Exact Expectation Analysis of the LMS Adaptive Identification of Nonlinear Systems

Pedro Lara, Filipe Igreja, Thiago T. P. Silva, Luís D. T. J. Tarrataca, and Diego B. Haddad

Most adaptive filtering schemes employ the tapped-delay line. In part, such a fact can be explained by the assumption that the plant they intend to estimate is linear. Although such a hypothesis can be reasonable if the input signal is constrained to a certain range, sometimes it may not be valid. In this last case, the performance and stability guarantees provided by stochastic models that presume linearity of the ideal system are no longer valid. This paper advances an analytic model of the least mean square learning capabilities when the ideal system is not linear, with the additive noise including nonlinear functions of the input samples. The proposed analysis does not assume neither that the excitation signal is statistically independent from the adaptive weights, nor that the additive noise is white and/or independent from input data. Furthermore, it can be applied to non-Gaussian and/or non-white input signals. Simulations show that the advanced model is more accurate than traditional approaches.

Introduction: The least-mean-square (LMS) algorithm emerged as an effective, yet simple, adaptive filtering scheme whose robustness against perturbations and implementation errors is well-know [1–3]. Despite its simplicity, the mathematical analysis of the LMS is complicated due to its stochastic and nonlinear nature [4]. Furthermore, most analyses assume that the ideal plant the adaptive filter intends to estimate[1] is linear [5]. Unfortunately, this may not be strictly valid in practice, due to a plethora of phenomena, such as operation near the saturation region [6], intermodulation distortion [7] and usage of high power amplifiers [8]. Since a small distortion engineered by a nonlinear component (NLC) in such systems may dominate the overall identification performance [9], analyses that do not take into account the NLC are not able to provide neither performance nor stability guarantees. This paper advances, for the first time, a stochastic model of the LMS performance that incorporates the effets of the NLC. The proposed analysis does not employ the almost ubiquitous *independence assumption* (IA), which states that the input samples $x(k)$ are statistically independent from the adaptive weights $w_i(k)$ (for $i \in \{0, 1, \dots, N\}$, where N denotes the adaptive filter length). IA, although widely used in the field of stochastic approximations [1], is recognized to provide accurate predictions only when the step size is small [10]. The avoidance of IA engineers the exact expectation technique [14–16], which is also able to provide a proper step-size upper bound that guarantees convergence [15]. Furthermore, it can be employed in the derivation of a *deterministic* theoretical step-size sequence that optimizes performance [17].

This paper is structured as follows. We start by describing the adopted nonlinear model of the plant to be identified. We then present the advanced analysis, which can be employed to predict a stability region. We then propose a method for deriving a deterministic step size sequence that optimizes performance and takes into account the stochastic coupling between the excitation data and the adaptive coefficients. This is followed by the results. We then present the main conclusions of the paper.

Nonlinear System Model (NSM) : Nonlinear discrete-time systems can be frequently modeled by Volterra series [13, 18–20], which is a stable functional series expansion of a nonlinear time-invariant system. Such series take into account memory effects, which can not be performed with static nonlinear models [21, 22]. Considering $N \in \mathbb{N}$ as the memory depth, $\overline{w}_p^\star(n_1, \cdots, n_p) \in \mathbb{R}$ as the coefficients of the p-th order polynomial basis function and $P \in \mathbb{N}$ as the maximum nonlinearity order, in this paper the reference signal $d(k)$ is modeled as the sum of three components

$$d(k) = (\boldsymbol{w}^\star)^T \boldsymbol{x}(k) + \nu_{\mathrm{M}}(k) + \overbrace{\sum_{p=2}^{P} \sum_{n_1=0}^{N-1} \cdots \sum_{n_P=0}^{N-1} \overline{w}_p^\star(n_1, \dots, n_P) \prod_{l=1}^{p} x(n - n_l)}^{\triangleq \nu_{\mathrm{NLC}}(k)}, \quad (1)$$

