On the Skewness of the LMS Adaptive Weights

Thiago T. P. Silva, Filipe Igreja, Pedro Lara, Luís Tarrataca, Asutosh Kar, *Senior Member, IEEE*, and Diego B. Haddad, *Member, IEEE*

***Abstract*—The adjustable weights of adaptive filtering algorithms are usually assumed to obey a Gaussian distribution. This is somewhat natural under maximal-entropy considerations, since most analyses in the open literature only take into account first- and second-order statistics. This work investigates the third-order statistical feature known as skewness of the least mean square parameters distribution. Two theoretical analyses for skewness estimation are proposed: *i*) one that employs the independence assumption, which states that the excitation data is statistically independent from the adaptive weights; *ii*) one derived from the exact expectation analysis, a method that is able to predict the learning capabilities of the least mean square algorithm even when the step size is not infinitesimally small. This paper shows that the skewness of the adaptive weights distribution may present a large deviation from the common Gaussian assumption, especially in the first phase of the learning. Furthermore, it is also demonstrated that the skewness may grow without limit even when adaptive weights present convergence in both average and mean square behaviours.**

***Index Terms*—Adaptive Filtering, Exact Expectation Analysis, LMS, Convergence Analysis, Skewness.**

I. Introduction

STOCHASTIC models of adaptive filtering algorithms usually adopt the Gaussian Assumption (GA), which states that the marginal distributions of elements of the weight vector[1] $\boldsymbol{w}(k) \in \mathbb{R}^N$ are Gaussian [1]–[3]. Relevant models that exceptionally address the non-Gaussianity of the adaptive weights can be observed in [4], [5], which focus on proportionate adaptive algorithms. A stationary setting is assumed in the following, so that each adaptive coefficient $w_i(k) \in \mathbb{R}$ ideally should have the optimal value $w_i^\star$. Consider $m_{3,i}(k)$ as the third central moment and $\sigma_i(k)$ as the standard deviation of the deviation $\tilde{w}_i(k) \triangleq w_i^\star - w_i(k)$ (for $i \in \{0, 1, \ldots, N-1\}$). Under GA, the skewness $\chi_i(k)$, defined as

$$\chi_i(k) \triangleq \frac{m_{3,i}(k)}{\sigma_i^3(k)}, \tag{1}$$

is zero, which is also coherent with the asymptotic symmetry results derived from the Fokker-Planck stochastic model [6]. Unfortunately, there is no theoretical guarantee that the coefficients skewness is approximately zero (which is most of time implicitly assumed), except by a non-rigorous employment of the central limit theorem [7].

This work was supported in part by CNPq, FAPERJ and CAPES (Code 001), Brazil. Thiago T. P. Silva, Filipe Igreja, Pedro Lara, Luís Tarrataca, and Diego B. Haddad are with Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Brazil (e-mails: thiagoteodoro501@gmail.com, figreja@gmail.com, pedro.lara@cefet-rj.br, luis.tarrataca@cefet-rj.br, diego.haddad@cefet-rj.br). Asutosh Kar is with ECE Dept. at IIITDM, Kancheepuram, Chennai, India (e-mail: asutoshkar@iiitdm.ac.in).

[1] All vectors in this paper are supposed to be of column-type.

Skewness is a measure of a probability density function (pdf) asymmetry. Its analysis brings additional statistical information about the distribution of the adaptive taps, which may allow better control of the adaptive filter learning abilities [4], [8]. Furthermore, this examination can be crucial when the weight vector is directly used in a test statistic, which occurs in adaptive line enhancer applications [9], adaptive detector structures [10], [11], distributed detection over adaptive networks [12]–[14], when the filtered output is used as a test statistic [15], [16], in time delay estimation [17] and when the least mean squares (LMS) algorithm is employed as a canceller in spread spectrum communications systems [18]–[20]. In these cases, the first- and second-order statistics are not sufficient to calculate error probabilities for the binary decisions, the receiver operating characteristics (ROCs) (which relates the detection and false alarm probabilities for the detectors) and to compute decision boundaries [1], [6].

In this paper, two theoretical models for the evolution of $\chi_i(k)$ are derived for the first time, considering the traditional LMS algorithm. The first makes use of the independence assumption (IA, or independence heuristic [21]), which states that the adjustable weights and the input data are statistically independent [22]. Although almost universal, IA-based models are known to generate inaccurate predictions when the step size $\beta \in \mathbb{R}_+$ assumes non-infinitesimal values. It is precisely in this range of values that some potential advantages of adaptive filtering algorithms may become apparent [23]. In order to circumvent this issue, the second advanced model employs the *exact expectation analysis* (EEA) [24]–[29], a sophisticated theoretical analysis that avoids the use of the IA. Adopting EEA allows one to theoretically analyze setups where the skewness significantly deviates from zero, a phenomenon more frequent for large values of β. Furthermore, the adoption of the ubiquitous IA tends to underestimate the deviation of the adaptive weights marginal distribution from the Gaussian one.

