Exact Expectation Analysis of the Deficient-Length LMS Algorithm

Pedro Lara, Luís D. T. J. Tarrataca, and Diego B. Haddad

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

Abstract

Stochastic models that predict adaptive filtering algorithms performance usually employ several assumptions in order to simplify the analysis. Although these simplifications facilitate the recursive update of the statistical quantities of interest, they by themselves may hamper the modeling accuracy. This paper simultaneously avoids for the first time the employment of two ubiquitous assumptions often adopted in the analysis of the least mean squares (LMS) algorithm. The first of them is the so-called independence assumption, which presumes statistical independence between adaptive coefficients and input data. The second one assumes a sufficient-order configuration, in which the lengths of the unknown plant and the adaptive filter are equal. State equations that characterize both the mean and mean square performance of the deficient-length configuration without using the independence assumption are provided. The devised analysis, encompassing both transient and steady-state regimes, is not restricted neither to white nor to Gaussian input signals and is able to provide a proper step size upper bound that guarantees stability.

Keywords: Adaptive Filtering, Exact Expectation Analysis, Deficient-Length Configuration

1. Introduction

Adaptive filtering algorithms are now a widespread technique for a plethora of applications, such as adaptive equalization, acoustic echo cancellation, and

system identification (the focus of this paper) [1]. Loosely speaking, they consist of recursive and nonlinear estimators of a set of parameters that extracts from the input (or excitation) signal $x(k)$ the information of interest by adjusting themselves to variations in their environments. The least mean squares (LMS) [2] is one of the most popular adaptive filters that often benchmarks others. Its robustness and low arithmetic complexity order (essentially obtained due to the fact that its update equation can be described in terms of inner products) make it suitable for hardware implementation, although some modifications are required in order to employ a pipeline architecture or to obtain a lower register complexity [3, 4, 5, 6]. Since the gradient noise of the stochastic LMS procedure increases the mean-squared error (MSE), one popular metric for the assessment of its learning performance is the MSE itself [7].

The establishment of deterministic or stochastic models for predicting the performance of adaptive filters is of primary concern, since they provide performance guarantees, guidelines for the designer, stability bounds, rate convergence estimates or even clarify in which sense they present robustness against perturbations [8, 9, 10]. Due to the nonlinear nature of the learning process, stochastic analyses often require lengthy manipulations, with most approaches relying on assumptions in order to maintain the mathematics tractable. One of the most employed is the so-called *independence assumption* (`IA`), which considers that adaptive weights are statistically independent from current input data. Alternatively, `IA` presumes that the sequence of input vectors[1] $\boldsymbol{x}(k) \in \mathbb{R}^N$ are statistically independent, a common assumption in the field of stochastic approximations [11] that is strictly valid in some specific cases (e.g., in the synchronous multiuser communication setting [12]).

In this paper, it is assumed that the adaptive algorithm employs a transversal structure, which imposes a deterministic coherence between successive input vectors. In such tapped-delay lines, `IA` is not even approximately true, although it provides agreement with actual performance in the case of small step sizes [13,

---

[1]All vectors of this paper are of column-type.

14]. In general terms, such a hypothesis cannot be invoked when the adaptive filter is present in the adaptation loop [15].

Note that the length of the ideal transfer function to identify can surpass the adaptive filter length N. In practice, system identification tasks may operate in such a deficient-length setting, especially when the unknown plant transfer function is long [16] or the designer intends to deal with computational limitations [17]. Since dimension adversely affects LMS performance [18, 19], such a configuration can also arise when an increase of the convergence rate is obtained by the usage of a time-variant adaptive filter length [20, 21, 22]. This realistic under-modeling configuration is not addressed by the majority of adaptive filtering analyses [16, 23, 24, 25]. This paper devises for the first time a comprehensive stochastic model that quantifies the statistical behavior of the LMS algorithm learning process under suboptimal operation. The proposed analysis does not employ the almost ubiquitous IA (which is invalid for input-shift data) and is able to generate time-independent linear state equations which recursively update the statistical quantities of interest. The devised procedure is able to perform analysis either to weight mean behavior or to mean squared evolution as well, by furnishing the necessary theoretical joint moments, besides providing a closed-form solution that describes asymptotic operation. Such an "exact expectation analysis" [26, 27, 28, 29, 30, 31] permits one to model the sophisticated learning capability of the deficient-length LMS, providing good adherence to experimental curves even when a non-infinitesimally small step size (or learning factor) β is adopted. Additionally, it is able to provide a more accurate upper bound on the learning factor in order to ensure convergence.

This paper is structured as follows. Section 2 describes the LMS algorithm operation in the deficient-length setting. Section 3 presents the classical analysis of such a configuration (i.e., one that employs IA), whereas Section 4 details the proposed exact expectation analysis. Section 5 depicts the results, performing comparisons between the advanced analysis method and the classical one. Finally, Section 6 presents the concluding remarks of this paper.

3

2. LMS Algorithm

The LMS is a stochastic gradient algorithm whose adjustment can be regarded as a feedback process driven by the error signal. It updates the signal-dependent coefficients vector

$$\boldsymbol{w}(k) \triangleq \begin{bmatrix} w_0(k) & w_1(k) & \ldots & w_{N-1}(k) \end{bmatrix}^T \tag{1}$$

by adding to the previous estimate a change proportional to the negative gradient of the instantaneous squared error signal:

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

where $e(k)$ denotes the prior error at the k-th iteration and β can also be regarded as a relaxation parameter. Note that (2) consists of a strategy that recursively converts an instantaneous performance assessment (i.e., the error signal $e(k)$) into a parameter adaptation that proceeds as more data becomes available.

The choice of a fixed step size imposes a trade-off between convergence rate and steady-state performance. Such a trade-off explicits a fundamental relationship between the amount of data used in obtaining the adaptive solution and its quality [32], which is related to the overall efficiency of an adaptive scheme [33]. Therefore, the choice of the step size cannot be overstated. It should be further noticed that the step size value also influences the divergence probability and strikes a balance between the amount of gradient noise and lag noise in nonstationary environments [34].

The error signal incorporates the discrepancy between a noisy measurement signal $d(k) \in \mathbb{R}$ and the filter output at the k-th iteration $y(k) \in \mathbb{R}$, which consists of a weighted sum of the elements of the input vector $\boldsymbol{x}(k)$:

$$e(k) \triangleq d(k) - y(k) = d(k) - \boldsymbol{w}^T(k)\boldsymbol{x}(k), \tag{3}$$

where $\boldsymbol{x}(k) \triangleq \begin{bmatrix} x(k) & x(k-1) & \ldots & x(k-N+1) \end{bmatrix}^T$ is the input vector at the k-th iteration.

Using (2) and (3), it is straightforward to derive the update equation of the LMS

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

whose related identification convergence rate is strongly dependent on the second-order moments of the input signal [18]. The LMS intends to estimate a set of parameters $w_i(k)$ (for $i \in \{0, 1, \ldots, N-1\}$) based on a single realization of the noisy stochastic process $\{d(j), x(j)\}_{j=0}^{k-1}$, since the involved statistics are assumed to be unknown [35]. Such an algorithm is able to operate satisfactorily without the intervention of the designer in an unknown and possibly time-varying environment [36]. The backpropagation algorithm, usually employed to the training of neural networks, can be regarded as a generalization of the LMS [37].

Note that the stochastic nature of (4) implements a sort of Brownian motion [9] and is the foundation of the analysis performed herein. Under some mild conditions, the standard LMS filter of sufficient order performs an unbiased estimation, although a weight-drift problem may occur when the input signal does not satisfy a persistence of excitation condition [38].

Update equation (4) may also be derived using another paradigm, one that understands the LMS as an exact solver of a deterministic optimization problem with a linear constraint:

$$\min_{\boldsymbol{w}(k+1)} \mathcal{F}[\boldsymbol{w}(k+1)] \triangleq \|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^2 \text{ s.t. } e_{\mathrm{p}}(k) = (1 - \beta\|\boldsymbol{x}(k)\|^2)e(k), \tag{5}$$

where $e_{\mathrm{p}}(k)$ is the posterior error, which is evaluated using the adaptive coefficients vectors with the pair of data $\{d(k), \boldsymbol{x}(k)\}$ *after* the update procedure:

$$e_{\mathrm{p}}(k) \triangleq d(k) - \boldsymbol{w}^T(k+1)\boldsymbol{x}(k). \tag{6}$$

This deterministic paradigm for the derivation of the LMS algorithm clarifies in which sense it makes use of the Minimal Disturbance Principle (MDP), which biases the estimation procedure in order to avoid new adaptive coefficients vectors $\boldsymbol{w}(k+1)$ located far from the previous solution [39]. Such an alternative approach has been revealed to be useful for the generation of new adaptive filtering algorithms [40, 41, 42, 43].