where the first right-side term corresponds to the affine-in-the-parameters linear component (LC, with $\boldsymbol{w}^\star \in \mathbb{R}^N$ denoting the unknown and ideal coefficients of the LC), $\nu_{\mathrm{M}}(k)$ is the measurement noise and $\nu_{\mathrm{NLC}}(k)$ corresponds to the nonlinear part (that is not captured by the linear adaptive model). Note that the consecutive samples of the overall noise $\nu(k) \triangleq \nu_{\mathrm{M}}(k) + \nu_{\mathrm{NLC}}(k)$ can be correlated, so that the noise whiteness (a very common hypothesis) is not assumed in this paper. The Volterra model (1) is assumed to be symmetric (i.e, $\overline{w}_p^\star(n_1, \cdots, n_p)$ is invariant w.r.t. the $p!$ possible permutations of the indices $n_1, \cdots, n_p$ [23, 24]). Finally, the input vector $\boldsymbol{x}(k) \in \mathbb{R}^N$ is defined as

$$\boldsymbol{x}(k) \triangleq \begin{bmatrix} x(k) & x(k-1) & \cdots & x(k-N+1) \end{bmatrix}^T, \quad (2)$$

so that there is a *deterministic coherence* between successive input vectors.

Exact Expectation Analysis of LMS-Based NLS Identification: The LMS update equation is a nonvanishing step-size version of a stochastic gradient algorithm that intends to minimize the mean-squared error (MSE) [25]:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \beta \nabla_{\boldsymbol{w}(k)} \left[\frac{1}{2} e^2(k)\right] = \boldsymbol{w}(k) + \beta e(k) \boldsymbol{x}(k), \quad (3)$$

where $\beta \in \mathbb{R}^+$ is the step size and the signal error $e(k) \in \mathbb{R}$ is defined by

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

In order to not restrict the analysis to white inputs, in the following the input signal is assumed to be colored according to an M-th order moving-average model:

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

where $u(k)$ is an i.i.d. signal with even probability density function (pdf) and $\gamma_n \triangleq \mathbb{E}[u^n(k)]$. Furthermore, its variance γ_2, without loss of generality, is supposed to be unitary.

The average performance of the adaptive identification depends on a set of joint moments (or *state variables*), such as $\mathbb{E}\left[\tilde{w}_0^2(k) u^2(k-1)\right]$, where $\tilde{w}_i(k)$ is the i-th component of the deviation vector $\tilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}^\star - \boldsymbol{w}(k)$. When IA is employed, the number R of state variables (or number of equations) is dramatically decreased, since under IA it is possible to perform approximations such as

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

whereas it should be noted that even if IA is not assumed the identity $\mathbb{E}\left[\tilde{w}_0^2(k) u^2(k)\right] = \gamma_2 \mathbb{E}\left[\tilde{w}_0^2(k)\right]$ is strictly valid, since $\boldsymbol{w}(k)$ does not depend on the sample $u(k)$ (see (3)). In the following, joint moments such as $\mathbb{E}\left[u(k) \nu_{\mathrm{M}}(k)\right]$ will be considered to be zero, due to the usage of the following *noise assumption* (NA):

NA. The measurement noise $\nu_{\mathrm{M}}(k)$ is a zero-mean i.i.d. stochastic process, which is statistically independent from the input signal.

Remark: NA is a standard assumption in analyses of adaptive filtering algorithms, and often is satisfied in practice [12, 26]. In this paper, the *overall* noise $\nu(k)$ is neither i.i.d. nor statistically independent from the excitation data.

For didactic purposes, consider the configuration $N = 2$, $M = 1$, $\nu_{\mathrm{NLC}}(k) = \alpha x^2(k-1)$ and a first-order stochastic analysis. In scalar

[1] This paper focuses on the system identification task.

terms, the application of recursion (3) in $\tilde{w}_0(k)$ can be expressed as

$$\begin{aligned}\tilde{w}_0(k+1) &= \tilde{w}_0(k) - b_0^2 u^2(k)\beta - b_0^2 u(k)u(k-1)\beta \\ &\quad - b_0 u(k)\nu(k)\beta - b_0^3 u(k)u^2(k-1)\alpha\beta \\ &\quad + b_0^2 u^2(k)\tilde{w}_0(k)\beta + b_0^2 u(k)\tilde{w}_1(k)u(k-1)\beta, \quad (6)\end{aligned}$$

which is a deterministic recursion that can be converted to a stochastic one by applying the expectation operator:

$$\mathbb{E}[\tilde{w}_0(k+1)] = (1 + b_0^2\beta\gamma_2)\mathbb{E}[\tilde{w}_0(k)] - b_0^2\beta\gamma_2. \quad (7)$$

