A Novel Bias-Variance Decomposition of the LMS Algorithm Learning Behavior

João Pedro Costa e Silva Mayworm, Flavio Henrique Origuela Meira, Luís Tarrataca, Diego Barreto Haddad

***Abstract*—This paper revisits the learning dynamics of the least-mean-squares (LMS) algorithm through a new bias–variance decomposition of the mean-square deviation that remains valid for colored and non-Gaussian inputs. Building on the radial-angular input model of Slock, which separates a continuous radial component from a discrete angular distribution, novel closed-form, modewise expressions for the transient and steady-state behavior are derived. Our framework unifies and extends recent results on bias-variance decomposition of the mean-square deviation of the LMS algorithm. The analysis clarifies when classical Gaussian-based predictions fail and offers a tractable pathway to performance prediction and parameter tuning of LMS under realistic, non-ideal input models.**

***Index Terms*—Adaptive Filtering, Stochastic Model, Least Mean Squares, Bias-variance.**

I. Introduction

THE LMS algorithm stands as the most fundamental approach in adaptive signal processing, having demonstrated its usefulness in a wide range of applications, from system identification and channel equalization to adaptive noise cancellation [1]. Despite its low computational cost, its learning dynamics remain remarkably intricate, still challenging advanced mathematical tools developed to provide a rigorous theoretical framework capable of explaining its notable success and flexibility [2]–[4].

A recent contribution in adaptive filtering introduced a different perspective for analyzing the LMS algorithm by decomposing its mean-square deviation into bias and variance components along the iterations [2]. This approach sheds new light on the transient performance of LMS filters, offers connections with machine learning literature, where the bias-variance trade-off is a central paradigm [5], and provides qualitative explanations for practical phenomena such as the impact of filter length mismatch and the effects of impulsive noise. Moreover, it suggests potential utility for guiding the design and analysis of future solutions, including variable step-size and variable tap-length algorithms.

João Pedro Costa e Silva Mayworm, Flavio Henrique Origuela Meira, Luís Tarrataca and Diego Barreto Haddad are with CEFET/RJ, Rio de Janeiro, RJ, Brazil (e-mails: joao.mayworm@aluno.cefet-rj.br, flaviohenrique-origuelameira@gmail.com, luis.tarrataca@cefet-rj.br, diego.haddad@cefet-rj.br). ORCID: 0009-0006-2698-6789 (João), 0009-0008-5755-5927 (Flavio), 0000-0001-9359-5143 (Luís), 0000-0002-7634-5481 (Diego).

This paper 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).

Submission: 2025-11-29, First decision: 2026-02-05, Acceptance: 2026-03-03, Publication: 2026-03-18.

Digital Object Identifier: 10.14209/jcis.2026.5

However, the analysis presented in [2] is overly restrictive in its statistical assumptions regarding the input signal, which is taken to be both white and Gaussian. In this work, we extend this paradigm to accomodate colored input signals (which engenders an important drawback on LMS performance [6]), without imposing Gaussianity. To this end, we adapt the theoretical formulation originally proposed in [7], hereafter referred to as the "Slock model." This paradigm introduced valuable qualitative insights into the learning dynamics of adaptive filtering algorithms, and has consequently inspired an extensive body of research [8]–[14]. The Slock model assumes that the input vector is composed of three independent stochastic components that respectively describe the vector signal and its radial and angular distributions. In concise terms, this paper argues that this framework enables the characterization of the distinct convergence modes of the bias and variance terms that govern the mean-square error dynamics of the LMS algorithm.

This paper is structured as follows: Section II provides the fundamental concepts about the LMS algorithm. The first- and second-order analysis of the LMS is addressed in Section III. Section IV, which elucidates the bias-variance behavior of the LMS learning process, contains the main theoretical contributions of this paper. Section V presents the simulation results of the paper and Section VI contains the conclusions of this paper.

Mathematical notation. Throughout this paper, $(\cdot)^\top$ denotes the transpose. All vectors are column type, written in bold lowercase letters. Matrices are written in bold uppercase letters. Scalars are written in non-bold lowercase letters. The symbol $\mathbb{E}[\cdot]$ denotes the expectation operator, $\boldsymbol{I}$ is the identity matrix, and $\text{Tr}[\boldsymbol{A}]$ is the trace of matrix $\boldsymbol{A}$. Finally, $\lfloor x \rceil$ denotes the nearest integer with respect to x.

