

Advances on the Exact Expectation Analyses of Adaptive Filtering Algorithms

Filipe Igreja

A thesis submitted in partial satisfaction of the requirements for the degree Doctor of Science in Instrumentation and Optics, Federal Education Center of Technology Celso Suckow da Fonseca, CEFET/RJ and Fluminense Federal University, UFF.

Advisor:
Diego Barreto Haddad

Rio de Janeiro,
November, 2023

Advances on the Exact Expectation Analysis of Adaptive Filtering Algorithms

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Instrumentação e Óptica, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ e Universidade Federal Fluminense, UFF, como parte dos requisitos necessários à obtenção do título de doutor.

Filipe Igreja

Aprovada por:

Documento assinado digitalmente
gov.br DIEGO BARRETO HADDAD
Data: 28/11/2023 09:23:59-0300
Verifique em https://validar.iti.gov.br

Prof. Diego Barreto Haddad, D.Sc. (orientador)

Documento assinado digitalmente
gov.br FELIPE DA ROCHA HENRIQUES
Data: 28/11/2023 14:15:30-0300
Verifique em https://validar.iti.gov.br

Prof. Felipe da Rocha Henriques, D.Sc.

Documento assinado digitalmente
gov.br GABRIEL MATOS ARAUJO
Data: 01/12/2023 21:30:18-0300
Verifique em https://validar.iti.gov.br

Prof. Gabriel Matos Araujo, D.Sc.

Documento assinado digitalmente
gov.br MARIANE REMBOLD PETRAGLIA
Data: 05/12/2023 15:09:56-0300
Verifique em https://validar.iti.gov.br

Prof. Mariane Rembold Petraglia, D.Sc.

Documento assinado digitalmente
gov.br TADEU NAGASHIMA FERREIRA
Data: 05/12/2023 15:33:51-0300
Verifique em https://validar.iti.gov.br

Prof. Tadeu Nagashima Ferreira, D.Sc.

Rio de Janeiro,

November, 2023

Ficha catalográfica elaborada pela Biblioteca Central do CEFET/RJ

I24 Igreja, Filipe
Advances on the exact expectation analyses of adaptive filtering algorithms / Filipe Igreja – 2023.
100f : il. (algumas color.) , enc.

Tese (Doutorado). Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, 2023.
Bibliografia : f. 90-100.
Orientador: Diego Barreto Haddad.

1. Processamento de sinais. 2. Algoritmos de filtragem adaptativa. 3. Sistemas não-lineares. 4. Análise de sistemas. I. Haddad, Diego Barreto (Orient.). II. Centro Federal de Educação Tecnológica Celso Suckow da Fonseca. Coordenadoria do Programa de Pós-Graduação em Instrumentação e Óptica Aplicada. III. Título.

CDD 621.3822

Elaborada pelo bibliotecário Leandro Mota de Menezes – CRB/7 nº 5281

Acknowledgments

First and foremost, I would like to express my deep gratitude to CEFET-RJ and UFF, for providing the doctoral program that made this study possible. Being a part of these esteemed institutions and collaborating with its faculty and fellow scholars was instrumental in shaping the journey of this work.

I cannot move forward without extending my heartfelt appreciation to my supervisor, DSc. Diego Barreto Haddad. His dedication, competence, and unwavering passion for his profession are truly inspiring. Throughout this academic odyssey, Diego has been an exceptional mentor who not only provided guidance but also instilled encouragement, presented challenges, and unequivocally believed in my potential. His sage counsel and relentless support were pivotal to the success of this project. His qualities as a teacher and researcher are admirable, and I am thankful for the opportunity to learn from him.

I want to express a special thank you to my wife, Lívia, and my children, Laura and Lucas. They were not only a source of unwavering support but also an endless source of inspiration during the most challenging moments of this journey.

Additionally, I would like to express my sincere gratitude to the esteemed members of my doctoral committee. Their valuable insights and feedback have been instrumental in shaping this work. I am truly honored to have had the opportunity to learn from their expertise and benefit from their guidance.

In closing, I wish to express my gratitude to all who, in diverse ways, contributed to the realization of this work. This is a milestone in my life, and I humbly acknowledge that it would not have been achievable without the unwavering support and inspiration of those mentioned above. I hope that this work can, in some way, give back to the academic community and society at large. I extend my heartfelt thanks to all.

Filipe Igreja

“The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge.”

Stephen Hawking

CONTENTS

1	**Introduction**	**11**
1.1	Objectives	14
1.2	Publications	16
1.3	Mathematical notation	16
1.4	Structure of the thesis	17
2	**LMS Fundamentals**	**19**
2.1	The Wiener solution	22
2.2	The steepest descent method	24
2.3	The LMS algorithm	26
2.4	Concluding remarks	30
3	**Second-Order Stochastic Analyses of the LMS: An Overview**	**31**
3.1	Introduction	31
3.2	Standard Mean-Square Analysis	31
3.3	Exact Expectation Analysis	35
3.4	Concluding Remarks	40
4	**Analysis of the LMS-based Identification of Nonlinear Plants**	**41**
4.1	Introduction	41
4.2	Nonlinear System Model (NSM)	42
4.3	Exact Expectation Analysis of LMS-Based NLS Identification	43
4.4	Results	46
4.4.1	Evaluation of Different Noise Nonlinear Terms	49
4.5	Conclusions	53
5	**An Exact Expectation Model for the LMS Tracking Abilities**	**55**
5.1	Introduction	55

5.2 Time-variant Plant Model 57
5.3 Standard LMS Tracking Analysis 58
5.4 Exact Expectation Tracking Analysis 60
5.4.1 Mean Weight Analysis 61
5.4.2 Mean Square Analysis 63
5.5 Theoretical Derivation of an Optimum Variable Step-Size Sequence 69
5.5.1 Theoretical VSS Sequence - Version I 70
5.5.2 Theoretical VSS Sequence - Version II 71
5.6 Results 72
5.7 Conclusions 82

6 Exact-Analysis Based Estimation of the Weight Error Covariance Matrix 84
6.1 Introduction 84
6.2 Scenarios 87
6.2.1 Scenario 1 88
6.2.2 Scenario 2 90
6.2.3 Scenario 3 91
6.2.4 Scenario 4 92
6.2.5 Scenario 5 94
6.2.6 Scenario 6 95
6.3 Concluding Remarks 96

7 Conclusions 97
References 99

LIST OF TABLES

Table 1 – Scenarios properties. 50

Table 2 – Computational performance comparison. 67
Table 3 – Amount of equations necessary to MSD computation. 68

Table 4 – Scenarios. 87
Table 5 – Number of equations for each scenario. 87
Table 6 – Results for Scenario 1. 89
Table 7 – Results for Scenario 2. 91
Table 8 – Results for Scenario 3. 92

LIST OF FIGURES

Figure 1 – Roadmap. 18

Figure 2 – Transversal filter structure. 20
Figure 3 – Adaptive filter block diagram. 21

Figure 4 – Evolution of the mean value of adaptive weight. 46
Figure 5 – LMS MSE evolution with $N = 3$ and $\beta = 0.085$. 47
Figure 6 – Steady-State MSE for the $N = 2$ configuration. 47
Figure 7 – Divergence probability for $N = 3$. 48
Figure 8 – MSE - Scenario 1 50
Figure 9 – MSE - Scenario 2 51
Figure 10 – MSE - Scenario 3 52
Figure 11 – MSE - Scenario 4 53

Figure 12 – Evolution of $\mathbb{E}[w_0(k)]$, with 10^6 independent Monte Carlo trials. . . 73
Figure 13 – Evolution of the MSE. 73
Figure 14 – Divergence probability as a function of β. 74
Figure 15 – Steady-state MSE as a function of the step size. 76
Figure 16 – Misadjustment evolution (in dB). 77
Figure 17 – Steady-state MSE as a function of the step size. 77
Figure 18 – Theoretical steady-state $\text{MSE}_{\text{ss},q}$. 78
Figure 19 – Theoretical steady-state $\text{MSE}_{\text{ss,std}}$. 78
Figure 20 – Optimal step size choice in function of σ_q^2. 79
Figure 21 – Theoretical sequences $\beta(k)$ obtained by `EEA` and `IA` analyses. . . 80
Figure 22 – MSE evolution comparing IA versus EEA. 81
Figure 23 – Theoretical sequences prescription comparison. 81
Figure 24 – MSD using step size sequences suggested by IA and EEA. . . . 82

Figure 25 – MSD (dB) for Scenario 1. 90

ADVANCES ON THE EXACT EXPECTATION ANALYSIS OF ADAPTIVE FILTERING ALGORITHMS

Filipe Igreja

Advisor:

Diego B. Haddad, D.Sc.

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ, como parte dos requisitos necessários à obtenção do título de doutor.

Adaptive filtering algorithms are flexible and low-cost solutions for many applications where prior knowledge about the statistics of the involved signals is absent. Their sophisticated learning dynamics challenge the design of accurate stochastic models. Most of these models, in the literature adopt the independence theory, which engenders discrepancies between theoretical predictions and actual results. This thesis avoids such a ubiquitous assumption, utilizing the so-called exact expectation analysis, a computationally intensive method that generates more accurate predictions. For the first time, this analysis is employed to model both the case of nonlinear plants and the tracking configuration. Further, it also provides a deterministic sequence of step sizes that optimize the mean-square performance of the LMS under time-variant settings.

Adaptive Filtering; Exact Expectation Analysis; Identification of nonlinear systems; Identification of time-variant systems.

Rio de Janeiro,

November, 2023

1- Introduction

The term “estimator” or “filter” is usually utilized when one intends to extract information of interest from (frequently noisy) data. Filtering procedures have a variety of applications in fields such as communications, radar, imaging, biomedical engineering, and financial time series forecasting [1; 2], among many others [3]. The task is often difficult, especially when the environment from where the information of interest is collected may change over time. Further complications arrive when prior knowledge about a dynamic process is limited. Adaptive filter algorithms are popular schemes tailored for these challenging configurations. They are indeed based on an approach that is somewhere between the parametric and nonparametric ones [4].

Adaptive filtering algorithms are particularly advantageous when the signals or the involved system vary over a period of time. Depending on the application and how the adaptive processor is arranged they may work on: prediction [5], noise-cancelling [6], system identification [7], adaptive beamforming [8], adaptive spectrum analysis [9], among several others tasks. The plethora of applications for adaptive filtering is unquestionably important to fields such as signal processing, communications, and control systems, which are as much part of military systems [10] as of electronic devices utilized by an average person on day-to-day life [11]. As an example, it is possible to mention cell phone noise cancellation, which utilizes adaptive filtering schemes. Therefore, when its loudspeaker is activated during a conversation, its microphone will be exposed to the sound produced by the loudspeaker. In this case, without the noise cancellation, it would reproduce echo *ad infinitum*, which harms the communication’s intelligibility.

A detailed list containing all adaptive filtering applications would be too extensive since it only grows over time. One factor that explains such a fact is that they rely on software and microprocessor units, which are becoming cheaper through the years. Therefore, one is able to implement adaptive filtering algorithms on virtually any engineering application, where it is possible to collect data and insert it back modified by an algorithm.

There is a multitude of adaptive filtering algorithms and many more will yet be invented. It is expected that all of them operate under environments in a stable manner,

which is not a trivial task due to the inherent feedback in their learning procedure. This means that many of these algorithms may not reach the purpose they were intended for, which requires tools to analyze their efficiency, tracking abilities, and stability.

Many performance metrics are commonly utilized for assessment purposes, such as the mean square deviation, the mean square error, and the convergence rate. In this thesis, the learning capabilities of the least mean square (LMS) algorithm will be the focus, since no other adaptive algorithm has been so successfully placed in commercial products [12].

Abstracting hardware idiosyncrasies, this thesis advances stochastic models that predict some operating characteristics of a specific design. Thus, it focuses on stochastic models of adaptive filtering algorithms rather than on their physical implementation. Such models usually rely on theoretical constructs and statistical hypotheses that are employed to *explain* and/or *predict* phenomena and are judged according to their ability to do so. Most of the time, modelers intend to derive practical guidelines from their analyses. In other words, usually the *explanation* ability of one model for an adaptive filter is emphasized at the expense of their *prediction* capabilities. Interestingly, the opposite is quite true in contemporary physics, where often the prediction ability of a theoretical model is the ultimate goal.

The discrepancy between the predictions of a specific stochastic model and the actual behavior of an adaptive filter occurs since analyses are conducted under some statistical hypotheses. Some hypotheses are too strong and are violated in effect. Since the learning dynamics of an adaptive filter is extremely sophisticated, the motivation for using such hypotheses is twofold since they *i)* avoid lengthy algebraic manipulations and *ii)* help to derive an easy-to-understand guidance to the designer. Due to *ii)*, one may conclude that there exists a conflict between the explanation and accuracy dimensions of a stochastic model. Due to this tension, the learning process of the LMS, the most simple adaptive filter, remains understood very poorly [13].

Although it is not our intention to disregard the significance of obtaining qualitative insights about one algorithm's behavior, this thesis is more unconventional, in the sense that it is committed to extracting more accurate predictions from these models. It should be clear to the reader that this goal is not an unfounded one, since questions such as "When one chooses this specific value of the step size, will the algorithm diverge?" are indeed very important. We should not necessarily understand prediction and explanation

capabilities in zero-sum terms. Learning that explanation capabilities are important is not a reason to think predictive capabilities are unimportant, since a realistic description of the learning dynamics still is fundamental. Thus, this thesis intends to offer several predictions about the learning characteristics of the LMS, to answer important questions like the above in a manner much more accurate than traditional approaches do. At any rate, it should be noted that divergence is a catastrophic event that renders the adaptive filter useless and even harmful to the intended application, and it is expected that at least some models might predict accurately whether it occurs or not.

Out of the hypotheses mentioned above, the most troublesome one is the widespread independence assumption, commonly encountered in the stochastic approximations field [14] and widely used in the dynamic convergence analysis of adaptive algorithms [15]. Such a strong stochastic hypothesis states that the adaptive weights are statistically independent of the input samples that currently feed the adaptive structure. In fact, such a hypothesis is clearly violated in practice in tapped-delay line structures, since they impose a *deterministic coherence* between consecutive input vectors. Due to this fact, some authors claim that the expression “independence *heuristic*” is more adequate than “independence assumption” [16].

It turns out that the independence assumption renders the resulting stochastic model inaccurate when a non-infinitesimally small step size factor is adopted, especially when non-normalized algorithms (such as the LMS) are utilized. However, some potential advantages of adaptive filtering schemes may become apparent only for *large* step sizes [17]. Furthermore, instability issues are commonly encountered when the step size presents a large magnitude. Thus, the traditional analyses cannot even accurately predict an upper bound for the step size that guarantees stability. Such an issue, widely recognized in standard textbooks about adaptive filtering [3; 18; 19], should raise concerns since divergence is a very harmful kind of failure.

This thesis skirts such limitations using the exact expectation analysis (EEA), in which references [20; 21; 22] pioneered. In loose terms, the EEA is a systematic constructive procedure that utilizes algebraic manipulations for the generation of a system of equations that model in a precise manner the learning behavior of an adaptive filter. The analysis is computationally intensive, requiring a judicious and efficient design of the programming code. It is also positioned at an extreme point of the explanation *versus* prediction dimensions, giving much more weight to the latter than to the former

one. Thus, its enhanced prediction capabilities come with a price, since the amount of qualitative factors that drive the algorithm performance is strongly reduced under the devised approach.

1.1- Objectives

Succinctly, the main goals of this thesis are:

1. To devise an accurate model of the LMS learning capabilities when the plant it intends to identify contains nonlinearities.
2. To provide an accurate model of the LMS learning behavior when the unknown ideal plant varies according to a first-order Markovian model, even when the unknown system presents a high nonstationary degree.
3. To provide an optimization problem whose solution is a deterministic sequence of step sizes that maximizes the LMS performance.
4. To analyze asymptotic properties of the weight error covariance matrix, without the usage of the widespread independence assumption.

It should be emphasized that all these contributions are under the umbrella of the exact expectation analysis so that their predictions are precise even when the learning step size presents large values. In the following, we present some motivations for the aforementioned goals.

The motivation for studying the LMS learning abilities in the case of a nonlinear ideal plant derives from the fact that frequently the linear model is not realistic, which is especially important when the input signal range is large or the nonlinearities' impact on the single-valued output is not negligible [23]. Several nonlinear phenomena might cause such nonlinearities, such as operation near the saturation region, intermodulation distortion, and diffusion capacitance, which are more important when cheap transducers are employed [24; 25; 26; 27]. Since a linear model is assumed for the plant to identify, it is important to obtain theoretical guarantees concerning the algorithm's performance when such a condition is not fulfilled.

One of the main advantageous features of adaptive filters is their tracking ability. Unfortunately, most published analyses focus on the stationary configuration, where the involved ideal plant is time-invariant. In practice, often this is not the case. Thus, this thesis for the first time extends the EEA approach to modeling the tracking scenario, where the ideal plant varies over time according to a random walk model. As will be seen, the adoption of the EEA brings an additional benefit, since it permits to model the case where the nonstationary degree of the ideal plant is large. In such a configuration, the predictions offered by traditional analyses are not precise.

The choice of a fixed step size engenders a trade-off between convergence rate and asymptotic performance. This trade-off can be circumvented by the adoption of variable step size schemes, whose assessment often is carried out by comparisons with theoretical optimal step size sequences [28]. In this thesis, such theoretically derived sequences are obtained without the ubiquitous independence assumption. Since this assumption is of little consequence only for sufficiently small step sizes does erode sever- al results already reported in [35–39; 41; 42] as far as the optimal selection of the step size is concerned. Further, the derived step size sequences are obtained for the non-stationary case, where the ideal plant is time-variant. In such a case, the step size should not converge asymptotically to zero, since it is necessary to preserve the algorithm tracking ability. Further, the computation of such sequences for each adaptive tap was carried out. After all, the usage of the EEA method avoids instability issues, which may be encountered in the first phase of the learning when the standard independence assumption is employed.

Lastly, computing the covariance matrix is one of the initial tasks in high-dimensional estimation analysis, as it offers insights into the uncertainties of the estimation process. In adaptive filtering, the weight error covariance matrix plays a vital role in determining how the filter adjusts to varying input statistics and noise levels. This matrix provides an objective assessment of the uncertainties associated with the estimated filter coefficients. This thesis presents an accurate modeling approach for this matrix in a steady-state regime by employing the exact expectation model, enabling precise predictions even with large step sizes.

1.2- Publications

The publications directly related to this thesis are

- Silva, Thiago Teodoro Pereira; Lara, Pedro; **Igreja, Filipe**; Oliveira, Fernanda; Tarrataca, Luís; Haddad, Diego B. “An Exact Expectation Model for the LMS Tracking Abilities”, IEEE Transactions on Signal Processing, v. 68, p. 1-1, 2020.
- Lara, Pedro; **Igreja, Filipe**, Silva, Thiago T. P.; Tarrataca, Luís; Haddad, Diego B. “Exact Expectation Analysis of the LMS Adaptive Identification of Nonlinear Systems,” Electronics Letters, v. 56, p. 45-48, 2019.

The following publication, although related to this thesis, derived from the master thesis of the author:

- Lara, Pedro; **Igreja, Filipe**; Tarrataca, Luís ; Haddad, Diego Barreto; Petraglia, Mariane R. “Exact Expectation Evaluation and Design of Variable Step-Size Adaptive Algorithms”, IEEE Signal Processing Letters, v. 26, p. 74-78, 2019.

It should be noted that there are additional references that cannot be considered a straightforward result of this thesis, since their main contributions (even in the signal-processing domain) were carried out by another researchers:

- T. T. P. Silva, **F. Igreja**, P. Lara, L. Tarrataca, A. Kar and D. B. Haddad, “On the Skewness of the LMS Adaptive Weights,” in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 68, no. 8, pp. 3022-3026, Aug. 2021.
- Resende, Leonardo C.; Pimenta, Rodrigo M. S.; **Igreja, Filipe**; Siqueira, Newton N.; Petraglia, Mariane R.; Haddad, Diego B. “Analysis of the Least Mean Squares Algorithm with Reusing Coeficient Vector”, Signal Processing, vol. 202, 2023.

1.3- Mathematical notation

Throughout this thesis, vector and matrices are represented with lowercase and uppercase bold fonts, respectively, while scalars are denoted by italics. The symbol $\otimes$

stands for the Kronecker product of matrices, and $\mathbb{E}[\cdot]$ is the expectation operator. $(\cdot)^T$ denotes transpose and the vec$(\mathbf{A})$ operator generates a column-type vector by successively stacking the columns of the matrix $\mathbf{A}$ (*e.g.*, vec$\left(\begin{bmatrix} a & c \\ b & d \end{bmatrix}\right) = [a\ b\ c\ d\ ]^T$). The symbol $\mathbf{I}_N$ means the $N \times N$ identity matrix, whereas $\mathbf{0}_N$ denotes a null column vector with dimensions $N \times 1$. Operator Tr$(\boldsymbol{A})$ denotes the trace of matrix $\boldsymbol{A}$.

1.4- Structure of the thesis

This thesis is structured as follows: Chapter 2 describes the fundamentals of the LMS algorithm. Chapter 3 presents some analyzes of this algorithm, with no claim to exhaust the topic. Chapter 4 discusses about the EEA-based modeling of the LMS learning capabilities when the plant it intends to estimate contains nonlinear effects. Chapter 5 describes the devised model for the LMS tracking abilities when the unknown transfer function varies according to a random walk model. Chapter 5 also presents some contributions regarding to the optimization of step size sequences based on the exact expectation analysis. Chapter 6 describes the devised stochastic model for the asymptotic weight error covariance matrix. At last, Chapter 7 contains the concluding remarks of the thesis.

For a better disposition of how the chapters were organized based on the published articles:

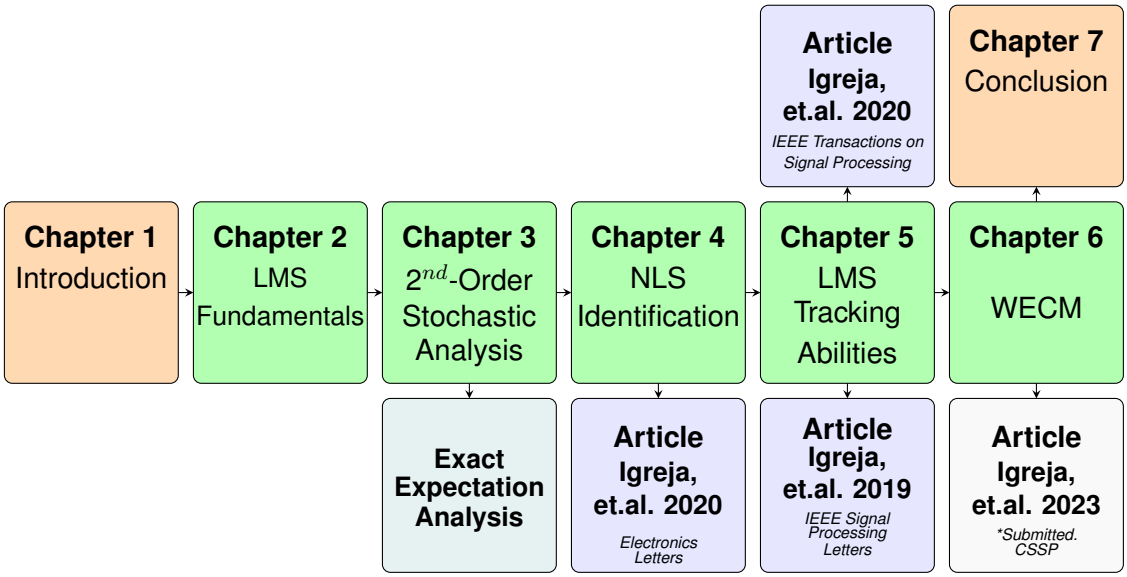


Figure 1 – In orange: chapter without original contribution. In green: are chapters with original contributions. In lilac: articles resulting from the research that have been published in journals. In light gray: article resulting from the research that has been submitted to journals.