Assuming a deficient-length scenario, the reference signal $d(k)$ is henceforth assumed to be related to $x(k)$ according to the following noisy and linear-in-the-parameters regression model:

$$d(k) = \boldsymbol{x}^T(k)\boldsymbol{w}^\star + \overline{\boldsymbol{x}}^T(k)\overline{\boldsymbol{w}}^\star + \nu(k), \tag{7}$$

where $\nu(k)$ accounts for an additive noise that incorporates measurement inaccuracies, error modeling, and talkers voice and/or background noise in echo cancellation applications. Vectors $\boldsymbol{w}^\star \in \mathbb{R}^N$ and $\overline{\boldsymbol{w}}^\star \in \mathbb{R}^P$ contain the unknown coefficients of the plant the adaptive filter intends to emulate.

The crucial parameter that influences the model (7) is the length of the unknown plant, assumed to be $N + P$ (whereas the adaptive filter presents a shorter length N), with the vectors $\overline{\boldsymbol{x}}(k)$, $\boldsymbol{w}^\star$ and $\overline{\boldsymbol{w}}^\star$ defined as

$$\overline{\boldsymbol{x}}(k) \triangleq \Big[x(k-N) \quad x(k-N-1) \quad \ldots \quad x(k-N-P+1) \Big]^T, \tag{8}$$

$$\boldsymbol{w}^\star \triangleq \Big[w_0^\star \quad w_1^\star \quad \ldots \quad w_{N-1}^\star \Big]^T, \tag{9}$$

$$\overline{\boldsymbol{w}}^\star \triangleq \Big[w_N^\star \quad w_{N+1}^\star \quad \ldots \quad w_{N+P-1}^\star \Big]^T. \tag{10}$$

Model (7) may be interpreted as the linearization of more general nonlinear models (which includes neural networks) around an operating point [9]. Henceforth, it is assumed that $P > 0$, which characterizes a suboptimal operation. Since the deficient-length LMS presents a learning behavior for correlated inputs that is distinct from that when the input signal is white [23, 44], the analysis put forth in this paper is not restricted to an uncorrelated excitation sequence $x(k)$.

3. Classical Statistical Analysis

In this section, an IA-based stochastic analysis of the LMS algorithm under the suboptimal configuration is concisely described. The seminal reference in this context is [23], whose formulation differs from the one presented in this section, although it can be shown that both are equivalent. Note that although

this section does not contain original contributions, it introduces some crucial concepts for the next section (such as the noise assumption and the state vectors). As well as the exact expectation analysis addressed in the next section, the classical analysis method makes use of the following assumption regarding the noise signal:

> *Noise Assumption* (`NA`). The zero-mean noise signal $\nu(k)$ sequence is i.i.d. (independent and identically distributed) and is statistically independent from the input signal.

Remark: note that `NA` is typical in the context of adaptive filtering analyses [45, 46] and is often satisfied in practice [47].

Both first- and second-order analyses make use of the deviation vector $\tilde{\boldsymbol{w}}(k)$ defined as

$$\tilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}^{\star} - \boldsymbol{w}(k), \tag{11}$$

whose energy, rigorously speaking, does not converge asymptotically to zero even in the sufficient-order case due to the ubiquitous presence of the stochastic additive noise $\nu(k)$. Using (3), (4), (7) and (11), the following recursion can be proven to be valid

$$\tilde{\boldsymbol{w}}(k+1) = \left[\boldsymbol{I} - \beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\tilde{\boldsymbol{w}}(k) - \beta\boldsymbol{x}(k)\overline{\boldsymbol{x}}^T(k)\overline{\boldsymbol{w}}^{\star} - \beta\boldsymbol{x}(k)\nu(k), \tag{12}$$

which is a time-varying forced or nonhomogeneous stochastic difference equation. The application of the expectation operator $\mathbb{E}[\cdot]$ on (12), combined with some manipulations and simplifications, is the foundation of the following statistical analyses. Namely, Section 3.1 presents the first-order analysis whilst Section 3.2 discusses the second-order analysis.

3.1. Mean Weight Behavior

Since the expectation is a linear operation, a recursive update of the average deviation $\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right]$ can be obtained using (12) combined with `IA` and `NA`, which leads to the following compact form

$$\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] = \left[\boldsymbol{I} - \beta\boldsymbol{R}_x\right]\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right] - \beta\boldsymbol{R}_{\overline{x}}\overline{\boldsymbol{w}}^{\star}, \tag{13}$$

7

where $\boldsymbol{R}_x \triangleq \mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]$ and $\boldsymbol{R}_{\overline{x}} \triangleq \mathbb{E}\left[\boldsymbol{x}(k)\overline{\boldsymbol{x}}^T(k)\right]$ are the input autocorrelation and cross-correlation matrices, respectively. Note that the statistical dependence between $\boldsymbol{x}(k)$ and $\tilde{\boldsymbol{w}}(k)$ is neglected by IA, implying that the adaptive coefficients behave on average like the coefficients of the steepest descent algorithm operating in the same configuration. Additionally, Eq. (13) may also prove that the deficient-length LMS algorithm is stable in the mean if the step size satisfies [32, 23]

$$0 < \beta < \frac{2}{\text{Tr}\left[\boldsymbol{R}_x\right]}, \tag{14}$$

where $\text{Tr}[\boldsymbol{X}]$ denotes the trace of matrix $\boldsymbol{X}$, which is an upper bound of its maximum absolute eigenvalue. The theoretical upper bound in (14) is inversely proportional to the energy of the input signal, a statement that remains true in the case of second-order classical analysis [39]. It is worth noting that recursion (13) can be rewritten according to the following linear state equations

$$\boldsymbol{y}^{(\text{IA},1)}(k+1) = \boldsymbol{A}^{(\text{IA},1)}\boldsymbol{y}^{(\text{IA},1)}(k) + \boldsymbol{b}^{(\text{IA},1)}, \tag{15}$$

where the superscript (IA, n) denotes a statistical analysis of n-th order moments based on IA, $\boldsymbol{A}^{(\text{IA},1)} \triangleq \boldsymbol{I} - \beta\boldsymbol{R}_x$ is the time-invariant transition matrix, $\boldsymbol{y}^{(\text{IA},1)}(k) \triangleq \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right]$ is the state vector containing the mean deviation elements $\tilde{w}_i(k)$ (for $i \in \{0, 1, \ldots, N-1\}$) and $\boldsymbol{b}^{(\text{IA},1)} \triangleq -\beta\boldsymbol{R}_{\overline{x}}\overline{\boldsymbol{w}}^\star$ is a constant vector. Assuming a step size choice that ensures stability, the steady-state solution of (15) can be found in a closed-form:

$$\boldsymbol{y}_\infty^{(\text{IA},1)} \triangleq \lim_{k \to \infty} \boldsymbol{y}^{(\text{IA},1)}(k) = \left[\boldsymbol{I} - \boldsymbol{A}^{(\text{IA},1)}\right]^{-1}\boldsymbol{b}^{(\text{IA},1)}, \tag{16}$$

which implies an unbiased estimation of $\boldsymbol{w}^\star$ in the case of an uncorrelated input signal, since in this case $\boldsymbol{R}_{\overline{x}} = \boldsymbol{0}$ and $\boldsymbol{b}^{(\text{IA},1)} = \boldsymbol{0}$. When the input signal is colored, the coefficient vector converge in the mean to

$$\mathbb{E}\left[\boldsymbol{w}(\infty)\right] = \boldsymbol{w}^\star + \boldsymbol{R}_x^{-1}\boldsymbol{R}_{\overline{x}}\overline{\boldsymbol{w}}^\star, \tag{17}$$

which consists of the first N elements of the unknown impulse response $\boldsymbol{w}^\star$ the adaptive filter intends to emulate plus a perturbation term.

