Transient Analysis of the Bias-Compensated LMS Algorithm

Rodrigo M. S. Pimenta, Newton N. Siqueira, Mariane R. Petraglia and Diego B. Haddad

Abstract—In most supervised adaptive filtering settings, only the additive noise of the reference signal is taken into account. However, in many practical situations the excitation data is also immersed in noise, which leads to a bias in the estimation procedure. In order to mitigate such issue, adaptive algorithms with bias compensation schemes have been proposed. This paper advances for the first time a stochastic model that predicts the average and mean-square learning behavior of the bias-compensated least mean squares algorithm in the transient region. Asymptotic predictions can also be obtained as a result of the devised analysis. Tracking capabilities and the impact of employing sub-optimal length adaptive filter are also considered, without restricting the input signal to be white. Results indicate that the proposed analysis reveals accurate agreement with simulation results.

Index Terms—Signal Processing, Adaptive Filtering, Transient Analysis, Bias-Compensated LMS.

I. INTRODUCTION

THE flexibility of adaptive filtering algorithms (AFAs) derives from the fact that they do not require prior statistical information about the involved signals. They also demand low computational effort, since they are restricted to the exploration of local data in order to feed the adaptive estimator. Such features make them appropriate for a plethora of important applications nowadays, such as system identification[1], temporal series prediction, adaptive control and active noise cancelling [1].

Recently, several papers have raised attention to the fact that the acquisition procedure of the input data is also prone to imperfections (such as quantization noise), so the input of the adaptive filter is also corrupted by noise signal. In this case, the emergence of a bias hampers the algorithm learning. In order to circumvent such an issue, bias-compensated schemes have been devised. The first of them proposed the BC-LMS (Bias-compensated LMS) [2], an algorithm obtained from a modification of the cost function in order to take into account the statistical properties of the input noise. An affine projection version of the BC-LMS was derived in [3]. The optimality in the mean square sense of the BC-LMS was

recently demonstrated in [4], which also proposes a closed-form equation for asymptotic performance for a white input, based on energy conservation arguments. The instability in the algorithm induced by the estimation of the input noise variance was addressed in [5].

This paper describes for the first time a comprehensive stochastic model that predicts the average learning behavior of the BC-LMS algorithm. The input signal is not restricted to be white, and both deficient-lenght and tracking configurations are addressed in a unified way. The paper presents the following structure: Section II describes the fundamentals of the BC-LMS algorithm. Section III provides an analysis in the mean sense, whereas Section IV offers a second-order (*i.e.*, mean square) theoretical analysis. Section V addresses the tracking setting, assuming a first-order Markovian perturbation on the ideal plant coefficients. Section VI models the BC-LMS learning behavior when the length of the adaptive filter is suboptimal. Section VII compares some simulated learning curves with the theoretical ones. At last, Section VIII contains the concluding remarks.

II. BC-LMS ALGORITHM

Usually, the design of supervised AFAs considers that only the reference signal $d(k) \in \mathbb{R}$ is noisy, that is:

$$d(k) = [\boldsymbol{w}^\star]^T \boldsymbol{u}(k) + \nu(k), \qquad (1)$$

where[2] $\boldsymbol{w}^\star \in \mathbb{R}^N$ contains the coefficients of the ideal (and unknown) transfer function, $\nu(k) \in \mathbb{R}$ denotes the additive noise and $\boldsymbol{u}(k) \triangleq \begin{bmatrix} u(k) & u(k-1) & \dots & u(k-N+1) \end{bmatrix}^T$ concatenates N successive samples of the input signal $u(k)$.

Considering that the measuring process of the input data is inaccurate, the input $x(k)$ of the adaptive filter is also corrupted by noise signal $\eta(k)$, that is,

$$x(k) = u(k) + \eta(k), \qquad (2)$$

where $\eta(k)$ is usually assumed to be a stationary process with zero-mean and variance σ_η^2. Fig. 1 depicts the points where noise is assumed to disturb the feedback process.