2- LMS Fundamentals

In the 1950s, online estimation or adaptive algorithms developed in two separate disciplines: control and signal processing [16]. More recently, the machine learning community has been involved with such methods, and indeed the original motivation for devising the LMS derived from pattern-recognition problems. This thesis concentrates on the signal processing aspects, which in general assume in an open loop setting, whereas in control discipline a closed looping setting is the norm [16]. Adaptive filtering techniques can be applied to several problems, such as temporal series prediction, channel equalization, adaptive control, and system identification. The following discussions are constrained to system identification tasks.

The operation of an adaptive filtering scheme can be characterized by four aspects:

1. The signal that will be processed by the filter;
2. The structure that defines the filter output from an input signal;
3. Parameters inside a mathematical structure that can be altered to modify input-output relation;
4. The adaptive algorithm itself describes the parameters adjusted to the next iteration.

In this thesis, the input signal is assumed to be wide-sense stationary, which is a common assumption in the open literature, with a few exceptions [29; 30; 31]. The whiteness of the input signal is a very strong assumption utilized by numerous references [32; 33; 34; 35; 36] and [37] claims that the theoretical results obtained under this assumption are suggestive of fundamental design issues. This simplification is avoided in this thesis, which permits to incorporate the coloring of the input signal in the model. It is an interesting feature of an analysis, since the input signal's autocorrelation level may decrease the expected convergence rate of the LMS, for example [30].

The most popular structure in adaptive filtering schemes is the tapped-delay line (also known as the transversal structure), which processes at each iteration the input

vector $\boldsymbol{x}(k)$ defined as

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

where N is the length of the adaptive filter.

The disposal of the adaptive parameters $w_i(k) \in \mathbb{R}$ (which should be adapted purposefully) in a proper order gives birth to the adaptive weight vector $\boldsymbol{w}(k) \in \mathbb{R}^N$, designated by

$$\boldsymbol{w}(k) \triangleq \begin{bmatrix} w_0(k) & w_1(k) & w_2(k) & \dots & w_{N-1}(k) \end{bmatrix}^T, \tag{2}$$

which leads to a concise formulation of the output $y(k) \in \mathbb{R}$ of the adaptive filter at the k-th iteration, as an inner product:

$$y(k) = \sum_{i=0}^{N-1} w_i(k)x(k-i) = \boldsymbol{w}^T(k)\boldsymbol{x}(k), \tag{3}$$

whose computation process is depicted by Figure 2.

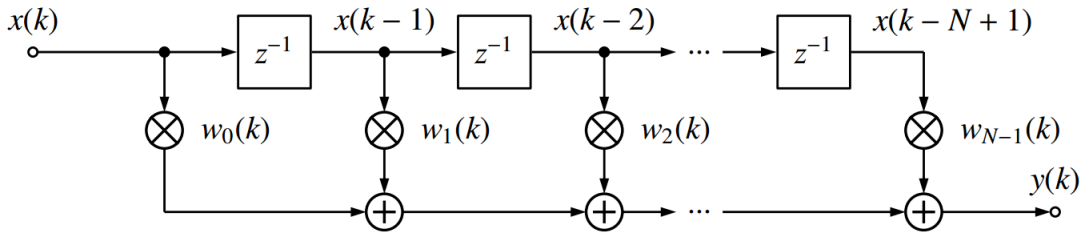


Figure 2 – Transversal filter structure.

In supervised settings, a reference signal $d(k) \in \mathbb{R}$ is available, and the main goal of the adaptive filter is that its output $y(k)$ approaches $d(k)$ as learning takes place. As mentioned earlier, throughout this thesis, a system identification scenario is assumed. In the case of a linear finite impulse response of the unknown plant, the reference signal is often assumed to have risen from a linear regression of the form

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

where $\boldsymbol{w}^\star \in \mathbb{R}^N$ contains the coefficients of the ideal plant (whose transfer function can be represented as $W^\star(z) = w_0^* + w_1^* z^{-1} + \ldots + w_{N-1}^* z^{-N+1}$) and $\nu(k)$ models the measurement noise. Note that it is assumed that the length of the adaptive

filter is equal to the length of the unknown ideal plant, a common assumption in the literature [22; 38; 39; 40; 41; 42], although some analyses are not constrained by this condition [43; 44].

Since a system identification scenario (depicted in Figure 3) is assumed, it is expected that the adaptive filter $\boldsymbol{w}(k)$ converges asymptotically to $\boldsymbol{w}^{\star}$. Due to the stochastic nature of the adaptive identification, such a convergence is not strict, and it is expected that the adaptive filter fluctuates around the optimal solution in steady-state, presenting a Brownian-like motion.

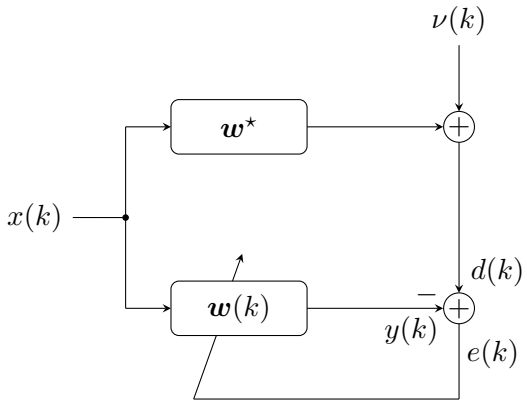


Figure 3 – Adaptive filtering block diagram structure applied on system identification.

Several adaptive algorithms have been proposed to train the considered structure. In general, they rely on the error signal $e(k) \in \mathbb{R}$:

$$e(k) \triangleq d(k) - y(k) = d(k) - \boldsymbol{w}^{T}(k)\boldsymbol{x}(k), \tag{5}$$

which contains information correlated with the quality of the current solution $\boldsymbol{w}(k)$.

The focus of this thesis is the most popular of the adaptive schemes: the LMS algorithm, whose relevance is emphasized by [18]:

> "The LMS filter and its variants are an important part of the adaptive signal processing kit of tools, not just for the past 40 years but for many years to come. Simply put, the LMS filter has withstood the test of time."

2.1- The Wiener solution

The instantaneous error signal carries information about how good the current adaptive weight vector $\boldsymbol{w}(k)$. Since it only establishes a *stochastic* kind of assessment, ideally one should evaluate the solution quality based on the *mean* behavior. The most popular strategy for this purpose begins with the square of the error:

$$e^2(k) = d^2(k) - 2\boldsymbol{x}^T(k)\boldsymbol{x}(k)d(k) + \boldsymbol{w}^T\boldsymbol{x}(k)\boldsymbol{x}^T(k)\boldsymbol{w}, \tag{6}$$

where $\boldsymbol{w} \in \mathbb{R}^N$ is considered to be time invariant. Due to the attractive mathematical properties of quadratic cost functions, often one considers the optimal solution the one that minimizes the mean square error (MSE)

$$\begin{aligned} \xi &\triangleq \mathbb{E}\left[e^2(k)\right] = \mathbb{E}\left[d^2(k)\right] - 2\mathbb{E}[d(k)\boldsymbol{x}^T(k)]\boldsymbol{w} + \boldsymbol{w}^T\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\boldsymbol{w} \\ &= \mathbb{E}\left[d^2(k)\right] - 2\boldsymbol{p}^T\boldsymbol{w} + \boldsymbol{w}^T\boldsymbol{R}\boldsymbol{w}, \end{aligned} \tag{7}$$

where

$$\boldsymbol{p} \triangleq \begin{bmatrix} d(k)x(k) \\ d(k)x(k-1) \\ \vdots \\ d(k)x(k-N+1) \end{bmatrix} \tag{8}$$

designates the cross-correlation between the input signal and the desired response and

$$\boldsymbol{R} \triangleq \mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] = \mathbb{E}\left\{\begin{bmatrix} x^2(k) & \dots & x(k)x(k-N+1) \\ x(k-1)x(k) & \dots & x(k-1)x(k-N+1) \\ \vdots & & \vdots \\ x(k-N+1)x(k) & \dots & x^2(k-N+1) \end{bmatrix}\right\} \tag{9}$$

denotes the input autocorrelation matrix.

The MSE is minimized when one sets its gradient w.r.t. $\boldsymbol{w}$ to zero, which leads to the *Wiener-Hopf* equation

$$\boldsymbol{R}\boldsymbol{w}_{\text{opt}} = \boldsymbol{p}, \tag{10}$$

whose solution $\boldsymbol{w}_{\text{opt}}$ can be obtained in a closed-form manner when $\boldsymbol{R}$ has an inverse:

$$\boldsymbol{w}_{\text{opt}} = \boldsymbol{R}^{-1}\boldsymbol{p}, \tag{11}$$

also termed as the “Wiener weight” vector or simply “Wiener solution”.

Replacing $\boldsymbol{w}$ by $\boldsymbol{w}_{\text{opt}}$ and $\boldsymbol{R}\boldsymbol{w}_{\text{opt}}$ by $\boldsymbol{p}$ in (7) leads to

$$\xi_{\text{min}} = \mathbb{E}\left[d^2(k)\right] - \boldsymbol{w}_{\text{opt}}^T\boldsymbol{p} = \mathbb{E}\left[d^2(k)\right] - \boldsymbol{w}_{\text{opt}}^T\boldsymbol{R}\boldsymbol{w}_{\text{opt}} = \mathbb{E}\left[d^2(k)\right] - \boldsymbol{w}_{\text{opt}}^T\boldsymbol{R}\boldsymbol{w}_{\text{opt}}, \tag{12}$$

which is the minimum mean-square error that can be attained by the Wiener solution.

Considering $e_0(k)$ as the estimation error when the filter weights are set equal to the Wiener solution, it can be proved that

$$\mathbb{E}\left[e_0(k)\boldsymbol{x}(k)\right] = \boldsymbol{0}. \tag{13}$$

Eq. (13) is a version of the widely known *orthogonality principle*, which states that the optimum estimation error is uncorrelated with the filter inputs. If one denotes as $y_0(k)$ the output obtained by the Wiener solution, this principle has a useful corollary:

$$\mathbb{E}[e_0(k)y_0(k)] = 0, \tag{14}$$

which shows that the estimation error and the output of the Wiener solution are uncorrelated with each other.

After all, the computation of the Wiener solution (11) requires the exact statistics of both $\boldsymbol{R}$ and $\boldsymbol{p}$. For a given data sample, it can be obtained after estimating both $\boldsymbol{R}$ and $\boldsymbol{p}$, which is the main strategy of the *exact* least-mean-squares family of solutions. An additional constraint for the practical implementation of the Wiener solution is that it considers that the input-output mapping is time-invariant, so there exists only one optimal solution for the entire dataset. Additionally, evaluating the Wiener solution is computationally intensive. All of these limitations can be circumvented by adaptive filtering approaches, which are low-cost solutions with tracking capabilities that do not demand prior knowledge about the statistics of the involved signals.

2.2- The steepest descent method

Since the MSE is a convex function of the set of parameters, its minimization can be obtained by gradient methods. The steepest descent is one of them. It makes recursive modifications of the weight vector along the direction of the negative gradient of the MSE [3], which utilizes the gradient of the MSE w.r.t. $\boldsymbol{w}$:

$$\nabla_{\boldsymbol{w}}\xi = 2\boldsymbol{R}\boldsymbol{w} - 2\boldsymbol{p} \tag{15}$$

in order to modify the current estimate $\boldsymbol{w}(k)$ by

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \beta\nabla_{\boldsymbol{w}(k)}\xi \tag{16}$$

where $\beta \in \mathbb{R}_+$ is an user-defined positive scalar termed as the “step size” or the “learning factor”.

Substituting Eq. (15) in (16) leads to

$$\boldsymbol{w}(k+1) = (\boldsymbol{I} - 2\beta\boldsymbol{R})\,\boldsymbol{w}(k) + 2\beta\boldsymbol{p}. \tag{17}$$

The convergence of the steepest descent can be better understood when recursion (17) is rewritten in terms of the deviation vector $\tilde{\boldsymbol{w}}(k) \in \mathbb{R}^N$, defined by

$$\tilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}_{\text{opt}} - \boldsymbol{w}(k). \tag{18}$$

The combination of (18) and (17) leads to

$$\tilde{\boldsymbol{w}}(k+1) = \tilde{\boldsymbol{w}}(k) + 2\beta\boldsymbol{R}\boldsymbol{w}(k) - 2\beta\boldsymbol{p}. \tag{19}$$

Since $\boldsymbol{w}(k)$ can be expressed as $\boldsymbol{w}_{\text{opt}} - \tilde{\boldsymbol{w}}(k)$ (see (18)), Eq. (19) may be converted into

$$\tilde{\boldsymbol{w}}(k+1) = \tilde{\boldsymbol{w}}(k) + 2\beta\boldsymbol{R}\boldsymbol{w}_{\text{opt}} - 2\beta\boldsymbol{R}\tilde{\boldsymbol{w}}(k) - 2\beta\boldsymbol{p}. \tag{20}$$

Further, as long as $\boldsymbol{R}\boldsymbol{w}_{\text{opt}} = \boldsymbol{p}$ (see (10)), one obtains the recursion

$$\tilde{\boldsymbol{w}}(k+1) = [\boldsymbol{I} - 2\beta\boldsymbol{R}]\,\tilde{\boldsymbol{w}}(k). \tag{21}$$

Equation (21) may be utilized to obtain a better understanding of the steepest descent's learning convergence in terms of the eigenvalues of the input autocorrelation matrix $\boldsymbol{R}$. Such a matrix can be decomposed by

$$\boldsymbol{R} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T, \tag{22}$$

where the i-th column of $\boldsymbol{Q}$ contains the orthonormal eigenvector $\boldsymbol{q}_i \in \mathbb{R}^N$, whose corresponding eigenvalue λ_i is positionated at the i-th element of the main diagonal of the diagonal matrix $\boldsymbol{\Lambda}$. By using standard manipulations [3], it is possible to establish the following recursion

$$\boldsymbol{v}(k+1) = (\boldsymbol{I} - 2\beta\boldsymbol{\Lambda})\,\boldsymbol{v}(k), \tag{23}$$

where

$$\boldsymbol{v}(k) \triangleq \boldsymbol{Q}^T\tilde{\boldsymbol{w}}(k) \tag{24}$$

designates the rotated deviation vector. Eq. (23) decouples the distinct converge modes of the steepest descent, which leads to independent scalar recursive equations

$$v_i(k+1) = (1 - 2\beta\lambda_i)v_i(k), \tag{25}$$

for $i \in \{0, 1, \ldots, N-1\}$, where $v_i(k)$ is the i-th element of $\boldsymbol{v}(k)$. In terms of the initial values $v_i(0)$, (25) may be employed to describe $v_i(k)$ in a closed-form manner:

$$v_i(k) = (1 - 2\beta\lambda_i)^k v_i(0), \tag{26}$$

which implies that the convergence of $\boldsymbol{w}(k)$ to $\boldsymbol{w}_{\text{opt}}$ occurs if $\boldsymbol{v}(k)$ convergences to the null vector. Such a convergence is guaranteed when

$$|1 - 2\beta\lambda_i| < 1 \Rightarrow -1 < 1 - 2\beta\lambda_i < 1 \Rightarrow 0 < \beta < \frac{1}{\lambda_i}, \tag{27}$$

for $i \in \{0, 1, \ldots, N-1\}$. The learning factor should satisfy the N constraints (27) in order to obtain asymptotic convergence. The more strict constraint is encountered when λ_i

attains its maximum value $\lambda_{\max}$. In other words, the N conditions (27) can be succinctly expressed as

$$0 < \beta < \frac{1}{\lambda_{\max}}. \tag{28}$$

Whenever condition (28) is fulfilled, the scalar parameters $v_i(k)$ decay exponentially as long as k increases. Note that (25) conditions the stability in terms of the input statistics. Such a relationship gives important insights for the designer and sometimes is provided by some stochastic analyses, especially under simple cases or when statistically strong hypotheses are adopted.

2.3- The LMS algorithm

The LMS algorithm can be obtained by applying the stochastic gradient technique to the instantaneous quadratic error:

$$\begin{aligned} \boldsymbol{w}(k+1) &= \boldsymbol{w}(k) - \beta \nabla_{\boldsymbol{w}(k)} \left[\frac{1}{2} e^2(k) \right] \\ &= \boldsymbol{w}(k) - \beta e(k) \nabla_{\boldsymbol{w}(k)} \left[e(k) \right] \\ \Rightarrow \boldsymbol{w}(k+1) &= \boldsymbol{w}(k) + \beta \boldsymbol{x}(k) e(k), \end{aligned} \tag{29}$$

a recursive equation that completely describes the LMS adaptive process, which requires $2N$ sums and $2N+1$ multiplications per iteration. Although the update process of the LMS is very simple (at least in computational terms), its stochastic and nonlinear features imply that its mathematical analysis is complicated. Due to the feedback caused by the usage of the error signal $e(k)$, the adaptive weights depend on input data and the system behavior differs from that of a linear system [18]. Note that we opted to utilize the half of the instantaneous quadratic error in (29) to ensure coherence with traditional treatments of the open literature. As a consequence, some minor differences w.r.t. our previous analyses of the steepest descent will occur.

In an algorithmic view, the LMS algorithm sequentially performs three steps:

1. Filtering: $y(k) = \boldsymbol{w}^T(k)\boldsymbol{x}(k)$
2. Error estimation: $e(k) = d(k) - y(k)$

3. Adaptive vector update: $\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta e(k)\boldsymbol{x}(k)$

The simplicity of the LMS explains why it is the most popular adaptive filtering scheme. Its first-order behavior can be studied with a very similar approach to the one previously utilized for the study of the steepest descent. Utilizing (4) as a valid model for the reference signal and assuming the system identification task (which implies that $\boldsymbol{w}_{\text{opt}} = \boldsymbol{w}^{\star}$), the error signal can be written as a function of the deviation vector:

$$e(k) = \tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k) + \nu(k). \tag{30}$$

By combining (30), (29) and (18), the following recursion on the deviation vector can be derived

$$\tilde{\boldsymbol{w}}(k+1) = \left[\boldsymbol{I} - \beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\tilde{\boldsymbol{w}}(k) + \beta\boldsymbol{x}(k)\nu(k), \tag{31}$$

which is a relation that presents some similarities with (21), especially when the driving force $-\beta\boldsymbol{x}(k)\nu(k)$ is not considered. Despite this, the fact that the samples of the input vector $\boldsymbol{x}(k)$ and the measurement noise $\nu(k)$ are indeed realizations of stochastic processes has profound implications. The first of them is that a first-order analysis (such as the one employed in the steepest descent algorithm) is no longer sufficient for the characterization of the algorithm since the variance of the adaptive estimator should be taken into account. Due to this, in the next chapter, second-order analyses of the LMS are described in more detail.

Another implication is that the trajectory of the adaptive coefficients differs from one realization of the learning process to another one. In fact, there are infinite possible trajectories, since both input data and measurement noise are sampled from distinct stochastic processes. In this context, the characterization of the *mean* behavior brings a measure of centrality about the *average* trajectory around which the coefficients evolve along the iterations. Such a first-order characterization can be carried out by the application of the expectation operator in (31), which leads to

$$\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] = \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right] - \beta\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\right] + \beta\mathbb{E}\left[\boldsymbol{x}(k)\nu(k)\right], \tag{32}$$

which is a complex recursion, especially due to the presence of the third-order moment vector $\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\right]$.

To maintain the mathematics tractable, LMS advanced models that can be found

in the literature often rely on some simplifying assumptions, that can provide insights about factors that enhance algorithmic performance even when the resulting predictions are not accurate [14; 45; 46]. The first of these hypotheses is the noise assumption (NA):

Noise assumption (NA): The measurement noise $\nu(k)$ is a zero-mean i.i.d. stochastic process, which is statistically independent from the input signal.

It is noteworthy that NA is a standard assumption in analyses of adaptive filtering algorithms, being often satisfied in practice [47; 48]. However, it is not valid in the presence of physical nonlinear phenomena, such as operation near the saturation region, diffusion capacitance, and intermodulation distortion [49]. In these cases the signal $\nu(k)$ in (30) contains nonlinear components of the excitation data.

Under NA, recursion (32) can be simplified:

$$\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] = \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right] - \beta\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\right], \tag{33}$$

although dealing with the joint moment $\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\right]$ be still a difficult task. Clearly $\tilde{\boldsymbol{w}}(k)$ presents a statistical dependency w.r.t. $\boldsymbol{x}(k)$, since $\boldsymbol{x}(k-1)$ is explicitly employed to obtain $\boldsymbol{w}(k)$:

$$\boldsymbol{w}(k) = \boldsymbol{w}(k-1) + \beta\boldsymbol{x}(k-1)e(k-1), \tag{34}$$

and there exists a deterministic coherence between $\boldsymbol{x}(k-1)$ and $\boldsymbol{x}(k)$ in a tapped-delay structure. In fact, such vectors have $N-1$ elements in common, although in distinct positions, since the transversal structure induces a shift between consecutive input vectors.

The most critical (and perhaps one of the most useful) statistical hypothesis is the *independence assumption*, which considers that the excitation data is statistically independent of the adaptive weights. Despite its wide usage in the stochastic approximations field [14], the independence assumption performs unreliable predictions when the step size parameter assumes large values. Nevertheless, some potential advantages of adaptive filtering schemes may become apparent only for large step sizes [17].

The independence assumption can be stated in the following manner:

$\star$ *Independence Assumption* (IA). The components of the instantaneous input vector $\boldsymbol{x}(k)$ are statistically independent from the adaptive weights $w_i(k)$, for $i \in \{0, \ldots, N-1\}$.

As mentioned before, although almost ubiquitous, as previously said, IA-based analyses[1] are accurate only when small step sizes are utilized [50]. The usage of IA partially neglects the data-dependent feature of the adaptive weight vector, converting Equation (33) into:

$$\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] = (\boldsymbol{I} - \beta\boldsymbol{R})\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right], \tag{35}$$

which is very similar to (21). This implies that many considerations about the convergence of the steepest descent algorithm can be adapted to characterize the mean behavior of the LMS adaptive weights. Hence, an upper bound for the learning factor that avoids divergence in the mean that is inversely proportional to the maximum eigenvalue of the input autocorrelation matrix (see (28)) still holds.

Nevertheless, some considerations should be undertaken. The first of them is that we have omitted factor 2 which multiplies the learning factor of the LMS, to utilize a formulation more common in the literature. This factor slightly alters the final results of the analysis. The second consideration derives from the stochastic nature of the LMS, which requires the modeling of second-order statistics of the adaptive weights, such as their variance. This observation is fundamental since in practice the variance of the adaptive weights can grow without limit, even when the first-order statistics predict convergence. It implies that the algorithm can diverge in practice even when its mean behavior is stable. Thus, a second-order analysis of the LMS is necessary in order to better characterize its learning abilities.

As a final observation, note that Equation (35) is a particular case of a linear time-invariant state space equation

$$\boldsymbol{y}(k+1) = \boldsymbol{A}\boldsymbol{y}(k) + \boldsymbol{b}, \tag{36}$$

where $\boldsymbol{y}(k) = \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right]$, $\boldsymbol{A} = \boldsymbol{I} - \beta\boldsymbol{R}$, and $\boldsymbol{b} = \boldsymbol{0}$. The state space formulation is crucial in the novel models proposed in this thesis, as will be clearer in the next chapter.

[1] In this thesis, the acronym IA refers to independence assumption, where IA denotes the analysis that relies on the independence assumption.

2.4- Concluding remarks

In this chapter, an overview of the Wiener solution, the steepest descent algorithm, and the LMS algorithm is presented. After this, the LMS can be seen in a more proper context. Some standard stochastic assumptions are discussed, and a first-order analysis of the LMS was described. In the next chapter, second-order analyses of the LMS are presented.