This work is structured as follows: Section II succinctly describes the LMS algorithm. An IA-based model for the evolution is presented in Section III. The EEA is described in Section IV. Section V discusses how the empirical results relate to the theoretical models, emphasizing the fact the contributions presented contribute for further clarity into two important theoretical problems concerning the behaviour of the adaptive coefficients: what are the conditions that guarantee that the learning process is stable and when does there exist an asymptotic distribution for them [30]. Section VI presents our concluding remarks.

II. Least Mean Squares

Assuming a supervised context, the LMS algorithm updates the weights of a filter (often with a tapped-delay-line structure) using the stochastic gradient of the instantaneous quadratic error:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \boldsymbol{x}(k) e(k), \tag{2}$$

where $\boldsymbol{x}(k) \triangleq \begin{bmatrix} x(k) & x(k-1) & \dots & x(k-N+1) \end{bmatrix}^T$ is the vector containing the N most recent input data samples and $e(k) \in \mathbb{R}$ is the error signal, which is computed using

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

where $d(k) \in \mathbb{R}$ is a reference signal (sometimes constructed through ingenious ways), commonly generated according to the following affine-in-the-parameters noisy regression model:

$$d(k) = [\boldsymbol{w}^\star]^T \boldsymbol{x}(k) + \nu(k), \tag{4}$$

where $\boldsymbol{w}^\star \in \mathbb{R}^N$ is the ideal (and unknown) plant the adaptive filter intends to emulate[2] and $\nu(k)$ is an additive noise, usually associated to measurement issues [31]. The adaptation process induces an overall system behavior that is distinct from that of a linear system, making it difficult to carry out a rigorous mathematical analysis of the LMS learning abilities.

In order to render the mathematics tractable, most stochastic models employ strong statistical hypotheses. The ubiquitous noise assumption (NA), for example, states that the measurement noise has zero mean and is statistically independent from the remaining random variables. A less physically plausible statistical hypothesis is IA, which can be stated as:

Independence assumption (IA): The adaptive weights $\boldsymbol{w}(k)$ and the input vector $\boldsymbol{x}(k)$ are statistically independent.

Since the statistical properties of the input signal have profound effects on filter performance [28], it is important to avoid restricting the analysis of this paper to white input signals. Thus, the following moving-average generation model is adopted

$$x(k) = \sum_{m=0}^{M-1} b_m u(k-m), \tag{5}$$

where $u(k)$ derives from a white stationary process that is statistically independent from the remaining random variables. The pdf of $u(k)$ is supposed to be symmetric, and the respective moments are described by $\gamma_n \triangleq \mathbb{E}\left[u^n(k)\right]$.

III. IA-based Skewness Model

Using (3)-(4), Eq. (2) may be converted into a recursion for the deviations:

$$\tilde{\boldsymbol{w}}(k+1) = \left[\boldsymbol{I}_N - \beta \boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\tilde{\boldsymbol{w}}(k) - \beta \boldsymbol{x}(k)\nu(k), \tag{6}$$

where $\boldsymbol{I}_N$ denotes the $N \times N$ identity matrix. Note that the distribution of adaptive weights can be easily recovered from the distributions of the deviations, since

$$\tilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}^\star - \boldsymbol{w}(k). \tag{7}$$

Consider the configuration $(N, M) = (1, 2)$, for pedagogical purposes. By applying the expectation operator in (6) and using the IA and NA, a first-order stochastic recursion for the deviation can be obtained

$$\mathbb{E}[\tilde{w}_0(k+1)] = a_{0,0}\mathbb{E}[\tilde{w}_0(k)] \tag{8}$$

[2]This paper focuses on system identification tasks.

where $a_{0,0} = (1 - b_0^2\beta\gamma_2 - b_1^2\beta\gamma_2)$. Assuming the algorithm converges (*i.e.*, $\lim_{k\to\infty}\mathbb{E}\left[\tilde{w}_0(k+1)\right] = \lim_{k\to\infty}\mathbb{E}\left[\tilde{w}_0(k)\right]$), recursion (8) means that, under the IA, the LMS consists of an asymptotic unbiased estimator.