Most adaptive filters based on a tapped-delay structure utilizes the input vector

$$\boldsymbol{x}(k) \triangleq \begin{bmatrix} x(k) & x(k-1) & \dots & x(k-N+1) \end{bmatrix} \qquad (3)$$

to update the coefficient vector $\boldsymbol{w}(k) \in \mathbb{R}^N$, where the correction term is usually proportional to the step size $\beta \in \mathbb{R}_+$, whose choice establishes the well-known trade-off between

[1]This paper focuses on system identification tasks.

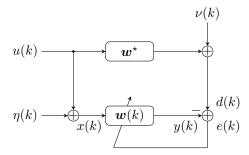[2]In this paper, all vectors are of column-type.

Fig. 1: Block diagram of an adaptive system identification, where the adaptive filter input is corrupted by noise $\eta(k)$.

convergence rate and asymptotic performance. In the supervised context assumed by this paper, these algorithms also employ error signal $e(k) \in \mathbb{R}$, which is computed through

$$e(k) \triangleq d(k) - y(k) = d(k) - \boldsymbol{w}^T(k)\boldsymbol{x}(k). \qquad (4)$$

Since the LMS derivation does not take into account the additive noise $\eta(k)$, its adoption in the configuration presented in Fig. 1 induces the emergence of a bias in the resulting estimate [2].

The bias compensated LMS (BC-LMS) algorithm [2] employs an estimate $\hat{\sigma}_\eta^2$ of the variance of $\eta(k)$ in order to mitigate the bias induced by the LMS algorithm by inserting an additional term in the LMS update equation, which gives place to the following update rule:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta\boldsymbol{x}(k)e(k) + \beta\hat{\sigma}_\eta^2\boldsymbol{w}(k). \qquad (5)$$

III. FIRST-ORDER ANALYSIS

The analyses of this paper are based on the *deviation vector* $\tilde{\boldsymbol{w}}(k) \in \mathbb{R}^N$, defined as

$$\tilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}^\star - \boldsymbol{w}(k). \qquad (6)$$

Using (5) and (6), one obtains the following exact relationship

$$\begin{aligned}
\tilde{\boldsymbol{w}}(k+1) &= \left[\boldsymbol{I} - \beta\boldsymbol{x}(k)\boldsymbol{x}^T(k) + \beta\hat{\sigma}_\eta^2\boldsymbol{I}\right]\tilde{\boldsymbol{w}}(k) \qquad (7)\\
&\quad - \beta\boldsymbol{x}(k)\nu(k) + \beta\left[\boldsymbol{x}(k)\boldsymbol{\eta}^T(k) - \hat{\sigma}_\eta^2\boldsymbol{I}\right]\boldsymbol{w}^\star.
\end{aligned}$$

The application of the expectation operator $\mathbb{E}[\cdot]$ in (8) demands the evaluation of complex joint moments. In order to make the mathematics tractable, the following hypotheses are adopted:

Independence assumption (IA). The input data (*i.e.*, $u(k)$ and $\eta(k)$) are statistically independent from $\tilde{\boldsymbol{w}}(k)$.

Noise Assumption (NA). The additive noise $\nu(k)$ is white and statistically independent from the remaining random variables.

Remarks: although it does not generate accurate predictions when the value of the step size is large, IA is an almost ubiquitous statistical hypothesis, due to the fact that overcoming it leads to cumbersome mathematical complexity [6]. In its turn, NA is a very popular simplification that, although physically plausible in some settings, was circumvented in recent analyses [7].

Using IA and NA, recursion (8) implies the following difference equation:

$$\begin{aligned}
\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] &= \left[\boldsymbol{I} - \beta\boldsymbol{R}_x + \beta\hat{\sigma}_\eta^2\boldsymbol{I}\right]\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right]\\
&\quad + \beta(\boldsymbol{R}_\eta - \hat{\sigma}_\eta^2\boldsymbol{I})\boldsymbol{w}^\star, \qquad (8)
\end{aligned}$$

where $\boldsymbol{R}_x \triangleq \mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]$ and $\boldsymbol{R}_\eta \triangleq \mathbb{E}\left[\boldsymbol{\eta}(k)\boldsymbol{\eta}^T(k)\right]$ are the autocorrelation matrices of signals $x(k)$ and $\eta(k)$, respectively.