By replicating the same steps to $\tilde{w}_1(k)$, one obtains

$$\mathbb{E}[\tilde{w}_1(k+1)] = \mathbb{E}[\tilde{w}_1(k)] - b_0^2\beta\gamma_2 + b_0^2\beta\mathbb{E}[u^2(k-1)\tilde{w}_1(k)], \quad (8)$$

where the *nuisance* term $\mathbb{E}[u^2(k-1)\tilde{w}_1(k)]$ appears. Such a state variable is termed as a nuisance one due to the fact that one is not primarily interested in it, but its computation is necessary in order to evaluate the quantities of interest [11]. A recursion on such term can be obtained if one multiplies both sides of recursion

$$\begin{aligned}\tilde{w}_1(k+1) &= \tilde{w}_1(k) - b_0^2 u(k-1)u(k)\beta - b_0^2 u^2(k-1)\beta \\ &\quad - b_0 u(k-1)\nu(k)\beta - b_0^3 u^3(k-1)\alpha\beta \\ &\quad + b_0^2 u(k-1)\tilde{w}_0(k)u(k)\beta + b_0^2 u^2(k-1)\tilde{w}_1(k)\beta \quad (9)\end{aligned}$$

by $u^2(k)$ before the application of operator $\mathbb{E}[\cdot]$, which provides the following recursion

$$\begin{aligned}\mathbb{E}[u^2(k)\tilde{w}_1(k+1)] &= \gamma_2\mathbb{E}[\tilde{w}_1(k)] - b_0^2\beta\gamma_2^2 \\ &\quad + b_0^2\beta\gamma_2\mathbb{E}[u^2(k-1)\tilde{w}_1(k)]. \quad (10)\end{aligned}$$

Note that recursions (7), (9) and (10) can be employed to construct a *linear* state equation system

$$\boldsymbol{y}(k+1) = \boldsymbol{A}\boldsymbol{y}(k) + \boldsymbol{b}, \quad (11)$$

in which state vector $\boldsymbol{y}(k)$ contains the state variables $\mathbb{E}[\tilde{w}_0(k)]$, $\mathbb{E}[\tilde{w}_1(k)]$ and $\mathbb{E}\left[u^2(k-1)\tilde{w}_1(k)\right]$, $\boldsymbol{A}$ is a time-invariant *transition matrix* and $\boldsymbol{b}$ is a vector that contains terms that do not depend on state variables (such as $\beta b_0^2\gamma_2^2$). In this first-order analysis of the considered setting, matrix $\boldsymbol{A}$ has dimensions 3×3. Since the mean behavior of the adaptive weights leads to estimates for the range of stable operation that are way off, a second-order analysis is desirable. In the considered configuration, mean-square analysis requires the computation of five state variables, since the MSE can be expressed (without employing `IA`) as

$$\begin{aligned}\text{MSE}(k) &\triangleq b_0^2\gamma_2 - 2b_0^2\gamma_2\mathbb{E}[\tilde{w}_0(k)] \\ &\quad + b_0^2\gamma_2 - 2b_0^2\mathbb{E}[u^2(k-1)\tilde{w}_1(k)] + \sigma_\nu^2 \\ &\quad + b_0^4\alpha^2\gamma_4 - 2b_0^3\alpha\mathbb{E}[u^3(k-1)\tilde{w}_1(k)] + b_0^2\gamma_2\mathbb{E}[\tilde{w}_0^2(k)] \\ &\quad + b_0^2\mathbb{E}[u^2(k-1)\tilde{w}_1^2(k)], \quad (12)\end{aligned}$$

and a recursion of these state variables requires 14 equations (due to the need of computing the recursions for 9 nuisance state variables). The number of state variables grows rapidly with the increase of N and M [27], so that an efficient C++-code was written in order to automatically derive the necessary recursions.