The variance $\sigma_i^2(k)$ of the adaptive weights depends on second-order statistics, which can be obtained by squaring both sides of (6) (combined with the application of both the IA and the NA), which leads to

$$\mathbb{E}[\tilde{w}_0^2(k+1)] = a_{1,1}\mathbb{E}[\tilde{w}_0^2(k)] + d_1, \tag{9}$$

where

$$a_{1,1} = (1 - 2b_0^2\gamma_2\beta - 2b_1^2\gamma_2\beta + b_0^4\beta^2\gamma_4 + 6b_0^2\beta^2 b_1^2\gamma_2^2 + b_1^4\beta^2\gamma_4),$$

and $d_1 = (b_0^2 + b_1^2)a_0^2\sigma_\nu^2\beta^2\gamma_2$, with σ_ν^2 denoting the additive noise variance. The mean-squared stability of the LMS is guaranteed, under the IA, when $|a_{1,1}| < 1$. In general, second-order stochastic models, such as the ones based on the energy conservation approach, are of great interest for the designer [32], [33].

The evaluation of $\chi_i(k)$ requires third-order statistics (see (1)). These quantities can be obtained by cubing (6), applying the expectation operator $\mathbb{E}\left[\cdot\right]$ and using the IA and the NA, so that one derives the following recursion

$$\mathbb{E}\left[\tilde{w}_0^3(k+1)\right] = a_{2,2}\mathbb{E}\left[\tilde{w}_0^3(k)\right] + a_{2,0}\mathbb{E}\left[\tilde{w}_0(k)\right], \tag{10}$$

where

$$\begin{aligned} a_{2,2} \triangleq\ & 1 - 3b_0^2\gamma_2\beta - 3b_1^2\gamma_2\beta + 3b_0^4\gamma_4\beta^2 + 18b_0^2\gamma_2^2\beta^2 b_1^2 \\ & + 3b_1^4\gamma_4\beta^2 - b_0^6\beta^3\gamma_6 - 15b_0^4\beta^3 b_1^2\gamma_2\gamma_4 \\ & - 15b_0^2\beta^3 b_1^4\gamma_4\gamma_2 - b_1^6\beta^3\gamma_6, \\ a_{2,0} \triangleq\ & 3b_0^2\gamma_2 a_0^2\sigma_\nu^2\beta^2 + 3b_1^2\gamma_2 a_0^2\sigma_\nu^2\beta^2 - 3b_0^4 a_0^2\sigma_\nu^2\beta^3\gamma_4 \\ & - 18b_0^2 a_0^2\sigma_\nu^2\beta^3 b_1^2\gamma_2^2 - 3b_1^4 a_0^2\sigma_\nu^2\beta^3\gamma_4. \end{aligned}$$

Recursions (8)-(10) can be concisely described as a state space equation system

$$\boldsymbol{y}^{\text{IA}}(k+1) = \boldsymbol{A}^{(\text{IA})}\boldsymbol{y}^{(\text{IA})}(k) + \boldsymbol{d}^{(\text{IA})}, \tag{11}$$

where the state vector $\boldsymbol{y}^{(\text{IA})}(k) \in \mathbb{R}^R$ contains the statistical quantities of interest (*i.e.*, $\mathbb{E}[\tilde{w}_0(k)]$, $\mathbb{E}[\tilde{w}_0^2(k)]$ and $\mathbb{E}[\tilde{w}_0^3(k)]$, which are the state variables of the considered setup). Note that the time-invariant transition matrix $\boldsymbol{A}^{(\text{IA})}$ depends on β, and that LMS third-order stability is guaranteed if the chosen step size value implies that $\rho\left[\boldsymbol{A}^{(\text{IA})}\right] < 1$, where $\rho\left[\boldsymbol{A}^{(\text{IA})}\right]$ stands for the *spectral radius* of $\boldsymbol{A}^{(\text{IA})}$ (*i.e.*, $\rho\left[\boldsymbol{A}^{(\text{IA})}\right] \triangleq \max_i\left|\lambda_i\left(\boldsymbol{A}^{(\text{IA})}\right)\right|$, where λ_i is the i-th eigenvalue of $\boldsymbol{A}^{(\text{IA})}$) [34]. It is noteworthy that the aforementioned stability concept also takes into account third-order statistics.

Note that the presented derivation of a state space equation system that models the dynamics of the LMS can be very lengthy in more complex configurations. As the next section demonstrates, avoiding the IA makes the associated mathematics even more difficult.