Remark: note that (8) does not assume that $\eta(k)$ is white, which makes it distinct from the previous literature (see, *e.g.*, [2], [5]).

If the algorithm indeed converges in the mean (*i.e.*, $\lim_{k\to\infty}\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] = \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right]$), Eq. (8) implies that in steady-state

$$\lim_{k\to\infty}\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right] = \left(\boldsymbol{R}_x - \hat{\sigma}_\eta^2\boldsymbol{I}\right)^{-1}\left(\boldsymbol{R}_\eta - \hat{\sigma}_\eta^2\boldsymbol{I}\right)\boldsymbol{w}^\star, \quad (9)$$

where the asymptotic bias does not depend on the value of the step size.

Remark: Note that when the traditional hypothesis of a white noise signal $\eta(k)$ is utilized, Eqs. (8) and (9) turn into:

$$\begin{aligned}
\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] &= \left[\boldsymbol{I} - \beta\boldsymbol{R}_x + \beta\hat{\sigma}_\eta^2\boldsymbol{I}\right]\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right]\\
&\quad + \beta(\sigma_\eta^2 - \hat{\sigma}_\eta^2)\boldsymbol{w}^\star \qquad (10)
\end{aligned}$$

and

$$\lim_{k\to\infty}\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right] = \left(\boldsymbol{R}_x - \hat{\sigma}_\eta^2\boldsymbol{I}\right)^{-1}\left(\sigma_\eta^2 - \hat{\sigma}_\eta^2\right)\boldsymbol{w}^\star, \qquad (11)$$

which implies that the BC-LMS is asymptotically unbiased whenever $\sigma_\eta^2 = \hat{\sigma}_\eta^2$.

Unfortunately, the average behavior of each deviation coefficient $\mathbb{E}\left[\tilde{w}_i(k+1)\right]$ (for $i \in \{0, 1, \dots, N-1\}$) depends on the set $\{\mathbb{E}\left[\tilde{w}_j(k)\right]\}$, for $j \in \{0, 1, \dots, N-1\}$. In its turn, matrix $\boldsymbol{R}_x$ can be decomposed according to [1]

$$\boldsymbol{R}_x = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T, \qquad (12)$$

where the i-th column of $\boldsymbol{Q} \in \mathbb{R}^N$ contains the i-th eigenvector $\boldsymbol{q}_i \in \mathbb{R}^N$ and the i-th element of the main diagonal of the diagonal matrix $\boldsymbol{\Lambda}$ contains the corresponding eigenvalue. Since $\boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_N$ [1], where I_N denotes the N-order identity matrix, and defining

$$\boldsymbol{v}(k) \triangleq \left[\boldsymbol{Q}^T\tilde{\boldsymbol{w}}(k)\right], \qquad (13)$$

where $v_i(k) \triangleq \boldsymbol{q}_i^T\tilde{\boldsymbol{w}}(k)$, by multiplying both sides of (9) by $\boldsymbol{Q}^T$, one may convert (8) to

$$\mathbb{E}\left[\boldsymbol{v}(k+1)\right] = \left[\boldsymbol{I} - \beta\boldsymbol{\Lambda} + \beta\hat{\sigma}_\eta^2\boldsymbol{I}\right]\mathbb{E}\left[\boldsymbol{v}(k)\right] + \boldsymbol{c}, \qquad (14)$$

where $\boldsymbol{c} \triangleq \beta\boldsymbol{Q}^T\left(\boldsymbol{R}_\eta - \hat{\sigma}_\eta^2\boldsymbol{I}\right)\boldsymbol{w}^\star$. From (14), one may conclude, for the i-th element of $\boldsymbol{v}(k+1)$, that the following recursion is valid

$$\mathbb{E}[v_i(k+1)] = (1 - \beta(\lambda_i + \hat{\sigma}_\eta^2))\mathbb{E}[v_i(k)] + c_i, \qquad (15)$$

so that the BC-LMS is stable in the average if

$$\left|1 - \beta\left(\lambda_i + \hat{\sigma}_\eta^2\right)\right| < 1 \Rightarrow 0 < \beta < \frac{2}{\lambda_i + \hat{\sigma}_\eta^2}, \qquad (16)$$

or, in an equivalent way,

$$0 < \beta < \frac{2}{\lambda_{\max} + \hat{\sigma}_\eta^2}, \tag{17}$$

where $\lambda_{\max}$ is the largest eigenvalue of $\boldsymbol{R}_x$.