3- Second-Order Stochastic Analyses of the LMS: An Overview

3.1- Introduction

In the previous chapter, the convergence of the LMS averaged weight-error vector is compared to the deterministic weight-error of the steepest-descent algorithm. Such analyses share some features with the ones described in this chapter, since they make some assumptions about the input and noise signals, engender difference equations that relate the stochastic quantities of interest in a recursive manner, apply the expectation operator in both sides of these equations and investigate the results.

Due to the stochastic feature of the LMS, its mean behavior is not sufficient to characterize its learning. This chapter introduces the standard second-order stochastic analysis of the LMS, which relies on the independence assumption. Additionally, the original exact expectation analysis is described. Such a method is generalized in Chapters 4, 5, and 6, which contain the contributions of this thesis.

3.2- Standard Mean-Square Analysis

In this section, some standard mean-square analyses that utilize the independence assumption are described.

The most popular performance measure of an adaptive filter are the MSE (see (7)) and the mean square deviation (MSD), defined by

$$\text{MSD}(k) \triangleq \mathbb{E}\left[\|\boldsymbol{w}(k) - \boldsymbol{w}^{\star}\|^2\right]. \tag{37}$$

Both of these quantities can be inferred if the autocorrelation matrix $\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}$ (also

termed as the weight error covariance matrix - WECM) of the deviation $\tilde{\boldsymbol{w}}(k)$ is available

$$\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)} \triangleq \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right], \tag{38}$$

since the following identities hold:

$$\begin{aligned} \mathsf{MSD}(k) &= \mathsf{Tr}[\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}], & (39)\\ \mathsf{MSE}(k) &= \sigma_\nu^2 + \mathsf{Tr}[\boldsymbol{R}\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}]. & (40) \end{aligned}$$

It should be noted that relation (40) requires the NA, which is not a very strong statistical assumption.

Equations (39)-(40) establish a bridge between the estimation of mean-square behavior and the computation of $\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}$. This fact explains why several second-order analyses utilize different methods to obtain a recursion for this matrix. In the following, a standard derivation of this recursion is presented.

After multiplying each side of (31) by its transpose, one has

$$\begin{aligned} \tilde{\boldsymbol{w}}(k+1)\tilde{\boldsymbol{w}}^T(k+1) &= \left(\boldsymbol{I}-\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\left(\boldsymbol{I}-\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right)\\ &\quad+\beta\nu(k)\left(\boldsymbol{I}-\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right)\tilde{\boldsymbol{w}}(k)\boldsymbol{x}^T(k)\\ &\quad+\beta\nu(k)\boldsymbol{x}(k)\tilde{\boldsymbol{w}}^T(k)\left(\boldsymbol{I}-\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right)\\ &\quad+\beta^2\nu^2(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k). \end{aligned} \tag{41}$$

After using the NA and applying the expectation operator in (41), one obtains

$$\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k+1)} = \mathbb{E}\left\{\left(\boldsymbol{I}-\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\left(\boldsymbol{I}-\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right)\right\}+\beta^2\sigma_\nu^2\boldsymbol{R}. \tag{42}$$

Expression (42) can be further simplified by the independence assumption (IA), which leads to

$$\begin{aligned} \boldsymbol{R}_{\tilde{\boldsymbol{w}}(k+1)} &= \boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}-\beta\left[\boldsymbol{R}\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}+\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}\boldsymbol{R}\right]+\beta^2\sigma_\nu^2\boldsymbol{R}\\ &\quad+\beta^2\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right], \end{aligned} \tag{43}$$

a recursion that still requires additional assumptions in order to establish the evolution of $\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}$.

Considering the transformed input vector as

$$\tilde{\boldsymbol{x}}(k) = \boldsymbol{Q}^T \boldsymbol{x}(k), \tag{44}$$

where $\boldsymbol{Q} \in \mathbb{R}^{N \times N}$ contains in its i-th column the i-th eigenvector or $\boldsymbol{R}$ (see (22)). Transformation (44) is convenient, since the covariance matrix of $\tilde{\boldsymbol{x}}(k)$ is diagonal:

$$\boldsymbol{R}_{\tilde{\boldsymbol{x}}} \triangleq \mathbb{E}\left[\tilde{\boldsymbol{x}}(k)\tilde{\boldsymbol{x}}^T(k)\right] = \boldsymbol{\Lambda}_{\tilde{\boldsymbol{x}}}. \tag{45}$$

Using (44), Eq. (43) can be rewritten as

$$\begin{aligned} \boldsymbol{R}_{\tilde{\boldsymbol{w}}(k+1)} &= \boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)} - \beta\left[\boldsymbol{R}\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)} + \boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}\boldsymbol{R}\right] + \beta^2\sigma_\nu^2 R \\ &\quad + \beta^2 \boldsymbol{Q}\mathbb{E}\left\{\tilde{\boldsymbol{x}}(k)\tilde{\boldsymbol{x}}^T(k)\boldsymbol{v}(k)\boldsymbol{v}^T(k)\tilde{\boldsymbol{x}}(k)\tilde{\boldsymbol{x}}^T(k)\right\}\boldsymbol{Q}^T, \end{aligned} \tag{46}$$

where $\boldsymbol{v}(k)$ was defined in (24). When the input signal is Gaussian, (46) can be rewritten as [3]

$$\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k+1)} = \boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)} - \beta\left[\boldsymbol{R}\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)} + \boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}\boldsymbol{R}\right] + 2\beta^2\boldsymbol{R}\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}\boldsymbol{R} + \beta^2\boldsymbol{R}\left\{\sigma_\nu^2 + \mathsf{Tr}[\boldsymbol{R}\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}]\right\}, \tag{47}$$

which is a closed-form recursion. Note that several restrictive assumptions were adopted in its derivation, which engenders a discrepancy between the theoretical model and empirical results.

Recursion (47) is not a state space system (see (36)). Since the exact expectation analysis adopts such a formulation, comparisons between the standard approach (*i.e.*, one that utilizes the independence assumption) and the "exact" one can be facilitated when both of them are expressed in terms of a state space equation. Such a formulation is derived in the following.

Consider vec$(\boldsymbol{A})$ as the operator that concatenates the successive columns of matrix $\boldsymbol{A}$, generating a column vector. Using the identity

$$\mathsf{vec}(\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Z}) = \left[\boldsymbol{Z}^T \otimes \boldsymbol{X}\right]\mathsf{vec}(\boldsymbol{Y}), \tag{48}$$

and the linearity of the expectation operator, Eq. (42) leads to the following dynamic *time-*

invariant state-space model of the deviation error covariance matrix:

$$\boldsymbol{y}^{(\mathtt{IA},2)}(k+1) = \boldsymbol{A}^{(\mathtt{IA},2)}\boldsymbol{y}^{(\mathtt{IA},2)}(k) + \boldsymbol{d}^{(\mathtt{IA},2)}, \tag{49}$$

where the superscript $(\mathtt{IA},2)$ means that the state space linear model is obtained by adopting IA in a mean square (*i.e.*, second-order) stochastic analysis and

$$\boldsymbol{y}^{(\mathtt{IA},2)}(k) \triangleq \mathbb{E}\left\{\mathsf{vec}\left[\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}\right]\right\}, \tag{50}$$

$$\begin{aligned}\boldsymbol{A}^{(\mathtt{IA},2)} \triangleq\ & \boldsymbol{I}_{N^2} - \beta\mathbb{E}\left[\boldsymbol{I}_N \otimes \boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] \\ & -\beta\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k) \otimes \boldsymbol{I}_N\right] \\ & +\beta^2\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k) \otimes \boldsymbol{x}(k)\boldsymbol{x}^T(k)\right],\end{aligned} \tag{51}$$

$$\boldsymbol{d}^{(\mathtt{IA},2)} \triangleq \beta^2\sigma_\nu^2\mathsf{vec}\left[\boldsymbol{R}_x\right]. \tag{52}$$

Note that the *transition matrix* $\boldsymbol{A}^{(\mathtt{IA},2)}$ depends on β, and the analysis predicts stable operation if the magnitude of its maximum eigenvalue is less than unity [50]. Also important is that such an eigenvalue can be efficiently computed, for a fixed step size, using the power method [51]. Moreover, if the algorithm does not diverge, the relevant steady-state statistical quantities may be estimated by [52]

$$\lim_{k\to\infty} \boldsymbol{y}^{(\mathtt{IA},2)}(k) = \left(\boldsymbol{I}_N - \boldsymbol{A}^{(\mathtt{IA},2)}\right)^{-1}\boldsymbol{d}^{(\mathtt{IA},2)}. \tag{53}$$

The exact expectation analysis also aims to construct a linear state space equation system. However, the entire procedure, as will be described in the next section, is much more complex, because IA is no longer employed.

3.3- Exact Expectation Analysis

The exact expectation analysis (EEA) [53; 54; 50; 55; 52; 44; 56] is a refined technique that recursively generates update equations for a set of joint moments (or *state variables*). The joint moments update equations govern the dynamics of the LMS learning. Since EEA does not adopt the independence assumption, its predictions are a better fit for experimental results than standard approaches (*e.g.*, in the specification of an upper bound for the learning factor that ensures convergence [50; 52; 44]).

The EEA technique was proposed in [57], in a configuration where the excitation data is assumed to be white. An unequal-mode convergence behaviour in the variances of the filter coefficients (which contradicts the independence assumption) was theoretically predicted and confirmed by simulations. Work [53] extended the approach to the sign-data LMS algorithm, without assuming a white input data. A more comprehensive analysis of the LMS algorithm (assuming colored input data) under EEA is the focus of [54; 50], which effectively established EEA as a powerful and alternative analysis method for predicting LMS performance. More recently, this technique has experienced some generalizations, namely: [55] removed the whiteness assumption of the additive noise; [56] has demonstrated that coloring the additive noise does not impact mean-square stability; [58] addressed the identification of nonlinear plants; [52] employed EEA to derive an optimal sequence of step size values that optimize performance; and [44] modeled the deficient case that occurs when the adaptive filter length is surpassed by the length of the unknown transfer function the adaptive filter intends to emulate.

In order to explain the EEA, let us consider a simple configuration, for pedagogical purposes. This setup is the same as the one presented in [22]. Even in this simple configuration, the EEA results in lengthy algebraic manipulations.

Suppose that the input signal is generated by the following moving-average process

$$x(k) = b_0 u(k) + b_1 u(k-1), \tag{54}$$

so that the analysis is not restricted to white input signals, although the signal $u(k)$ is modeled as a zero-mean iid random process, whose probability density function is even,

so that

$$\mathbb{E}[u^n(k)] = \begin{cases} 0, & \text{for odd } n \\ \gamma_n, & \text{for even } n \end{cases}, \tag{55}$$

where γ_n depends on the specific considered distribution.

Note that the correlation between $x(k_1)$ and $x(k_2)$, which is supposed to be zero whenever $|k_1 - k_2| > M$, where M is finite. This in fact is a restriction imposed by the EEA analysis (otherwise, its recursion procedure never halts), which is guaranteed by model (54).

Assuming that the adaptive filter has only one adaptive tap (*i.e.*, $N = 1$), recursion (31) degenerates into

$$\tilde{w}_0(k+1) = (1 - \beta x^2(k))\tilde{w}_0(k) + \beta\nu(k)x(k), \tag{56}$$

where $\tilde{w}_i(k) \in \mathbb{R}$ denotes the i-th element of vector $\tilde{\boldsymbol{w}}(k)$ (see (18)). From (30), the error can be written as

$$e(k) = \tilde{w}_0(k)x(k) + \nu(k) = b_0\tilde{w}_0(k)u(k) + b_1\tilde{w}_0(k)u(k-1) + \nu(k), \tag{57}$$

so that the square of the error, in terms of signal $u(k)$, is

$$\begin{aligned} e^2(k) &= b_0^2u^2(k)\tilde{w}_0^2(k) + 2b_0b_1u(k)u(k-1)\tilde{w}_0^2(k) + 2b_0\nu(k)u(k)\tilde{w}_0(k) \\ &\quad + b_1^2u^2(k-1)\tilde{w}_0^2(k) + 2b_1\nu(k)u(k-1)\tilde{w}_0(k) + \nu^2(k). \end{aligned} \tag{58}$$

The MSE can be obtained by the application of the expectation operator in (58), which leads to

$$\begin{aligned} \xi(k) = \mathbb{E}\left[e^2(k)\right] &= b_0^2\mathbb{E}\left[u^2(k)\tilde{w}_0^2(k)\right] + 2b_0b_1\mathbb{E}\left[u(k)u(k-1)\tilde{w}_0^2(k)\right] \\ &\quad + 2b_0\mathbb{E}\left[\nu(k)u(k)\tilde{w}_0(k)\right] + b_1^2\mathbb{E}\left[u^2(k-1)\tilde{w}_0^2(k)\right] \\ &\quad + 2b_1\mathbb{E}\left[\nu(k)u(k-1)\tilde{w}_0(k)\right] + \mathbb{E}\left[\nu^2(k)\right]. \end{aligned} \tag{59}$$

Utilizing NA[1], terms $\mathbb{E}\left[\nu(k)u(k)\tilde{w}_0(k)\right]$ and $\mathbb{E}\left[\nu(k)u(k-1)\tilde{w}_0(k)\right]$ can be canceled, which

[1]The NA is not a very strong assumption that even can be avoided when EEA is employed (see [55]). It will be avoided in Chapter 4.

simplifies (59):

$$\begin{aligned}\xi(k) = \mathbb{E}\left[e^2(k)\right] &= b_0^2\mathbb{E}\left[u^2(k)\tilde{w}_0^2(k)\right] + 2b_0b_1\mathbb{E}\left[u(k)u(k-1)\tilde{w}_0^2(k)\right] \\ &\quad + b_1^2\mathbb{E}\left[u^2(k-1)\tilde{w}_0^2(k)\right] + \mathbb{E}\left[\nu^2(k)\right]. \end{aligned} \tag{60}$$

Note that $\tilde{w}_0(k)$ depends directly on $x(k-1)$, since (see (56)):

$$\tilde{w}_0(k) = (1-\beta x^2(k-1))\tilde{w}_0(k-1) + \beta\nu(k-1)x(k-1), \tag{61}$$

which implies that $\tilde{w}_0(k)$ depends statistically on sample $x(k)$, due to the fact that $x(k)$ is colored (see (54)) and $x(k)$ is correlated with $x(k-1)$. Due to this fact, (60) was written in terms of $u(k)$. Being $u(k)$ an i.i.d. process, $\tilde{w}_0(k)$ *depends on* $u(k-1)$ and is indeed statistically *independent* from $u(k)$, since $u(k)$ is statistically independent from $u(k-1)$. This implies that

$$\mathbb{E}\left[u^2(k)\tilde{w}_0^2(k)\right] = \mathbb{E}\left[u^2(k)\right]\mathbb{E}\left[\tilde{w}_0^2(k)\right] = \gamma_2\mathbb{E}\left[\tilde{w}_0^2(k)\right], \tag{62}$$

$$\mathbb{E}\left[u(k)u(k-1)\tilde{w}_0^2(k)\right] = \overbrace{\mathbb{E}\left[u(k)\right]}^{=0}\mathbb{E}\left[u(k-1)\tilde{w}_0^2(k)\right] = 0. \tag{63}$$

The combination of (62) and (63) with (60) leads to

$$\xi(k) = \sigma_\nu^2 + \mathbb{E}\left[\tilde{w}_0(k)x^2(k)\right] = \sigma_\nu^2 + b_0^2\gamma_2\mathbb{E}\left[\tilde{w}_0^2(k)\right] + b_1^2\mathbb{E}\left[\tilde{w}_0^2(k)u^2(k-1)\right], \tag{64}$$

which means that the estimation of the MSE can be equivalently converted in the estimation of the evolution of the statistical quantities $\mathbb{E}\left[\tilde{w}_0^2(k)\right]$ and $\mathbb{E}\left[\tilde{w}_0^2(k)u^2(k-1)\right]$. Such quantities are also termed as *state variables*. Further, when `IA` is adopted, Equation (64) can be simplified, since

$$\mathbb{E}\left[\tilde{w}_0^2(k)u^2(k-1)\right] \approx \mathbb{E}\left[\tilde{w}_0^2(k)\right]\mathbb{E}\left[u^2(k-1)\right] = \gamma_2\mathbb{E}\left[\tilde{w}_0^2(k)\right], \tag{65}$$

although such an approximation will not be utilized, since we are interested in the `EEA`.

The `EEA` is a systematic procedure that intends to obtain recursive equations that describe the evolution of the state variables. At the moment, there are two state variables: $\mathbb{E}\left[\tilde{w}_0^2(k)\right]$ and $\mathbb{E}\left[\tilde{w}_0^2(k)u^2(k-1)\right]$. The first step to obtain a recursion for the first of them

requires the squaring of both sides of (56):

$$\begin{aligned}\tilde{w}_0^2(k+1) &= b_0^4\beta^2u^4(k)\tilde{w}_0^2(k) + 4b_0^3b_1\beta^2u^3(k)u(k-1)\tilde{w}_0^2(k) - 2b_0^3\beta^2\nu(k)u^3(k)\tilde{w}_0(k)\\ &+6\beta^2b_0^2b_1^2u^2(k)u^2(k-1)\tilde{w}_0^2(k) - 6b_0^2b_1\beta^2\nu(k)u^2(k)u(k-1)\tilde{w}_0(k)\\ &+b_0^2\beta^2\nu^2(k)u^2(k) - 2b_0^2\beta u^2(k)\tilde{w}_0^2(k) + 4b_0b_1^3\beta^2u(k)u^3(k-1)\tilde{w}_0^2(k)\\ &-6b_0b_1^2\beta^2\nu(k)u(k)u^2(k-1)\tilde{w}_0(k) + 2b_0b_1\beta^2\nu^2(k)u(k)u(k-1)\\ &-4b_0b_1\beta u(k)u(k-1)\tilde{w}_0^2(k) + 2b_0\beta\nu(k)u(k)\tilde{w}_0(k) + b_1^4\beta^2u^4(k-1)\tilde{w}_0^2(k)\\ &-2b_1^3\beta^2\nu(k)u^3(k-1)\tilde{w}_0 + b_1^2\beta^2\nu^2(k)u^2(k-1) - 2b_1^2\beta u^2(k-1)\tilde{w}_0^2(k)\\ &+2b_1\beta\nu(k)u(k-1)\tilde{w}_0(k) + \tilde{w}_0^2(k). \end{aligned} \quad (66)$$

After applying the expectation operator in (66) and utilizing the simplifications that arrive from the combination of both NA and the whiteness of $u(k)$, one has

$$\begin{aligned}\mathbb{E}\left[\tilde{w}_0^2(k+1)\right] &= (1 - 2b_0^2\beta\gamma_2 + b_0^4\beta^2\gamma_4)\mathbb{E}\left[\tilde{w}_0^2(k)\right] + b_1^4\beta^2\mathbb{E}\left[\tilde{w}_0^2(k)u^4(k-1)\right]\\ &+(6b_0^2b_1^2\beta^2\gamma_2 - 2b_1^2\beta)\mathbb{E}\left[\tilde{w}_0^2(k)u^2(k-1)\right]\\ &+(b_0^2\beta^2\sigma_\nu^2\gamma_2 + b_1^2\beta^2\sigma_\nu^2\gamma_2), \end{aligned} \quad (67)$$

where one observes the emergence of a novel state variable: $\mathbb{E}\left[\tilde{w}_0^2(k)u^4(k-1)\right]$. Note that we are not essentially interested in estimating such a quantity to obtain the MSE (see Eq. (64)). Due to this, this new term is termed as a *nuisance state variable*, because its estimation is necessary for the estimation of one statistical quantity of interest [59]. Unfortunately, in more complex configurations, the nuisance parameters may be the large majority of the state variables.

After obtaining the recursion for $\mathbb{E}[\tilde{w}_0^2(k)]$ in (67), two state variables remains to be addressed: the term $\mathbb{E}\left[\tilde{w}_0^2(k)u^2(k-1)\right]$, which is originally encountered in the MSE expression (59) and the *nuisance* state variable $\mathbb{E}\left[\tilde{w}_0^2(k)u^4(k-1)\right]$. The recursions for these state variables can be derived by multiplying both sides of (66) by convenient terms before the application of the expectation operator and the subsequent simplification by both NA and the whiteness of $u(k)$. Due to the length of the necessary manipulations, the

derivation is omitted and the final recursions are:

$$\begin{aligned}\mathbb{E}\left[\tilde{w}_0^2(k+1)u^2(k)\right] &= (\gamma_2 - 2b_0^2\beta\gamma_4 + b_0^4\beta^2\gamma_6)\mathbb{E}\left[\tilde{w}_0^2(k)\right] + b_1^4\beta^2\gamma_2\mathbb{E}\left[\tilde{w}_0^2(k)u^4(k-1)\right] \\ &\quad +(6b_0^2b_1^2\beta^2\gamma_4 - 2b_1^2\beta\gamma_2)\mathbb{E}\left[\tilde{w}_0^2(k)u^2(k-1)\right] \\ &\quad +(b_0^2\beta^2\sigma_\nu^2\gamma_4 + b_1^2\beta^2\sigma_\nu^2\gamma_2^2), \end{aligned} \tag{68}$$

$$\begin{aligned}\mathbb{E}\left[\tilde{w}_0^2(k+1)u^4(k)\right] &= (\gamma_4 - 2b_0^2\beta\gamma_6 + b_0^4\beta^2\gamma_8)\mathbb{E}\left[\tilde{w}_0^2(k)\right] + b_1^4\beta^2\gamma_4\mathbb{E}\left[\tilde{w}_0^2(k)u^4(k-1)\right] \\ &\quad +(6b_0^2b_1^2\beta^2\gamma_6 - 2b_1^2\beta\gamma_4)\mathbb{E}\left[\tilde{w}_0^2(k)u^2(k-1)\right] \\ &\quad +(b_0^2\beta^2\sigma_\nu^2\gamma_6 + b_1^2\beta^2\sigma_\nu^2\gamma_2\gamma_4). \end{aligned} \tag{69}$$

Since recursions (67)-(69) are self-contained, they can be used to create a state space equation like (36), where the state vector is

$$\boldsymbol{y}(k) = \begin{bmatrix} \mathbb{E}\left[\tilde{w}_0^2(k)\right] \\ \mathbb{E}\left[\tilde{w}_0^2(k)u^2(k-1)\right] \\ \mathbb{E}\left[\tilde{w}_0^2(k)u^4(k-1)\right] \end{bmatrix}, \tag{70}$$

the *transition matrix* is

$$\boldsymbol{A} = \begin{bmatrix} 1 - 2b_0^2\beta\gamma_2 + b_0^4\beta^2\gamma_4 & 6b_0^2b_1^2\beta^2\gamma_2 - 2b_1^2\beta & b_1^4\beta^2 \\ \gamma_2 - 2b_0^2\beta\gamma_4 + b_0^4\beta^2\gamma_6 & 6b_0^2b_1^2\beta^2\gamma_4 - 2b_1^2\beta\gamma_2 & b_1^4\beta^2\gamma_2 \\ \gamma_4 - 2b_0^2\beta\gamma_6 + b_0^4\beta^2\gamma_8 & 6b_0^2b_1^2\beta^2\gamma_6 - 2b_1^2\beta\gamma_4 & b_1^4\beta^2\gamma_4 \end{bmatrix} \tag{71}$$

and

$$\boldsymbol{b} = \begin{bmatrix} b_0^2\beta^2\sigma_\nu^2\gamma_2 + b_1^2\beta^2\sigma_\nu^2\gamma_2 \\ b_0^2\beta^2\sigma_\nu^2\gamma_4 + b_1^2\beta^2\sigma_\nu^2\gamma_2^2 \\ b_0^2\beta^2\sigma_\nu^2\gamma_6 + b_1^2\beta^2\sigma_\nu^2\gamma_2\gamma_4 \end{bmatrix}. \tag{72}$$

For the considered setup, the resulting state space model contains $R = 3$ equations. Unfortunately, R can be very large, even for simple configurations. To perform the demanded algebraic operations, efficient C++-based codes were implemented. Thus, commercial general-purpose software for symbolic manipulations was avoided, which has permitted us to model configurations that have never been considered in the EEA.