8

3.2. Mean-Square Convergence

The previous first-order analysis presents a restricted significance in terms of stability, since it is widely known that stable-in-the-mean adaptive filters can diverge in practice due to an unbounded variance of the weight vector [33]. Such a fact demands a second-order statistical analysis, in order to derive a theoretical model for the elements of deviation autocorrelation matrix $\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right]$. In order to accomplish such a task, one may multiply (12) by its transpose, which leads to

$$
\begin{aligned}
\tilde{\boldsymbol{w}}(k+1)\tilde{\boldsymbol{w}}^T(k+1) \quad = \quad & \tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k) - \beta\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k) - \beta\tilde{\boldsymbol{w}}(k)\left(\overline{\boldsymbol{w}}^\star\right)^T\overline{\boldsymbol{x}}(k)\boldsymbol{x}^T(k) \\
& -\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}(k) + \beta^2\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k) \\
& +\beta^2\boldsymbol{x}(k)\overline{\boldsymbol{x}}(k)\overline{\boldsymbol{w}}^\star\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k) + \beta^2\boldsymbol{x}(k)\boldsymbol{x}^T(k)\nu^2(k) \qquad (18) \\
& +\beta^2\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\left(\overline{\boldsymbol{w}}^\star\right)^T\overline{\boldsymbol{x}}(k)\boldsymbol{x}^T(k) - \beta\boldsymbol{x}(k)\overline{\boldsymbol{x}}^T(k)\overline{\boldsymbol{w}}^\star\tilde{\boldsymbol{w}}^T(k) \\
& +\beta^2\boldsymbol{x}(k)\overline{\boldsymbol{x}}^T(k)\overline{\boldsymbol{w}}^\star\left(\overline{\boldsymbol{w}}\right)^T\overline{\boldsymbol{x}}(k)\boldsymbol{x}^T(k) + \mathcal{O}[\nu(k)],
\end{aligned}
$$

where $\mathcal{O}[\nu(k)]$ contains first-order noise related terms. A popular approach for performing mean-square analyses consists of deriving a recursion of the elements of matrix $\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) \triangleq \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right]$ or, alternatively, by constructing recursive equations that update the vector $\boldsymbol{v}(k) \triangleq \mathbb{E}\left\{\text{vec}\left[\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right]\right\}$, where $\text{vec}(\boldsymbol{X})$ is an operator (whose output is a column vector) that stacks the columns of $\boldsymbol{X}$. Consider $\boldsymbol{A}\otimes\boldsymbol{B}$ as the Kronecker product between matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. In the case of a white input signal[2] and using the identity $\text{vec}[\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Z}] = \left(\boldsymbol{Z}^T \otimes \boldsymbol{X}\right)\text{vec}(\boldsymbol{Y})$, the classical analysis (i.e., a stochastic model that combines `IA` and `NA`) generates the following recursion:

$$
\boldsymbol{y}^{(\text{IA},2)}(k+1) = \boldsymbol{A}^{(\text{IA},2)}\boldsymbol{y}^{(\text{IA},2)}(k+1) + \boldsymbol{b}^{(\text{IA},2)}, \qquad (19)
$$

where $\boldsymbol{y}^{(\text{IA},2)}(k) \triangleq \boldsymbol{v}(k)$, $\boldsymbol{b}^{(\text{IA},2)} \triangleq \beta^2\sigma_\nu^2\mathbb{E}\left\{\text{vec}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\right\}$ (where σ_ν^2 is the additive noise variance), with the transition matrix $\boldsymbol{A}^{(\text{IA},2)}$ described as

$$
\boldsymbol{A}^{(\text{IA},2)} \triangleq \boldsymbol{I} - \beta\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k) \otimes \boldsymbol{I}\right] - \beta\mathbb{E}\left[\boldsymbol{I} \otimes \boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] + \beta^2\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k) \otimes \boldsymbol{x}(k)\boldsymbol{x}^T(k)\right].
$$

$$(20)$$

Note that it is possible to infer from (19) a closed-form estimate for steady-state regime similar to the one presented in (16). Furthermore, if the absolute

---

[2]For the colored input signal configuration, see [23].

eigenvalues values of matrix are $\boldsymbol{A}^{(\texttt{IA},2)}$ are upper bounded by the unity, the classical analysis predicts second-order stability, a much more informative criterion than convergence in the mean. It is worth noting that the designer has partial control on these eigenvalues, since matrix $\boldsymbol{A}^{(\texttt{IA},2)}$ depends both on the step size β as on the filter length as well. The observed discrepancy between performance predictions derived from (19) and empirical results for large step sizes can be minored by the employment of the exact expectation analysis, which is the focus of the next section.

4. Exact Expectation Analysis

Due to their assumptions, classical stochastic analyses focus on second-order characteristics of the excitation signal, especially on the eigenvalues spread of the autocorrelation matrix $\boldsymbol{R}_x$ and on its trace. When $\texttt{IA}$ is not employed, joint moments between input signal samples and adaptive coefficients should be taken into account, which incorporates more statistical information into the analysis. Consider in the following that the input $x(k)$ is a finite-time-correlated stationary signal generated through an M-th order moving average process:

$$x(k) = \sum_{m=0}^{M-1} b_m u(k-m), \tag{21}$$

where $u(k)$ is a white stationary signal that presents an even-symmetric distribution. The usage of model (21) implies that the proposed stochastic analysis is not restricted neither to a white nor to a Gaussian input signal $x(k)$, which are common limitations of most analyses (e.g., [48, 49, 50, 51, 52, 53]).

The exact expectation analysis systematically employs symbolic manipulations of mathematical expressions in order to construct a set of linear update equations that describe the dynamics of the statistical quantities of interest. Unfortunately, the derivation of the recursion for a specific state variable may generate new terms, which by themselves will require the generation of new recursions. The construction procedure eventually halts if the input data presents a finite-time correlation, which is guaranteed by (21) [26].

10

In order to illustrate the identification process of the stochastic state variables performed by the procedure, consider in the following the configuration $N = P = 1$, and $M = 2$. It is noteworthy to emphasize that this example showcases the ability of the exact expectation analysis technique to model non-white input signals. Both mean and mean-square exact expectation analyses are carried out for this particular setting.

4.1. Mean Convergence

Since in the considered case there is only one adaptive coefficient, recursion (12) degenerates into a scalar identity

$$
\begin{aligned}
\tilde{w}_0(k+1) \;=\; & \tilde{w}_0(k) - b_0^2 u^2(k)\tilde{w}_0(k)\beta - 2b_0 u(k)b_1 u(k-1)\tilde{w}_0(k)\beta \\
& - b_1^2 u^2(k-1)\tilde{w}_0(k)\beta - b_0^2 u(k-1)\overline{w}_0^\star \beta u(k) \\
& - b_0 u^2(k-1)\overline{w}_0^\star \beta b_1 - b_1 u(k-2)\overline{w}_0^\star \beta b_0 u(k) \\
& - b_1^2 u(k-2)\overline{w}_0^\star \beta u(k-1) - b_0 u(k)a_0\nu(k)\beta \\
& - b_1 u(k-1)a_0\nu(k)\beta,
\end{aligned}
\tag{22}
$$

which is a difference equation that does not describe the desired average deviation weight behavior. To proceed further, it is necessary to apply the expectation operator in (22), which, combined with the employment of NA, leads to

$$
\mathbb{E}[\tilde{w}_0(k+1)] = (1 - b_0^2\beta\gamma_2)\mathbb{E}[\tilde{w}_0(k)] - b_1^2\beta\mathbb{E}[u^2(k-1)\tilde{w}_0(k)] - \beta\overline{w}_0^\star\gamma_2 b_0 b_1, \tag{23}
$$

where $\gamma_n \triangleq \mathbb{E}\left[u^n(k)\right]$ and the expected product between the weight error coefficient and input data is not approximated as

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

because IA is no longer assumed to be valid. Due to this fact, recursion (23) is not self-contained, due to the emergence of the state variable $\mathbb{E}[u^2(k-1)\tilde{w}_0(k)]$, a nuisance term that requires by itself a specific recursion. This new parameter is termed as a nuisance element because we are not primarily interested in it (at

11

least in this first-order analysis), even though its estimation is a necessary step to the update of the statistical quantity of interest [54]. Note that in more complex configurations, the nuisance parameters may compose the large majority of the state variables. The recursion of the term of (24) can be obtained by multiplying both sides of (22) by $u^2(k)$ before the application of operator $\mathbb{E}[\cdot]$, which gives rise to

$$\mathbb{E}[u^2(k)\tilde{w}_0(k+1)] = (\gamma_2 - b_0^2\beta\gamma_4)\mathbb{E}[\tilde{w}_0(k)] - b_1^2\beta\gamma_2\mathbb{E}[u^2(k-1)\tilde{w}_0(k)] - \beta\overline{w}_0^\star b_0 b_1 \gamma_2^2.\tag{25}$$

Since Eqs. (23) and (25) provide the recursions for all required statistical quantities, they may be used to construct a state space linear model for the convergence in the mean that does not employ IA:

$$\boldsymbol{y}^{(1)}(k+1) = \boldsymbol{A}^{(1)}\boldsymbol{y}^{(1)}(k) + \boldsymbol{b}^{(1)},\tag{26}$$

where

$$\boldsymbol{y}^{(1)}(k) \triangleq \begin{bmatrix} \mathbb{E}[\tilde{w}_0(k)] \\ \mathbb{E}[u^2(k-1)\tilde{w}_0(k)] \end{bmatrix},\tag{27}$$

$$\boldsymbol{A}^{(1)} \triangleq \begin{bmatrix} 1 - \beta b_0^2\gamma_2 & -\beta b_1^2 \\ \gamma_2 - \beta b_0^2\gamma_4 & -\beta b_1^2\gamma_2 \end{bmatrix},\tag{28}$$

$$\boldsymbol{b}^{(1)} \triangleq \begin{bmatrix} -\beta b_0 b_1\overline{w}_0^\star\gamma_2 \\ -\beta b_0 b_1\overline{w}_0^\star\gamma_2^2 \end{bmatrix}.\tag{29}$$

Since we assumed that the adaptive coefficients are initialized to zero, $\tilde{w}_0(0)$ is deterministic and $\boldsymbol{y}^{(1)}(0) = \begin{bmatrix} w_0^\star & w_0^\star\gamma_2 \end{bmatrix}^T$. As a result, recursion (26) allows the calculation of $\boldsymbol{y}^{(1)}(k)$ for all k values.