II. LMS Fundamental Concepts

This work considers an adaptive estimator based on the LMS algorithm, which adjusts the weights $\boldsymbol{w}(k) \in \mathbb{R}^N$. The output of the adaptive filter at the k-th iteration is given by

$$y(k) \triangleq \boldsymbol{w}^\top(k)\boldsymbol{x}(k), \tag{1}$$

where $\boldsymbol{x}(k) \in \mathbb{R}^N$ contains N consecutive input samples:

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

Consider that a noisy linear-in-the-parameters reference signal $d(k)$ is available, so that

$$d(k) = [\boldsymbol{w}^\star]^\top \boldsymbol{x}(k) + \nu(k), \tag{3}$$

where $\boldsymbol{w}^\star \in \mathbb{R}^N$ is unknown and $\nu(k) \in \mathbb{R}$ accounts for measurement noise. The LMS weight update recursion with step size β has the form

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

where $e(k) \triangleq d(k) - y(k)$ is the error signal. Equation (4) falls into the class of equation-error identification methods. The LMS aims to identify in an adaptive manner the optimal parameter vector $\boldsymbol{w}^\star$ that minimizes the performance index $\mathbb{E}[d(k) - y(k)]^2$.

III. Stochastic Model for the LMS

A. Input vector statistics

Consider the eigendecomposition of the input vector auto-correlation matrix $\boldsymbol{R}$

$$\boldsymbol{R} \triangleq \mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^\top(k)\right] = \sum_{i=1}^{N} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^\top, \tag{5}$$

where $\boldsymbol{v}_i$ denotes the i-th eigenvector of $\boldsymbol{R}$, with $\|\boldsymbol{v}_i\| = 1$, and $\lambda_i \in \mathbb{R}$ is its corresponding eigenvalue.

It is assumed that the input vector $\boldsymbol{x}(k)$ is generated by the product of three independent random components, each being i.i.d., *viz.*, [7]

$$\boldsymbol{x}(k) = s(k) r(k) \boldsymbol{\mathcal{V}}(k), \tag{6}$$

where $s(k)$ is a random variable that takes values in $\{+1, -1\}$ with equal probability, encoding the sign of the resulting vector, $r(k)$ is a nonnegative random variable distributed as the norm of the input vector (*i.e.*, $r(k) \sim \|\boldsymbol{x}(k)\|$) and $\boldsymbol{\mathcal{V}}(k)$ is a *discrete* random vector that takes the value of eigenvector $\boldsymbol{v}_i$ with probability

$$\Pr\{\boldsymbol{\mathcal{V}}(k) = \boldsymbol{v}_i\} = \frac{\lambda_i}{\mathrm{Tr}[\boldsymbol{R}]}. \tag{7}$$

Note that model (6) is consistent with first- and second-order statistics of the input vector, namely $\mathbb{E}[\boldsymbol{x}(k)] = \boldsymbol{0}$ and $\mathbb{E}[\boldsymbol{x}(k)\boldsymbol{x}^\top(k)] = \boldsymbol{R}$. In this formulation, $\boldsymbol{\mathcal{V}}(k)$ discretizes the angular distribution into N possible directions, while the radial component $r(k)$ governs the input magnitude. As will be shown, the distribution of $r(k)$ plays a key role in shaping the second-order learning behavior of the LMS algorithm. Moreover, the proposed framework naturally accommodates a wide range of input vector distributions (with distinct radial distributions), beyond the Gaussian case assumed by [2].

B. First-order (bias) component

Consider the deviation vector $\tilde{\boldsymbol{w}}(k)$,

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

whose dynamics is governed by the following stochastic difference equation (see Equations (3) and (4)):

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

For analytical convenience, consider the following statistical assumptions: *(i)*`IA` (*independence assumption*). The stationary input vector $\boldsymbol{x}(k)$ is statistically independent from vector $\boldsymbol{w}(k)$; and *(ii)* `NA` (*noise assumption*). The zero-mean stationary noise signal is i.i.d. and statistically independent from the remaining random signals. Both assumptions (*i.e.*, `IA` and `NA`) are widely used and discussed in papers that address stochastic models of adaptive filtering algorithms (see, *e.g.*, [1], [4]).