In short, the exact expectation analysis consists of a recursive procedure that obtains the update equations of the involved state variables. Since the input data is correlated only in a finite horizon (see (5)), the generation of equations eventually halts. Note that the algorithm convergence can be inferred from matrix $\boldsymbol{A}$ (which depends on β). The stability is predicted if the maximum absolute eigenvalue of $\boldsymbol{A}$ (which can be evaluated by the Power Method [28]) is not greater than unity [15].

Results: In the following simulations, the input signal is obtained by filtering an unitary-variance white Gaussian noise by the transfer function $H(z) = 1 - 0.8z^{-1}$. The reference signal is described by

$$d(k) = [\boldsymbol{w}^\star]^T \boldsymbol{x}(k) + 0.005x^2(k) + \nu(k), \quad (13)$$

where $\nu(k)$ is an additive Gaussian noise with variance $\sigma_\nu^2 = 10^{-3}$. The unknown vector $\boldsymbol{w}^\star$ is composed by ones (i.e., $w_i^\star = 1$, for $i \in \{0, 1, \ldots, N-1\}$).

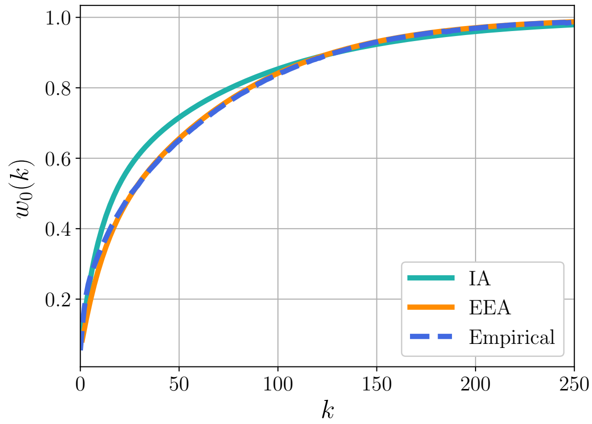

Fig. 1 *Evolution of the mean value of adaptive weight $w_0(k)$ of the LMS with $N = 6$ and $\beta = 0.065$. The empirical curve was obtained with 10^7 independent Monte Carlo trials.*

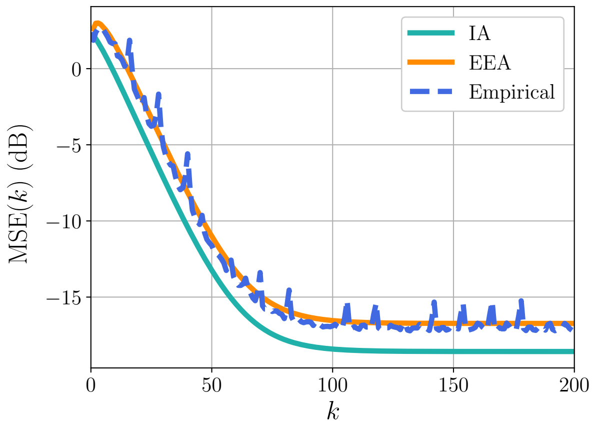

Fig. 2 *Evolution of the mean square error of the LMS with $N = 3$ and $\beta = 0.085$. The empirical curve was obtained with 10^8 independent Monte Carlo trials.*

Fig. 1 compares the first-order evolution of the adaptive coefficient $w_0(k)$ for both exact and standard theoretical analyses. Note that the former is able to accurately predict the first-order dynamics of the considered configuration. Fig. 2 permits one to conclude that the exact expectation analysis also adheres better to experimental second-order ensemble learning curves. Fig. 3 depicts the divergence between IA-based predictions for steady-state performance with respect to the empirical one. In order to empirically assess the probability of divergence, the LMS is executed for 10^5 independent Monte Carlo trials for 500 iterations. A specific realization is counted 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 $i \in \{0, \ldots, N-1\}$). The maximum value of β that implies algorithm stability can be inferred from the transition matrices $\boldsymbol{A}$, which can be constructed for both `IA`- and `EEA`-based models. Fig. 4 depicts the probability of divergence for different step-size values. Note that the upper bound established by the `EEA` specifies the region in which the algorithm converges and that it may indeed diverge in some cases where the `IA`-based model predicts stable operation.