Generally, the eigenvalues of matrix $\boldsymbol{A}^{(1)}$ can be explicitly found:

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

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

where

$$\Delta = \beta^2\gamma_2^2(b_0^4 + b_1^4) + b_0^2 b_1^2 \beta^2(4\gamma_4 - 2\gamma_2^2) - 2\beta\gamma_2(b_0^2 - b_1^2) + 1. \qquad (32)$$

Note that a choice of β that ensures $|\lambda_n| < 1$ (for $n \in \{1, 2\}$) implies that the algorithm is stable in the mean (under the exact expectation sense). Assuming such a condition, in the steady state regime, vector $\boldsymbol{y}^{(1)}(k)$ can be obtained by

$$\lim_{k\to\infty} \boldsymbol{y}^{(1)}(k) = \left(\boldsymbol{I} - \boldsymbol{A}^{(1)}\right)^{-1} \boldsymbol{b}^{(1)}. \qquad (33)$$

Then, assuming stability, we can obtain the steady-state mean value of $w_0(k)$ using Eq. (33) as

$$\lim_{k\to\infty} \mathbb{E}[w_0(k)] = \frac{b_0 b_1 \overline{w}_0^\star \gamma_2}{\beta\, b_0{}^2 b_1{}^2 \gamma_2{}^2 + \left(b_0{}^2 + b_1{}^2\right)\gamma_2 - \beta\, b_0{}^2 b_1{}^2 \gamma_4} + w_0^\star. \qquad (34)$$

In this generic setup, $w_0(k)$ converges to a different value of $w_0^\star$. It is worth noting that when the input signal is white (i.e., $b_1 = 0$), it can be proved (see Eq. (34)) that the adaptive coefficient $w_0(k)$ converges in the mean at steady-state to $w_0^\star$, a result that coincides with the one derived by IA [23].

4.2. Mean-Square Convergence

The prediction of second-order statistics of the deviation coefficient is more demanding than the previous mean weight analysis, but it is necessary both for performance and for stability prediction purposes. In the considered setting and avoiding the simplifications imposed by IA, the MSE can be computed from

$$
\begin{aligned}
\mathbb{E}\left[e^2(k)\right] &= b_0^2\gamma_2\mathbb{E}[\tilde{w}_0^2(k)] + b_1^2\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] + 2b_1\overline{w}_0^\star b_0\mathbb{E}[u^2(k-1)\tilde{w}_0(k)] \\
&\quad + 2b_1^2\overline{w}_0^\star\mathbb{E}[u(k-1)u(k-2)\tilde{w}_0(k)] + (b_0^2 + b_1^2)\overline{w}_0^{\star 2}\gamma_2 + \sigma_\nu^2, \qquad (35)
\end{aligned}
$$

which requires four stochastic state variables (more than what is necessary in the sufficient-order case). The recursion of term $\tilde{w}_0^2(k)$ may be obtained by squaring both sides of (22). Since the result is lengthy, it is omitted here. The application of the operator $\mathbb{E}[\cdot]$ in this result permits one to establish the following identity:

13

$$\mathbb{E}[\tilde{w}_0^2(k+1)] = (b_0^4\beta^2\gamma_4 + 1 - 2b_0^2\gamma_2\beta)\mathbb{E}[\tilde{w}_0^2(k)]$$

$$+(6b_0^2\beta^2 b_1^2\gamma_2 - 2b_1^2\beta)\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)]$$

$$+(6b_0^3\beta^2\overline{w}_0^\star b_1\gamma_2 - 2b_0\overline{w}_0^\star\beta b_1)\mathbb{E}[u^2(k-1)\tilde{w}_0(k)]$$

$$+(6b_0^2\beta^2 b_1^2\overline{w}_0^\star\gamma_2 - 2b_1^2\overline{w}_0^\star\beta)\mathbb{E}[u(k-1)u(k-2)\tilde{w}_0(k)]$$

$$+b_1^4\beta^2\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)] + 2b_1^3\beta^2 b_0\overline{w}_0^\star\mathbb{E}[u^4(k-1)\tilde{w}_0(k)] \qquad (36)$$

$$+2b_1^4\beta^2\overline{w}_0^\star\mathbb{E}[u^3(k-1)u(k-2)\tilde{w}_0(k)] + b_0^4\overline{w}_0^{\star 2}\beta^2\gamma_2^2$$

$$+b_0^2\overline{w}_0^{\star 2}\beta^2 b_1^2\gamma_4 + b_1^2\overline{w}_0^{\star 2}\beta^2 b_0^2\gamma_2^2 + b_1^4\overline{w}_0^{\star 2}\beta^2\gamma_2^2 + b_0^2\sigma_\nu^2\beta^2\gamma_2 + b_1^2\sigma_\nu^2\beta^2\gamma_2.$$

Eq. (36) introduces new state variables, whose recursion should be derived. Multiplying the square of $\tilde{w}_0(k+1)$ (see (22)) by judiciously chosen terms (such as performed in the derivation of the recursion of $\mathbb{E}[u^2(k-1)\tilde{w}_0(k)]$ in Eq. (25)) and applying the expectation operator, one may derive the following relationships:

$$\mathbb{E}[u^2(k)\tilde{w}_0^2(k+1)] = (b_0^4\beta^2\gamma_6 + \gamma_2 - 2b_0^2\gamma_4\beta)\mathbb{E}[\tilde{w}_0^2(k)]$$

$$+(6b_0^2\beta^2 b_1^2\gamma_4 - 2b_1^2\beta\gamma_2)\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)]$$

$$+(6b_0^3\beta^2\overline{w}_0^\star b_1\gamma_4 - 2b_0\overline{w}_0^\star\beta b_1\gamma_2)\mathbb{E}[u^2(k-1)\tilde{w}_0(k)]$$

$$+(6b_0^2\beta^2 b_1^2\overline{w}_0^\star\gamma_4 - 2b_1^2\overline{w}_0^\star\beta\gamma_2)\mathbb{E}[u(k-1)u(k-2)\tilde{w}_0(k)]$$

$$+b_1^4\beta^2\gamma_2\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)] + 2b_1^3\beta^2 b_0\overline{w}_0^\star\gamma_2\mathbb{E}[u^4(k-1)\tilde{w}_0(k)]$$

$$+2b_1^4\beta^2\overline{w}_0^\star\gamma_2\mathbb{E}[u^3(k-1)u(k-2)\tilde{w}_0(k)] + b_0^4\overline{w}_0^{\star 2}\beta^2\gamma_4\gamma_2$$

$$+2b_0^2\overline{w}_0^{\star 2}\beta^2 b_1^2\gamma_2\gamma_4 + b_1^4\overline{w}_0^{\star 2}\beta^2\gamma_2^3 + b_0^2\sigma_\nu^2\beta^2\gamma_4 + b_1^2\sigma_\nu^2\beta^2\gamma_2^2, \, (37)$$

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

$$+(\gamma_2 - b_0^2\beta\gamma_4)\mathbb{E}[\tilde{w}_0(k)] - b_0\overline{w}_0^\star\beta b_1\gamma_2^2, \qquad (38)$$

$$\mathbb{E}[u(k)u(k-1)\tilde{w}_0(k+1)] = -2b_0 b_1\beta\gamma_2\mathbb{E}[u^2(k-1)\tilde{w}_0(k)] - b_0^2\overline{w}_0^\star\beta\gamma_2^2, \qquad (39)$$