The first-order behavior of $\tilde{\boldsymbol{w}}(k)$ can be characterized by the application of the expectation operator in (9), in conjunction with `IA` and `NA`:

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

We are particularly interested in the projections of the deviation vector onto the different eigenvectors of $\boldsymbol{R}$. Using (10) and (5), the following *first-order* recursion can be derived

$$\underbrace{\mathbb{E}\left[\boldsymbol{v}_j^\top \tilde{\boldsymbol{w}}(k+1)\right]}_{\triangleq z_j(k+1)} = \left[\boldsymbol{v}_j^\top - \beta \lambda_j \boldsymbol{v}_j^\top\right] \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right], \tag{11}$$

where, for simplicity, $z_j(k)$ is referenced as the j-th *mode*. Since $\mathbb{E}[\tilde{\boldsymbol{w}}(k)]$ corresponds to a bias of the adaptive estimator, the state variables $z_j(k)$ (for $j \in \{1, 2, \ldots, N\}$) are directly related to the bias of the LMS. Using (11), one can establish the following recursion for $z_j(k)$:

$$z_j(k+1) = (1 - \beta \lambda_j) z_j(k). \tag{12}$$

Since $z_j(k)$ depends on the initial conditions and its square directly affects the second-order statistics of the learning process, it is particularly useful to rewrite (12) in the form

$$z_j^2(k) = (1 - \beta \lambda_j)^{2k} z_j^2(0), \tag{13}$$

which makes clear that the convergence time depends strongly on the product of the step size β and the eigenvalues of the data covariance matrix $\boldsymbol{R}$.

Remark: In the case of a white input signal with variance σ_x^2, the eigenvectors of $\boldsymbol{R}$ coincide with the canonical basis. Consequently, one has $\sum_{j=0}^{N-1} z_j^2(0) = \|\boldsymbol{w}^\star\|^2$, $\lambda_j = \sigma_x^2$, for all $j \in \{0, 1, \ldots, N-1\}$, $\sum_{j=0}^{N-1} z_j^2(k) = \|\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right]\|^2$. In this scenario, the summation over all modes yields

$$\|\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right]\|^2 = (1 - \beta \sigma_x^2)^{2k} \|\boldsymbol{w}^\star\|^2, \tag{14}$$

which coincides with Equation (13) of [2].

C. Second-order analysis

After computing the outer product of both terms in (9), applying the expectation operator, and using `IA` and `NA`, one obtains:

$$\begin{aligned} \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k+1) \;=\; & \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) - \beta \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^\top(k)\boldsymbol{x}(k)\boldsymbol{x}^\top(k)\right] \\ & - \beta \mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^\top(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^\top(k)\right] + \beta^2 \sigma_\nu^2 \boldsymbol{R} \\ & + \beta^2 \mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^\top(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^\top(k)\boldsymbol{x}(k)\boldsymbol{x}^\top(k)\right] \end{aligned} \tag{15}$$

where $\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) \triangleq \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^\top(k)\right]$.

While $z_j(k)$ is a first-order moment of the projections of the deviation vector onto the different eigenvectors $\boldsymbol{v}_j$, we define the second-order moment $\rho_j(k)$:

$$\rho_j(k) \triangleq \mathbb{E}\left\{\left[\boldsymbol{v}_j^\top \tilde{\boldsymbol{w}}(k)\right]^2\right\} = \boldsymbol{v}_j^\top \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) \boldsymbol{v}_j. \tag{16}$$

By multiplying (15) on the left by $\boldsymbol{v}_j^\top$ and on the right by $\boldsymbol{v}_j$, and employing the model (6), one obtains, after several algebraic manipulations [7]:

$$\rho_j(k+1) = \alpha_j \rho_j(k) + \gamma_j, \tag{17}$$

where

$$\alpha_j \triangleq \left(1 - 2\beta\lambda_j + \frac{\beta^2 \lambda_j \mathbb{E}[r^4]}{\mathrm{Tr}[\boldsymbol{R}]}\right) \tag{18}$$

and

$$\gamma_j \triangleq \beta^2 \sigma_\nu^2 \lambda_j. \tag{19}$$

It is worth emphasizing that second-order convergence requires prior knowledge of the input power level to properly tune the step size β, ensuring both stability and convergence. Furthermore, the fourth-order moment $\mathbb{E}[r^4]$ of the input vector norm depends on the underlying probability distribution of the input signal. Table I summarizes the resulting expressions of α_j for several representative zero-mean i.i.d. distributions with variance σ_x^2. Notably, when $x(k)$ is Gaussian (first row of Table I), the expression reduces to Equation (20) in [2], which specifically addresses the case of white Gaussian input signals. Since the condition $|\alpha_j| < 1$ guarantees second-order stability, the maximum admissible step size β_{max} can be directly derived for each distribution. In particular, for Gaussian $\mathcal{N}(0, \sigma_x^2)$ inputs, $\beta_{\text{max}} = \frac{2}{(N+2)\sigma_x^2}$; for uniform $[-a, a]$, $\beta_{\text{max}} = \frac{2}{\left(N + \frac{4}{5}\right)\sigma_x^2}$.