By comparing (17) to the upper bound that guarantees convergence in the mean for the LMS [1], that is,

$$0 < \beta_{\mathrm{LMS}} < \frac{2}{\lambda_{\max}}, \tag{18}$$

it is possible to infer that the theoretical upper bound on the value of β that guarantees stability in the mean of the BC-LMS is *always lower* than the one of the LMS.

Unfortunately, such upper bound has restricted usefulness, since an adaptive algorithm may diverge even when its first-order statistics do not, due to the fact that the variance of its coefficients may grow without limit [8]. Such fact motivates the study of the mean-square learning behavior of the BC-LMS, which is the goal of the next section.

IV. MEAN SQUARE ANALYSIS

Note that Eq. (8) can be rewritten as

$$\tilde{\boldsymbol{w}}(k+1) = \boldsymbol{A}(k)\tilde{\boldsymbol{w}}(k) - \beta\boldsymbol{x}(k)\nu(k) + \beta\boldsymbol{B}(k)\boldsymbol{w}^\star, \tag{19}$$

where

$$\boldsymbol{A}(k) \triangleq \boldsymbol{I} - \beta\boldsymbol{x}(k)\boldsymbol{x}^T(k) + \beta\hat{\sigma}_\eta^2\boldsymbol{I} = \boldsymbol{A}^T(k), \tag{20}$$

$$\boldsymbol{B}(k) \triangleq \boldsymbol{x}(k)\boldsymbol{\eta}^T(k) - \hat{\sigma}_\eta^2\boldsymbol{I} = \boldsymbol{B}^T(k). \tag{21}$$

By multiplying (19) by its transpose, one obtains

$$
\begin{aligned}
\boldsymbol{\mathcal{W}}(k+1) = \ & \boldsymbol{A}\boldsymbol{\mathcal{W}}(k)\boldsymbol{A}^T - \beta\nu(k)\boldsymbol{A}\tilde{\boldsymbol{w}}(k)\boldsymbol{x}^T(k) \\
& + \beta\boldsymbol{A}\tilde{\boldsymbol{w}}(k)\left[\boldsymbol{w}^\star\right]^T\boldsymbol{B} - \beta\nu(k)\boldsymbol{x}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{A} \\
& + \beta^2\nu^2(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k) - \beta^2\nu(k)\boldsymbol{x}(k)\left[\boldsymbol{w}^\star\right]^T\boldsymbol{B} \\
& + \beta\boldsymbol{B}\boldsymbol{w}^\star\tilde{\boldsymbol{w}}^T(k)\boldsymbol{A}^T - \beta^2\nu(k)\boldsymbol{B}\boldsymbol{w}^\star\boldsymbol{x}^T(k) \\
& + \beta^2\boldsymbol{B}\boldsymbol{w}^\star\left[\boldsymbol{w}^\star\right]^T\boldsymbol{B},
\end{aligned}
\tag{22}
$$

where $\boldsymbol{\mathcal{W}}(k) \triangleq \tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)$ is a quantity whose average evolution is of interest. Consider $\mathrm{vec}[\boldsymbol{A}]$ as the operator that concatenates the columns of matrix $\boldsymbol{A}$ in a single column vector. Denoting the Kronecker product by $\otimes$, one has

$$\mathrm{vec}\left(\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Z}\right) = \left[\boldsymbol{Z}^T \otimes \boldsymbol{X}\right]\mathrm{vec}(\boldsymbol{Y}). \tag{23}$$