14

$$
\begin{aligned}
\mathbb{E}[u^4(k)\tilde{w}_0^2(k+1)] \;=\; & (b_0^4\beta^2\gamma_8 + \gamma_4 - 2b_0^2\gamma_6\beta)\mathbb{E}[\tilde{w}_0^2(k)] \\
& + (6b_0^2\beta^2 b_1^2\gamma_6 - 2b_1^2\beta\gamma_4)\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\
& + (6b_0^3\beta^2\overline{w}_0^\star b_1\gamma_6 - 2b_0\overline{w}_0^\star\beta b_1\gamma_4)\mathbb{E}[u^2(k-1)\tilde{w}_0(k)] \\
& + (6b_0^2\beta^2 b_1^2\overline{w}_0^\star\gamma_6 - 2b_1^2\overline{w}_0^\star\beta\gamma_4)\mathbb{E}[u(k-1)u(k-2)\tilde{w}_0(k)] \\
& + b_1^4\beta^2\gamma_4\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)] \\
& + 2b_1^3\beta^2 b_0\overline{w}_0^\star\gamma_4\mathbb{E}[u^4(k-1)\tilde{w}_0(k)] \\
& + 2b_1^4\beta^2\overline{w}_0^\star\gamma_4\mathbb{E}[u^3(k-1)u(k-2)\tilde{w}_0(k)] \\
& + b_0^4\overline{w}_0^{\star\,2}\beta^2\gamma_6\gamma_2 + b_0^2\overline{w}_0^{\star\,2}\beta^2 b_1^2\gamma_4^2 + b_1^2\overline{w}_0^{\star\,2}\beta^2 b_0^2\gamma_6\gamma_2 \\
& + b_1^4\overline{w}_0^{\star\,2}\beta^2\gamma_2^2\gamma_4 + b_0^2\sigma_\nu^2\beta^2\gamma_6 + b_1^2\sigma_\nu^2\beta^2\gamma_4\gamma_2,
\end{aligned}
\tag{40}
$$

$$
\begin{aligned}
\mathbb{E}[u^4(k)\tilde{w}_0(k+1)] \;=\; & -b_1^2\beta\gamma_4\mathbb{E}[u^2(k-1)\tilde{w}_0(k)] \\
& + (\gamma_4 - b_0^2\beta\gamma_6)\mathbb{E}[\tilde{w}_0(k)] - b_0\overline{w}_0^\star\beta b_1\gamma_4\gamma_2,
\end{aligned}
\tag{41}
$$

$$
\mathbb{E}[u^3(k)u(k-1)\tilde{w}_0(k+1)] = -2b_0 b_1\beta\gamma_4\mathbb{E}[u^2(k-1)\tilde{w}_0(k)] - b_0^2\overline{w}_0^\star\beta\gamma_4\gamma_2.
\tag{42}
$$

From initial values of state variable quantities, Eqs. (23) and (36)-(42) characterize the mean square learning behavior of the LMS, which can be concisely described by a state equations system

$$
\boldsymbol{y}^{(2)}(k+1) = \boldsymbol{A}^{(2)}\boldsymbol{y}^{(2)}(k) + \boldsymbol{b}^{(2)},
\tag{43}
$$

where $\boldsymbol{A}^{(2)}$ is a sparse transition matrix (please refer to Appendix A) with

dimensions 8×8 responsible for updating the state vector $\boldsymbol{y}^{(2)}(k)$

$$\boldsymbol{y}^{(2)}(k) \triangleq \begin{bmatrix} \mathbb{E}[\tilde{w}_0^2(k)] \\ \mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\ \mathbb{E}[u^2(k-1)\tilde{w}_0(k)] \\ \mathbb{E}[\tilde{w}_0(k)] \\ \mathbb{E}[u(k-1)u(k-2)\tilde{w}_0(k)] \\ \mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)] \\ \mathbb{E}[u^4(k-1)\tilde{w}_0(k)] \\ \mathbb{E}[u^3(k-1)u(k-2)\tilde{w}_0(k)] \end{bmatrix}, \tag{44}$$

which also contains all statistical quantities of interest of the first-order analysis (26) and

$$\boldsymbol{b}^{(2)} =$$

$$\begin{bmatrix} b_0^4\overline{w}_0^{\star\,2}\beta^2\gamma_2^2 + b_0^2\overline{w}_0^{\star\,2}\beta^2 b_1^2\gamma_4 + b_1^2\overline{w}_0^{\star\,2}\beta^2 b_0^2\gamma_2^2 + b_1^4\overline{w}_0^{\star\,2}\beta^2\gamma_2^2 + b_0^2\sigma_2\beta^2\gamma_2 + b_1^2\sigma_2\beta^2\gamma_2 \\ b_0^4\overline{w}_0^{\star\,2}\beta^2\gamma_4\gamma_2 + 2b_0^2\overline{w}_0^{\star\,2}\beta^2 b_1^2\gamma_2\gamma_4 + b_1^4\overline{w}_0^{\star\,2}\beta^2\gamma_2^3 + b_0^2\sigma_2\beta^2\gamma_4 + b_1^2\sigma_2\beta^2\gamma_2^2 \\ -b_0\overline{w}_0^{\star}\beta b_1\gamma_2^2 \\ -b_0\overline{w}_0^{\star}\beta b_1\gamma_2 \\ -b_0^2\overline{w}_0^{\star}\beta\gamma_2^2 \\ b_0^4\overline{w}_0^{\star\,2}\beta^2\gamma_6\gamma_2 + b_0^2\overline{w}_0^{\star\,2}\beta^2 b_1^2\gamma_4^2 + b_1^2\overline{w}_0^{\star\,2}\beta^2 b_0^2\gamma_6\gamma_2 + b_1^4\overline{w}_0^{\star\,2}\beta^2\gamma_2^2\gamma_4 + b_0^2\sigma_2\beta^2\gamma_6 + b_1^2\sigma_2\beta^2\gamma_4\gamma_2 \\ -b_0\overline{w}_0^{\star}\beta b_1\gamma_4\gamma_2 \\ -b_0^2\overline{w}_0^{\star}\beta\gamma_4\gamma_2 \end{bmatrix} . \tag{45}$$

Similarly to the first-order analysis, in order to calculate recursion (43), terms such as $\mathbb{E}[u^4(k-1)\tilde{w}_0(k)]$ and $\mathbb{E}[u(k-1)u(k-2)\tilde{w}_0(k)]$ are initialized to, respectively, $w_0^{\star}\gamma_4$ and 0.

Model (43) summarizes the second-order learning behavior of the deficient-length LMS algorithm, and predicts its mean square convergence if the maximum absolute eigenvalue $|\lambda_{\max}|$ of matrix $\boldsymbol{A}^{(2)}$ is less than unity. Note that $|\lambda_{\max}|$ depends on the adjustable step size β and can be efficiently computed using the power method [55]. Since model (43) takes into account the shift-structure of the excitation data, it may provide a more accurate step size bound that guarantees convergence, especially when the input signal is colored or is

16

distributed according to a "heavy-tailed" probability density function [26]. Additionally, the stationary operation point (i.e., the steady-state regime) can be computed by a closed-form equation:

$$\lim_{k \to \infty} \boldsymbol{y}^{(2)}(k) = \left(\boldsymbol{I} - \boldsymbol{A}^{(2)}\right)^{-1} \boldsymbol{b}^{(2)}. \tag{46}$$

The following section describes the computational framework developed to perform the advanced exact expectation analysis. Comparisons against `IA`-based predictions and empirical results are also provided.

5. Results

In this manuscript we chose to develop a C++ code responsible for generating the required equations of the exact expectation analysis and for simulating the empirical MSE evolution. Such an option requires a labor-intensive development but allows for significant performance gains. Some of the most tangible advantages include the ability to generate (i) millions of equations, which permits one to model more complex configurations; and (ii) a high number of Monte Carlo trials (e.g., 10^9), usually required for computing empirical learning curves. The use of symbolic/numerical algebra softwares such as Maple™ would impose drastic reductions on the previous parameters.

Assuming that the resulting state space equation can be stored in computer memory, the C++ code automatically performs the required algebraic symbolic operations for generic configurations, i.e., arbitrary values of N, M, and P. For instance, the code can be employed to generate over 8×10^6 equations for the first-order exact expectation analysis (for the configuration $N = 12$ and $M = P = 1$). Table 1 presents the number of recursive equations required for generating the exact mean-weight behavior as a function of different values of N and M. Note that the number of equations of this analysis rapidly increases with N and M and remains unaltered as a function of P. Table 2 shows the number of state variables required for the second-order exact expectation analysis.

The additive noise is assumed to be white Gaussian with variance $\sigma_\nu^2 = 0.01$. The input signal is colored, obtained by filtering a white signal $u(k)$ by the

N	M	# Eqs.	N	M	# Eqs.
1	1	1	3	5	10928
1	2	2	3	6	61178
1	3	7	4	1	50
1	4	31	4	2	451
1	5	152	4	3	2505
1	6	790	4	4	13859
1	7	4271	4	5	77997
1	8	23767	5	1	217
1	9	135221	5	2	2766
2	1	3	5	3	16332
2	2	12	5	4	93561
2	3	55	6	1	954
2	4	273	6	2	17060
2	5	1428	6	3	105927
2	6	7752	7	1	4245
2	7	43263	7	2	105848
3	1	12	8	1	19085
3	2	74	9	1	86528
3	3	379	10	1	395066
3	4	2003	11	1	8373252

Table 1: Number of state equations of the first-order exact expectation analysis.

