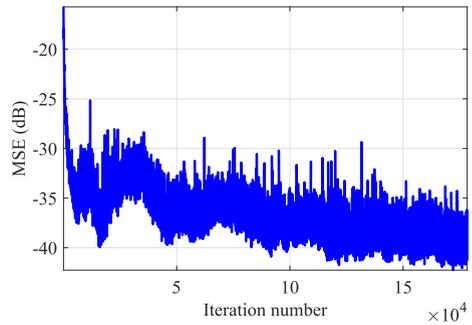
$$\mathbf{w}^* = [0.05377 \quad 0.18339 \quad -0.22588 \quad 0.08622], \quad (55)$$

whose output is corrupted by a white additive Gaussian noise whose variance is 10^{-6} . The evolution of the computed MSE is depicted in Figure 1 for the LMS algorithm with $\beta = 10^{-4}$, which reflects a less smooth learning curve, as expected since the input signal is non-stationary and quite rich in statistical terms. The computed asymptotic WECM in this experiment is given by

$$\begin{bmatrix} 0.0010 & -0.0032 & 0.0041 & -0.0020 \\ -0.0032 & 0.0207 & -0.0302 & 0.0128 \\ 0.0041 & -0.0302 & 0.0448 & -0.0189 \\ -0.0020 & 0.0128 & -0.0189 & 0.0081 \end{bmatrix}. \quad (56)$$

Note that the matrix in (56) exhibits no symmetries, and the quantities $\mathbb{E}[\tilde{w}_i^2(k)]$ vary significantly for different values of i . Also, observe that the correlations $\mathbb{E}[\tilde{w}_i(k)\tilde{w}_j(k)]$ are quite distinct from each other and, in magnitude, occasionally exceed the values observed on the main diagonal of the WECM. Thus, only the exact expectation analysis

Fig. 1 MSE evolution of the experiment with speech signals as inputs of the LMS adaptive filter



technique has the sophistication necessary to extract from simulated signals such phenomena that occur in practice, demonstrating a significant limitation in the use of the ubiquitous independence hypothesis.

6 Conclusions

In this paper, EEA was employed to predict asymptotic characteristics of the steady-state deviation autocorrelation matrix that classical analysis is unable to accurately predict. The WECM provides statistical information such as mean square error and mean square deviation, which can be useful in situations where the adaptive filter is adopted to predict a narrowband component or for some statistical test. It was observed that exact analysis outlines a much richer asymptotic behavior of the adaptive estimator than traditional analyses assume.

With the presented results, one can conclude that several symmetries found in WECM prediction through classical analysis are actually artifacts produced by the IA. It was also observed that the stochastic coupling between the adaptive weights can be quite different from that predicted by the classical analysis. The experiments confirmed the indications of the exact analysis that the cross-correlations between these deviations seem to decrease as the filter size increases. Lastly, such couplings (or, more precisely, the joint moments) do not appear to be sensitive to the optimal value of these coefficients, which aligns with what is expected from traditional stochastic models. Many of these findings are also observed through experiments with real signals, demonstrating that only the exact expectation analysis technique has the necessary sophistication to infer from simulated signals phenomena that occur in practical scenarios.

In a sense, the results presented are a tribute to the sophistication of the learning process of the LMS algorithm, as they show that the actual behavior of the algorithm can deviate considerably from that predicted by classical analyses, even in simple configurations. The theoretical analysis proposed in this paper can be extended to other configurations, such as deficient length and tracking, as well as to other adaptive filtering algorithms (e.g., adaptive control methods with time delays [8, 25]), which are promising future investigations.

Acknowledgements This work was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES, Finance Code 001) and in part by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