By defining $\tilde{\boldsymbol{v}}(k) \triangleq \mathrm{vec}\left[\boldsymbol{\mathcal{W}}(k)\right]$, (22) can be rewritten as

$$
\begin{aligned}
\tilde{\boldsymbol{v}}(k+1) = \ & \left[\boldsymbol{A}(k) \otimes \boldsymbol{A}(k)\right]\tilde{\boldsymbol{v}}(k) \\
& + \beta\left[\boldsymbol{B}(k) \otimes \boldsymbol{A}(k)\right]\mathrm{vec}\left\{\tilde{\boldsymbol{w}}(k)\left[\boldsymbol{w}^\star\right]^T\right\} \\
& + \beta^2\nu^2(k)\mathrm{vec}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] \\
& + \beta\left[\boldsymbol{A}(k) \otimes \boldsymbol{B}(k)\right]\mathrm{vec}\left[\boldsymbol{w}^\star\tilde{\boldsymbol{w}}^T(k)\right] \\
& + \beta^2\left[\boldsymbol{B}(k) \otimes \boldsymbol{B}(k)\right]\mathrm{vec}\left\{\boldsymbol{w}^\star\left[\boldsymbol{w}^\star\right]^T\right\} + \mathcal{O}[\nu(k)],
\end{aligned}
\tag{24}
$$

where $\mathcal{O}[\nu(k)]$ contains quantities that are $\mathcal{O}[1]$ w.r.t. $\nu(k)$.

The use of the expectation operator in (25) along with IA and NA leads to

$$
\begin{aligned}
\mathbb{E}\left[\tilde{\boldsymbol{v}}(k+1)\right] = \ & \boldsymbol{C}\mathbb{E}\left[\tilde{\boldsymbol{v}}(k)\right] + \beta\boldsymbol{D}\boldsymbol{g} + \beta^2\sigma_\nu^2\mathrm{vec}\left[\boldsymbol{R}_x\right] \\
& + \beta\boldsymbol{E}\bar{\boldsymbol{g}} + \beta^2\boldsymbol{F},
\end{aligned}
\tag{25}
$$

where $\boldsymbol{g} \triangleq \mathbb{E}\left\{\mathrm{vec}\left[\tilde{\boldsymbol{w}}(k)\left[\boldsymbol{w}^\star\right]^T\right]\right\}$, $\bar{\boldsymbol{g}} \triangleq \mathbb{E}\left\{\mathrm{vec}\left[\boldsymbol{w}^\star\tilde{\boldsymbol{w}}^T(k)\right]\right\}$ can be inferred from (8), and

$$\boldsymbol{C} \triangleq \mathbb{E}\left[\boldsymbol{A}(k) \otimes \boldsymbol{A}(k)\right], \boldsymbol{D} \triangleq \mathbb{E}\left[\boldsymbol{B}(k) \otimes \boldsymbol{A}(k)\right], \tag{26}$$

$$\boldsymbol{E} \triangleq \mathbb{E}\left[\boldsymbol{A}(k) \otimes \boldsymbol{B}(k)\right], \tag{27}$$

$$\boldsymbol{F} \triangleq \mathbb{E}\left[\boldsymbol{B}(k) \otimes \boldsymbol{B}(k)\right]\mathrm{vec}\left\{\boldsymbol{w}^\star\left[\boldsymbol{w}^\star\right]^T\right\}, \tag{28}$$

are input data-dependent matrices. In order to theoretically compute these matrices, the input signal is henceforth assumed to be Gaussian (although not necessarily white), so that an additional assumption is necessary:

Gaussianity assumption (GA). Signals $u(k)$ and $\eta(k)$ are samples from a Gaussian distribution.

Remark: Using GA, fourth-order moments of these random variables can be computed using the formulas presented in [9].

It is noteworthy that (25) can be employed to predict the evolution of the mean square distortion (MSD)

$$\mathrm{MSD}(k) \triangleq \mathbb{E}\left[\|\boldsymbol{w}^\star - \boldsymbol{w}(k)\|^2\right], \tag{29}$$

since the MSD can be estimated through

$$\mathrm{MSD}(k) \approx \mathrm{Tr}\left\{\mathrm{unvec}\left[\mathbb{E}\left[\tilde{\boldsymbol{v}}(k)\right]\right]\right\}, \tag{30}$$

where the operator $\mathrm{unvec}[\cdot]$ reverses the $\mathrm{vec}[\cdot]$ operation and $\mathrm{Tr}[\boldsymbol{X}]$ denotes the trace of matrix $\boldsymbol{X}$.