N	M	P	# Eqs.	N	M	P	# Eqs.
1	5	8	10202	3	3	7	63197
1	6	1	33752	3	3	8	69927
1	6	2	42412	4	2	1	30468
1	6	3	51072	4	2	2	39840
1	6	4	59732	4	2	3	49212
1	6	5	68392	4	2	4	58584
1	6	6	77052	4	2	5	67956
1	6	7	85712	4	2	6	77328
1	6	8	94372	4	2	7	86700
1	7	1	327868	4	2	8	96072
1	7	2	411310	5	1	2	13863
1	7	3	494752	5	1	3	18091
2	4	1	13091	5	1	4	22319
2	4	2	16995	5	1	5	26547
2	4	3	20899	5	1	6	30775
2	4	4	24803	5	1	7	35003
2	4	5	28707	5	1	8	39231
2	4	6	32611	6	1	1	87099
2	4	7	36515	6	1	2	125018
2	4	8	40419	6	1	3	162810
2	5	1	123642	6	1	4	200602
3	3	1	22817	6	1	5	238394
3	3	2	29547	6	1	6	276186
3	3	3	36277	6	1	7	313978
3	3	4	43007	6	1	8	351770
3	3	5	49737	7	1	1	801096
3	3	6	56467	7	1	2	1148761

Table 2: Number of state equations of the second-order exact expectation analysis. Due to lack of space, only setups that yield more then 10^4 state equations are considered.
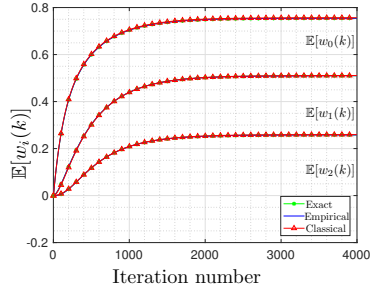
(a) Configuration 1 with $\beta = 0.004$ and an input signal obtained by filtering a unitary-variance white Gaussian signal by the transfer function $B(z)$.
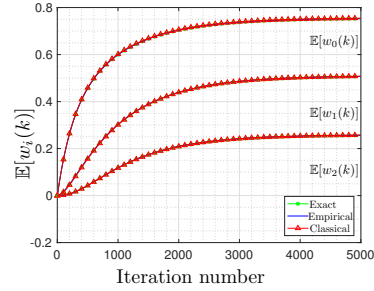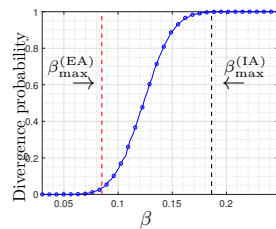
(b) Configuration 1 with $\beta = 0.002$ and an input signal obtained by filtering a unitary-variance white Laplacian signal by the transfer function $B(z)$.

Figure 1: Mean-weight behavior of the adaptive coefficients for the configuration $(N, M, P) = (3, 2, 2)$ as a function of the number of iterations for relatively small β values.

transfer function $B(z) = 1 - 0.9z^{-1}$. The ideal transfer function has two possible configurations, namely:

⋆ Configuration 1

$$w_i^\star = 1 \text{ for } i \in \{0, 1, \cdots, N + P - 1\}$$

⋆ Configuration 2

$$w_i^\star = \begin{cases} 1, & \text{for } i \in \{0, 1, \cdots, N - 1\} \\ 0.01, & \text{for } i \in \{N, N + 1, \cdots, N + P - 1\} \end{cases}$$

Note that Configuration 1 depicts a more challenging undermodeled setting, related to a worse steady-state mean square error of the adaptive filter.

The remainder of this section is organized as follows: Section 5.1 presents the results for the first-order analysis; Section 5.2 describes the data gathered for analyzing stability; and Section 5.3 describes the transient and steady-state analyses.

5.1. First-Order Analysis

For small β values the theoretical performance curve obtained from the classical analysis is close to the empirical one. This fact is widely recognized in the literature, since a slower dynamic of the adaptive weights reduces their statistical dependence on recent input samples [56]. It is noteworthy that [13] employs the matrix perturbation theory in order to show that the actual performance of an stochastic adaptive filter converges with the one predicted by IA when β is small. In this configuration, it should be clear that the proposed analysis also provides an accurate prediction for the first-order coefficients evolution. This behavior is illustrated in Figure 1, which presents the mean-weight behavior for the adaptive coefficients when using Configuration 1. The number of independent Monte Carlo trials employed was 10^6.

However, for bigger β values the classical and exact curves diverge, as is exemplified in Figure 2, where the β value for Figure 2a is 0.08 and 0.035 for Figure 2b. The rest of the parameters remain equal. It is important to mention that the proposed analysis adheres well to the simulated curve.
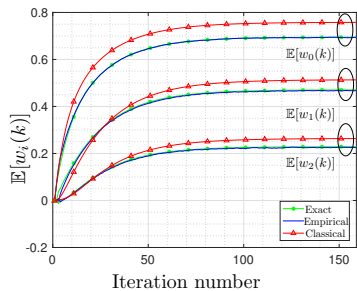
(a) Configuration 1 with $\beta = 0.08$ and an input signal obtained by filtering a unitary-variance white Gaussian signal by the transfer function $B(z)$.
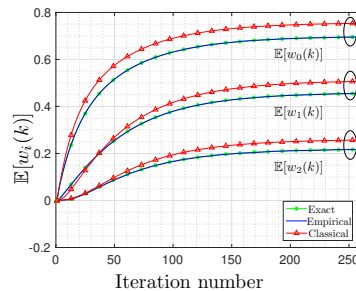
(b) Configuration 1 with $\beta = 0.035$ and an input signal obtained by filtering a unitary-variance white Laplacian signal by the transfer function $B(z)$.

Figure 2: Mean-weight behavior of the adaptive coefficients for the configuration $(N, M, P) = (3, 2, 2)$ as a function of the number of iterations for bigger β values.

5.2. Stability Analysis

After the computation of transition matrices $\boldsymbol{A}^{(\mathtt{IA},2)}$ and $\boldsymbol{A}^{(2)}$, it is possible to theoretically obtain the upper bound value of parameter β that ensures stable operation (i.e., one guaranteeing that the maximum absolute eigenvalue of the transition matrix is upper bounded by the unity).

For this specific set of results we counted a realization as divergent if the absolute value of any adaptive coefficient surpasses 10 (i.e., if there exists at least a single k for which $|w_i(k)| > 10$, for $i \in \{0,\ldots,N-1\}$ we consider that the trial being evaluated has diverged). Figure 3 presents the results for Configuration 1 when distinct input signal distributions are employed. Using the standard analysis, the state space model (19) predicts stability when the step size is below $\beta_{\max}^{(\mathtt{IA})}$. This upper bound has value 0.186279 for Fig. 3a and 0.129639 for Fig. 3b. The advanced exact expectation analysis, in its turn, provides tighter upper bounds $\beta_{\max}^{(\mathtt{EA})}$: respectively, 0.0850143 and 0.0398865, for Figures 3a and 3b. Note that the exact analysis accurately indicates a range for the values of β that guarantees a negligible probability of divergence.

(a) Configuration 1 and an input signal obtained by filtering a unitary-variance white Gaussian signal by the transfer function $B(z)$.
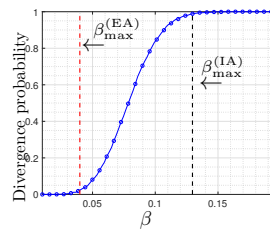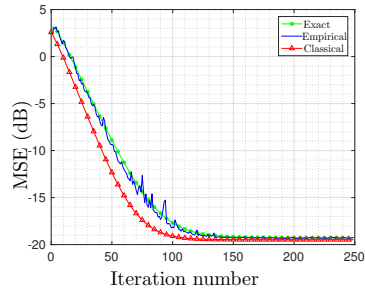
(b) Configuration 1 and an input signal obtained by filtering a unitary-variance white Laplacian signal by the transfer function $B(z)$.

Figure 3: Divergence probability for $(N, M, P) = (3, 2, 2)$ as a function of β for 10^5 Monte Carlo trials and with 1000 iterations for each realization.
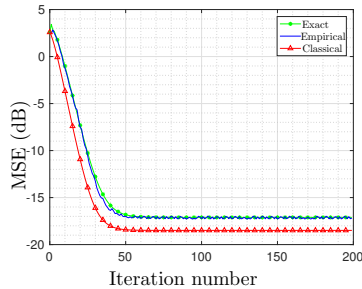
(a) The input signal was obtained by filtering a unitary-variance white Gaussian signal by the transfer function $B(z)$ and $\beta = 0.115$.

(b) The input signal was obtained by filtering a unitary-variance white Laplacian signal by the transfer function $B(z)$ and $\beta = 0.045$.

Figure 4: MSE evolution (in dB) with $(N, M, P) = (2, 2, 2)$ as a function of the number of iterations. The ideal transfer function was chosen according to Configuration 2.

5.3. Transient and Steady-State Analysis

Figure 4 illustrates the results obtained for the MSE evolution for different input signal distributions (with 10^9 independent Monte Carlo trials). The ideal transfer function utilized was the one described in Configuration 2. The MSE evolution for a more complex setup is depicted in Figure 5. It is also important to reemphasize that our model also presents better adherence to the empirical simulations, and that some discrepancies may occur due to the usage of a finite number of Monte Carlo trials, as elucidated in [57].