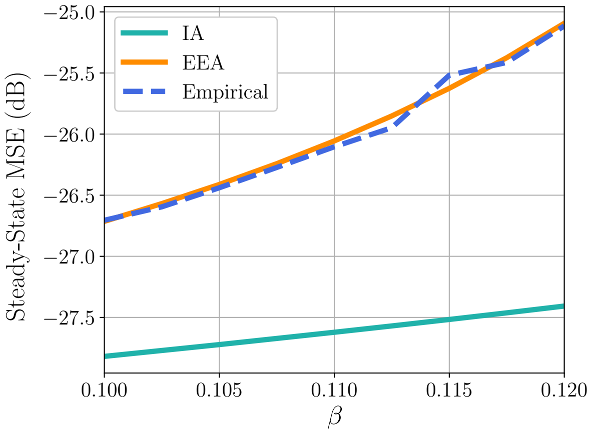

Fig. 3 *Steady-State MSE for the $N = 2$ configuration. The empirical results were obtained by the usage of 10^6 independent Monte Carlo trials.*

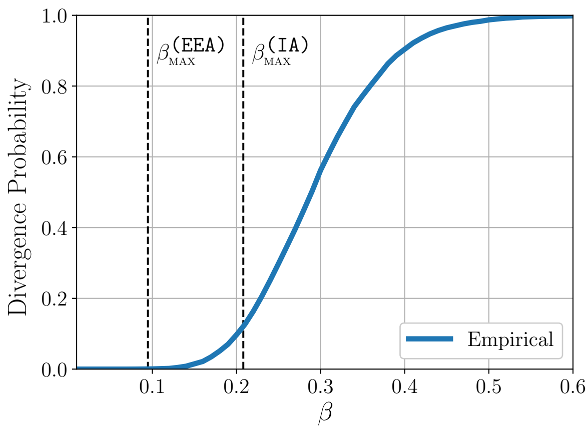

Fig. 4 *Divergence probability for $N = 3$. For this configuration, $\beta_{MAX}^{(IA)} = 0.208149$ and $\beta_{MAX}^{(EEA)} = 0.0947366$.*

Conclusions: In this paper, a comprehensive model that predicts the learning behaviour of the LMS algorithm when the ideal plant is nonlinear is devised. The advanced analysis does not assumes neither input signal Gaussianity, nor its whiteness. Simulations demonstrate that the proposed stochastic model is able to accurately characterize both transient and steady-state regimes. Furthermore, it provides a proper step size upper bound that effectively avoids stability issues.

Acknowledgment: The authors would like to thank the Brazilian Federal Agencies CAPES, CNPq, and FAPERJ for supporting this research.

Pedro Lara (*CEFET/RJ, Brazil*)

Email: pedro.lara@cefet-rj.br

Filipe Igreja (*CEFET/RJ, Brazil*)

Thiago T. P. Silva (*CEFET/RJ, Brazil*)

Luís D. T. J. Tarrataca (*CEFET/RJ, Brazil*)

Diego B. Haddad (*CEFET/RJ, Brazil*)

References

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

2 B. Hassibi, A. H. Sayed, and T. Kailath. LMS is $\mathcal{H}^\infty$ optimal. In *Proceedings of 32nd IEEE Conference on Decision and Control*, pages 74–79 vol.1, Dec 1993.

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

4 Simon S Haykin, Bernard Widrow, and Bernard Widrow. *Least-mean-square adaptive filters*, volume 31. Wiley Online Library, 2003.

5 Simon Haykin. *Adaptive filter theory*. Prentice Hall, Upper Saddle River, NJ, 4th edition, 2002.

6 W. Saabe, J. Sombrin, E. Ngoya, G. Soubercaze-Pun, and L. Lapierre. Volterra-based modeling of traveling wave tube amplifier for system level simulation. In *2017 Eighteenth International Vacuum Electronics Conference (IVEC)*, pages 1–2, April 2017.