IV. Exact Expectation Analysis for the Skewness

Under large step size configurations, it is advisable to avoid the IA in order to avoid modelling inaccuracies. In this case, simplifications such as

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

are no longer valid, although $\mathbb{E}\left[u^2(k)\tilde{w}_i^2(k)\right] = \gamma_2\mathbb{E}\left[\tilde{w}_i^2(k)\right]$ remains true, since $u(k)$ is i.i.d. by hypothesis. Assuming the same configuration of the last section (*i.e.*, $(N, M) = (1, 2)$), under EEA Eq. (8) should be rewritten as

$$\mathbb{E}[\tilde{w}_0(k+1)] = (1-b_0^2\beta\gamma_2)\mathbb{E}[\tilde{w}_0(k)] - b_1^2\beta\mathbb{E}[u^2(k-1)\tilde{w}_0(k)], \quad (12)$$

where one may note the emergence of the *nuisance* term $\mathbb{E}[u^2(k-1)\tilde{w}_0(k)]$. This is considered as a nuisance because one is not primarily interested in it, even though its estimation is a required step to the update of one statistical quantity of interest [35]. The EEA method is a systematic procedure that should generate recursions for nuisance state variables, whose structure may give rise to new nuisance terms. Due to the limited time-lag correlation of the input sequence (see (5)), the recursive generation of update equations eventually halts, although it can require millions of equations in some setups [36].

In the pedagogical example of this paper, the recursion of the nuisance term $\mathbb{E}[u^2(k-1)\tilde{w}_0(k)]$ is derived by multiplying both sides of (6) by $u^2(k)$ before the application of the expectation operator. Using the NA, one obtains

$$\begin{aligned}\mathbb{E}[u^2(k)\tilde{w}_0(k+1)] &= (\gamma_2 - b_0^2\beta\gamma_4)\mathbb{E}[\tilde{w}_0(k)] \\ &\quad -b_1^2\beta\gamma_2\mathbb{E}[u^2(k-1)\tilde{w}_0(k)]. \quad (13)\end{aligned}$$

The derivation of (13) illustrates the recursive feature of the EEA, which systematically constructs a state space system that describes the statistical learning dynamics of the LMS. It can be proved that the evolution of $\mathbb{E}\left[\tilde{w}_0^3(k)\right]$ fits into the recursion

$$\begin{aligned}\mathbb{E}\left[\tilde{w}_0^3(k+1)\right] &= \overline{a}_{0,0}\mathbb{E}\left[\tilde{w}_0^3(k)\right] + \overline{a}_{0,1}\mathbb{E}\left[u^2(k-1)\tilde{w}_0^3(k)\right] \\ &\quad +\overline{a}_{0,2}\mathbb{E}\left[\tilde{w}_0(k)\right] + \overline{a}_{0,3}\mathbb{E}\left[u^2(k-1)\tilde{w}_0(k)\right] \\ &\quad +\overline{a}_{0,4}\mathbb{E}\left[u^4(k-1)\tilde{w}_0^3(k)\right] \\ &\quad +\overline{a}_{0,5}\mathbb{E}\left[u^4(k-1)\tilde{w}_0(k)\right] \\ &\quad +\overline{a}_{0,6}\mathbb{E}\left[u^6(k-1)\tilde{w}_0^3(k)\right], \quad (14)\end{aligned}$$

where

$$\begin{aligned}\overline{a}_{0,0} &= 1 - 3b_0^2\gamma_2\beta + 3b_0^4\gamma_4\beta^2 - b_0^6\beta^3\gamma_6, \\ \overline{a}_{0,1} &= -3b_1^2\beta + 18b_0^2\gamma_2\beta^2 b_1^2 - 15b_0^4\beta^3 b_1^2\gamma_4, \\ \overline{a}_{0,2} &= 3b_0^2\gamma_2 a_0^2\sigma_\nu^2\beta^2 - 3b_0^4 a_0^2\sigma_\nu^2\beta^3\gamma_4, \\ \overline{a}_{0,3} &= 3b_1^2 a_0^2\sigma_\nu^2\beta^2 - 18b_0^2 a_0^2\sigma_\nu^2\beta^3 b_1^2\gamma_2, \\ \overline{a}_{0,4} &= 3b_1^4\beta^2 - 15b_0^2\beta^3 b_1^4, \\ \overline{a}_{0,5} &= -3b_1^4 a_0^2\sigma_\nu^2\beta^3, \quad \overline{a}_{0,6} = -b_1^6\beta^3.\end{aligned}$$