Figure 6 describes the data collected for the steady-state MSE when employing the transfer function described in Configuration 2. Again, our model shows similar behavior to the empirical data. As expected, for small β values the three curves coincide. Notice also that for bigger β's the classical model underestimates the MSE. The number of Monte Carlo trials was 10^9 for each individual data point of both input signal distributions.

6. Conclusions

In this paper, a theoretical stochastic model that avoids the high-level statistical description performed by classical analyses of the LMS algorithm is
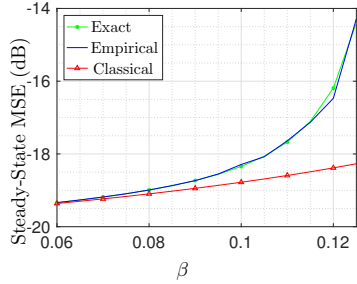
23
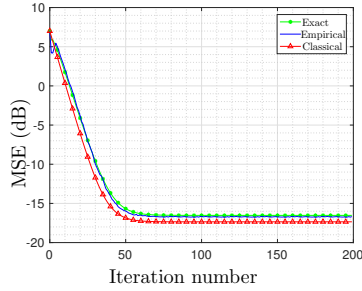
(a) The input signal was obtained by filtering a unitary-variance white Gaussian signal and $\beta = 0.145$ with tuple $(N, M, P) = (5, 1, 6)$.
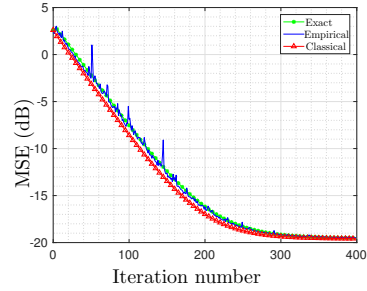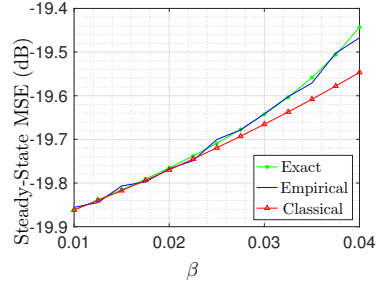

(b) The input signal was obtained by filtering a unitary-variance white Laplacian signal by the transfer function $B(z)$ and $\beta = 0.025$ with $(N, M, P) = (3, 2, 4)$.

Figure 5: MSE evolution (in dB) as a function of the number of iterations. The ideal transfer function was chosen according to Configuration 2.

(a) The input signal was obtained by filtering a unitary-variance white Gaussian signal by the transfer function $B(z)$.

(b) The input signal was obtained by filtering a unitary-variance white Laplacian signal by the transfer function $B(z)$.

Figure 6: Steady-state MSE (in dB) for the configuration $(N, M, P) = (2, 2, 2)$ as a function of β. The ideal transfer function was chosen according to Configuration 2.

advanced. The proposed analysis predicts both learning behavior and stability operation more accurately than state-of-the-art approaches that employ the ubiquitous independence assumption, and is not restricted neither to white nor to Gaussian input signal distributions.

In general, the advanced method demands a high computational burden,

except for small-size adaptive filters. Furthermore, we suspect that some statistical couplings are more significant than others. This implies that some random variables could be considered as independent, whereas the statistical dependence between others should be accounted for. We expect that such a strategy may provide a "semi-exact" expectation analysis able to address more complex configurations. Please note that this alternative could offer a reasonable trade-off between the simpler standard approach (i.e., one based on `IA`) and the complexity of the exact expectation analysis method.

The devised model is tailored for configurations in which a large step size is adopted, a crucial setting for applications where faster convergence is required. In this case, a high β value should be adopted and the standard analysis method becomes at best unreliable (since its predictions are very distant from the reality) or even catastrophic, since it may predict stability when in fact the adaptive filter may diverge most of the time. In such a situation, there is no trade-off between the alternatives, because only the exact expectation analysis method will provide accurate results.

References

[1] A. H. Sayed, Adaptive filters, John Wiley & Sons, 2011.

[2] B. Widrow, P. E. Mantey, L. J. Griffiths, B. B. Goode, Adaptive antenna systems, Proceedings of the IEEE 55 (12) (1967) 2143–2159.

[3] B. Hassibi, A. H. Sayed, T. Kailath, $\mathcal{H}^\infty$ optimality of the LMS algorithm, IEEE Transactions on Signal Processing 44 (2) (1996) 267–280.

[4] W. Liu, P. P. Pokharel, J. C. Principe, The kernel least-mean-square algorithm, IEEE Transactions on Signal Processing 56 (2) (2008) 543–554.

[5] K. Matsubara, K. Nishikawa, H. Kiya, Pipelined LMS adaptive filter using a new look-ahead transformation, IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing 46 (1) (1999) 51–55.

[6] P. K. Meher, S. Y. Park, Critical-path analysis and low-complexity implementation of the LMS adaptive algorithm, IEEE Transactions on Circuits and Systems I: Regular Papers 61 (3) (2014) 778–788.

[7] M. Reuter, K. Quirk, J. Zeidler, L. Milstein, Non-linear effects in LMS adaptive filters, in: Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373), 2000, pp. 141–146.

[8] A. H. Sayed, M. Rupp, Error-energy bounds for adaptive gradient algorithms, IEEE Transactions on Signal Processing 44 (8) (1996) 1982–1989.

[9] S. Haykin, B. Widrow, Least-mean-square adaptive filters, Vol. 31, John Wiley & Sons, Hoboken, New Jersey, 2003.

[10] D. B. Haddad, M. R. Petraglia, Transient and steady-state MSE analysis of the IMPNLMS algorithm, Digital Signal Processing 33 (2014) 50 – 59.

[11] B. Widrow, J. M. McCool, M. G. Larimore, C. R. Johnson, Stationary and nonstationary learning characteristics of the LMS adaptive filter, Proceedings of the IEEE 64 (8) (1976) 1151–1162.

[12] A. Kapur, M. K. Varanasi, D. Das, Noncoherent MMSE multiuser receivers and their blind adaptive implementations, IEEE Transactions on Communications 50 (3) (2002) 503–513.

[13] J. E. Mazo, On the independence theory of equalizer convergence, The Bell System Technical Journal 58 (5) (1979) 963–993.

[14] L. Dogariu, S. Ciochina, C. Paleologu, On the properties of the system mismatch covariance matrix in the LMS adaptive algorithm, in: 2018 International Conference on Communications, 2018, pp. 39–44.

[15] O. J. Tobias, R. Seara, Stochastic analysis for the leaky delayed LMS algorithm: A new model without invoking the independence assumption, in: 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, 2002, pp. II–1093–II–1096.

[16] M. Wu, J. Yang, A step size control method for deficient length FBLMS algorithm, IEEE Signal Processing Letters 22 (9) (2015) 1448–1451.

[17] S. J. M. de Almeida, M. H. Costa, J. C. M. Bermudez, A stochastic model for the deficient length pseudo affine projection adaptive algorithm, in: 2009 17th European Signal Processing Conference, 2009, pp. 1715–1719.

[18] J. Homer, R. R. Bitmead, I. Mareels, Quantifying the effects of dimension on the convergence rate of the LMS adaptive FIR estimator, IEEE Transactions on Signal Processing 46 (10) (1998) 2611–2615.

[19] M. R. Petraglia, D. B. Haddad, New adaptive algorithms for identification of sparse impulse responses - analysis and comparisons, in: 2010 7th International Symposium on Wireless Communication Systems, 2010, pp. 384–388.

[20] Y. Gu, K. Tang, H. Cui, W. Du, Convergence analysis of a deficient-length LMS filter and optimal-length sequence to model exponential decay impulse response, IEEE Signal Processing Letters 10 (1) (2003) 4–7.

[21] Y. Zhang, J. A. Chambers, Convex combination of adaptive filters for a variable tap-length LMS algorithm, IEEE Signal Processing Letters 13 (10) (2006) 628–631.

[22] D. Xu, B. Yin, W. Wang, W. Zhu, Variable tap-length LMS algorithm based on adaptive parameters for tdl structure adaption, IEEE Signal Processing Letters 21 (7) (2014) 809–813.

[23] K. Mayyas, Performance analysis of the deficient length LMS adaptive algorithm, IEEE Transactions on Signal Processing 53 (8) (2005) 2727–2734.

[24] Y. Xia, S. C. Douglas, D. P. Mandic, Performance analysis of the deficient length augmented CLMS algorithm for second order noncircular complex signals, Signal Processing 144 (2018) 214 – 225.

[25] C. Paleologu, S. Ciochina, J. Benesty, Variable step-size NLMS algorithm for under-modeling acoustic echo cancellation, IEEE Signal Processing Letters 15 (2008) 5–8.

[26] S. C. Douglas, W. Pan, Exact expectation analysis of the LMS adaptive filter, IEEE Transactions on Signal Processing 43 (12) (1995) 2863–2871.