TABLE I
VALUES OF α_j FOR COMMON I.I.D. ZERO-MEAN DISTRIBUTIONS WITH VARIANCE σ_x^2.

Distribution	α_j
Gaussian $\mathcal{N}(0, \sigma_x^2)$	$1 - 2\beta\sigma_x^2 + \beta^2(N+2)\sigma_x^4$
Uniform $[-a, a]$	$1 - 2\beta\sigma_x^2 + \beta^2\left(N + \frac{4}{5}\right)\sigma_x^4$
Symmetric Bernoulli $\{\pm a\}$	$1 - 2\beta\sigma_x^2 + \beta^2 N \sigma_x^4$
Laplace (double-exponential)	$1 - 2\beta\sigma_x^2 + \beta^2(N+5)\sigma_x^4$
Student-t ($\nu > 4$)	$1 - 2\beta\sigma_x^2 + \beta^2\left(N + 2 + \frac{6}{\nu-4}\right)\sigma_x^4$

Equation (17) admits the following analytical solution for $\rho_j(k)$:

$$\rho_j(k) = \alpha_j^k \rho_j(0) + \gamma_j \frac{1 - \alpha_j^k}{1 - \alpha_j}. \tag{20}$$

The first- and second-order Equations (13) and (20) are utilized for a bias variance description in the following.

IV. Bias-variance Decomposition

Using NA and IA, the learning curve of the mean-squared error (MSE) can be written as [7]

$$\xi(k) \triangleq \mathbb{E}\left[e^2(k)\right] = \sigma_\nu^2 + \mathrm{Tr}[\boldsymbol{R}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)] = \sigma_\nu^2 + \sum_{j=1}^{N} \lambda_j \rho_j(k), \tag{21}$$

where σ_ν^2 is the variance of the $\nu(k)$. From (16) and (21), one may conclude that $\rho_j(k)$ contains second-order statistics about the adaptive estimator. Further, $\rho_j(k)$ can be decomposed as

$$\rho_j(k) = \left\{\underbrace{\mathbb{E}\left[\boldsymbol{v}_j^\top \tilde{\boldsymbol{w}}(k+1)\right]}_{\text{bias}}\right\}^2 + \underbrace{\sigma_j^2(k)}_{\text{variance}}, \tag{22}$$

where $\sigma_j^2(k)$ is formally defined as

$$\sigma_j^2(k) \triangleq \mathbb{E}\Big[\big(\boldsymbol{v}_j^\top \tilde{\boldsymbol{w}}(k)\big)^2\Big] - \big(\mathbb{E}\big[\boldsymbol{v}_j^\top \tilde{\boldsymbol{w}}(k)\big]\big)^2 = \rho_j(k) - z_j^2(k),$$

that is, the variance of the projection of the deviation vector onto the j-th eigenmode.

After applying (13) and (20) to (22), and substituting α_j and γ_j by their definitions in (18) and (19), respectively, one obtains after some algebraic manipulations:

$$\sigma_j^2(k) = \rho_j(0)\left(\alpha_j^k - (1 - \beta\lambda_j)^{2k}\right) + \chi\left(1 - \alpha_j^k\right), \tag{23}$$

where

$$\chi \triangleq \frac{\beta\sigma_\nu^2}{2 - \frac{\beta\mathbb{E}[r^4]}{\mathrm{Tr}[\boldsymbol{R}]}}. \tag{24}$$

Moreover, we recall that the bias term in (22) corresponds to $z_j^2(k)$. In addition, under the usual initialization $\tilde{\boldsymbol{w}}(0) = \boldsymbol{w}^\star - \boldsymbol{w}(0)$ deterministic (e.g., $\boldsymbol{w}(0) = \boldsymbol{0}$), one has $\rho_j(0) = \mathbb{E}\Big[\big(\boldsymbol{v}_j^\top \tilde{\boldsymbol{w}}(0)\big)^2\Big] = \big(\mathbb{E}\big[\boldsymbol{v}_j^\top \tilde{\boldsymbol{w}}(0)\big]\big)^2 = z_j^2(0)$, since $\mathrm{Var}\big(\boldsymbol{v}_j^\top \tilde{\boldsymbol{w}}(0)\big) = 0$.