7 S. Micheal Serunjogi, M. A. Sanduleanu, and M. S. Rasras. Volterra series based linearity analysis of a phase-modulated microwave photonic link. *Journal of Lightwave Technology*, 36(9):1537–1551, May 2018.

8 D. Zhou and V. E. DeBrunner. Novel adaptive nonlinear predistorters based on the direct learning algorithm. *IEEE Transactions on Signal Processing*, 55(1):120–133, Jan 2007.

9 Yong Hoon Lim, Yong Soo Cho, Il Whan Cha, and Dae Hee Youn. An adaptive nonlinear prefilter for compensation of distortion in nonlinear systems. *IEEE Transactions on Signal Processing*, 46(6):1726–1730, June 1998.

10 J. E. Mazo. On the independence theory of equalizer convergence. *The Bell System Technical Journal*, 58(5):963–993, May 1979.

11 J.-François Cardoso. Blind signal separation: statistical principles. *Proceedings of the IEEE*, 86(10):2009–2025, Oct 1998.

12 Karen da S. Olinto, Diego B. Haddad, and Mariane R. Petraglia. Transient analysis of l0-LMS and l0-NLMS algorithms. *Signal Processing*, 127:217 – 226, 2016.

13 Felipe B. da Silva and Wallace A. Martins. Data-selective volterra adaptive filters. *Circuits, Systems, and Signal Processing*, 37(10):4651–4664, Oct 2018.

14 S. C. Douglas and T. H. Meng. Exact expectation analysis of the LMS adaptive filter without the independence assumption. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 61–64, March 1992.

15 S. C. Douglas and Weimin Pan. Exact expectation analysis of the LMS adaptive filter. *IEEE Transactions on Signal Processing*, 43(12):2863–2871, Dec 1995.

16 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. *Electronics Letters*, 54(24):1401–1403, 2018.

17 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*, 26(1):74–78, Jan 2019.

18 R. A. d. Prado, F. d. R. Henriques, and D. B. Haddad. Sparsity-aware distributed adaptive filtering algorithms for nonlinear system identification. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, July 2018.

19 P. Śliwiński, A. Marconato, P. Wachel, and G. Birpoutsoukis. Non-linear system modelling based on constrained Volterra series estimates. *IET Control Theory Applications*, 11(15):2623–2629, 2017.

20 R.M. Lin and T.-Y. Ng. Identification of Volterra kernels for improved predictions of nonlinear aeroelastic vibration responses and flutter. *Engineering Structures*, 171:15 – 28, 2018.

21 Guang Feng, Hengjian Li, Jiwen Dong, and Jiashu Zhang. Face recognition based on Volterra kernels direct discriminant analysis and effective feature classification. *Information Sciences*, 441:187 – 197, 2018.

22 Thomas Kamalakis and Georgia Dede. Nonlinear degradation of a visible-light communication link: A Volterra-series approach. *Optics Communications*, 417:46 – 53, 2018.

23 M. Rudko and D. Weiner. Volterra systems with random inputs: A formalized approach. *IEEE Transactions on Communications*, 26(2):217–227, Feb 1978.

24 L. Yao and C. c. Lin. Identification of nonlinear systems by the genetic programming-based Volterra filter. *IET Signal Processing*, 3(2):93–105, March 2009.

25 D. T. M. Slock. On the convergence behavior of the LMS and the normalized LMS algorithms. *IEEE Transactions on Signal Processing*, 41(9):2811–2825, Sep. 1993.

26 M. R. Petraglia, D. B. Haddad, and E. L. Marques. Normalized subband adaptive filtering algorithm with reduced computational complexity. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 62(12):1164–1168, Dec 2015.

27 Pedro Lara, Luís D.T.J. Tarrataca, and Diego B. Haddad. Exact expectation analysis of the deficient-length LMS algorithm. *Signal Processing*, 162:54 – 64, 2019.

28 Gene H. Golub and Charles F. Van Loan. *Matrix Computations (3rd Ed.)*. Johns Hopkins University Press, Baltimore, MD, USA, 1996.