V. TRACKING ANALYSIS

It is important to obtain performance guarantees in non-stationary settings, which occur when the ideal plant to be identified is time variant. In this section, the tracking capability of the BC-LMS is analyzed when the coefficients of ideal plant vary slowly, according to a first-order Markovian model [10], [11]:

$$\boldsymbol{w}^\star(k+1) = \boldsymbol{w}^\star(k) + \boldsymbol{\vartheta}(k), \tag{31}$$

where $\boldsymbol{\vartheta}(k) \in \mathbb{R}^N$ is a zero-mean i.i.d. stochastic perturbation whose (diagonal) autocorrelation matrix is denoted by $\boldsymbol{R}_\vartheta$. Note that time-varying channels introduce a "lag" in the adaptive process to the emulation of the optimal and unknown vector $\boldsymbol{w}^\star(k)$.

In order to simplify the second-order analysis of the tracking setting, consider an additional stochastic hypothesis:

Tracking Assumption (TA). The random vector $\boldsymbol{\vartheta}(k)$ is statistically independent from the remaining random variables.

Using IA, NA and TA, and following steps similar to those that led to Eq. (25), one may demonstrate that, under the considered tracking scenario, the recursion that describes the second-order quantity can be written as [12]

$$
\begin{aligned}
\mathbb{E}\left[\tilde{\boldsymbol{v}}(k+1)\right] = \ & \boldsymbol{C}\mathbb{E}\left[\tilde{\boldsymbol{v}}(k)\right] + \beta\boldsymbol{D}\boldsymbol{g} + \beta^2\sigma_\nu^2\mathrm{vec}\left[\boldsymbol{R}_x\right] \\
& + \beta\boldsymbol{E}\bar{\boldsymbol{g}} + \beta^2\boldsymbol{F} + \mathrm{vec}\left[\boldsymbol{R}_\vartheta\right],
\end{aligned}
\tag{32}
$$

which implies that, under the adopted Markovian model, the tracking configuration does not modify the range of values of the step size that guarantees stability.

TABLE I: Transfer functions responsible for the coloring of signals $u(k)$ and $\eta(k)$ in the considered scenarios.

Scenario	$B_u(z)$	$B_\eta(z)$
1	$1 - 0.8z^{-1}$	$1 - 0.1z^{-1}$
2	$1 - 0.8z^{-1} + 0.2z^{-2}$	$1 - 0.3z^{-1} + 0.1z^{-1}$
3	$1 - 0.8z^{-1} + 0.2z^{-1}$	$1 - 0.3z^{-1} + 0.1z^{-2}$
4	1	1

VI. DEFICIENT-LENGTH ANALYSIS

Sometimes system identification procedures may operate in a deficient-length setting, which occurs when the length of the ideal transfer function surpass the adaptive filter length [13]. This may happen when the designer intends to deal with computational limitations or if the unknown transfer function is long [8], [14]. In such a suboptimal scenario, suppose that the reference signal can be written as

$$d_{\text{def}}(k) = [\boldsymbol{w}^\star]^T \boldsymbol{u}(k) + [\overline{\boldsymbol{w}}^\star]^T \overline{\boldsymbol{u}}(k) + \nu(k), \qquad (33)$$

where $\overline{\boldsymbol{w}} \in \mathbb{R}^L$ contains the additional coefficients of the ideal plant (which the adaptive filter does not cover up) and

$$\overline{\boldsymbol{u}}(k) \triangleq \begin{bmatrix} u(k-N) & u(k-N-1) & \dots & u(k-N-L+1) \end{bmatrix}^T. \qquad (34)$$

In order to simplify the mathematics, the following stochastic assumption was assumed in the following second-order analysis:

Whiteness Assumption (WA). The input signal $u(k)$ is white.