Remark: Equation (23) provides an analytical expression for the evolution of the variance of the j-th mode $\rho_j(k)$ in the second-order dynamics of the LMS algorithm.

Assuming that the adaptive coefficients are initialized at zero, the initial bias term is determined by the projection of the optimal filter onto the eigenbasis of the input covariance matrix, namely $z_j^2(k) = \left[\boldsymbol{v}_j^\top \boldsymbol{w}^\star\right]^2$ (see Equations (8), (11), and (13)). It should be emphasized that, during the initial stage of the learning process, the bias term prevails, as the variance $\sigma_j^2(k)$ vanishes at $k = 0$. In contrast, the variance term $\sigma_j^2(k)$ exhibits a more intricate behavior. Rather than decreasing steadily, it initially grows, reaches a peak at a certain iteration, and only then starts to decay towards its steady-state value. This transient peaking effect is of particular interest, as it characterizes the point at which stochastic fluctuations dominate the learning process. The following theorem formalizes the iteration index at which this maximum occurs.

Theorem 1. The variance $\sigma_j^2(k)$ reaches its maximum value around iteration $k_{\mathrm{p},j}$, where

$$k_{\mathrm{p},j} = \left\lfloor \frac{\log\left(\frac{\rho_j(0)}{\rho_j(0) - \chi}\right) + \log\left[\frac{2\log(1-\beta\lambda_j)}{\log\alpha_j}\right]}{\log\alpha_j - 2\log(1-\beta\lambda_j)} \right\rceil. \tag{25}$$

Proof. By calculating the derivative of $\sigma_j^2(k)$ with respect to k (see Equation (23)), one obtains:

$$\begin{aligned} \frac{\partial\sigma_j^2(k)}{\partial k} &= -2\rho_j(0)\log(1-\beta\lambda_j)(1-\beta\lambda_j)^{2k} \\ &\quad + \rho_j(0)\alpha_j^k \log\alpha_j - \alpha_j^k \chi \log(\alpha_j). \end{aligned} \tag{26}$$

After setting the derivative equal to zero, performing several algebraic manipulations, and applying the logarithm function, one arrives at:

$$\begin{aligned} &\log\left(\frac{\rho_j(0)}{\rho_j(0) - \chi}\right) + \log\left[\frac{2\log(1-\beta\lambda_j)}{\log\alpha_j}\right] \\ &\qquad = k_{\mathrm{p},j}\left[\log\alpha_j - 2\log(1-\beta\lambda_j)\right], \end{aligned} \tag{27}$$

which leads to the desired result (Equation (25)). □

Remark: Unlike equation (25) in [2], which yields a single peak iteration index under white Gaussian inputs, the expression in (25) depends on the mode j. This mode dependence results from input coloration, as each eigenvalue λ_j induces distinct transient dynamics. The variance of the estimates tends to peak at an intermediate stage of adaptation, when they have already moved significantly away from the initial guess but have not yet reached convergence to the true solution.

From a certain iteration onward, the j-th mode is dominated by the variance component, since the bias term decreases monotonically. The iteration k_a at which the two terms alternate in relative relevance can be determined by equating the right-hand sides of Equations (13) and (23), which leads to:

$$\left[2(1-\beta\lambda_j)^{2k_\text{a}} - \alpha_j^{k_\text{a}}\right]\rho_j(0) + \chi\alpha_j^{k_\text{a}} = \chi, \qquad (28)$$

an expression that does not admit a closed-form solution but allows k_a to be obtained numerically. Note that (28) generalizes Equation (29) of [2].

V. Results

The simulation setup considers an adaptive filtering scenario where the optimal system to be identified by the LMS algorithm is represented by vector $\boldsymbol{w}^\star = [1, -0.7, 0.5, -0.2, 0.3]^\top$. The algorithm uses a learning step size of $\beta = 10^{-2}$, while the additive measurement noise in the error signal is modeled as zero-mean Gaussian with variance $\sigma_\nu^2 = 10^{-6}$. To ensure statistical reliability, the performance results are averaged over $n_\text{repeats} = 10^4$ independent Monte Carlo trials. The input signal is generated by filtering a zero-mean white Gaussian sequence through the FIR filter $B(z) = 1 + 0.3z^{-1} - 0.1z^{-2}$, which induces correlation and thus yields a colored excitation process.