[27] 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 vol.3.

[28] S. C. Douglas, T. H. . Meng, Exact expectation analysis of the LMS adaptive filter without the independence assumption, in: [Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 4, 1992, pp. 61–64 vol.4.

[29] 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, 1992, pp. 566–570 vol.1.

[30] P. Lara, F. Igreja, L. D. T. J. Tarrataca, D. B. Haddad, M. R. Petraglia, Exact expectation evaluation and design of variable step-size adaptive algorithms, IEEE Signal Processing Letters 26 (1) (2019) 74–78.

[31] P. Lara, K. da Silva Olinto, F. R. Petraglia, D. B. Haddad, Exact analysis of the LMS algorithm with coloured measurement noise, Electronics Letters 54 (24) (2018) 1401–1403.

[32] B. Widrow, E. Walach, On the statistical efficiency of the LMS algorithm with nonstationary inputs, IEEE Transactions on Information Theory 30 (2) (1984) 211–221.

[33] A. Feuer, E. Weinstein, Convergence analysis of LMS filters with uncorrelated gaussian data, IEEE Transactions on Acoustics, Speech, and Signal Processing 33 (1) (1985) 222–230.

[34] W.-P. Ang, B. Farhang-Boroujeny, A new class of gradient adaptive step-size LMS algorithms, IEEE Transactions on Signal Processing 49 (4) (2001) 805–810.

[35] O. Dabeer, E. Masry, Analysis of mean-square error and transient speed of the LMS adaptive algorithm, IEEE Transactions on Information Theory 48 (7) (2002) 1873–1894.

[36] D. G. Manolakis, V. K. Ingle, S. M. Kogon, Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering, and array processing, McGraw-Hill Boston, 2000.

[37] S. Haykin, Neural networks: a comprehensive foundation, Prentice Hall PTR, 1994.

[38] V. H. Nascimento, A. H. Sayed, Unbiased and stable leakage-based adaptive filters, IEEE Transactions on Signal Processing 47 (12) (1999) 3261–3276.

[39] S. S. Haykin, Adaptive filter theory, Pearson Education India, 2008.

[40] K. A. Lee, W. S. Gan, Improving convergence of the NLMS algorithm using constrained subband updates, IEEE Signal Processing Letters 11 (9) (2004) 736–739.

[41] D. B. Haddad, M. R. Petraglia, A. Petraglia, A unified approach for sparsity-aware and maximum correntropy adaptive filters, in: 2016 24th European Signal Processing Conference (EUSIPCO), 2016, pp. 170–174.

[42] M. R. Petraglia, D. B. Haddad, E. L. Marques, Affine projection subband adaptive filter with low computational complexity, IEEE Transactions on Circuits and Systems II: Express Briefs 63 (10) (2016) 989–993.

[43] L. Resende, R. Pimenta, N. Norat, D. B. Haddad, M. R. Petraglia, LMS algorithm with reuse of coefficients and robustness against impulsive noise, XXII Congresso Brasileiro de Automática, 2018, pp. 1–5.

[44] Z. Zhang, T. Bose, L. Xiao, R. Thamvichai, Performance analysis of the deficient length EDS adaptive algorithm, in: APCCAS 2006 - 2006 IEEE Asia Pacific Conference on Circuits and Systems, 2006, pp. 222–226.

[45] E. Eweda, Transient performance degradation of the LMS, RLS, sign, signed regressor, and sign-sign algorithms with data correlation, IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing 46 (8) (1999) 1055–1062.

[46] M. R. Petraglia, D. B. Haddad, Mean-square error and stability analysis of a subband structure for the rapid identification of sparse impulse responses, Digital Signal Processing 22 (6) (2012) 1068 – 1072.

[47] S. Marcos, O. Macchi, Tracking capability of the least mean square algorithm: Application to an asynchronous echo canceller, IEEE Transactions on Acoustics, Speech, and Signal Processing 35 (11) (1987) 1570–1578.

[48] E. Masry, F. Bullo, Convergence analysis of the sign algorithm for adaptive filtering, IEEE Transactions on Information Theory 41 (2) (1995) 489–495.

[49] G. Su, J. Jin, Y. Gu, J. Wang, Performance analysis of ℓ_0-norm constraint least mean square algorithm, IEEE Transactions on Signal Processing 60 (5) (2012) 2223–2235.

[50] M. H. Costa, J. C. M. Bermudez, N. J. Bershad, Statistical analysis of the LMS algorithm with a zero-memory nonlinearity after the adaptive filter, in: 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258), Vol. 3, 1999, pp. 1661–1664 vol.3.

[51] N. J. Bershad, J. C. M. Bermudez, J. . Tourneret, On performance bounds for an affine combination of two LMS adaptive filters, in: 2008 IEEE In-

ternational Conference on Acoustics, Speech and Signal Processing, 2008, pp. 3297–3300.

[52] E. Eweda, N. J. Bershad, Stochastic analysis of a stable normalized least mean fourth algorithm for adaptive noise canceling with a white gaussian reference, IEEE Transactions on Signal Processing 60 (12) (2012) 6235–6244.

[53] K. da S. Olinto, D. B. Haddad, M. R. Petraglia, Transient analysis of ℓ_0-LMS and ℓ_0-NLMS algorithms, Signal Processing 127 (2016) 217 – 226.

[54] J.-F. Cardoso, Blind signal separation: statistical principles, Proceedings of the IEEE 86 (10) (1998) 2009–2025.

[55] G. H. Golub, C. F. Van Loan, Matrix Computations (3rd Ed.), Johns Hopkins University Press, Baltimore, MD, USA, 1996.

[56] E. Eweda, Tracking analysis of the normalized lms algorithm without the independence and small step size assumptions, in: 2009 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2009, pp. 129–134.

[57] V. H. Nascimento, A. H. Sayed, On the learning mechanism of adaptive filters, IEEE Transactions on Signal Processing 48 (6) (2000) 1609–1625.

$$\mathbf{A}^{(2)} = \begin{bmatrix} b_0^4\beta^2\gamma_4 + 1 - 2b_0^2\gamma_2\beta & 6b_0^2\beta^2 b_1^2\gamma_2 - 2b_1^2\beta & 6b_0^3\beta^2\overline{w}_0^\star b_1\gamma_2 - 2b_0\overline{w}_0^\star\beta b_1 & 0 & 6b_0^2\beta^2 b_1^2\overline{w}_0^\star\gamma_2 - 2b_1^2\overline{w}_0^\star\beta & b_1^4\beta^2 & 2b_1^3\beta^2 b_0\overline{w}_0^\star & 2b_1^4\beta^2\overline{w}_0^\star \\[4pt] b_0^4\beta^2\gamma_6 + \gamma_2 - 2b_0^2\gamma_4\beta & 6b_0^2\beta^2 b_1^2\gamma_4 - 2b_1^2\beta\gamma_2 & 6b_0^3\beta^2\overline{w}_0^\star b_1\gamma_4 - 2b_0\overline{w}_0^\star\beta b_1\gamma_2 & 0 & 6b_0^2\beta^2 b_1^2\overline{w}_0^\star\gamma_4 - 2b_1^2\overline{w}_0^\star\beta\gamma_2 & b_1^4\beta^2\gamma_2 & 2b_1^3\beta^2 b_0\overline{w}_0^\star\gamma_2 & 2b_1^4\beta^2\overline{w}_0^\star\gamma_2 \\[4pt] 0 & 0 & -b_1^2\beta\gamma_2 & \gamma_2 - b_0^2\beta\gamma_4 & 0 & 0 & 0 & 0 \\[4pt] 0 & 0 & -b_1^2\beta & 1 - b_0^2\beta\gamma_2 & 0 & 0 & 0 & 0 \\[4pt] 0 & 0 & -2b_0 b_1\beta\gamma_2 & 0 & 0 & 0 & 0 & 0 \\[4pt] b_0^4\beta^2\gamma_8 + \gamma_4 - 2b_0^2\gamma_6\beta & 6b_0^2\beta^2 b_1^2\gamma_6 - 2b_1^2\beta\gamma_4 & 6b_0^3\beta^2\overline{w}_0^\star b_1\gamma_6 - 2b_0\overline{w}_0^\star\beta b_1\gamma_4 & 0 & 6b_0^2\beta^2 b_1^2\overline{w}_0^\star\gamma_6 - 2b_1^2\overline{w}_0^\star\beta\gamma_4 & b_1^4\beta^2\gamma_4 & 2b_1^3\beta^2 b_0\overline{w}_0^\star\gamma_4 & 2b_1^4\beta^2\overline{w}_0^\star\gamma_4 \\[4pt] 0 & 0 & -b_1^2\beta\gamma_4 & \gamma_4 - b_0^2\beta\gamma_6 & 0 & 0 & 0 & 0 \\[4pt] 0 & 0 & -2b_0 b_1\beta\gamma_4 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$