Finally, the analysis up to the third-order of the considered setting leads to $R = 10$ state variables:

$$\begin{aligned}\{&\mathbb{E}\left[\tilde{w}_0(k)\right], \mathbb{E}\left[u^2(k-1)\tilde{w}_0(k)\right], \mathbb{E}\left[\tilde{w}_0^3(k)\right], \mathbb{E}\left[u^2(k-1)\tilde{w}_0^3(k)\right], \\ &\mathbb{E}\left[u^4(k-1)\tilde{w}_0^3(k)\right], \mathbb{E}\left[u^4(k-1)\tilde{w}_0(k)\right], \mathbb{E}\left[u^6(k-1)\tilde{w}_0^3(k)\right], \\ &\mathbb{E}\left[\tilde{w}_0^2(k)\right], \mathbb{E}\left[u^2(k-1)\tilde{w}_0^2(k)\right], \mathbb{E}\left[u^4(k-1)\tilde{w}_0^2(k)\right]\},\end{aligned}$$

which may be used to construct a state vector $\boldsymbol{y}^{(\mathrm{EEA})}(k) \in \mathbb{R}^{10}$ whose evolution is governed by

$$\boldsymbol{y}^{(\mathrm{EEA})}(k+1) = \boldsymbol{A}^{(\mathrm{EEA})}\boldsymbol{y}^{(\mathrm{EEA})}(k) + \boldsymbol{d}^{(\mathrm{EEA})}, \quad (15)$$

TABLE I

Number R of state variables for both IA and EEA methods in different configurations (N, M).

N	M	R (IA)	R (EEA)	N	M	R (IA)	R (EEA)
1	2	3	10	2	1	5	12
1	3	3	80	2	2	9	215
1	4	3	842	2	3	9	2342
1	5	3	10022	3	1	7	173
2	1	5	12	3	2	19	4611

where matrix $\boldsymbol{A}^{(\mathrm{EEA})}$ and vector $\boldsymbol{d}^{(\mathrm{EEA})}$ are not entirely described, due to lack of space. The number R of equations demanded for the theoretical analyses in some configurations is presented in Table I. Unfortunately, such a quantity grows rapidly when N and M are increased, which may impose a prohibitively high computational burden for the EEA method, even for values of N lower than 10.

V. Results

In the following simulations, the coefficients of the ideal plant are $w_i^\star = 1$, for $i \in \{0, 1, \ldots, N-1\}$, the additive noise is a white Gaussian signal and $u(k)$ is an unitary-variance signal, which can sampled from a Gaussian or a Laplacian distribution. The computation of the empirical curves was performed using K independent Monte Carlo trials, assuming that the adaptive weights are initialized as zeros. An efficient C++-based code was written to perform the required algebraic operations and simplifications.

In the first simulation, the colored input signal $x(k)$ is obtained by filtering $u(k)$ by the transfer function $B_1(z) = 1 + 0.8z^{-1}$, and $M = 2$. Figures 1 and 2 depict the evolution of both theoretical and empirical skewness of the first adaptive coefficient. Note that the standard (*i.e.*, based on IA) model underestimates the deviation of the skewness from zero, whereas the empiric curve fits well with the EEA-based prediction. This result emphasizes that avoiding the ubiquitous IA may enhance the theoretical understanding of the third-order behaviour of the adaptive weights. More specifically, in the configuration of Fig. 1, the skewness converges asymptotically to zero, which coincides with both IA and EEA predictions and indicates that GA may be reasonable after convergence in this setup[3]. Fig. 2 shows that it is not always the case, since under this setting the steady-state skewness deviates from zero in a significant manner.

Fig. 3 compares the empirical distribution of $w_0(k)$ at the point where the theoretical skewness (under EEA) attains its maximum (*i.e.*, $k = 16$ in Fig. 1) with the usual Gaussian model. The plot reveals that the actual marginal distribution of an adaptive weight can be very distinct from a Gaussian one (and much more peaky), which demonstrates that the ever-present Gaussian model is not adequate for the adaptive parameters actual distribution in every setup.

In the second configuration, the colored input signal $x(k)$ is obtained by filtering $u(k)$ (sampled from a Gaussian distribution) by the transfer function $B_2(z) = 1 + 0.8z^{-1} + 0.8z^{-2}$.

[3]It should be noted, however, that the non-Gaussianity can be in effect even when the skewness is zero.