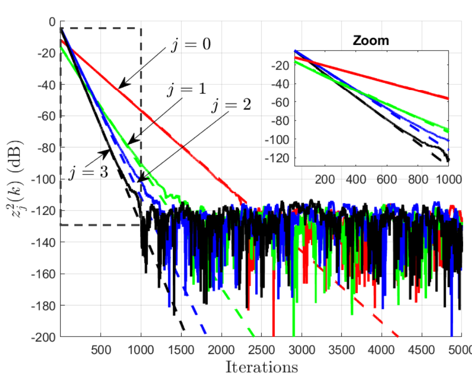

Fig. 1. Evolution of the expected value of $z_j^2(k)$ for $j \in \{0,1,2,3\}$. Theoretical curves (obtained from Equation (13)) shown with dashed lines. Simulated curves (obtained via the Monte Carlo method) shown with solid lines.

Figure 1 depicts the theoretical and simulated evolution of $z_j^2(k)$. As predicted by Equation (13), the exponential decay of each mode results in a straight line on a logarithmic scale. Due to numerical inaccuracies inherent to the computational environment, the simulated curves of $z_j^2(k)$ converge to a very small but nonzero value. It is observed that Equation (13) accurately predicts the bias decay for distinct values of j. Despite the simplicity of the underlying hypotheses, the model also reproduces the transient dynamics of $\rho_j(k)$ with surprising precision (see Fig. 2). Such correspondence underscores that the apparent randomness in the learning curve can be traced back to structured, mode-dependent dynamics.

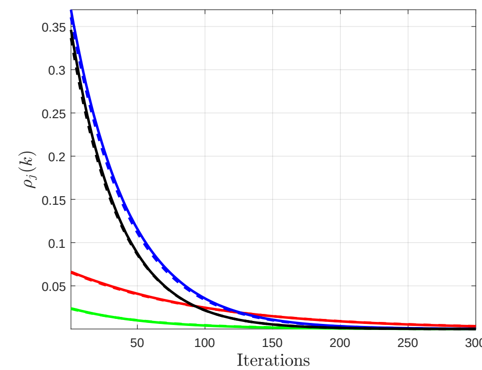

Fig. 2. Evolution of $\rho_j(k)$ for $j \in \{0,1,2,3\}$. Theoretical curves (see Equation (17)) shown with dashed lines. Simulated curves (obtained via the Monte Carlo method) shown with solid lines.

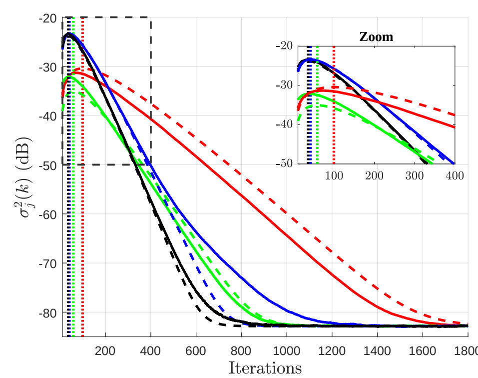

Fig. 3. Evolution of $\sigma_j^2(k)$ for $j \in \{0,1,2,3\}$. Theoretical curves shown with dashed lines. Simulated curves (obtained via the Monte Carlo method) shown with solid lines.

Regarding the evolution of the variances $\sigma_j^2(k)$, Fig. 3 compares the theoretical model (Equation (23)) with simulation results. A deviation between simulated and theoretical peak values is observed, mainly due to the simplifying assumptions (mainly the stochastic model (6) and IA, which becomes less valid under colored input [7], [15]). Nevertheless, the proposed model uniquely captures, for the first time to our knowledge, the occurrence of mode-dependent peaks at distinct time instants.

VI. Conclusions

This paper presented a novel bias–variance decomposition for analyzing the learning behavior of the LMS algorithm, separating deterministic bias dynamics from stochastic variance evolution. The proposed framework yields theoretical predictions that closely match simulations, accurately capturing mode-dependent peaks and eigenmode evolution, thereby offering a unified perspective on learning dynamics in stochastic adaptive filters. By advancing conventional analyses, the formulation provides a concise yet insightful tool for understanding trade-offs in adaptive filter learning, and future work will extend this decomposition to normalized and proportionate LMS algorithms.