3.4- Concluding Remarks

In this chapter, traditional and EEA-based approaches for modeling the learning of the LMS are described. The following chapter contains the first contribution of this thesis: the generalization of the EEA technique to a scenario where the plant the adaptive filter intends to identify is nonlinear.

4- Analysis of the LMS-based Identification of Nonlinear Plants

As we have seen, 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 chapter, the performance and stability guarantees provided by stochastic models that presume the linearity of the ideal system are no longer valid. This chapter 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.

4.1- Introduction

As already outlined in Chapter 2, the least-mean-square (LMS) algorithm emerged as an effective, yet simple, adaptive filtering scheme whose robustness against perturbations and implementation errors is well-known [14; 60; 61]. Despite its simplicity, the mathematical analysis of the LMS is complicated due to its stochastic and nonlinear nature [62]. Furthermore, most analyses assume that the ideal plant the adaptive filter intends to estimate[1] is linear [3]. Unfortunately, this may not be strictly valid in practice, due to a plethora of phenomena, such as operation near the saturation region [24], intermodulation distortion [25] and usage of high power amplifiers [63]. Since a small distortion engineered by a nonlinear component (NLC) in such systems may dominate the overall identification performance [64], analyses that do not take into account the NLC

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

are not able to provide either performance or stability guarantees. This chapter advances, for the first time, a stochastic model of the LMS performance that incorporates the effects of the NLC. The proposed analysis does not employ the ubiquitous independence assumption (IA), which states that the input samples $x(k)$ are statistically independent from the adaptive weights $w_i(k)$ (for $i \in \{0, 1, \ldots, N\}$. As discussed in Chapter 2, the IA, although widely used in the field of stochastic approximations [14], is recognized to provide accurate predictions only when the step size is small [65]. The avoidance of IA engineers the exact expectation technique, which is also able to provide a proper step-size upper bound that guarantees convergence [50]. Furthermore, it can be employed in the derivation of a *deterministic* theoretical step-size sequence that optimizes performance [52], a topic that is explored in Chapter 5.

This chapter is structured as follows. We start by describing the adopted nonlinear model of the plant to be identified, in Section 4.2. We then present the advanced analysis in Section 4.3, which can be employed to accurately predict the stability region. This is followed by the results, in Section 4.4. We then present the main conclusions of the contributions in Section 4.5.

4.2- Nonlinear System Model (NSM)

Nonlinear discrete-time systems can be frequently modeled by Volterra series [66; 67; 68; 69; 27], 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 [70; 71]. 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 thesis the reference signal $d(k)$ is modeled as the sum of three components

$$d(k) = (\boldsymbol{w}^{\star})^T \boldsymbol{x}(k) + \nu_{\mathsf{M}}(k) + \overbrace{\sum_{p=2}^{P} \sum_{n_1=0}^{N-1} \cdots \sum_{n_P=0}^{N-1} \overline{w}_p^{\star}(n_1, \ldots, n_P) \prod_{l=1}^{p} x(n - n_l)}^{\triangleq \nu_{\mathsf{NLC}}(k)}, \tag{73}$$

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), $\nu_{\mathsf{M}}(k)$ is the measurement noise and $\nu_{\mathsf{NLC}}(k)$ corresponds to the nonlinear part (that is not captured by the linear adaptive model). Note that the consecutive samples of the overall noise $\nu(k) \triangleq \nu_{\mathsf{M}}(k) + \nu_{\mathsf{NLC}}(k)$ can be correlated, so that the noise whiteness (a very common hypothesis) is not assumed in this chapter.

The Volterra model (73) 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 $n_1, \cdots, n_p$ [72; 73]).

4.3- Exact Expectation Analysis of LMS-Based NLS Identification

As discussed in Chapter 2, the LMS update equation is a nonvanishing step-size version of a stochastic gradient algorithm that intends to minimize the mean-squared error (MSE):

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \beta \nabla_{\boldsymbol{w}(k)} \left[\frac{1}{2} e^2(k)\right] = \boldsymbol{w}(k) + \beta e(k) \boldsymbol{x}(k), \tag{74}$$

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

$$e(k) \triangleq d(k) - \boldsymbol{w}^T(k) \boldsymbol{x}(k). \tag{75}$$

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:

$$x(k) = \sum_{m=0}^{M-1} b_m u(k-m), \tag{76}$$

where $u(k)$ is an i.i.d. signal with even probability density function (pdf) and $\gamma_n \triangleq \mathbb{E}\left[u^n(k)\right]$. 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 the 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 (74)). In the following, joint moments such as $\mathbb{E}\left[u(k)\nu_{\mathsf{M}}(k)\right]$ will be considered to be zero, due to the usage of the *noise assumption* (NA) (see Section 2.3).

Remark: In this chapter, the *overall* noise $\nu(k)$ is neither i.i.d. nor statistically independent from the excitation data.

The exact expectation analysis is a systematic procedure that aims at the construction of a state space equation that describes in a comprehensive manner the adaptive filter learning. Following similarly the pedagogical steps presented in Chapter 3, consider the configuration $N = 2$, $M = 1$, $\nu_{\mathsf{NLC}}(k) = \alpha x^2(k-1)$ and a first-order stochastic analysis. In scalar terms, the application of recursion (74) in $\tilde{w}_0(k)$ can be expressed as

$$\begin{aligned}\tilde{w}_0(k+1) &= \tilde{w}_0(k) - b_0^2u^2(k)\beta - b_0^2u(k)u(k-1)\beta - b_0u(k)\nu(k)\beta - b_0^3u(k)u^2(k-1)\alpha\beta \\ &\quad + b_0^2u^2(k)\tilde{w}_0(k)\beta + b_0^2u(k)\tilde{w}_1(k)u(k-1)\beta, \end{aligned} \tag{77}$$

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

$$\mathbb{E}[\tilde{w}_0(k+1)] = (1 + b_0^2\beta\gamma_2)\mathbb{E}[\tilde{w}_0(k)] - b_0^2\beta\gamma_2. \tag{78}$$

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

$$\mathbb{E}[\tilde{w}_1(k+1)] = \mathbb{E}[\tilde{w}_1(k)] - b_0^2\beta\gamma_2 + b_0^2\beta\mathbb{E}[u^2(k-1)\tilde{w}_1(k)], \tag{79}$$

where the *nuisance* term $\mathbb{E}[u^2(k-1)\tilde{w}_1(k)]$ appears, whose computation is necessary in order to evaluate the quantities of interest [59]. A recursion on such a term can be

obtained if one multiplies both sides of recursion

$$\begin{aligned}\tilde{w}_1(k+1) &= \tilde{w}_1(k) - b_0^2 u(k-1)u(k)\beta - b_0^2 u^2(k-1)\beta - b_0 u(k-1)\nu(k)\beta \\ &\quad -b_0^3 u^3(k-1)\alpha\beta + b_0^2 u(k-1)\tilde{w}_0(k)u(k)\beta + b_0^2 u^2(k-1)\tilde{w}_1(k)\beta \quad (80)\end{aligned}$$

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

$$\mathbb{E}[u^2(k)\tilde{w}_1(k+1)] = \gamma_2 \mathbb{E}[\tilde{w}_1(k)] - b_0^2\beta\gamma_2^2 + b_0^2\beta\gamma_2 \mathbb{E}[u^2(k-1)\tilde{w}_1(k)]. \quad (81)$$

Recursions (78), (79) and (81) can be employed to construct a *linear* state equation system

$$\boldsymbol{y}(k+1) = \boldsymbol{A}\boldsymbol{y}(k) + \boldsymbol{b}, \quad (82)$$

in which state vector $\boldsymbol{y}(k)$ contains the state variables $\mathbb{E}\left[\tilde{w}_0(k)\right]$, $\mathbb{E}\left[\tilde{w}_1(k)\right]$ 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 operations 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

$$\begin{aligned}\mathrm{MSE}(k) &\triangleq b_0^2\gamma_2 - 2b_0^2\gamma_2\mathbb{E}[\tilde{w}_0(k)] + b_0^2\gamma_2 - 2b_0^2\mathbb{E}[u^2(k-1)\tilde{w}_1(k)] + \sigma_\nu^2 + b_0^4\alpha^2\gamma_4 \\ &\quad -2b_0^3\alpha\mathbb{E}[u^3(k-1)\tilde{w}_1(k)] + b_0^2\gamma_2\mathbb{E}[\tilde{w}_0^2(k)] + b_0^2\mathbb{E}[u^2(k-1)\tilde{w}_1^2(k)], \quad (83)\end{aligned}$$

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 [44], so 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 updated equations of the involved state variables. Since the input data is correlated only in a finite horizon (see (76)), the generation of equations eventually halts. Note that the algorithm convergence can be inferred from matrix $\boldsymbol{A}$ (which depends on β).

4.4- Results

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

$$d(k) = [\boldsymbol{w}^\star]^T \boldsymbol{x}(k) + 0.005x^2(k) + \nu(k), \tag{84}$$

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 $i \in \{0, 1, \ldots, N-1\}$).

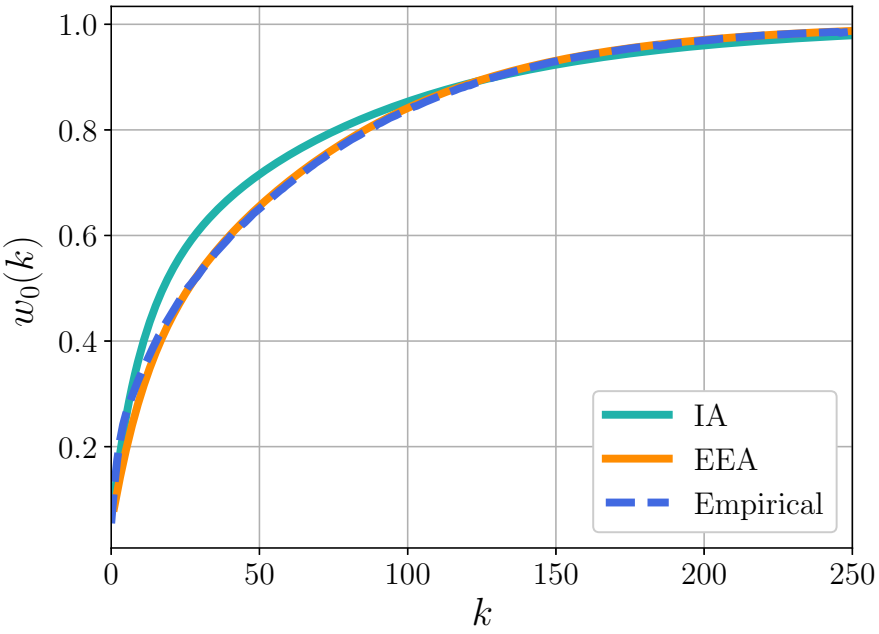


Figure 4 – Evolution of the mean value of adaptive weight $w_0(k)$ of the LMS with $N = 6$ and $\beta = 0.065$. The empirical curve was obtained with 10^7 independent Monte Carlo trials.

Fig. 4 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. This is a remarkable result since commonly one assumes that the lack of precision of the traditional approaches is encountered more relevant only in the second-order statistics. In the first phase of the learning, the EEA method predicts a slower convergence rate than the IA-based

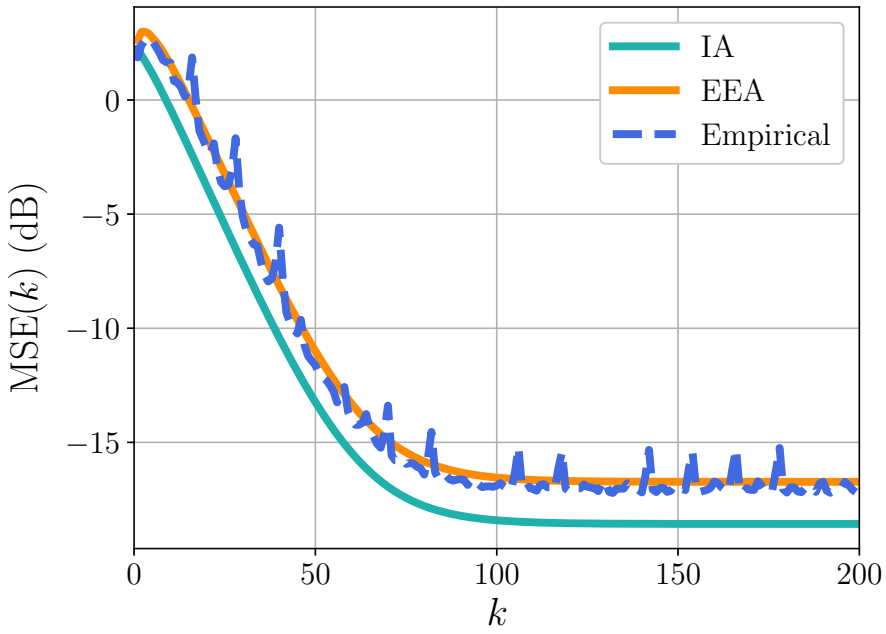


Figure 5 – Evolution of the mean squared error with $N = 3$ and $\beta = 0.085$. The empirical curve was obtained with 10^8 independent Monte Carlo trials.

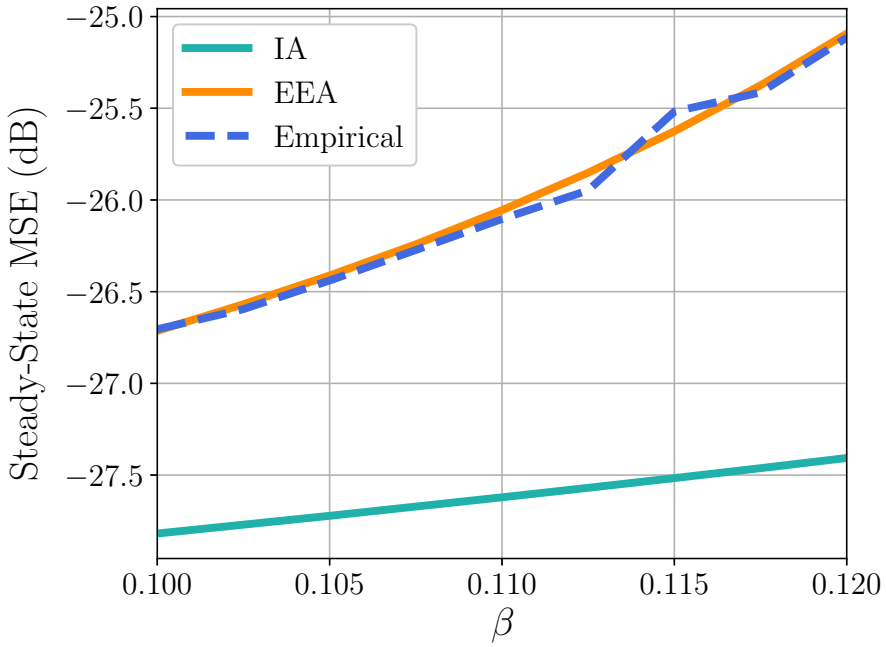


Figure 6 – Steady-State MSE for the $N = 2$ configuration. The empirical results were obtained by the usage of 10^6 independent Monte Carlo trials.

approach, which is confirmed by simulations.

Fig. 5 shows that under the considered configuration the exact expectation method predicts a curious nonconventional phenomenon: a non-monotonic behavior of

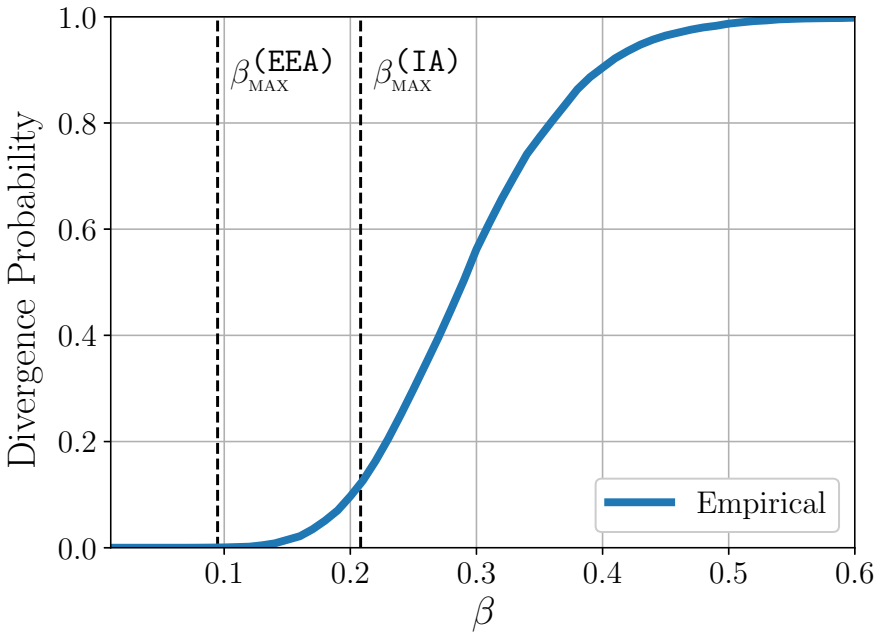


Figure 7 – Divergence probability for $N = 3$. For this configuration, $\beta_{\text{MAX}}^{(\text{IA})} = 0.208149$ and $\beta_{\text{MAX}}^{(\text{EEA})} = 0.0947366$.

the learning curve. In the first phase of the learning, the algorithm performs worse than at the initialization. This a somewhat counterintuitive phenomenon, which is not even suggested by the traditional approach. Such behavior is confirmed by the performed simulations. It is observed only after a huge number of Monte Carlo independent trials (10^8), probably due to low probability events that are not well-captured by the conventional analyses. Normally, in the filtering adaptive literature, the number of independent Monte Carlo trials stays around tens, hundreds, or at most thousands of executions, and such phenomenons are not discussed. Unfortunately, due to the cumbersome complexity of the model, it is very difficult to extract qualitative insights concerning the reasons for this atypical behavior, although it surely deserves more attention.

Fig. 6 depicts the divergence between IA-based predictions for steady-state performance with respect to the empirical one. In this figure, it is possible to observe that the prediction error of the standard analysis is over 1 dB. Further, such an error diminishes as long as β assumes small values, coherent with the widely known fact that the independence assumption implies better predictions when small learning factors are adopted.

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. 7 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. This implies that traditional models are unfit to accurately predict a theoretical upper bound for the step size that guarantees stable behavior of the algorithm. Fig. 7 shows for the first time that the exact expectation method can be used to provide such a guarantee *even when the plant to identify has nonlinear components*.

4.4.1 Evaluation of Different Noise Nonlinear Terms

To analyze whether IA or EEA has better adhesion to empirical curves, four scenarios were created where the first scenario has additive noise, with no contribution of nonlinear disturbances, therefore its reference signal $d(k) = [\boldsymbol{w}^\star]^T \boldsymbol{x}(k) + \nu(k)$. To evaluate how different nonlinear disturbances affect the MSE, three other simulations were plotted. These four simulations are summarized as per Table 4, where it is possible to observe that they are all similar, to each other, except for the nonlinear contribution which progressively increases. For these simulations $N = 3$ and coloring filter $M = 2$, with Gaussian distribution. The noise variance is $\sigma_\nu^2 = 10^{-3}$ and nonlinear noise contribution varies in each simulation, starting with no contribution at all in scenario 1 (Fig. 8), progressing to $\alpha x^2(k)$ on scenario 2 (Fig. 9), $x(k)$ is cubed on scenario 3 (Fig. 10) and finally simulated $\alpha x^4(k)$ on scenario 4 (Fig. 11). For scenarios 2,3 and 4 $\alpha = 0.005$

# Scenario	N	M	L	β	σ_ν^2	N_{samples}	$H(z)$	$\nu_{\text{NLC}}(k)$
1	3	2	1	0.075	10^{-3}	10^6	$1-0.8z^{-1}$	0
2	3	2	1	0.075	10^{-3}	10^6	$1-0.8z^{-1}$	$\alpha x^2(k)$
3	3	2	1	0.075	10^{-3}	10^6	$1-0.8z^{-1}$	$\alpha x^3(k)$
4	3	2	1	0.075	10^{-3}	10^8	$1-0.8z^{-1}$	$\alpha x^4(k)$

Table 1 – Scenarios properties. N_{samples} denotes the number of independent Monte Carlo trials. $H(z)$ is the coloring filter of the input signal.