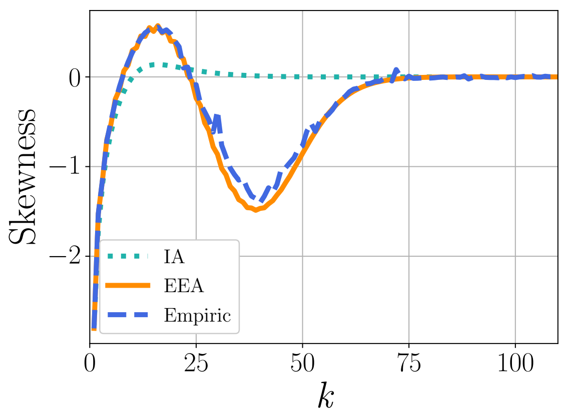

Fig. 1. Evolution of the skewness of the first adaptive tap of the first configuration, with $N = 2$, $M = 2$, $\beta = 0.09$, $\sigma_\nu^2 = 10^{-3}$, $K = 10^{10}$, and $u(k)$ obeying a Gaussian distribution.

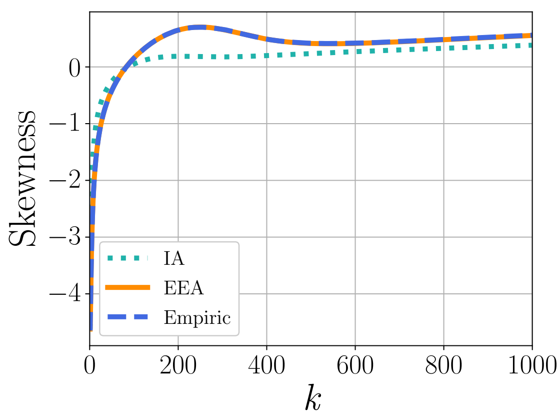

Fig. 2. Evolution of the skewness of the first adaptive tap of the first configuration, with $N = 3$, $M = 2$, $\beta = 0.005$, $\sigma_\nu^2 = 10^{-11}$, $K = 10^6$, with $u(k)$ obeying a Laplacian distribution.

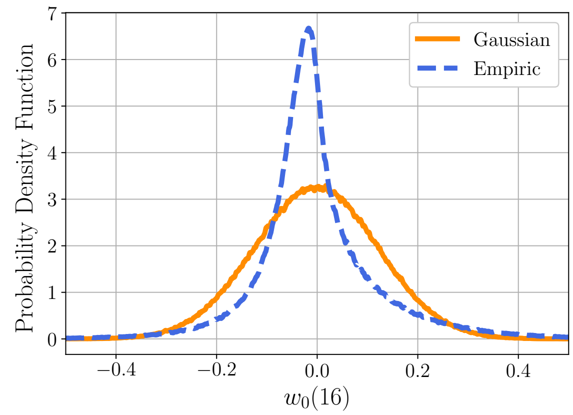

Fig. 3. Probability density function of $w_0(16)$, under the configuration of Fig. 1. Solid line: analytic Gaussian model, whose mean and variance coincides with empirical data. Dashed line: estimated distribution from empirical data, computed with a Gaussian kernel whose kernel size is $\sigma = 10^{-3}$. The empirical quantities were obtained using $K = 10^6$.

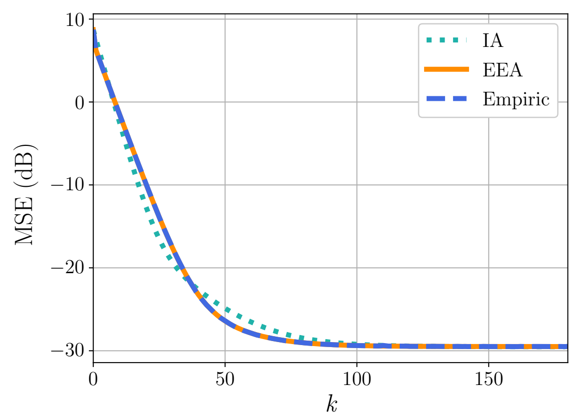

Fig. 4. Empirical and theoretical (*i.e.*, from IA and EEA) MSE of the second configuration with $\beta = 0.04$. The empirical curve was computed with $K = 10^{11}$.

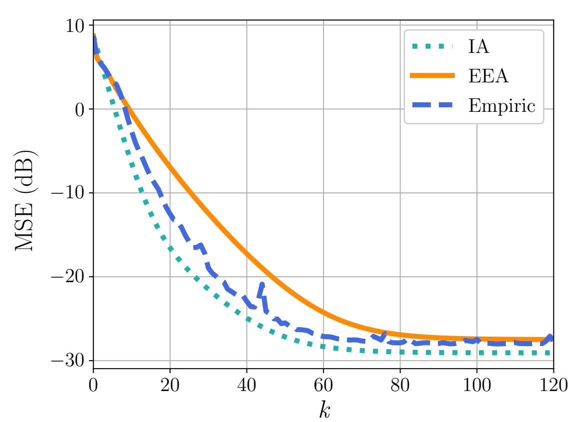

Fig. 5. Empirical and theoretical (*i.e.*, from IA and EEA) MSE of the second configuration with $\beta = 0.0665$. The empirical curve was computed with $K = 10^{11}$.