1. A.Q.J. Althahab, A new robust adaptive algorithm based adaptive filtering for noise cancellation. *Analog Integr. Circuit. Signal. Process* **94**(2), 217–231 (2018). <https://doi.org/10.1007/s10470-017-1091-3>
2. S. Becker, J. Vielhaben, M. Ackermann et al., Audiomnist: exploring explainable artificial intelligence for audio analysis on a simple benchmark. *J. Frankl. Inst.* **1**, 1 (2023). <https://doi.org/10.1016/j.jfranklin.2023.11.038>
3. N. Bershad, L. Qu, on the probability density function of the lms adaptive filter weights, in *ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing* (1987), pp. 109–112. <https://doi.org/10.1109/ICASSP.1987.1169750>
4. A.H. Bukhari, M.A.Z. Raja, M. Sulaiman et al., Fractional neuro-sequential arfima-lstm for financial market forecasting. *IEEE Access* **8**, 71326–71338 (2020). <https://doi.org/10.1109/ACCESS.2020.2985763>
5. H. Butterweck, An approach to lms adaptive filtering without use of the independence assumption, in *1996 8th European Signal Processing Conference (EUSIPCO 1996)* (1996), pp. 1–4
6. M.L.R. de Campos, G. Strang, QR decomposition an annotated bibliography, in *QRD-RLS Adaptive Filtering*. (Springer US, Boston, 2009), pp.1–22
7. J.F. Cardoso, Blind signal separation: statistical principles. *Proc. IEEE* **86**(10), 2009–2025 (1998)
8. H. Chen, X. Long, Y. Tang et al., Passive and h_{∞} control based on non-fragile observer for a class of uncertain nonlinear systems with input time-delay. *J. Vib. Control* (2023). <https://doi.org/10.1177/10775463231193458>
9. Y.S. Choi, Subband adaptive filtering with-norm constraint for sparse system identification. *Math. Probl. Eng.* **2013**, 1–7 (2013)
10. P.S.R. Diniz, *Adaptive Filtering* (Springer International Publishing, Berlin, 2020). <https://doi.org/10.1007/978-3-030-29057-3>
11. S. Douglas, T.Y. Meng, Exact expectation analysis of the LMS adaptive filter without the independence assumption, in *1992 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1992. ICASSP-92* (IEEE, 1992), pp. 61–64
12. S. Douglas, T.Y. Meng, Stochastic gradient adaptation under general error criteria. *IEEE Trans. Signal Process.* **42**(6), 1335–1351 (1994). <https://doi.org/10.1109/78.286951>
13. 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*, vol. 1 (1992), pp. 566–570
14. 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 (1993), pp. 519–522
15. 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 (1993), pp. 519–522
16. S.C. Douglas, 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* (1992), pp. 61–64

17. S.C. Douglas, W. Pan, Exact expectation analysis of the LMS adaptive filter. *IEEE Trans. Signal Process.* **43**(12), 2863–2871 (1995)
18. S.C. Douglas, W. Pan, Exact expectation analysis of the LMS adaptive filter. *IEEE Trans. Signal Process.* **43**(12), 2863–2871 (1995)
19. E. Eweda, A new approach for analyzing the limiting behavior of the normalized LMS algorithm under weak assumptions. *Signal Process.* **89**(11), 2143–2151 (2009)
20. J. Foley, F. Boland, A note on the convergence analysis of lms adaptive filters with Gaussian data. *IEEE Trans. Acoust. Speech Signal Process.* **36**(7), 1087–1089 (1988). <https://doi.org/10.1109/29.1632>
21. S. Gazor, Prediction in lms-type adaptive algorithms for smoothly time varying environments. *IEEE Trans. Signal Process.* **47**(6), 1735–1739 (1999). <https://doi.org/10.1109/78.765152>
22. S. Guan, Z. Li, Normalised spline adaptive filtering algorithm for nonlinear system identification. *Neural Process. Lett.* **46**(2), 595–607 (2017). <https://doi.org/10.1007/s11063-017-9606-6>
23. Z. Habibi, H. Zayyani, Markovian adaptive filtering algorithm for block-sparse system identification. *IEEE Trans. Circuits Syst. II Express Briefs* **68**(8), 3032–3036 (2021). <https://doi.org/10.1109/TCSII.2021.3069879>
24. S. Haykin, B. Widrow, *Least-Mean-Square Adaptive Filters*. Wiley Online Library (2003)
25. J. Huang, H. Chen, C. Shen, Event-triggered model-free adaptive control for wheeled mobile robot with time delay and external disturbance based on discrete-time extended state observer. *J. Dyn. Syst. Meas. Control* (2023). <https://doi.org/10.1115/1.4063996>
26. S.Z. Islam, S.Z. Islam, R. Jidin et al, Performance study of adaptive filtering algorithms for noise cancellation of ecg signal, in *2009 7th International Conference on Information, Communications and Signal Processing (ICICS)* (2009), pp. 1–5. <https://doi.org/10.1109/ICICS.2009.5397744>
27. T. Kailath, *Lectures on Wiener and Kalman Filtering* (Springer, Vienna, 1981). <https://doi.org/10.1007/978-3-7091-2804-6>
28. C.L. Keppenne, M. Ghil, Adaptive filtering and prediction of the southern oscillation index. *J. Geophys. Res.: Atmos.* **97**(D18), 20449–20454 (1992)
29. R. Kwong, E. Johnston, A variable step size lms algorithm. *IEEE Trans. Signal Process.* **40**(7), 1633–1642 (1992). <https://doi.org/10.1109/78.143435>
30. P. Lara, F. Igreja, L.D. Tarrataca et al., Exact expectation evaluation and design of variable step-size adaptive algorithms. *IEEE Signal Process. Lett.* **26**(1), 74–78 (2018)
31. P. Lara, K.S. Olinto, F.R. Petraglia et al., Exact analysis of the least-mean-square algorithm with coloured measurement noise. *Electron. Lett.* **54**(24), 1401–1403 (2018)
32. P. Lara, D.B. Haddad, L. Tarrataca, Advances on the analysis of the LMS algorithm with a colored measurement noise. *SIViP* **14**, 529–536 (2019)
33. P. Lara, L.D. Tarrataca, D.B. Haddad, Exact expectation analysis of the deficient-length LMS algorithm. *Signal Process.* **162**, 54–64 (2019)
34. P. Lara, F. Igreja, T.T.P. Silva et al., Exact expectation analysis of the LMS adaptive identification of non-linear systems. *Electron. Lett.* **56**(1), 45–48 (2020)
35. J. Li, P. Stoica, An adaptive filtering approach to spectral estimation and sar imaging. *IEEE Trans. Signal Process.* **44**(6), 1469–1484 (1996)
36. Y. Li, K.R. Liu, Static and dynamic convergence behavior of adaptive blind equalizers. *IEEE Trans. Signal Process.* **44**(11), 2736–2745 (1996). <https://doi.org/10.1109/78.542180>
37. D. Margaris, A. Kobusińska, D. Spiliotopoulos et al., An adaptive social network-aware collaborative filtering algorithm for improved rating prediction accuracy. *IEEE Access* **8**, 68301–68310 (2020). <https://doi.org/10.1109/ACCESS.2020.2981567>
38. V. Mathews, Z. Xie, A stochastic gradient adaptive filter with gradient adaptive step size. *IEEE Trans. Signal Process.* **41**(6), 2075–2087 (1993). <https://doi.org/10.1109/78.218137>
39. R. Morrison, R. Baptista, E. Basor, Diagonal nonlinear transformations preserve structure in covariance and precision matrices. *J. Multivar. Anal.* **190**, 104983 (2022). <https://doi.org/10.1016/j.jmva.2022.104983>
40. B.B. Nair, V. Mohandas, N. Sakthivel et al, Application of hybrid adaptive filters for stock market prediction, in *2010 International Conference on Communication and Computational Intelligence (INCOCCI)* (IEEE, 2010), pp. 443–447
41. A.H. Sayed, *Adaptive Filters*. Wiley (2008). <https://doi.org/10.1002/9780470374122>
42. H. Simon, *Adaptive Filter Theory*, vol. 2 (Prentice Hall, Hoboken, 2002), pp.478–481
43. V. Solo, LMS: past, present and future, in *ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2019), pp. 7740–7744