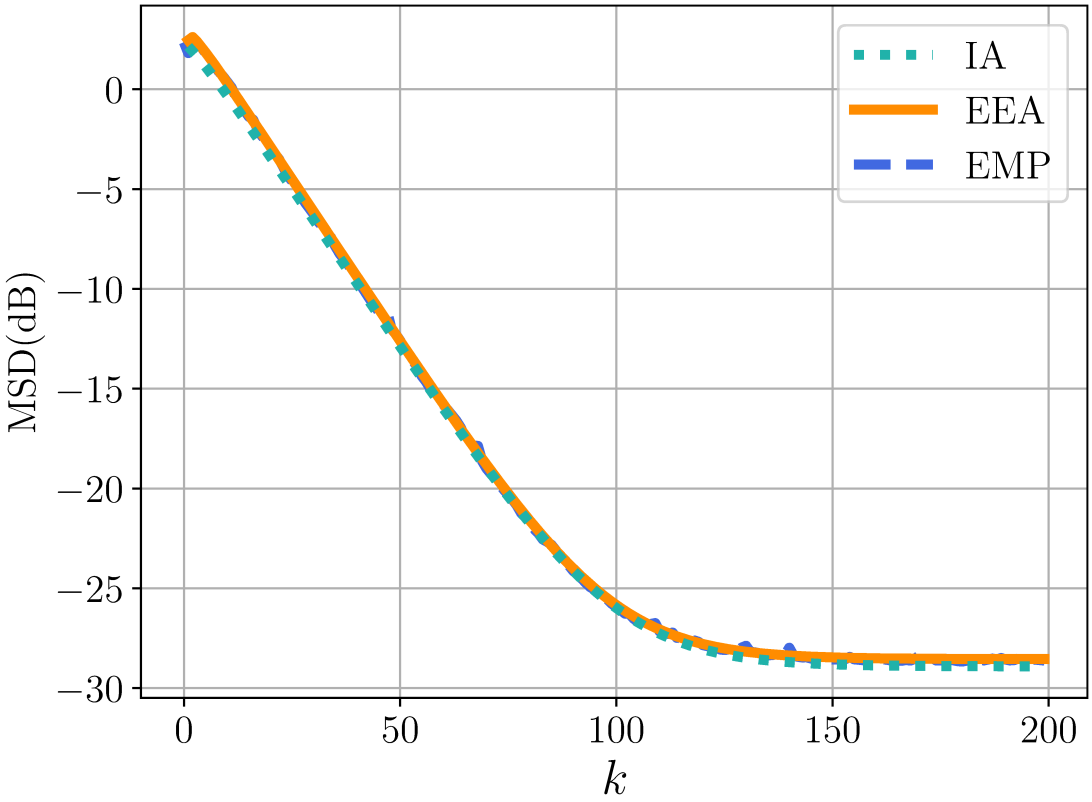


Figure 8 – MSE - Scenario 1

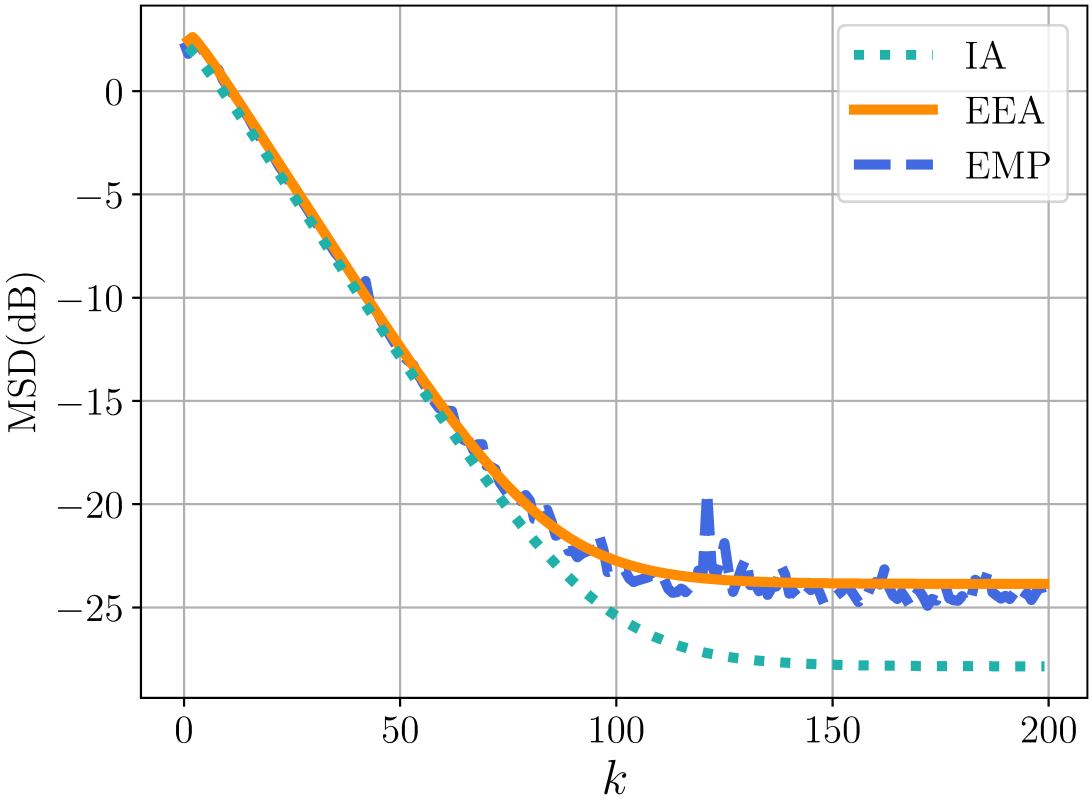


Figure 9 – MSE - Scenario 2

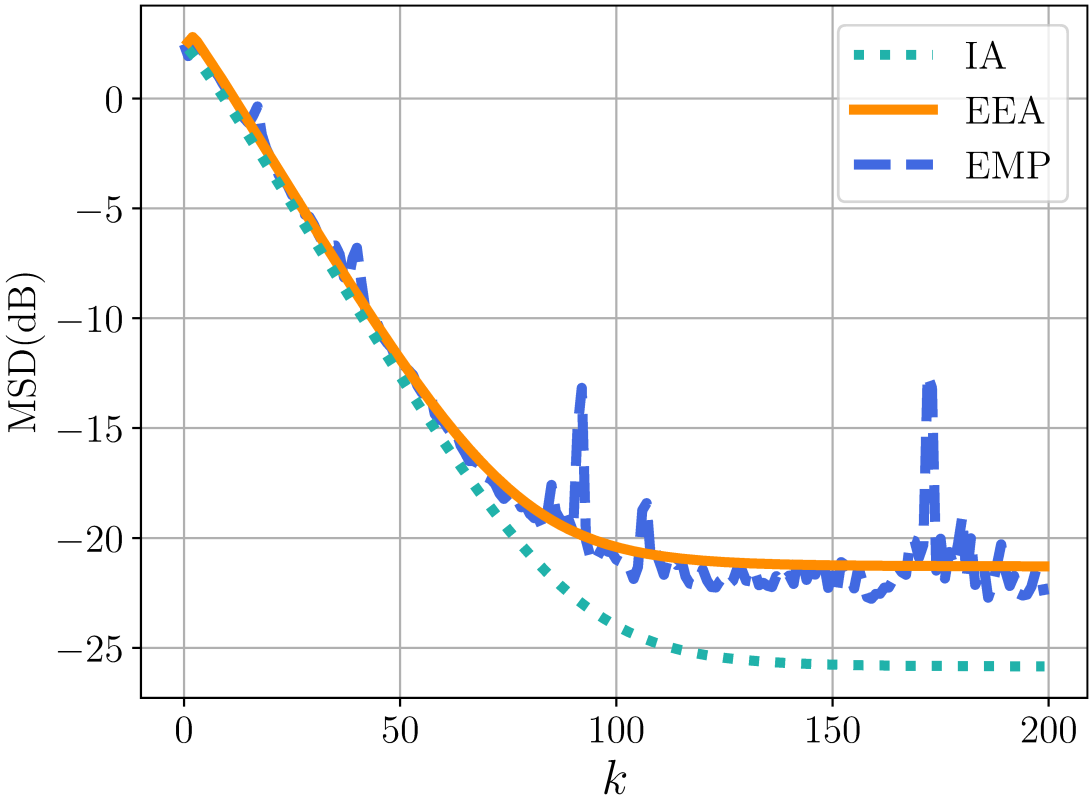


Figure 10 – MSE - Scenario 3

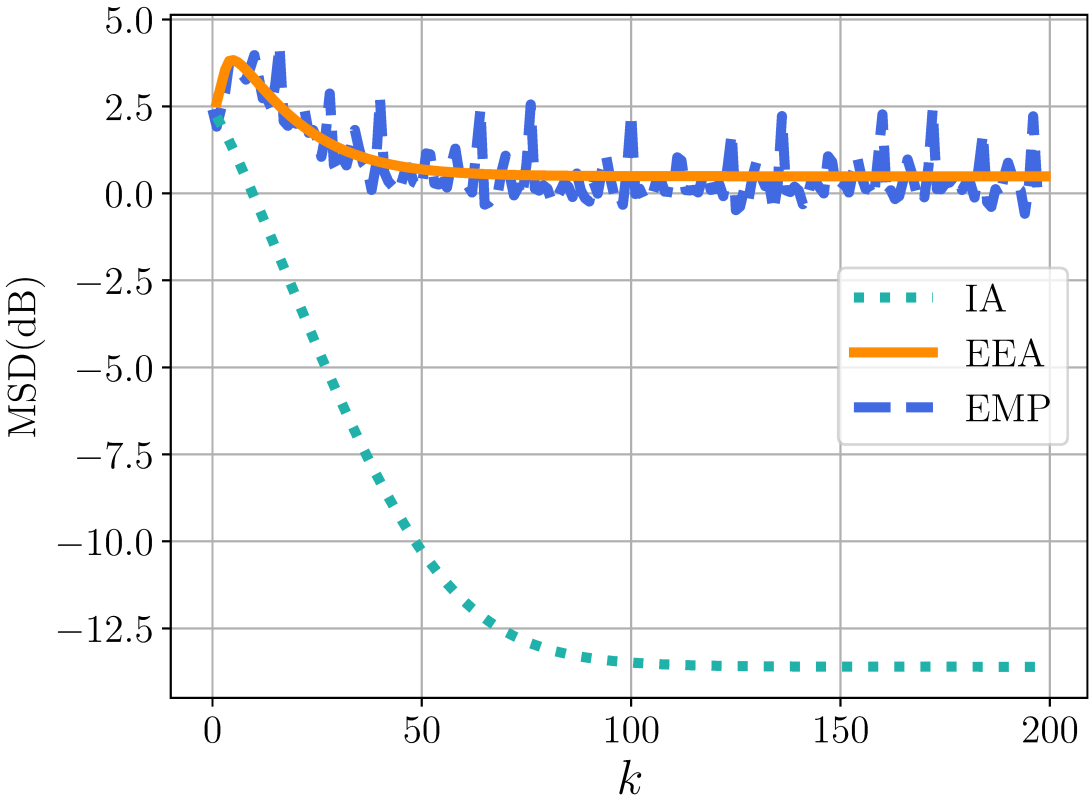


Figure 11 – MSE - Scenario 4

As it is possible to observe on Figure 8, classical analysis has a good approximation especially, as before mentioned, for being a small learning step that propitiates a better adhesion, but even for scenarios such as this, when it is added a nonlinear disturbance (figures 9, 10 and 11), it is possible to observe distancing from classical analysis to the EALC curves. On the other hand, exact expectation analysis adhered properly to the empirical curves, in particular on scenario 4 (Figure 11), where the nonlinear additive noise disturbance was $\alpha x^4(k)$, and classical analysis differs in approximately 14 dB, on steady state, from both empirical and exact expectation analysis.

4.5- Conclusions

In this chapter, 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 assume 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.

5- An Exact Expectation Model for the LMS Tracking Abilities

Nonstationary environments are ubiquitous in communications and acoustic systems. The ability to track their dynamics is one of the most desirable features of adaptive processing algorithms. The designers of these algorithms often employ guidelines derived from stochastic analyses to adjust user-defined parameters to maximize performance or avoid stability issues. From the discussions from the previous chapters, it is therefore important that analyzes of adaptive processing algorithms take into account their built-in sophisticated learning capabilities.

This chapter presents a stochastic model of the performance of the least mean square algorithm, operating under Markovian time-varying channels. Our advanced analysis considers both transient and steady-state regimes. Furthermore, in our analysis, the popular *independence assumption* is not adopted (as in Chapter 4), resulting in a stochastic model which is accurate even when: (i) the step size is not infinitesimally small; or (ii) when the unknown system presents a high nonstationary degree. In addition, our evaluation is also able to provide a deterministic theoretical step-size sequence that optimizes algorithmic performance, as well as an accurate step size upper bound that guarantees algorithm stability.

5.1- Introduction

In applications such as channel equalization, hearing aids and echo cancellation, it is crucial that adaptive algorithms present a low asymptotic mean-squared error (MSE) [74]. It is sometimes argued that the choice of a fixed learning factor (or step size) β is influenced by a trade-off between high convergence rate and good steady-state performance [75]. In practice, such a trade-off is more subtle, since time-varying channels introduce a “lag” in the adaptive process when emulating the optimal and unknown vector $\boldsymbol{w}^{\star}(k) \in \mathbb{R}^N$ [14]. As a consequence, reducing β does not necessarily imply that better steady-state performance will be attained [76].

The LMS algorithm demands a computational burden that is only proportional to the filter tap-length, since its update equation may be expressed in terms of an inner product of the current input vector. Additionally, the LMS performance dynamics depend on a feedback control mechanism, in which the error $e(k)$ drives the update of the adaptive coefficients. As discussed in Chapter 2, this feedback procedure is responsible for the complex behaviour of the LMS, which makes the development of analytical stochastic models challenging. Moreover, the fact that a proper adjustment of its learning factor remains a difficult task (since the optimal step size choice is an intricate function of the input statistics [77; 78]) is a driving force for the significant amount of effort invested in the development of new theoretical approaches.

In this chapter, EEA is for the first time employed in order to perform an in-depth analysis of asymptotic and non-asymptotic behaviour of the LMS under the time-varying plant setting. Popular assumptions (such as the *independence assumption*) are not used, which yields an accurate and more complex model. One may argue that our approach trades off computational cost for adherence with experiments. This point of view may be misleading, since standard approaches may incorrectly predict that the algorithm is stable in cases where it is not, as we have seen in Chapter 4. This problem does not exist when the proposed stochastic model is used, and in these configurations there is no trade-off at all.

This chapter is structured as follows. Section 5.2 concisely describes the LMS algorithm, whereas Section 5.3 depicts a standard stochastic model (*i.e.*, one based on the ubiquitous independence assumption) for its performance abilities. The devised exact expectation method is explained in Section 5.4, which is later utilized for constructing two versions of optimal and *deterministic* step-size sequences in Section 5.5. Although the first version can be seen as an adaptation of the method devised in [52] to the tracking setting, the second version is an entirely new contribution. Comparisons between theoretical predictions and experimental results are presented and discussed in Section 5.6. Concluding remarks are presented in Section 5.7 which outlines some possible promising future works that could be derived from it.

5.2- Time-variant Plant Model

Suppose that the output of the ideal system (excited by a stationary stochastic input signal $x(k)$) is generated by the following noisy *time-variant* affine-in-the-parameters regression model:

$$d(k) = [\boldsymbol{w}^{\star}(k)]^{T} \boldsymbol{x}(k) + \nu(k). \tag{85}$$

Linear observation models similar to (85) are popular in the literature, since they are apt to capture many practical cases of interest [79].

In the following, it is assumed a random walk time-variant plant, whose optimal set of coefficients $\boldsymbol{w}^{\star}(k)$ is updated according to a first-order stochastic multivariate Markovian model:

$$\boldsymbol{w}^{\star}(k+1) = \boldsymbol{w}^{\star}(k) + \boldsymbol{q}(k), \tag{86}$$

where $\boldsymbol{q}(k) \in \mathbb{R}^{N}$ is a driving white perturbation random vector whose components are statistically independent. It is noteworthy that first-order Markovian variations are encountered in several realistic applications such as audio signal processing, acoustic echo cancellation and transmission systems [80; 81; 82; 83]. In addition, LMS may outperform the Recursive Least Square (RLS) under slow-fading environments [84].

In a system identification setup, the LMS algorithm intends to emulate $\boldsymbol{w}^{\star}(k)$ from the observable data $\{\boldsymbol{x}(k), d(k)\}$. In order to do that, the LMS uses the stochastic approximation of the iterative steepest descent algorithm. By adopting the instantaneous quadratic error as the cost function, its update equation is given by (29), which uses a stochastic gradient of the cost function $\mathbb{E}\left[\frac{1}{2}e^{2}(k)\right]$. Alternatively, the LMS update equation (29) can be interpreted as an exact solver of the following *deterministic* problem:

$$\min_{\boldsymbol{w}(k+1)} \|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^{2} \text{ s.t. } e_{\mathsf{p}}(k) = \left(1 - \beta\|\boldsymbol{x}(k)\|^{2}\right) e(k), \tag{87}$$

where $e_{\mathsf{p}}(k) \triangleq d(k) - \boldsymbol{w}^{T}(k+1)\boldsymbol{x}(k)$ denotes the posterior error. The term $\|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^{2}$ in (87) incorporates the minimum distortion principle, which adopts a conservative guideline that prevents the adaptive coefficients from suffering abrupt changes. Deterministic analyses may elucidate relevant features of the LMS algorithm, as, for example the presence of amplified components of high-frequency disturbances in the error output [85]. Standard statistical assumptions, which are usually applied on LMS

stochastic models are discussed in the next section for the time-variant case.

5.3- Standard LMS Tracking Analysis

This section presents a concise theoretical tracking analysis of the LMS, since the main derivations were covered in Chapter 3. There are some differences here, since the time-variant setup is considered. The following discussion also helps to emphasize by contrast some import features of the EEA. Among the reasons that motivate EEA, one may highlight that adaptive filters exhibiting good learning behavior in stationary environments do not necessarily present good tracking ability when identifying time-variant plants [86]. Considering the deviation vector $\tilde{\boldsymbol{w}}(k)$ as

$$\tilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}^{\star}(k) - \boldsymbol{w}(k), \tag{88}$$

which explicitly considers the variations of the ideal plant and using (86) and (29), the following nonhomogeneous stochastic difference equation can be derived:

$$\tilde{\boldsymbol{w}}(k+1) = \left[\boldsymbol{I}_N - \beta \boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] \tilde{\boldsymbol{w}}(k) - \beta \boldsymbol{x}(k)\nu(k) + \boldsymbol{q}(k), \tag{89}$$

where $\boldsymbol{I}_N$ is the identity matrix of size N. Rigorously speaking, the driving forces $-\beta \boldsymbol{x}(k)\nu(k)$ and $\boldsymbol{q}(k)$ in (89) precludes asymptotic convergence of the deviation vector energy to zero, even if the algorithm operates in the stability region. The average behavior of the deviation vector can be studied by applying the expectation operator in (89). Some complicated joint moments usually are not computed due to the use of IA and NA, as we have seen. Using these assumptions, the mean weight tracking behavior of the LMS can be written as

$$\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] = \left[\boldsymbol{I}_N - \beta \boldsymbol{R}\right] \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right], \tag{90}$$

which implies that the LMS performs an unbiased asymptotic estimation and that the adaptive coefficients, despite their Brownian motion, behave on average like the weights of the LMS with a fixed plant. Moreover, under mild conditions, it may be stated that the

expected value of $\boldsymbol{w}(k)$ converges to $\boldsymbol{w}^\star$ (in a stationary setting) if the step size satisfies

$$\beta < \frac{2}{\rho\left[\boldsymbol{R}\right]}, \tag{91}$$

with $\rho\left[\boldsymbol{R}\right]$ standing for the *spectral radius* of $\boldsymbol{R}$, *i.e.*, $\rho\left[\boldsymbol{R}\right] \triangleq \max_i \left|\lambda_i\left(\boldsymbol{R}\right)\right|$, where λ_i is the i-th eigenvalue of $\boldsymbol{R}$.

Since stable-in-the-mean adaptive filters can diverge in practice due to an unbounded variance of the adaptive weight vector (see Chapter 3), the first-order analysis (90) is not very useful. Alternatively, a second-order (or mean square) analysis may be employed to estimate either the MSD or the MSE, as discussed in Chapter 3, whose equations are again presented for convenience:

$$\text{MSD}(k) \;=\; \varsigma(k) = \mathbb{E}\left[\|\tilde{\boldsymbol{w}}(k)\|^2\right], \tag{92}$$

$$\text{MSE}(k) \;=\; \xi(k) \approx \sigma_\nu^2 + \mathbb{E}\left[\boldsymbol{R}_x \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right], \tag{93}$$

where $\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) \triangleq \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right]$ is the system mismatch covariance matrix (also known as the weight error covariance matrix) and σ_ν^2 is the additive noise variance. Approximation (93) employs the NA, which remains as an almost universal hypothesis. Although it often is approximately true when $\nu(k)$ derives directly from quantization issues, one cannot assure that it will be always the case, since a colored interference (*e.g.*, a speech signal) may be captured by the transducers. Papers [55; 87] are important references that do not assume whiteness of the additive noise. Furthermore, approximation (93) helps to elucidate the major role that $\boldsymbol{R}$ plays in the standard statistical analyses of adaptive filtering algorithms. As will be clearer in the next section, such a role is less pronounced under the EEA paradigm.

Multiplying (89) by its transpose, one obtains

$$\begin{aligned}\boldsymbol{\Theta}(k+1) \;=\; & \boldsymbol{\Theta}(k) - \beta\boldsymbol{\Theta}(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k) + \mathcal{O}\left[\nu(k), \boldsymbol{q}(k)\right] \\ & -\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\boldsymbol{\Theta}(k) + \beta^2\nu^2(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k) \\ & +\beta^2\boldsymbol{x}(k)\boldsymbol{x}^T(k)\boldsymbol{\Theta}(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k) + \boldsymbol{q}(k)\boldsymbol{q}^T(k),\end{aligned} \tag{94}$$

where $\boldsymbol{\Theta}(k) \triangleq \tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)$ and $\mathcal{O}\left[\nu(k), \boldsymbol{q}(k)\right]$ contains first-order components of $\nu(k)$ and $\boldsymbol{q}(k)$ that do not impact the following derivations.

5.4- Exact Expectation Tracking Analysis

Model (86) is the most commonly employed in the study of the effects of nonstationarity in adaptive filtering systems [88]. In this section, an analysis of the LMS tracking abilities is performed without IA for the first time. Several theoretical analyses also assume independence between samples of the input sequence, which can be stated as follows:

$\star$ *White Input Assumption* (WIA). The excitation data $x(k)$ is a stationary sequence of independent random variables.

Remark: WIA is a very popular assumption that renders the analysis more tractable, but can be incoherent with practical settings [89; 90]. In this chapter, WIA is circumvented by modeling $x(k)$ as an output sequence of an M-length moving average system $B(z)$ applied to an unitary-variance white signal with an even probability density function (pdf) $u(k)$:

$$x(k) = \sum_{i=0}^{M-1} b_i u(k-i). \tag{95}$$

Remark: Note that model (95) does not assume a Gaussian $x(k)$, which is another popular hypothesis [91; 92; 93; 45; 54].

Loosely speaking, the EEA is a systematic procedure that recursively generates update equations for the statistical quantities of interest [50]. The method constructs a linear state space model similar to (49), which contains the necessary information of the algorithm performance dynamics.

Our proposal does not restrict the measurement noise to be white, which can present a finite memory character, since it is assumed to be generated by an L-length moving average model:

$$\nu(k) \triangleq \sum_{i=0}^{L-1} a_i v(k-i), \tag{96}$$

where $v(k)$ is an unitary-variance i.i.d. process with a symmetric distribution. Also, (96) does not assume that $\nu(k)$ is Gaussian.

Consider the configuration $(N, M, L) = (1, 2, 2)$, which maintains equation lengths tractable, in the following derivations, where N is the filter length and parameters M and L are related to the length of the filters responsible for the coloring of $x(k)$ and $\nu(k)$

(see (95)-(96)). Although the validity of the equations presented next is restricted to the considered configuration, the construction procedure is the same for more complex settings.

5.4.1 Mean Weight Analysis

Recursion (89) allows for the writing of the following recursion:

$$\tilde{w}_0(k+1) = (1 - \beta x^2(k))\tilde{w}_0(k) - \beta x(k)\nu(k) + q_0(k), \tag{97}$$

where $\tilde{w}_i(k)$ (resp. $q_i(k)$) is the i-th element of vector $\tilde{\boldsymbol{w}}(k)$ (resp. $\boldsymbol{q}(k)$), for $i \in \{0, 1, \ldots, N-1\}$. In a standard stochastic approach (*i.e.*, an IA-based one), the application of the expectation operator in (97) together with NA and (95) leads to

$$\mathbb{E}\left[\tilde{w}_0(k+1)\right] = \left[1 - \beta(b_0^2 + b_1^2)\gamma_2\right]\mathbb{E}\left[\tilde{w}_0(k)\right], \tag{98}$$

where $\gamma_n \triangleq \mathbb{E}\left[u^n(k)\right]$ and b_i (for $i \in \{0, 1\}$) derives from model (95). Eq. (98) is a simpler recursion than the one obtained with EEA, since in the EEA case one cannot make use of the IA-based approximation[1] $\mathbb{E}\left[u^n(k-j)\tilde{w}_i^m(k)\right] \approx \gamma_n \mathbb{E}\left[\tilde{w}_i^m(k)\right]$, for $j > 0$. Therefore, EEA generates the following update equation for the state variable $\mathbb{E}\left[\tilde{w}_0(k)\right]$:

$$\mathbb{E}[\tilde{w}_0(k+1)] = (1 - b_0^2\beta\gamma_2)\mathbb{E}[\tilde{w}_0(k)] - b_1^2\beta\mathbb{E}[u^2(k-1)\tilde{w}_0(k)], \tag{99}$$

where one may note the emergency of a *nuisance* state variable $\mathbb{E}[u^2(k-1)\tilde{w}_0(k)]$. Since (99) is no longer a self-contained recursion (like (98)), an additional recursion for the nuisance term $\mathbb{E}[u^2(k-1)\tilde{w}_0(k)]$ is necessary. This recursion may be obtained by multiplying both sides of (97) by $u^2(k)$, before the application of the expectation operator, which leads to:

$$\begin{aligned} \mathbb{E}[u^2(k)\tilde{w}_0(k+1)] &= (\gamma_2 - b_0^2\beta\gamma_4)\mathbb{E}[\tilde{w}_0(k)] \\ &\quad - b_1^2\beta\gamma_2\mathbb{E}[u^2(k-1)\tilde{w}_0(k)]. \end{aligned} \tag{100}$$

[1] Random variable $\tilde{w}_i(k)$ is indeed statistically dependent from $u(k)$, although it is not the case with $x(k)$, unless when $x(k)$ is a white input signal (*i.e.*, $M = 1$ in (95)).