Using IA, NA and WA and after some manipulations, one obtains the following recursion for the mean square analysis:

$$\begin{aligned} \mathbb{E}\left[\tilde{\boldsymbol{v}}(k+1)\right] &= \boldsymbol{C}\mathbb{E}\left[\tilde{\boldsymbol{v}}(k)\right] + \beta \boldsymbol{D}\boldsymbol{g} + \beta^2 \sigma_\nu^2 \text{vec}\left[\boldsymbol{R}_x\right] \\ &\quad + \beta \boldsymbol{E}\overline{\boldsymbol{g}} + \beta^2 \boldsymbol{F} \\ &\quad + \beta^2 \boldsymbol{H} \text{vec}\left\{\overline{\boldsymbol{w}}^\star [\overline{\boldsymbol{w}}^\star]^T\right\}, \end{aligned} \qquad (35)$$

where $\boldsymbol{H} \triangleq \mathbb{E}\left[\boldsymbol{x}(k)\overline{\boldsymbol{u}}^T(k) \otimes \boldsymbol{x}(k)\overline{\boldsymbol{u}}^T(k)\right]$.
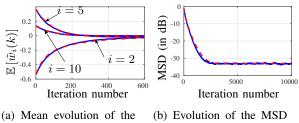
VII. RESULTS

In the following simulations, the following transfer function was utilized:

$$w_i^\star = \cos\left[0.4\pi(i-1)\right]\exp(-0.2(i-1)), \text{ for } 0 \le i \le N + L - 1, \qquad (36)$$

where $L > 0$ in deficient-length scenarios and $N = 20$ for all considered simulations. The noise signal $\nu(k)$ was sampled from a white Gaussian process. Signal $u(k)$ (resp. $\eta(k)$) is obtained by filtering a unitary-variance white Gaussian signal (resp. white Gaussian signal with variance σ_η^2) by a coloring filter $B_u(z)$ (resp. $B_\eta(z)$). The coloring filters employed in the four considered scenarios are presented in Tab. I. Note that the devised analysis is not restricted neither for white $u(k)$ nor for white $\eta(k)$, which avoids common statistical assumptions in the literature.

Fig. 2a depicts the averaged evolution of the deviations in the first scenario, with $\beta = 10^{-2}$, $\sigma_\nu^2 = 10^{-3}$, $\sigma_\eta^2 = 10^{-2}$ and $\hat{\sigma}_\eta^2 = 0.008$. The theoretical curves, evaluated using (8), reveal good agreement with the empirical ones, even when the estimated variance of η is different from its actual value.

The prediction of the MSD evolution is the main goal of the second scenario, in which $\sigma_\eta^2 = 0.007$, $\hat{\sigma}_\eta^2 = 0.01$, $\beta =$

(a) Mean evolution of the i-th coefficient deviation, for 100 independent Monte Carlo trials.

(b) Evolution of the MSD (in dB), with 1000 Monte Carlo trials.

Fig. 2: Results for the first scenario (a) and second scenario (b). Blue: empirical curve. Dashed red: theoretical curve.

(a) Evolution of the MSD (in dB) in the deficient-length setting.

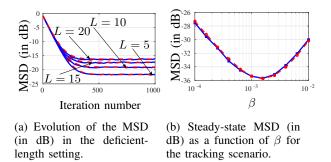(b) Steady-state MSD (in dB) as a function of β for the tracking scenario.

Fig. 3: Results for the third scenario (a) and fourth scenario (b). Blue: empirical curve, computed with 100 Monte Carlo trials. Dashed red: theoretical curve.

5×10^{-3}, and $\sigma_\nu^2 = 10^{-3}$. From Fig. 2b, one may note that the advanced stochastic model also accurately predicts the actual mean square learning curve.

The prediction of the performance loss induced by a suboptimal adaptive filter is studied in the third scenario, for which $\hat{\sigma}_\eta^2 = \sigma_\eta^2 = \beta = 10^{-2}$, and $\sigma_\nu^2 = 10^{-3}$. Fig. 3a depicts the comparison between the theoretical and empirical MSD curves for different values of L, where $\overline{w}_i^\star = 0.1$, for $i \in \{0, 1, \dots, L-1\}$. One observes that Eq. (35) accurately describes the empirical curves. Note that the adoption of the same value of β for distinct ideal plant length does not influence the algorithm stability, as predicted by the proposed stochastic model.

The last (fourth) scenario addresses the time-variant configuration, in which $\sigma_\nu^2 = 10^{-6}$ and $\beta = \sigma_\eta^2 = \hat{\sigma}_\eta^2 = 10^{-2}$. Fig. 3b depicts the the asymptotic mean square distortion for different values of β. The steady-state MSD was computed by averaging the last 10^4 iterations of a execution with 3×10^5 iterations. It can also be observed that there indeed exists an optimal value of β (as usual in tracking scenarios [10]), and that optimal value is well approximated by the advanced model.