44. J. Tan, J. Zhang, An optimal adaptive filtering algorithm with a polynomial prediction model. *Sci. China Inf. Sci.* **54**(1), 153–162 (2011). <https://doi.org/10.1007/s11432-010-4141-3>
45. N.V. Thakor, Y.S. Zhu, Applications of adaptive filtering to ecg analysis: noise cancellation and arrhythmia detection. *IEEE Trans. Biomed. Eng.* **38**(8), 785–794 (1991)
46. M.D.S. Vieitos, M.P. Tcheou, D.B. Haddad et al., Improved proportionate constrained normalized least mean square for adaptive beamforming. *Circuits Syst. Signal Process.* **42**(12), 7651–7665 (2023). <https://doi.org/10.1007/s00034-023-02459-3>
47. B. Widrow, J.M. McCool, M.G. Larimore et al., Stationary and nonstationary learning characteristics of the LMS adaptive filter. *Proc. IEEE* **64**(8), 1151–1162 (1976). <https://doi.org/10.1109/PROC.1976.10286>
48. A. Zaknich, *Principles of Adaptive Filters and Self-Learning Systems* (Springer, London, 2005)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Filipe Igreja¹  · Pedro Lara²  · Luís Tarrataca²  · Laura S. de Assis²  ·
Fernanda D. V. R. Oliveira³  · Ana L. F. de Barros¹  · Diego B. Haddad^{1,2} 

Pedro Lara
pedro.lara@cefet-rj.br

Luís Tarrataca
luis.tarrataca@cefet-rj.br

Laura S. de Assis
laura.assis@cefet-rj.br

Fernanda D. V. R. Oliveira
fernanda.dvro@poli.ufrj.br

Ana L. F. de Barros
ana.barros@cefet-rj.br

Diego B. Haddad
diego.haddad@cefet-rj.br

- 1 PhD in Instrumentation and Applied Optics (PPGIO), CEFET/RJ, Rua General Canabarro, Rio de Janeiro, RJ 20271-204, Brazil
- 2 Computer Engineering Department, CEFET/RJ, Rua do Imperador, 971, Petrópolis, RJ 25620-003, Brazil
- 3 Electrical Engineering Program, Universidade Federal do Rio de Janeiro (UFRJ), Av. Athos da Silveira Ramos, 149, Centro de Tecnologia, Bloco H, Rio de Janeiro, RJ 21941-972, Brazil