Equations (99) and (100) may be concisely rewritten as a self-contained linear state space equation system:

$$\boldsymbol{y}^{(1)}(k+1) = \boldsymbol{A}^{(1)}\boldsymbol{y}^{(1)}(k), \tag{101}$$

where

$$\boldsymbol{y}^{(1)}(k) = \begin{bmatrix} \mathbb{E}[\tilde{w}_0(k)] \\ \mathbb{E}[u^2(k-1)\tilde{w}_0(k)] \end{bmatrix} \tag{102}$$

contains the *state* of the statistical quantities the theoretical analysis is interested in at the k-th iteration and the time-invariant transition matrix is

$$\boldsymbol{A}^{(1)} = \begin{bmatrix} 1 - b_0^2\beta\gamma_2 & -b_1^2\beta \\ \gamma_2 - b_0^2\beta\gamma_4 & -b_1^2\beta\gamma_2 \end{bmatrix}. \tag{103}$$

The two eigenvalues of matrix $\boldsymbol{A}^{(1)}$ can be written as

$$\lambda_1 = \frac{1 + \sqrt{\Delta} + \left(-b_0^2 - b_1^2\right)\beta\gamma_2}{2}, \tag{104}$$

$$\lambda_2 = \frac{1 - \sqrt{\Delta} + \left(-b_0^2 - b_1^2\right)\beta\gamma_2}{2}, \tag{105}$$

where

$$\Delta = \beta^2 b_1^4\gamma_2^2 - 2\beta\left(\beta\left(\gamma_2^2 - 2\gamma_4\right)b_0^2 + \gamma_2\right)b_1^2 + \left(b_0^2\beta\gamma_2 - 1\right)^2. \tag{106}$$

The eigenvalues of the transition matrices depend on β in a complex way, which indicates the higher sophistication of taking into account additional stochastic couplings. One may state that in the considered configuration the LMS is first-order stable in the exact expectation sense if β is chosen so that $\max\{|\lambda_1|, |\lambda_2|\} \leq 1$. Note that neither the additive noise variance σ_ν^2 nor σ_q^2 (the variance of $q_i(k)$, for $i \in \{0, 1, \ldots, N-1\}$) influence the stability in the mean sense, which is a remarkable result, since it was firstly established under IA hypothesis. Here, at least for the considered configuration, such a result is proved to be valid even when IA is not employed.

5.4.2 Mean Square Analysis

Under the considered setting and circumventing the employment of the ubiquitous IA, the MSE at the k-th iteration may be expressed as

$$\begin{aligned} \mathrm{MSE}(k) &= \mathbb{E}\left[e^2(k)\right] = \mathbb{E}\left[\left(\tilde{w}_0(k)x(k)+\nu(k)\right)^2\right] \\ &= (a_0^2+a_1^2)\sigma_v^2 + b_0^2\gamma_2\mathbb{E}[\tilde{w}_0^2(k)] + b_1^2\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)]. \end{aligned} \tag{107}$$

Thus, a theoretical EEA-based predicion of the MSE requires two statistical quantities:

$$\mathbb{E}[\tilde{w}_0^2(k)] \text{ and } \mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)]. \tag{108}$$

Finding a recursion for the former quantity may be done by squaring both sides of (97) and applying the expectation operator, resulting in

$$\begin{aligned} \mathbb{E}[\tilde{w}_0^2(k+1)] &= (1-2b_0^2\gamma_2\beta+b_0^4\beta^2\gamma_4)\mathbb{E}[\tilde{w}_0^2(k)] + \sigma_q^2 \\ &\quad +(6b_0^2\beta^2b_1^2\gamma_2 - 2b_1^2\beta)\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\ &\quad +(a_0^2+a_1^2)b_0^2\sigma_\nu^2\beta^2\gamma_2 + (a_0^2+a_1^2)b_1^2\sigma_v^2\beta^2\gamma_2 \\ &\quad +b_1^4\beta^2\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)], \end{aligned} \tag{109}$$

where $\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)]$ is a nuisance state variable, whereas $\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)]$ is not, because it is required for computing the MSE (see (107)). Eq. (109) is not self-contained, since it requires the evaluation of two state variables, namely: $\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)]$ and $\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)]$. By carrying out similar steps to those that lead to (100), one may conclude that

$$\begin{aligned} \mathbb{E}[u^2(k)\tilde{w}_0^2(k+1)] &= (\gamma_2 - 2b_0^2\gamma_4\beta + b_0^4\beta^2\gamma_6)\mathbb{E}[\tilde{w}_0^2(k)] + \gamma_2\sigma_q^2 \\ &\quad +(6b_0^2\beta^2b_1^2\gamma_4 - 2b_1^2\beta\gamma_2)\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\ &\quad +b_0^2\left(a_0^2+a_1^2\right)\sigma_v^2\beta^2\gamma_4 \\ &\quad +b_1^2\left(a_0^2+a_1^2\right)\sigma_v^2\beta^2\gamma_2^2 \\ &\quad +b_1^4\beta^2\gamma_2\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)], \end{aligned} \tag{110}$$

and

$$\begin{aligned}\mathbb{E}[u^4(k)\tilde{w}_0^2(k+1)]=&(\gamma_4-2b_0^2\gamma_6\beta+b_0^4\beta^2\gamma_8)\mathbb{E}[\tilde{w}_0^2(k)]+\gamma_4\sigma_q^2\\&+(6b_0^2\beta^2b_1^2\gamma_6-2b_1^2\beta\gamma_4)\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)]\\&+b_0^2\left(a_0^2+a_1^2\right)\sigma_v^2\beta^2\gamma_6\\&+b_1^2\left(a_0^2+a_1^2\right)\sigma_v^2\beta^2\gamma_4\gamma_2\\&+b_1^4\beta^2\gamma_4\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)].\end{aligned}\tag{111}$$

A linear time-domain state space equation may concisely represent (109)-(111):

$$\boldsymbol{y}^{(2)}(k+1)=\boldsymbol{A}^{(2)}\boldsymbol{y}^{(2)}(k)+\boldsymbol{d}^{(2)},\tag{112}$$

where

$$\boldsymbol{A}^{(2)}=\begin{bmatrix}1-2b_0^2\gamma_2\beta+b_0^4\beta^2\gamma_4 & -2b_1^2\beta+6b_0^2\beta^2b_1^2\gamma_2 & b_1^4\beta^2\\ \gamma_2-2b_0^2\gamma_4\beta+b_0^4\beta^2\gamma_6 & -2b_1^2\beta\gamma_2+6b_0^2\beta^2b_1^2\gamma_4 & b_1^4\beta^2\gamma_2\\ \gamma_4-2b_0^2\gamma_6\beta+b_0^4\beta^2\gamma_8 & -2b_1^2\beta\gamma_4+6b_0^2\beta^2b_1^2\gamma_6 & b_1^4\beta^2\gamma_4\end{bmatrix},\tag{113}$$

$$\boldsymbol{y}^{(2)}(k)=\begin{bmatrix}\mathbb{E}[\tilde{w}_0^2(k)]\\ \mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)]\\ \mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)]\end{bmatrix},\tag{114}$$

and

$$\boldsymbol{d}^{(2)}=\begin{bmatrix}\sigma_q^2+b_0^2(a_0^2+a_1^2)\sigma_v^2\beta^2\gamma_2+b_1^2(a_0^2+a_1^2)\sigma_v^2\beta^2\gamma_2\\ \gamma_2\sigma_q^2+b_0^2(a_0^2+a_1^2)\sigma_v^2\beta^2\gamma_4+b_1^2(a_0^2+a_1^2)\sigma_v^2\beta^2\gamma_2^2\\ \gamma_4\sigma_q^2+b_0^2(a_0^2+a_1^2)\sigma_v^2\beta^2\gamma_6+b_1^2(a_0^2+a_1^2)\sigma_v^2\beta^2\gamma_4\gamma_2\end{bmatrix}.\tag{115}$$

The model under consideration requires $R=3$ internal state variables, of which only one is a nuisance parameter. The increase of N and M implies a substantial increment of R, and for most configurations the majority state variables are nuisance joint moments (*e.g.*, for $(N,M,L)=(8,1,1)$, $R=2,438,009$, whereas the number of non-nuisance terms is only 29). The large amount of equations necessary for performing EEA means that commercially available algebraic symbolic packages are not useful, and we utilized our own efficient C++-based code responsible for tackling the issue.

Vector $\boldsymbol{d}^{(2)}$ can be decomposed into two parts:

$$\boldsymbol{d}^{(2)}=\boldsymbol{d}_{\text{std}}^{(2)}+\boldsymbol{d}_q^{(2)},\tag{116}$$

where $\boldsymbol{d}_q^{(2)} \in \mathbb{R}^R$ only contains terms that depends on statistic features of vector $\boldsymbol{q}(k)$. For example, by applying the decomposition on $\boldsymbol{d}^{(2)}$ from (115), $\boldsymbol{d}_q^{(2)}$ can be written as:

$$\boldsymbol{d}_q^{(2)} = \begin{bmatrix} \sigma_q^2 & \gamma_2 \sigma_q^2 & \gamma_4 \sigma_q^2 \end{bmatrix}^T. \tag{117}$$

Decomposition (116) is important because it can describe asymptotic operation (assuming algorithm stability) as a sum of two terms:

$$\lim_{k\to\infty} \boldsymbol{y}^{(2)}(k) = \underbrace{\left(\boldsymbol{I}_R - \boldsymbol{A}^{(2)}\right)^{-1} \boldsymbol{d}_{\text{std}}^{(2)}}_{\triangleq \boldsymbol{y}_{\text{ss,std}^{(2)}}} + \underbrace{\left(\boldsymbol{I}_R - \boldsymbol{A}^{(2)}\right)^{-1} \boldsymbol{d}_q^{(2)}}_{\triangleq \boldsymbol{y}_q^{(2)}}, \tag{118}$$

where $\boldsymbol{y}_q^{(2)}$ isolates the contribution of the tracking task to steady-state statistical quantities. Obviously, such a decoupling is also possible with IA-based analyses.

Employing EEA, the upper bound β_{max} on the step size that guarantees that the state space realization is stable (*i.e.*, the algorithm converges) can be written as

$$\beta_{\text{max}} \triangleq \sup \left\{ \beta \text{ such that } \rho\left[\boldsymbol{A}^{(2)}\right] < 1 \right\}. \tag{119}$$

Remarks: Note that (119) is able to provide a more accurate upper bound, guaranteeing stability of the LMS, than the one provided by IA-based standard approaches, because instability issues occur for large step sizes, when such approaches are not precise [44]. The number of equations of even a slightly complex configuration (*i.e.* $(N, M, L) = (2, 2, 1)$) is too large to present here because it requires 48 recursions, which can be obtained by following the same steps described in this section. In this case, the length of the recursions (which are not the larger ones of the considered configuration) of the state variables $\mathbb{E}[\tilde{w}_0^2(k)]$ and $\mathbb{E}[\tilde{w}_1^2(k)]$ are somewhat cumbersome:

$$
\begin{aligned}
\mathbb{E}[\tilde{w}_0^2(k+1)] &= (1-2b_0^2\gamma_2\beta+b_0^4\beta^2\gamma_4)\mathbb{E}[\tilde{w}_0^2(k)]+(-2b_1^2\beta+6b_0^2\beta^2b_1^2\gamma_2)\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\
&\quad +(-2b_1b_0\beta+6b_0^3\beta^2b_1\gamma_2)\mathbb{E}[u^2(k-1)\tilde{w}_1(k)\tilde{w}_0(k)] \\
&\quad +\vartheta\mathbb{E}[u(k-2)u(k-1)\tilde{w}_1(k)\tilde{w}_0(k)]+\sigma_q^2 \\
&\quad +b_0^2a_0^2\sigma_\nu^2\beta^2\gamma_2+b_0^4\beta^2\gamma_2\mathbb{E}[u^2(k-1)\tilde{w}_1^2(k)] \\
&\quad +2b_0^3\beta^2b_1\gamma_2\mathbb{E}[u(k-2)u(k-1)\tilde{w}_1^2(k)] \\
&\quad +b_0^2b_1^2\beta^2\gamma_2\mathbb{E}[u^2(k-2)\tilde{w}_1^2(k)]+b_1^2a_0^2\sigma_\nu^2\beta^2\gamma_2 \\
&\quad +b_1^4\beta^2\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)]+2b_1^3\beta^2b_0\mathbb{E}[u^4(k-1)\tilde{w}_0(k)\tilde{w}_1(k)] \\
&\quad +2b_1^4\beta^2\mathbb{E}[u^3(k-1)u(k-2)\tilde{w}_0(k)\tilde{w}_1(k)]+b_1^2b_0^2\beta^2\mathbb{E}[u^4(k-1)\tilde{w}_1^2(k)] \\
&\quad +2b_1^3b_0\beta^2\mathbb{E}[u^3(k-1)u(k-2)\tilde{w}_1^2(k)] \\
&\quad +b_1^4\beta^2\mathbb{E}[u^2(k-1)u^2(k-2)\tilde{w}_1^2(k)], \qquad (120)
\end{aligned}
$$

where $\vartheta \triangleq -2b_1^2\beta+6b_0^2\beta^2b_1^2\gamma_2$, and

$$
\begin{aligned}
\mathbb{E}[\tilde{w}_1^2(k+1)] &= \mathbb{E}[\tilde{w}_1^2(k)]-2b_0b_1\beta\mathbb{E}[u^2(k-1)\tilde{w}_0(k)\tilde{w}_1(k)]-2b_0^2\beta\mathbb{E}[u^2(k-1)\tilde{w}_1^2(k)] \\
&\quad -4b_0b_1\beta\mathbb{E}[u(k-2)u(k-1)\tilde{w}_1^2(k)] \\
&\quad -2b_1^2\beta\mathbb{E}[u(k-1)u(k-2)\tilde{w}_0(k)\tilde{w}_1(k)]-2b_1^2\beta\mathbb{E}[u^2(k-2)\tilde{w}_1^2(k)] \\
&\quad +\sigma_q^2+b_0^2a_0^2\sigma_\nu^2\beta^2\gamma_2+b_0^4\gamma_2\beta^2\mathbb{E}[u^2(k-1)\tilde{w}_0^2(k)] \\
&\quad +2b_0^3\gamma_2\beta^2b_1\mathbb{E}[u(k-1)u(k-2)\tilde{w}_0^2(k)]+b_0^2b_1^2\beta^2\mathbb{E}[u^4(k-1)\tilde{w}_0^2(k)] \\
&\quad +2b_0^3b_1\beta^2\mathbb{E}[u^4(k-1)\tilde{w}_0(k)\tilde{w}_1(k)] \\
&\quad +6b_0^2b_1^2\beta^2\mathbb{E}[u^3(k-1)u(k-2)\tilde{w}_0(k)\tilde{w}_1(k)] \\
&\quad +2b_0b_1^3\beta^2\mathbb{E}[u^3(k-1)u(k-2)\tilde{w}_0^2(k)] \\
&\quad +6b_0b_1^3\beta^2\mathbb{E}[u^2(k-1)u^2(k-2)\tilde{w}_0(k)\tilde{w}_1(k)] \\
&\quad +b_0^4\beta^2\mathbb{E}[u^4(k-1)\tilde{w}_1^2(k)]+4b_0^3\beta^2b_1\mathbb{E}[u^3(k-1)u(k-2)\tilde{w}_1^2(k)] \\
&\quad +6b_0^2\beta^2b_1^2\mathbb{E}[u^2(k-1)u^2(k-2)\tilde{w}_1^2(k)]+4b_0b_1^3\beta^2\mathbb{E}[u(k-1)u^3(k-2)\tilde{w}_1^2(k)] \\
&\quad +b_1^2a_0^2\sigma_\nu^2\beta^2\gamma_2+b_1^2b_0^2\gamma_2\beta^2\mathbb{E}[u^2(k-2)\tilde{w}_0^2(k)] \\
&\quad +b_1^4\beta^2\mathbb{E}[u^2(k-2)u^2(k-1)\tilde{w}_0^2(k)]+2b_1^4\beta^2\mathbb{E}[u^3(k-2)u(k-1)\tilde{w}_0(k)\tilde{w}_1(k)] \\
&\quad +b_1^4\beta^2\mathbb{E}[u^4(k-2)\tilde{w}_1^2(k)]. \qquad (121)
\end{aligned}
$$

In order to gain a better grasp of the computational burden of EEA, Table 2 and

Table 2 – Computational performance comparison for $(N, M, L) = (6, 1, 2)$ and $(N, M, L) = (2, 4, 6)$.

	$(N, M, L) = (6, 1, 2)$		$(N, M, L) = (2, 4, 6)$	
	EEA	IA	EEA	IA
CPU Cycles	106×10^9	23×10^6	961×10^9	1×10^9
Instructions	163×10^9	53×10^6	1.2×10^{12}	894×10^6
Memory	583×10^6	5×10^6	896×10^6	10×10^6
Equations	28181	6	3517	3
Time	$33s$	$0.02s$	$291s$	$0.33s$

Table 3 present a comparison of some key computational metrics between the `EEA` and `IA` methods. The `C++` program was compiled using the `-Ofast` optimization directive and we employed the `perf` performance analyzing tool in Linux to get the results. The table describes the number of instructions, cycles, amount of memory used (peak), equations generated and time elapsed for two different configurations. Namely, the number of CPU cycles and instructions is, roughly, between an order and two orders of magnitude worse for the `EEA`. Memory usage is increased by a constant factor. The number of equations required by the `IA` is in the single digits, and the execution time is reduced to less than one second. Overall, in all observed measures, the `EEA` method presents a significantly worse computational performance than the `IA` method.

Table 3 – Amount of equations necessary to MSD computation.

(N,M,L)		Equations	(N,M,L)		Equations
$(1,1,1)$	IA	1	$(2,2,1)$	IA	3
	EEA	1		EEA	48
$(1,1,2)$	IA	1	$(2,2,2)$	IA	3
	EEA	1		EEA	66
$(1,1,3)$	IA	1	$(2,2,3)$	IA	3
	EEA	1		EEA	91
$(1,2,1)$	IA	1	$(2,3,1)$	IA	3
	EEA	3		EEA	394
$(1,2,2)$	IA	1	$(2,3,2$	IA	3
	EEA	5		EEA	532
$(1,2,3)$	IA	1	$(2,3,3)$	IA	3
	EEA	7		EEA	731
$(1,3,1)$	IA	1	$(3,1,1)$	IA	6
	EEA	19		EEA	61
$(1,3,2)$	IA	1	$(3,1,2)$	IA	6
	EEA	33		EEA	68
$(1,3,3)$	IA	1	$(3,1,3)$	IA	6
	EEA	47		EEA	94
$(2,1,1)$	IA	3	$(3,2,1)$	IA	6
	EEA	7		EEA	698
$(2,1,2)$	IA	3	$(3,2,2)$	IA	6
	EEA	8		EEA	830
$(2,1,3)$	IA	3	$(3,2,3)$	IA	6
	EEA	11		EEA	1109

5.5- Theoretical Derivation of an Optimum Variable Step-Size Sequence

Once a constant step size has been chosen, a trade-off between convergence rate, steady-state performance, and tracking ability is established. This trade-off may be circumvented by the adoption of a time-variant step size $\beta(k)$ [94; 75], whose adjustment relies on a data-dependent control policy [95]. This mechanism typically employs observable quantities, *e.g.:* the instantaneous quadratic error, estimated correlation between adjacent error samples or even the estimated correlation between the error and the input vector [96]. Paper [28] argues that the design (*i.e.*, the parameters tuning for better performance) of a variable step size (VSS) scheme depends on the evaluation of its efficiency, which may be performed by comparing it with an optimal theoretical sequence. Unfortunately, the use of IA-based derivation of such a *deterministic optimal* sequence may induce divergence in practice, because standard analyses overestimate the β upper bound that guarantees stability [52]. In the following, a novel contribution is presented. It generalizes the EEA-based derivation of an optimal step-size sequence (proposed in [52]) in order to address the MSD minimization in the tracking setting. In this context, the involved optimum step sizes do not asymptotically converge to zero. Two versions of such an extension are described:

$\star$ *Version I*: at each iteration, the step size $\beta(k)$ is the same for all adaptive weights;

$\star$ *Version II*: each adaptive coefficient $w_i(k)$ is updated by the utilization of a particular step size $\beta_i(k)$.

Remark: The second version has the potential to enhance the design of proportionate adaptive algorithms, which distribute the "adaptation energy" in order to provide faster convergence of high-magnitude coefficients [97; 45; 98]. Since the updating process of each adaptive coefficient is independent from the other coefficients, the adaptation gains are distributed in order to speed up their convergence [99]. Proportionate algorithms are practical mainly because they determine the direction update based on current filter estimates without demanding any prior information [100].

5.5.1 Theoretical VSS Sequence - Version I

From (92), the MSD at the k-th iteration can be written as:

$$\varsigma(k+1) = \sum_{i=0}^{N-1} \mathbb{E}\left[\tilde{w}_i^2(k)\right] = \sum_{i=0}^{N-1} y_{\mathcal{I}_i}^{(2)}(k+1), \tag{122}$$

where $\mathcal{I}_i$ is the position of vector $\boldsymbol{y}^{(2)}(k+1)$ containing state variable $\mathbb{E}\left[\tilde{w}_i^2(k)\right]$. Using (112), Eq. (122) can be explicitly rewritten as a function of elements of vector $\boldsymbol{y}^{(2)}(k)$:

$$\varsigma(k+1) = \sum_{i=0}^{N-1}\sum_{j=0}^{R-1} a_{\mathcal{I}_i,j}^{(2)}(k) y_j^{(2)}(k) + \sum_{i=0}^{N-1} d_{\mathcal{I}_i}^{(2)}(k), \tag{123}$$

where $a_{\mathcal{I}_i,j}^{(2)}(k)$ is the (I_i, j) element of matrix $\boldsymbol{A}^{(2)}(k)$ and $d_{\mathcal{I}_i}^{(2)}(k)$ is the $\mathcal{I}_i$-th element of vector $\boldsymbol{d}^{(2)}(k)$. Note that these quantities are no longer time-invariant, since they now depend on a time-variant step size $\beta(k)$. Since $\beta(k)$ is an user-defined parameter, it can be adjusted in order to minimize the posterior MSD $\varsigma(k+1)$ by zeroing the following derivative:

$$\frac{\partial \varsigma(k+1)}{\partial \beta(k)} = \sum_{i=0}^{N-1}\sum_{j=0}^{R-1} \frac{\partial a_{\mathcal{I}_i,j}^{(2)}(k)}{\partial \beta(k)} y_j^{(2)}(k) + \sum_{i=0}^{N-1} \frac{\partial d_{\mathcal{I}_i}^{(2)}(k)}{\beta(k)} = 0, \tag{124}$$

which results in a linear equation w.r.t. $\beta(k)$, because both $a_{\mathcal{I}_i,j}^{(2)}(k)$ and $d_{\mathcal{I}_i}^{(2)}(k)$ depend quadratically on $\beta(k)$, see (113) and (115).

Remarks: The sequence $\{\beta(k)\}$ that solves (124) is *deterministic*. Furthermore, unlike traditional approaches, the optimization that provides the sequence construction procedure takes into account the stochastic coupling between the input data and the adaptive coefficients. This reduces the probability of divergence, especially in the initial phase of the learning curve [52]. The optimal sequence is necessary to plot the optimal trajectory in the *learning plane*, which provides useful guidelines for the designer of practical VSS strategies that rely on observable data [28]. It should be emphasized that in experiments, the deterministic sequence $\{\beta(k)\}$, chosen through (124), tends to assume larger values in the initial learning phase, and provides small steps when the error magnitude is small, as expected. This type of behaviour avoids the trade-off between steady-state misadjusment and convergence rate that exists when a fixed learning factor is employed [101].