REFERENCES

[1] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*. Springer International Publishing, 2020. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-29057-3

[2] D. G. Tiglea, R. Candido, and M. T. M. Silva, "Bias-Variance Decomposition of the Mean-Square Deviation of the LMS Algorithm: Transient and Steady-State Analysis," *Circuits, Systems, and Signal Processing*, vol. 44, no. 6, p. 4101–4132, Feb. 2025. [Online]. Available: http://dx.doi.org/10.1007/s00034-025-03005-z

[3] T. T. P. Silva, F. Igreja, P. Lara, L. Tarrataca, A. Kar, and D. B. Haddad, "On the Skewness of the LMS Adaptive Weights," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 8, pp. 3022–3026, 2021. [Online]. Available: http://dx.doi.org/10.1109/TCSII.2021.3068857

[4] T. T. P. Silva, P. Lara, F. Igreja, F. D. V. R. Oliveira, L. Tarrataca, and D. B. Haddad, "An Exact Expectation Model for the LMS Tracking Abilities," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5882–5893, 2020. [Online]. Available: http://dx.doi.org/10.1109/TSP.2020.3028993

[5] Y. Hu, X. Chen, and N. He, "On the bias-variance-cost tradeoff of stochastic optimization," in *Proceedings of the 35th International Conference on Neural Information Processing Systems*, ser. NIPS '21. Red Hook, NY, USA: Curran Associates Inc., 2021. [Online]. Available: http://dx.doi.org/10.48550/arXiv.2106.15502

[6] M. Ferrer, M. de Diego, and A. Gonzalez, "Low Cost Variable Step-Size LMS With Maximum Similarity to the Affine Projection Algorithm," *IEEE Open Journal of Signal Processing*, vol. 5, pp. 82–91, 2024. [Online]. Available: http://dx.doi.org/10.1109/OJSP.2023.3340106

[7] D. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Transactions on Signal Processing*, vol. 41, no. 9, pp. 2811–2825, 1993. [Online]. Available: http://dx.doi.org/10.1109/78.236504

[8] J. Shin, B. Y. Park, W. I. Lee, J. Yoo, and J. Cho, "A novel normalized subband adaptive filter algorithm based on the joint-optimization scheme," *IEEE Access*, vol. 10, pp. 9868–9876, 2022. [Online]. Available: http://dx.doi.org/10.1109/ACCESS.2022.3143136

[9] N. N. Siqueira, M. R. Petraglia, and D. B. Haddad, "A novel heuristic for the analysis of block sign LMS algorithm," *Electronics Letters*, vol. 59, no. 7, Apr. 2023. [Online]. Available: http://dx.doi.org/10.1049/ell2.12772

[10] T. R. P. Gonzaga, V. d. S. N. Junior, A. L. F. de Barros, F. d. R. Henriques, and D. B. Haddad, "Transient Analysis of the Block Least Mean Squares Algorithm," *IEEE Communications Letters*, vol. 25, no. 2, pp. 608–612, 2021. [Online]. Available: http://dx.doi.org/10.1109/LCOMM.2020.3026610

[11] T. K. Paul and T. Ogunfunmi, "On the Convergence Behavior of the Affine Projection Algorithm for Adaptive Filters," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 8, pp. 1813–1826, 2011. [Online]. Available: http://dx.doi.org/10.1109/TCSI.2011.2106091

[12] J. Apolinario, M. Campos, and P. Diniz, "Convergence analysis of the binormalized data-reusing LMS algorithm," *IEEE Transactions on Signal Processing*, vol. 48, no. 11, pp. 3235–3242, 2000. [Online]. Available: http://dx.doi.org/10.1109/78.875480

[13] S. Sankaran and A. Beex, "Convergence behavior of affine projection algorithms," *IEEE Transactions on Signal Processing*, vol. 48, no. 4, pp. 1086–1096, 2000. [Online]. Available: http://dx.doi.org/10.1109/78.827542

[14] B. K. Das, G. V. Chakravarthi, and M. Chakraborty, "A Convex Combination of NLMS and ZA-NLMS for Identifying Systems With Variable Sparsity," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 64, no. 9, pp. 1112–1116, 2017. [Online]. Available: http://dx.doi.org/10.1109/TCSII.2017.2651388

[15] S. S. Haykin, *Adaptive Filter Theory*, 2nd ed. London, England: Prentice-Hall, Feb. 1991.