The remaining parameters are $N = 2$, $M = 3$, and $\sigma_\nu^2 = 10^{-3}$. Note that the maximum *theoretical* value of β that guarantees stability can be computed using the power method [37], since knowing the largest absolute eigenvalue is sufficient to characterize convergence. In the second configuration, when considering up to second-order statistics (see, *e.g.*, [28]), such an upper bound (under EEA) is $\beta_{\max}^{(2)} = 0.07055$. Since this paper extends EEA for third-order statistics, it is now possible to compute this limit when such a statistic is taken into account: $\beta_{\max}^{(3)} = 0.04616$.

Fig. 4 allows one to compare the empirical MSE evolution with the theoretical predictions provided by IA and EEA in the second configuration, for a value of $\beta = 0.04 < \beta_{\max}^{(3)}$, where the MSE presents a stable behaviour, as expected. Fig. 5 depicts the MSE evolution for a value of β that lies on the interval $\left(\beta_{\max}^{(3)}, \beta_{\max}^{(2)}\right)$, which implies that third-order statistics (usually overlooked by theoretical analyses) is no longer convergent. Fig. 5 shows that the empirical MSE fits well the EEA-based curve at the initial learning phase, and that, after some iterations, the algorithm indeed converges *faster* than predicted by the exact analysis. This phenomenon was already pointed out by [38], which utilized almost-sure analysis tools for explaining it. Since the difference between Fig. 4 and Fig. 5 basically relies on the third-order convergence of the former, one may observe that third-order instability is not necessarily harmful for the mean-square learning dynamics of the LMS.

VI. Conclusions

This paper models for the first time third-order statistics of the LMS adaptive weights, which allows predicting deviations from the usual Gaussianity assumption. The often employed IA has revealed to be inaccurate for predicting the skewness evolution. The devised framework permits the computation of proper theoretical upper bounds for the step size that guarantee *third-order stability* of the algorithm.

References

[1] N. J. Bershad and Lian Zuo Qu, "On the probability density function of the LMS adaptive filter weights," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 43–56, Jan 1989.

[2] Guang Deng and Wai-Yin Ng, "A model-based approach for the development of LMS algorithms [adaptive filter applications]," in *2005 IEEE International Symposium on Circuits and Systems*, pp. 2267–2270 Vol. 3, 2005.

[3] G. Su, J. Jin, Y. Gu, and J. Wang, "Performance analysis of l_0 norm constraint least mean square algorithm," *IEEE Transactions on Signal Processing*, vol. 60, pp. 2223–2235, May 2012.

[4] K. T. Wagner and M. I. Doroslovački, "Proportionate-type NLMS algorithms based on maximization of the joint conditional pdf for the weight deviation vector," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3738–3741, 2010.

[5] K. Wagner and M. Doroslovacki, *Proportionate-Type Normalized Least Mean Square Algorithms*. Wiley-IEEE Press, 1st ed., 2013.

[6] S. T. Alexander and V. L. Stonic, "Analytical probability density functions for LMS adaptive filters using the Fokker-Planck equation," in *[Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing*, pp. 2117–2120 vol.3, April 1991.

[7] K. T. Wagner and M. I. Doroslovacki, "Towards analytical convergence analysis of proportionate-type NLMS algorithms," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3825–3828, March 2008.

[8] K. T. Wagner and M. I. Doroslovacki, "Proportional-type NLMS algorithm with gain allocation providing maximum one-step conditional pdf for true weights," in *2009 43rd Annual Conference on Information Sciences and Systems*, pp. 61–66, 2009.

[9] B. Fisher and N. Bershad, "The complex LMS adaptive algorithm–transient weight mean and covariance with applications to the ale," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, no. 1, pp. 34–44, 1983.

[10] J. R. Zeidler, "Performance analysis of LMS adaptive prediction filters," *Proceedings of the IEEE*, vol. 78, no. 12, pp. 1781–1806, 1990.

[11] N. Bershad and Lian Qu, "On the joint characteristic function of the complex scalar LMS adaptive weight," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1166–1175, 1984.

[12] F. Cattivelli and A. H. Sayed, "Diffusion LMS-based distributed detection over adaptive networks," in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, pp. 171–175, 2009.