5.5.2 Theoretical VSS Sequence - Version II

It is widely recognized that the use of different step sizes for each adaptive weight may enhance the convergence rate of the adaptive filter [97; 45; 98]. In this case, the LMS update equation is

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \boldsymbol{\Gamma}(k)\boldsymbol{x}(k)e(k), \tag{125}$$

where $\boldsymbol{\Gamma}(k) \in \mathbb{R}^{N\times N}$ is a diagonal matrix, with the i-th element of its main diagonal denoted as $\beta_i(k) \in \mathbb{R}_+$ (*i.e.*, the step size associated with the adaptive coefficient $w_i(k)$). Using similar steps to those presented in Section 5.4, it is possible to write a state space linear *time-variant* equation system:

$$\boldsymbol{y}^{(2)}(k+1) = \boldsymbol{A}^{(2,\text{II})}(k)\boldsymbol{y}^{(2)}(k) + \boldsymbol{d}^{(2,\text{II})}(k), \tag{126}$$

where $\boldsymbol{A}^{(2,\text{II})}(k)$ and $\boldsymbol{d}^{(2,\text{II})}(k)$ depend on $\{\beta_i(k)\}$ (for $i \in \{0, 1, \ldots, N-1\}$), and no longer on $\beta(k)$. This fact is indicated by the superscript "II". The optimal step sizes $\{\beta_i(k)\}$ for the k-th iteration can be evaluated by zeroing the following gradient

$$\nabla_{\boldsymbol{\beta}(k)}\left[\sum_{i=0}^{N-1}\sum_{j=0}^{R-1} a^{(2,\text{II})}_{\mathcal{I}_i,j}(k)y^{(2)}_j(k) + \sum_{i=0}^{N-1} d^{(2,\text{II})}_{\mathcal{I}_i}(k)\right] = \mathbf{0}_N, \tag{127}$$

where $\boldsymbol{\beta}(k) \triangleq \begin{bmatrix}\beta_0(k) & \beta_1(k) & \ldots & \beta_{N-1}(k)\end{bmatrix}^T$.

Remark: Usually, it is assumed that the step size sequence should asymptotically approach zero, in order to enhance steady-state performance by reducing the variance of the adaptive estimator. This is not true for the tracking scenario, which is the one considered in this chapter.

5.6- Results

This section presents simulation results to confirm the theoretical analysis presented earlier. In these simulations we compare the results of the proposed analysis (EEA) and of the standard analysis (IA-based) with empirical results derived from computer-based LMS simulations, that mirror the actual behavior of the LMS. The analysis that better matches the empirical result presents a more accurate prediction of the LMS performance. The empirical curves were obtained by computing the LMS behavior for K independent Monte Carlo trials.

In the following experiments, the elements of the initial vector $\boldsymbol{w}^{\star}(0)$ are equal to unity, and the adaptive coefficient vector are initialized with zeros. The zero-mean perturbation vector $\boldsymbol{q}(k)$ is white and Gaussian. Furthermore, their elements are statistically independent from the remaining random variables, which is coherent with the adopted Markovian model. Unless stated otherwise, $u(k)$ and the noise $\nu(k)$ are zero-mean, white and Gaussian. The input signal is generated by filtering a zero-mean unitary Gaussian signal using the filter $B(z) = 1 - 0.8z^{-1}$, in order to demonstrate that the advanced analysis does not require a white input signal.

Fig. 12 depicts the evolution of the expected value $\mathbb{E}[w_0(k)]$ along the iterations, for a scenario where $N = 5$, $M = 2$, $\beta = 0.075$, and $\sigma_{\nu}^2 = \sigma_q^2 = 10^{-2}$. Since it is related to first-order statistics, one usually does not expect significant divergence between simulated and theoretical curves, even under IA. Fig. 12 demonstrates that this is not always the case, because the large value of β in this configuration (relative to the stability upper bound) emphasizes the discrepancy of empirical results w.r.t. the standard model (*i.e.*, one based on IA). From Fig. 12, we can see that larger discrepancies happen in the beginning of the simulation, which was also observed in Chapter 4, which considers nonlinearities in the ideal plant. At iteration 25, the IA-based simulation expected value is around 15% higher than the EEA and empirical expected values.

Under stable operation, the MSE curve usually monotonically decreases up to some point, after which the algorithm reaches the steady-state operation. For some scenarios, though, as the one presented in Fig. 13, this description is inaccurate, but EEA can predict the correct behavior. Again, this phenomenon was also observed in Chapter 4, and it is also encountered in tracking setups.

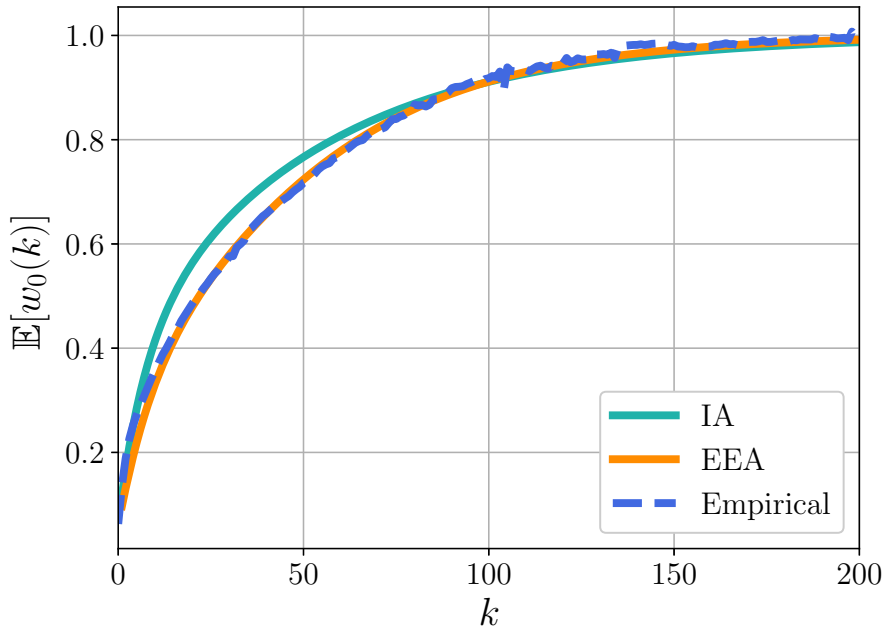


Figure 12 – Evolution of $\mathbb{E}[w_0(k)]$, with 10^6 independent Monte Carlo trials.

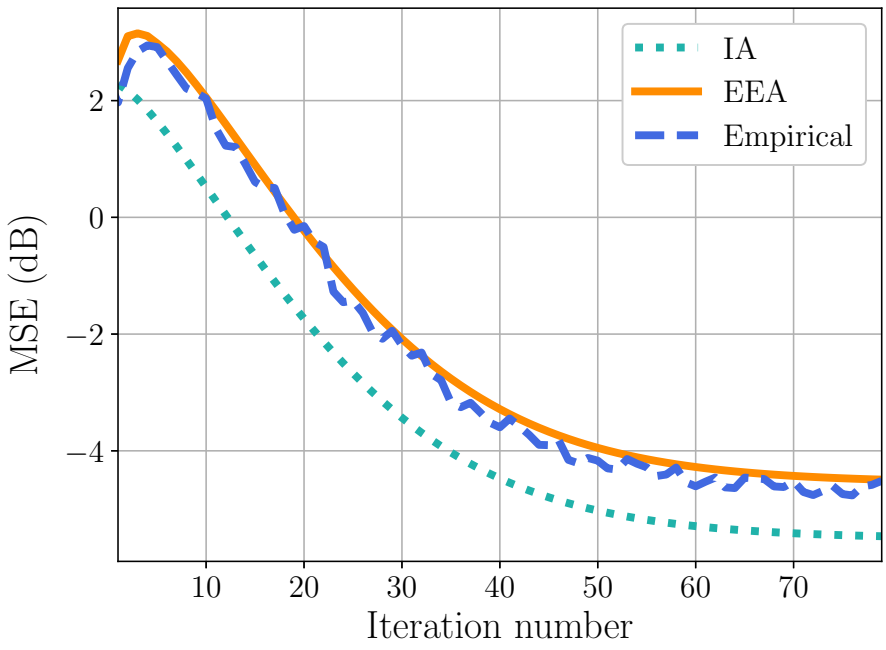


Figure 13 – Evolution of the MSE (in dB). The empirical results were obtained by the usage of $K = 10^{10}$ independent Monte Carlo trials.

As an example, consider the configuration $N = 3$, $\sigma_\nu^2 = 10^{-3}$ and $\sigma_q^2 = 10^{-2}$. Fig. 13 presents the MSE variation through the iterations for this configuration. In this scenario, EEA predicts an initial stage in which the algorithms experiment a performance

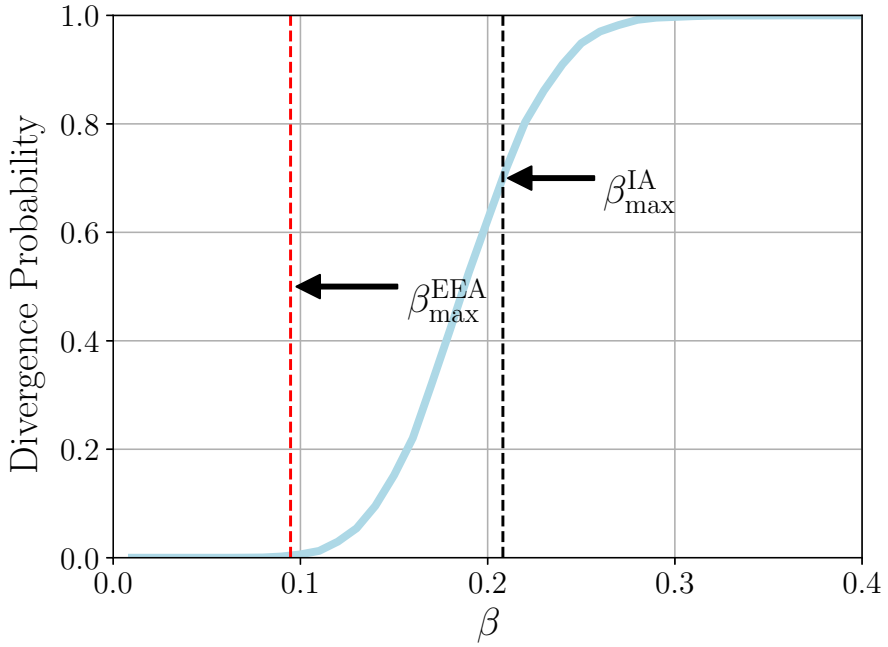


Figure 14 – Divergence probability as a function of β. The empirical curve (in blue) was obtained by the use of $K = 10^4$ independent Monte Carlo trials.

worse than the one obtained during initialization. Fig. 13 shows that this theoretical EEA forecast matches with the experimental curve, since the empirical curve shows an MSE increase in the first iterations. The standard analysis, on the other hand, is not capable of predicting the MSE increase and instead shows a more conventional behaviour where MSE always decreases.

Consider a scenario with $N = 3$, $M = 2$, $\sigma_q^2 = 10^{-2}$, $\sigma_\nu^2 = 10^{-3}$, which illustrates the advantage of our method in the determination of an accurate step size upper bound that avoids divergence. For each step size value, the LMS runs for 1000 iterations. The realization is assumed to experience instability if, in a specific realization, at least one adaptive parameter exceeds the threshold of 10 (*i.e.*, $|w_i(k)| > 10$). Using $K = 10^4$ independent trials, it is possible to estimate the divergence probability for each step size in a experimental manner. Fig. 14 shows the result of this estimation, that is, the probability of divergence as a function of β for this scenario, in solid line, along with the theoretical upper bound on β advocated by both IA and EEA analyses. The upper bound is obtained by the larger β value for which the magnitude of the maximum eigenvalue of the transition matrix is less than unity [50]. Note that the EEA upper bound actually avoids divergence, when it is not violated, whereas the `IA`-based upper bound on β is not trustworthy, since it results in approximately 70% of of divergence probability.

The asymptotic MSE for different values of the step size β is shown in Fig. 15, for a case where $N = 6$, $M = 1$, $\sigma_\nu^2 = 10^{-8}$ and $L = 5$. The colored additive noise is generated by filtering an white Gaussian noise by the following filter

$$A(z) = 1 - 0.8z^{-1} + 0.8z^{-2} - 0.7z^{-3} + 0.6z^{-4}. \tag{128}$$

As expected, Fig. 15 shows that larger step sizes lead to lower accuracy of IA-based analysis. That is, with the IA-based analysis we are not able to precisely predict the asymptotic MSE for β higher than 0.05.

The misadjustment, given by

$$\mathcal{M}(k) \triangleq \frac{\mathbb{E}[e^2(k)] - \sigma_\nu^2}{\sigma_\nu^2}, \tag{129}$$

is another important criterion used to asses the filter performance. The misadjustment is directly related to the excess MSE. In Fig. 16 we plot the evolution of the misadjustment considering $N = 3$, $M = 2$, $L = 1$, $\beta = 0.082$, $\sigma_\nu^2 = 10^{-2}$, and $\sigma_q^2 = 10^{-8}$. In this scenario, IA does not predict the misadjustment increase, in the first iterations which is a behavior similar to the one seen in Fig. 13. Using EEA, on the other hand, we are capable of correctly computing the misadjustment. Even in steady-state, the IA-based analysis predicts that the misadjustment is 0.5 dB smaller than actually is.

As a final observation, Equation (35) is a particular case of a linear time-invariant state space equation

$$\boldsymbol{y}(k+1) = \boldsymbol{A}\boldsymbol{y}(k) + \boldsymbol{b}, \tag{130}$$

where $\boldsymbol{y}(k) = \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right]$, $\boldsymbol{A} = \boldsymbol{I} - \beta\boldsymbol{R}$, and $\boldsymbol{b} = \boldsymbol{0}$. As has been clarified, the state space formulation is crucial in the novel models proposed in this thesis. IA-based analysis predicts that the misadjustment is 0.5 dB smaller than actually is.

Fig. 17 presents the steady-state MSE, but for a scenario with $N = 2$, $\sigma_\nu^2 = 10^{-3}$, and $\sigma_q^2 = 10^{-2}$. The curves of both Figs. 15 and 17 are important because they convey information that can be used as a guideline for the choice of the best step size (*i.e.*, one that yields the lower asymptotic MSE). Note that in Fig. 17 the adoption of the IA-based analysis leads to the misconception that the best step size is close to 0.14, but from the empirical results one can see that this step size value is actually located near the instability region. The EEA, on the other hand, predicts precisely the steady-state MSE for

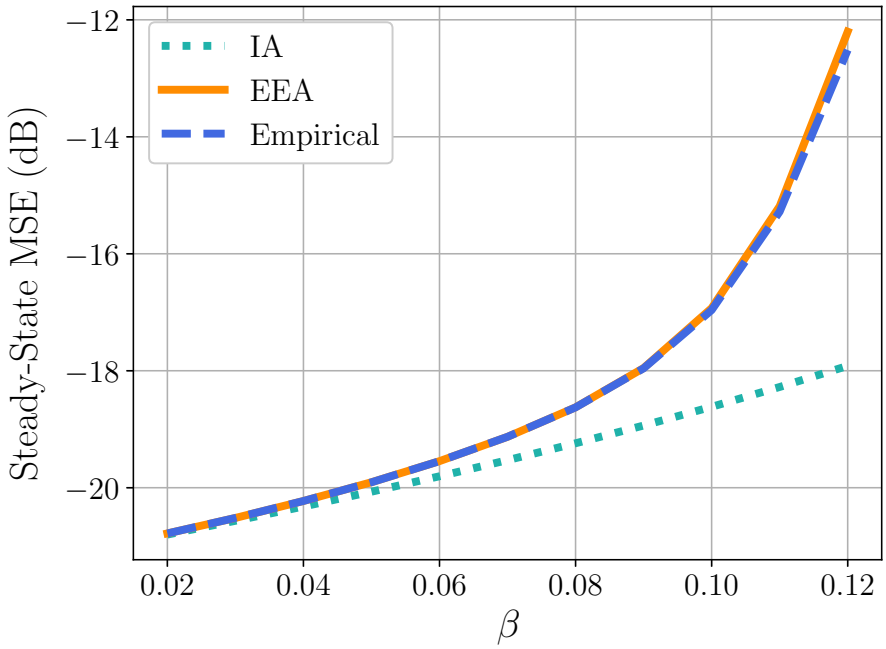


Figure 15 – Steady-state MSE as a function of the step size. The empirical curve was generated with $K = 10^5$ independent Monte Carlo trials.

each step size, allowing for the designer to choose a β that results in the optimal MSE in practice. Additionally, the impact of the Markovian noise $\boldsymbol{q}(k)$ can be assessed through decomposition (118), because the asymptotic MSE can be decoupled by a sum of two easily distinguishable contributions:

$$\lim_{k\to\infty} \text{MSE}(k) = \text{MSE}_{\text{ss,std}} + \text{MSE}_{\text{ss},q}, \tag{131}$$

where $\text{MSE}_{\text{ss},q}$ (*i.e.*, the tracking delay [102]) isolates the influence of the perturbation $\boldsymbol{q}(k)$, whereas $\text{MSE}_{\text{ss,std}}$ coincides with the steady-state mean squared error of the stationary case. The decomposition results are presented in Figs. 18 and 19. Fig. 18 depicts $\text{MSE}_{\text{ss},q}$ with respect to different β values. From Fig. 18 we can see that the standard analysis is not able to accurately model the impact of large step sizes, which is also valid for $\text{MSE}_{\text{ss,std}}$, as depicted in Fig. 19. It is interesting to note that such a lag contribution is not β-decreasing, as argued by some papers (*e.g.*, [102]). Furthermore, notice, by comparing Figs. 17, 18 and 19, that the contribution of $\text{MSE}_{\text{ss},q}$ is much more significant to the overall MSE (Fig. 17), under the considered setting.

The ability of the EEA to determine the optimal step size even in a high nonstationary scenario (*i.e.*, one with a large σ_q^2) can be seen in Fig. 20, in which

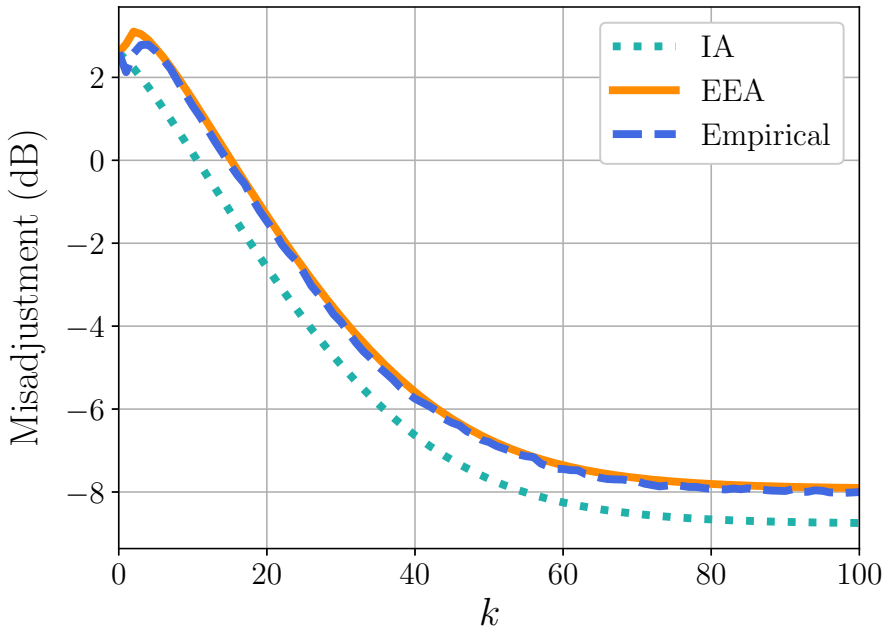


Figure 16 – Misadjustment evolution (in dB). The empirical results were obtained by the usage of $K = 10^7$ independent Monte Carlo Trials.

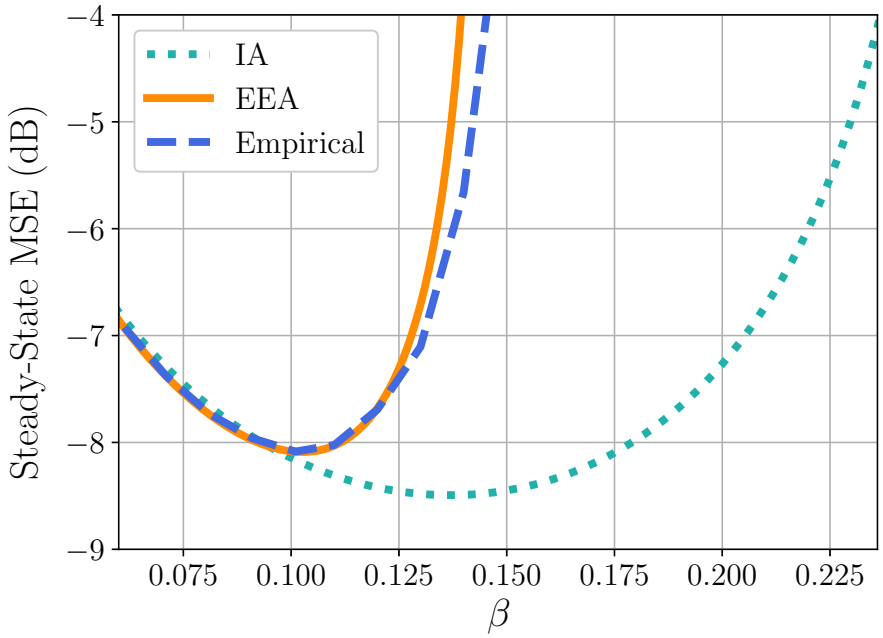


Figure 17 – Steady-state MSE as a function of the step size. The empirical curve was generated with $K = 10^6$ independent Monte Carlo trials.

the distribution of the signal $u(k)$ is Laplacian[2]. This experiment was performed with

[2]The Laplacian distribution was chosen to demonstrate that the advanced model does not require a Gaussian input signal

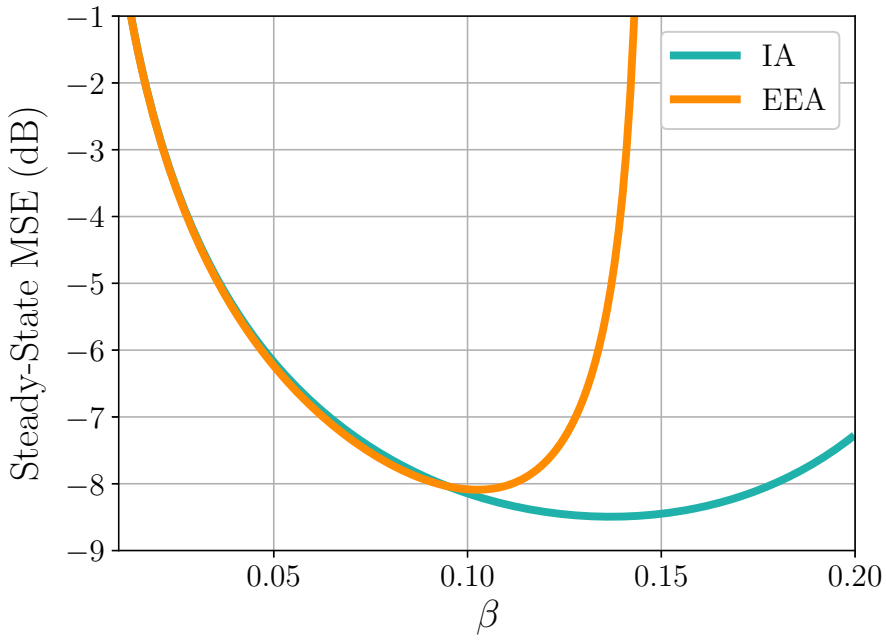


Figure 18 – Theoretical steady-state $\text{MSE}_{\text{ss},q}$ w.r.t. β using the configuration of Fig. 17.