VIII. CONCLUSION

This paper presented a comprehensive analysis of the learning behavior of the BC-LMS algorithm. Theoretical predictions for both first- and second-order statistics were provided, and the resulting model was successfully extended in order to address both tracking and deficient-length configurations.

JOURNAL OF COMMUNICATION AND INFORMATION SYSTEMS, VOL. 36, NO.1, 2021.

118

REFERENCES

[1] S. Haykin. Adaptive filter theory. Prentice-Hall, Inc., 1996.

[2] B. Kang, J. Yoo and P. Park, "Bias-compensated normalised LMS algorithm with noisy input," in Electronics Letters, vol. 49, no. 8, pp. 538-539, April 2013, doi: 10.1049/el.2013.0246.

[3] Zhao, H. and Zheng, Z., Bias-compensated affine-projection-like algorithms with noisy input. Electron. Lett., 52: 712-714, 2016, doi: 10.1049/el.2015.3550.

[4] Pimenta, R.; Resende, L.; Petraglia, M.R. and Haddad, D. B, On the steady-state performance of bias-compensated LMS algorithm. Electron. Lett., 57: 85-88, 2021, doi: 10.1049/ell2.12044.

[5] S. M. Jung and P. Park, "Stabilization of a Bias-Compensated Normalized Least-Mean-Square Algorithm for Noisy Inputs," in IEEE Transactions on Signal Processing, vol. 65, no. 11, pp. 2949-2961, June, 2017, doi: 10.1109/TSP.2017.2675865.

[6] P. Lara, F. Igreja, L. D. T. J. Tarrataca, D. B. Haddad and M. R. Petraglia, "Exact Expectation Evaluation and Design of Variable Step-Size Adaptive Algorithms," in IEEE Signal Processing Letters, vol. 26, no. 1, pp. 74-78, Jan. 2019, doi: 10.1109/LSP.2018.2880084.

[7] P. Lara, K. d. S. Olinto, F. R. Petraglia and D. B. Haddad, "Exact analysis of the least-mean-square algorithm with coloured measurement noise," in Electronics Letters, vol. 54, no. 24, pp. 1401-1403, 2018, doi: 10.1049/el.2018.6675.

[8] Pedro Lara, Luís D.T.J. Tarrataca, and Diego B. Haddad, "Exact expectation analysis of the deficient-length LMS algorithm," Signal Processing, Vol. 162, pp. 54-64, Sept. 2019, doi: 10.1016/j.sigpro.2019.04.009.

[9] P. Ahrendt. The Multivariate Gaussian Probability Distribution. Technical Report, DTU Library, 2005. Link: http://www2.imm.dtu.dk/pubdb/p.php?3312

[10] M. Turki-Hadj Alouane, M. Jaıdane-Saıdane, "A new nonstationary LMS algorithm for tracking Markovian time varying systems," Signal Processing, Volume 86, Issue 1, pp. 50-70, Jun. 2006, doi: 10.1016/j.sigpro.2005.04.010.

[11] 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," in IEEE Transactions on Signal Processing, vol. 68, pp. 5882-5893, 2020, doi: 10.1109/TSP.2020.3028993.

[12] B. Farhang-Boroujeny and S. Gazor, "Performance of LMS-based adaptive filters in tracking a time-varying plant," in IEEE Transactions on Signal Processing, vol. 44, no. 11, pp. 2868-2871, Nov. 1996, doi: 10.1109/78.542445.

[13] K. Mayyas, "Performance analysis of the deficient length LMS adaptive algorithm," in IEEE Transactions on Signal Processing, vol. 53, no. 8, pp. 2727-2734, Aug. 2005, doi: 10.1109/TSP.2005.850347.

[14] S. J. M. de Almeida, M. H. Costa and J. C. M. Bermudez, "A stochastic model for the deficient length Pseudo Affine Projection adaptive algorithm," 17th EUSIPCO, Glasgow, pp. 1715-1719, Aug. 2009.