[13] S. Al-Sayed, J. Plata-Chaves, M. Muma, M. Moonen, and A. M. Zoubir, "Node-specific diffusion LMS-based distributed detection over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 66, no. 3, pp. 682–697, 2018.

[14] A. E. Feitosa, V. H. Nascimento, and C. G. Lopes, "Adaptive detection in distributed networks using maximum likelihood detector," *IEEE Signal Processing Letters*, vol. 25, no. 7, pp. 974–978, 2018.

[15] X. Wang, G. Li, and P. K. Varshney, "Detection of sparse signals in sensor networks via locally most powerful tests," *IEEE Signal Processing Letters*, vol. 25, no. 9, pp. 1418–1422, 2018.

[16] H. Zayyani, M. Korki, and F. Marvasti, "Bayesian hypothesis testing detector for one bit diffusion LMS with blind missing samples," *Signal Processing*, vol. 146, pp. 61 – 65, 2018.

[17] F. Reed, P. Feintuch, and N. Bershad, "Time delay estimation using the LMS adaptive filter–static behavior," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 3, pp. 561–571, 1981.

[18] Wang Shi lian, Zhang Chen, Gao Kai, and Zhang Er yang, "Noncoherent PN code acquisition based on the test of I/Q LMS adaptive filters' tap weights," in *Proceedings 7th International Conference on Signal Processing, 2004. Proceedings. ICSP '04. 2004.*, vol. 1, pp. 13–17 vol.1, 2004.

[19] R. Iltis and L. Milstein, "An approximate statistical analysis of the Widrow LMS algorithm with application to narrow-band interference rejection," *IEEE Transactions on Communications*, vol. 33, no. 2, pp. 121–130, 1985.

[20] T. Yu, C. Kang, and D. Hong, "Limitations in the code acquisition performance for adaptive filter versus correlator," *IEEE Signal Processing Letters*, vol. 13, no. 11, pp. 653–656, 2006.

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

[22] M. Rupp and H. . Butterweck, "Overcoming the independence assumption in LMS filtering," in *The Thirty-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, vol. 1, pp. 607–611 Vol.1, Nov 2003.

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

[24] T. T. P. Silva, P. Lara, T. Gonzaga, D. Brandão, L. Tarrataca, and D. B. Haddad, "Recent advances in overcoming the independence assumption in adaptive filtering analyses," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 317–322, 2020.

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

[26] S. C. Douglas and Weimin Pan, "Exact expectation analysis of the LMS adaptive filter," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2863–2871, 1995.

[27] S. C. Douglas and 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, pp. 61–64 vol.4, 1992.

[28] P. Lara, F. Igreja, L. D. T. J. Tarrataca, D. B. Haddad, and M. R. Petraglia, "Exact expectation evaluation and design of variable step-size adaptive algorithms," *IEEE Signal Processing Letters*, vol. 26, pp. 74–78, Jan 2019.

[29] T. T. P. Silva, P. Lara, F. Igreja, F. D. V. R. Oliveira, L. Tarrataca, and D. B. Haddad, "An exact expectation model for the LMS tracking abilities," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5882–5893, 2020.

[30] J. A. Bucklew, T. G. Kurtz, and W. A. Sethares, "Weak convergence and local stability properties of fixed step size recursive algorithms," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 966–978, 1993.

[31] L. C. Resende, D. B. Haddad, and M. R. Petraglia, "A variable step-size NLMS algorithm with adaptive coefficient vector reusing," in *2018 IEEE International Conference on Electro/Information Technology (EIT)*, pp. 0181–0186, May 2018.

[32] S. Radhika, F. Albu, and A. Chandrasekar, "Steady state mean square analysis of standard maximum versoria criterion based adaptive algorithm," *IEEE Transactions on Circuits and Systems II: Express Briefs*, pp. 1–1, 2020.

[33] H. Yazdanpanah, P. S. Diniz, and M. V. Lima, "Improved simple set-membership affine projection algorithm for sparse system modelling: Analysis and implementation," *IET Signal Processing*, vol. 14, no. 2, pp. 81–88, 2020.

[34] S. C. Douglas and Weimin Pan, "Exact expectation analysis of the LMS adaptive filter," *IEEE Transactions on Signal Processing*, vol. 43, pp. 2863–2871, Dec 1995.

[35] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 86, pp. 2009–2025, Oct 1998.

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

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

[38] V. H. Nascimento and A. H. Sayed, "On the learning mechanism of adaptive filters," *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1609–1625, 2000.