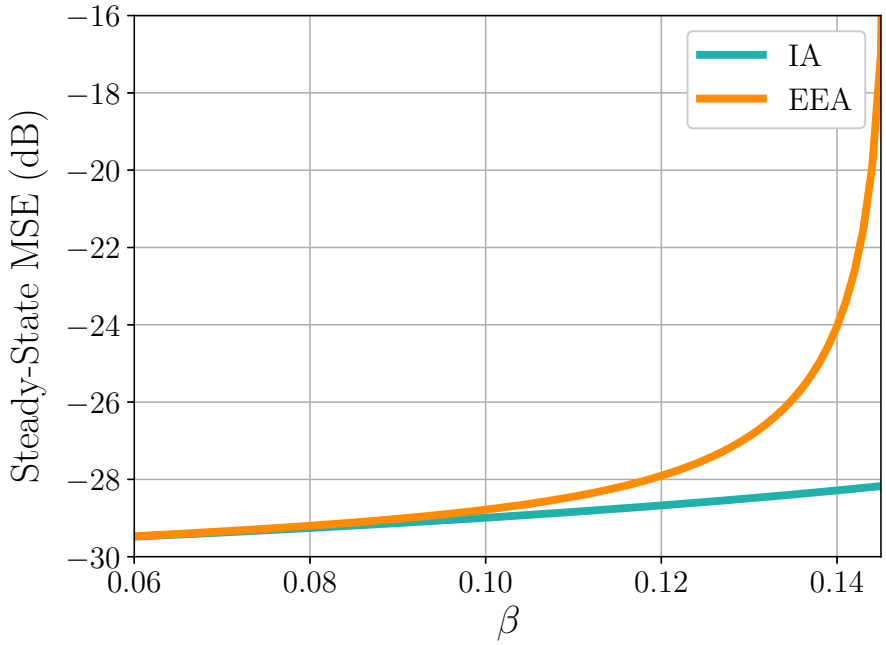


Figure 19 – Theoretical steady-state $\text{MSE}_{\text{ss,std}}$ w.r.t. β using the same configuration of Fig. 17.

$N = M = 2$ and $\sigma_\nu^2 = 10^{-3}$. In view of the unimodality of the steady-state MSE as a function of β (see Fig. 17), for each value of σ_q^2 a ternary search was used to obtain the β that minimizes the empirical steady-state value of the MSE. In addition, EEA indicates

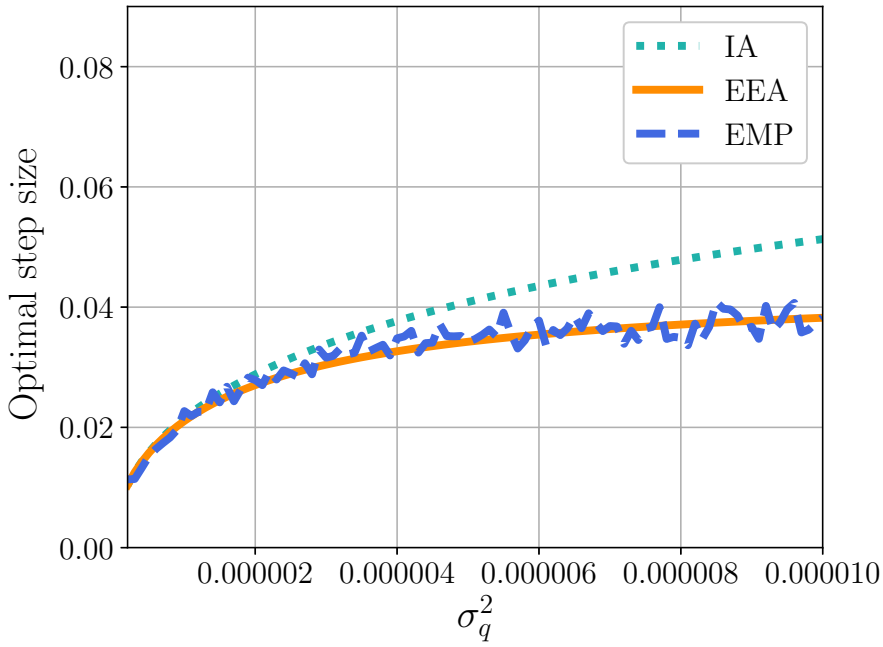


Figure 20 – This plot shows the optimal step size choice in function of σ_q^2. A total of 10^5 independent Monte Carlo trials were used to achieve the empirical results.

a proper step size value that optimizes asymptotic performance.

As stated in Section 5.5, the employment of a fixed step size value is often not an optimal choice, due to the inherent trade-off between convergence rate, asymptotic performance and instability issues. The impact of using the deterministic variable step size (VSS) derived in Section 5.5 can be seen in Fig. 21, which presents the evolution of a theoretical sequence $\beta(k)$, obtained using Version I of the VSS method described in Section 5.5. The considered setup employs $N = 3$, $M = 2$, $\sigma_\nu^2 = 10^{-4}$ and $\sigma_q^2 = 10^{-4}$. It should be emphasized that the standard model suggests a larger step size for enhanced performance, whereas our proposal indicates a more conservative step size magnitude, which is coherent with the discussion presented in[3] [52]. Furthermore, the different sequences do not converge asymptotically to zero, due to the fact that the use of infinitesimally small step sizes hampers the tracking capabilities of the adaptive algorithm.

The effectiveness of the optimal deterministic step size sequence obtained using both theoretical models is assessed in Fig. 22. This figure depicts the resulting empirical MSE evolution obtained by the utilization of the referred sequences. The employment of the sequence suggested by the standard model tends to induce divergence in practice, especially in the first phase of learning, which is again coherent with the results of [52].

[3]However [52] does not address the LMS tracking abilities.

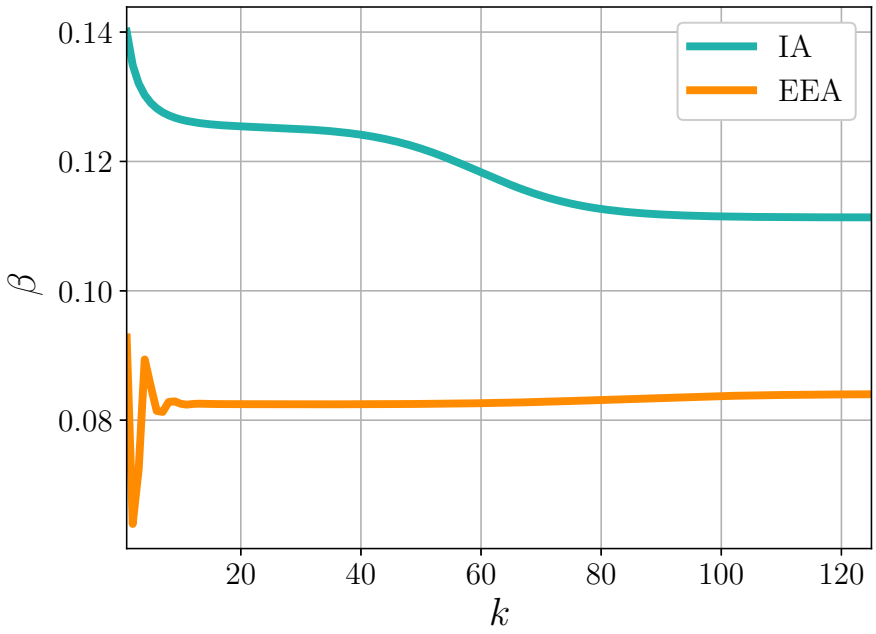


Figure 21 – Theoretical sequences $\beta(k)$ obtained by `EEA` and `IA` analyses.

Although the divergence occurs with low probability (which explains the elevated number of independent Monte Carlo trials used in the experiment in order to capture its effects), this is due to a catastrophic failure whose avoidance is crucial in practice. We can also see from Fig. 22 that by using the EEA sequence, MSE reaches steady-state approximately 50 iterations before the solution that uses the IA sequence.

Fig. 23 shows deterministic step size sequences (for both standard and advanced models) for Version II of the VSS technique derived in Section 5.5, which allows for distinct step size values for each adaptive coefficient. The figure setup has the following specifications: $N = 3$, $M = 2$, $\sigma_\nu^2 = 10^{-4}$, and $\sigma_q^2 = 10^{-4}$. Once again, the standard model suggests a more aggressive step size policy, in order to maximize the convergence rate. In theory, this type of step sizes may induce divergence, since an `IA`-based technique does not infer correctly the upper bound of β that avoids instability, as demonstrated in Fig. 14. The sequences suggested by both analyses can be compared by the empirical evaluation of the different step size sequences. This scenario, the same of Fig. 23, is shown in Fig. 24 which reveals that the theoretical step size sequence engineered by an `IA`-based stochastic model behaves in a far-from-optimal manner. On the other hand, with the EEA sequence the MSD, shown in Fig. 24, decreases throughout the iterations, as desired. By using the EEA sequence, MSD reaches steady-state

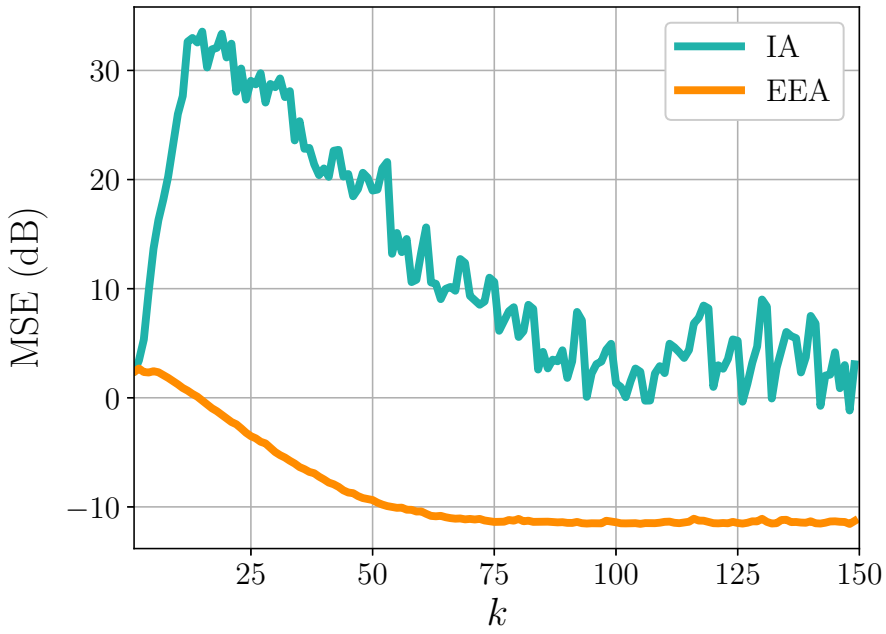


Figure 22 – Evolution of the empirical MSE using theoretical step size sequences from the standard (IA) and from the proposed (EEA) analyses, for $K = 10^7$ independent Monte Carlo trials.

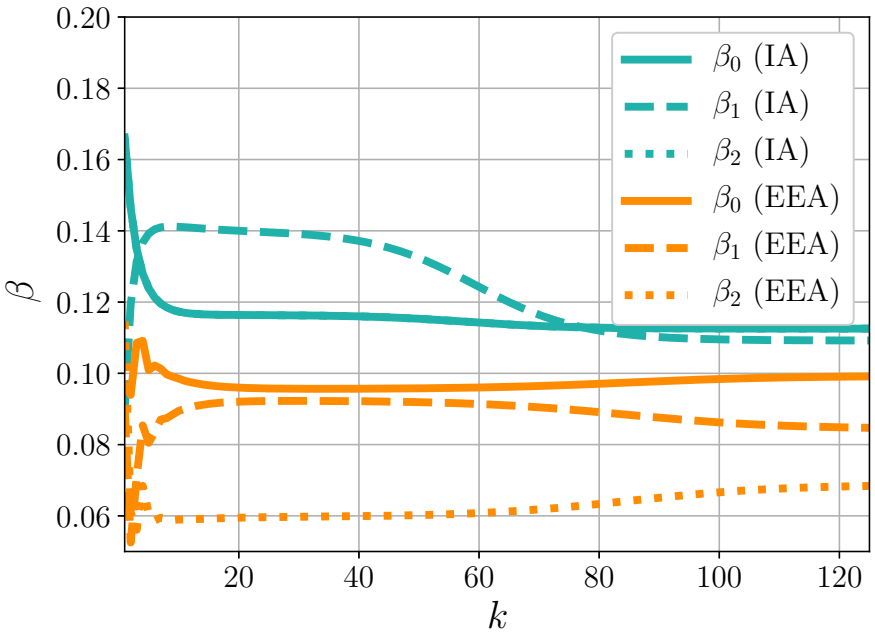


Figure 23 – Theoretical sequences $\beta_i(k)$ (for $i \in \{0, 1, \ldots, N-1\}$ prescribed by standard (IA) versus advanced (EEA) analyses.

performance around the 80-th iteration mark, by contrast, when utilizing IA generated sequence, 140 iterations are necessary to reach a similar MSD value.

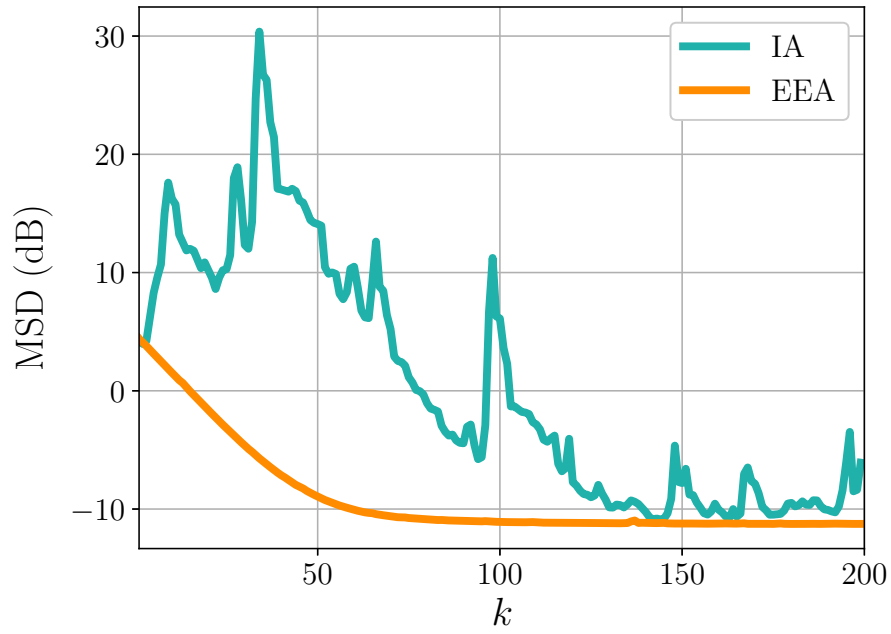


Figure 24 – Evolution of the empirical MSD (in dB) in the same scenario of Fig. 23, using the step size sequences suggested by IA- and EEA-based analyses. The empirical results were obtained by the usage of $K = 10^7$ independent Monte Carlo trials.

5.7- Conclusions

This chapter presents a comprehensive model of the tracking capabilities of the LMS algorithm, using an EEA-based approach that does not assume neither a white nor a Gaussian input signal. The model is able to accurately predict first-order and second-order learning behaviours of the LMS, as well as to provide proper step size upper bounds that effectively avoid stability issues. Furthermore, it provides proper guidelines for a step size value choice that optimizes performance, even when a “condition of slow variations” [102] of the ideal plant is not satisfied.

Additionally, the advanced stochastic model is capable of engineering a theoretical step size sequence that optimizes the overall LMS performance, and even the restriction of an equal step size for each adaptive coefficient can be removed, in a similar way to what proportionate algorithms do. The evaluation of such sequences in order to avoid divergence issues in practice is an important feature, since it may help the design of data-dependent variable step size algorithms.

The avoidance of the independence assumption is what lends EEA both its

significance and limitations. The former derives from the ability to accurately predict algorithm performance when the step size attains large values or when the unknown plant has a significant non-stationary characteristic. EEA main constraint derives from its high algebraic complexity, which limits the range of scenarios the advanced analysis can model.

6- Exact-Analysis Based Estimation of the Weight Error Covariance Matrix

6.1- Introduction

One of the first tasks in the analysis of high-dimensional estimation is to compute the covariance matrix [103], which provides a representation of the uncertainties of the estimation procedure [104]. In the context of adaptive filtering, the weight error covariance matrix (WECM) is a key parameter that determines how the filter adapts to changing input statistics and noise levels. This is because it provides an objective estimate of the uncertainties in the estimated filter coefficients. Thus, the WECM, defined as (see (38))

$$\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) \triangleq \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^{T}(k)\right] \tag{132}$$

is very important for the characterization of the learning abilities of an adaptive filter (see, *e.g.*, Eqs. (39)-(40)).

In this context, [105] states:

> "There are many situations where additional statistical information about the weight vector would be useful. These cases include the detection of a narrow-band line component in background noise using the weight vector as a test statistic, using the filtered output as a test statistic, and time delay estimation. The algorithm has been used as a canceller as part of a spread spectrum communication system. The output of the canceller acts as the input to a matched filter binary decision device. Knowledge of the statistics of the canceller output is crucial to predicting error probabilities for the system. Most often, via a central limit argument, it is assumed that the test statistic and/or the weights are Gaussian".

This excerpt emphasizes the common assumption that the adaptive weights are sampled according to a multivariate normal distribution. Under this scenario, a possible

sparseness of the WECM provides valuable information about independence and conditional independence properties. Thus, the WECM encode complete information about such properties, whereas for general distributions, they merely reveal correlations and partial correlations between variables. The second moment matrix and its inverse for multivariate normal distributions also contain complete information about independence and conditional independence properties, which is another attractive feature of these distributions. Specifically, a zero entry in the covariance matrix's ij-th position indicates that variables i and j are independent, while a zero entry in the precision (inverse covariance) matrix's ij-th position indicates that the two are conditionally independent [106].

Furthermore, the WECM provides information about the convergence behavior of the filter and the trade-off between convergence speed and steady-state error. Therefore, a more accurate estimate of the WECM can provide insights into the behavior of the algorithm being modeled. Obtaining a more accurate estimate of the weight error covariance matrix can help allocate the computational resources more efficiently by providing information about the degree of uncertainty in the estimated filter coefficients.

An accurate estimate of the weight error covariance matrix can also be used to infer statistical properties of the adaptive estimator. For example, by examining the estimated covariance matrix one can see if the variances of the deviation terms are equal.

The ubiquitous independence assumption provide some theoretical predictions w.r.t. the WECM. For example, most models assume that the off-diagonal elements of the WECM decrease monotonically to 0, so that the deviations are supposedly asymptotically uncorrelated [107]. Convergence studies often concentrate on the asymptotic properties of the diagonal elements of the WECM, as the excess MSE is a linear combination of these elements [108].

The adoption of the independence assumption usually introduces some symmetry in the theoretical asymptotic WECM. Thus, for example, assuming gaussianity of the input signal and traditional assumptions, (47) (reproduced here for didactic reasons) presents the following recursion for the WECM:

$$\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k+1)} = \boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)} - \beta \left[\boldsymbol{R}\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)} + \boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}\boldsymbol{R}\right] + 2\beta^2 \boldsymbol{R}\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}\boldsymbol{R} + \beta^2 \boldsymbol{R}\left\{\sigma_\nu^2 + \mathsf{Tr}[\boldsymbol{R}\boldsymbol{R}_{\tilde{\boldsymbol{w}}(k)}]\right\}, \tag{133}$$

where, it should be noted, matrix $\boldsymbol{R}$ is symmetrical.

Another standard result states that the WECM can be recursively computed accord-

ing to [3]

$$\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k+1) = (\boldsymbol{I} - \beta\boldsymbol{R})\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)(\boldsymbol{I} - \beta\boldsymbol{R}) + \beta^2 J_{\text{min}}\boldsymbol{R}, \tag{134}$$

where J_{min} denotes the minimum obtainable MSE. The symmetries presented in (133) and (134) engender structured asymptotic solutions for the WECM. Such a fact could be seen more clearly in the asymptotic identity [109]:

$$\lim_{k\to\infty} \boldsymbol{R}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) + \boldsymbol{R}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\boldsymbol{R} = \boldsymbol{F}, \tag{135}$$

where $\boldsymbol{F}$ is the following "excitation matrix":

$$\boldsymbol{F} = \sum_{l=-\infty}^{\infty} \mathbb{E}\left[\nu(k)\nu(k-l)\right] \mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k-l)\right], \tag{136}$$

which, in the case of white process $\nu(k)$, leads to the following closed-form solution for the WECM [109]:

$$\lim_{k\to\infty} \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) = \frac{1}{2}\sigma_\nu^2 \boldsymbol{I}. \tag{137}$$

Result (137) helps to explain in a intuitive manner some traditional theoretical findings. For example, frequently the WECM is assumed to be approximately uncorrelated from tap to tap [110]. Further, under a white input signal, the WECM is supposed to be a diagonal matrix, with all of its diagonal elements supposed to be equal [111]. Such predictions are obtained from usual stochastic approximations, and they could be refined with the more sophisticated tools brought about by the exact expectation analysis.

Thus, given the importance of the WECM, in this chapter the exact expectation analysis is employed to obtain a more accurate estimate of the LMS algorithm's WECM. Since the elements of the WECM are a subset of the state variables required for the MSE computation of the exact expectation method, the adaptation of the EEA is rather straightforward. It should be noted that the number of equations of the analysis can be smaller than the traditional setting, since there are less state variables to be recursively computed.

The theoretical analysis presents some findings, which are confirmed by simulations. In a sense, the results to be presented are a tribute to the sophistication of the learning process of the LMS algorithm, as they show that the actual behavior of the algorithm can deviate from that predicted by classical analyses, even in simple configurations.

6.2- Scenarios

In the following, the steady-state WECM (SS-WECM) of several scenarios are considered. Three distinct SS-WECM are computed: *(i)* the empirical one (EMP-SS-WECM, obtained through simulations); *(ii)* the classical one (IA-SS-WECM, computed with the independence assumption); and *(iii)* the WECM obtained with the exact expectation analysis (termed as EEA-SS-WECM).

The settings of the distinct scenarios are summarized on Table 4. In all scenarios, the optimum filter is $w_i^\star = 1$, for $i \in \{0, 1, \ldots, N-1\}$, apart from scenario 6, where all settings are similar to scenario 5, but $w_i^\star$ which in this case is $w_i^\star = 10$, for $i \in \{0, 1, \ldots, N-1\}$. Since we are interested in asymptotic operation, it is not expected a significant influence of the optimal solution in the WECM. The number of equations demanded for each scenario is presented in Table 5.

# Scenario	N	M	L	β	σ_ν^2	N_{samples}	$H(z)$	$w_i^\star$
1	3	2	1	0.075	10^{-6}	10^7	$1-0.9z^{-1}$	1
2	3	2	2	0.075	10^{-6}	10^7	$1-0.9z^{-1}$	1
3	4	2	1	0.05	10^{-6}	10^6	1	1
4	5	1	1	0.075	10^{-6}	10^7	1	1
5	6	1	1	0.055	10^{-6}	10^6	1	1
6	6	1	1	0.055	10^{-6}	10^7	1	10

Table 4 – Properties of the scenarios. N_{samples} denotes the number of independent Monte Carlo trials. $H(z)$ is the coloring filter of the input signal.

# Scenario	# Equations (IA)	# Equations (EEA)	Steady-state iteration
1	6	698	299
2	6	830	299
3	10	9578	649
4	15	5313	199
5	21	49695	399
6	21	49695	399

Table 5 – Number of equations required for the linear state space equation for the considered scenarios, and the number of the iteration considered for the asymptotic operation.

6.2.1 Scenario 1

In Scenario 1, the IA-SS-WECM is given by:

$$\begin{bmatrix} 5.826 \times 10^{-8} & -4.5 \times 10^{-9} & 4.091 \times 10^{-10} \\ -4.5 \times 10^{-9} & 5.867 \times 10^{-8} & -4.5 \times 10^{-9} \\ 4.091 \times 10^{-10} & -4.5 \times 10^{-9} & 5.826 \times 10^{-8} \end{bmatrix} \tag{138}$$

Note that $\mathbb{E}\left[\tilde{w}_0^2(k)\right]$ and $\mathbb{E}\left[\tilde{w}_2^2(k)\right]$ (the initial and the final element of the main diagonal) are supposed to be equal. This symmetry, as discussed in Section 6.1, is frequent when the independence assumption is adopted.

In the same scenario, the EEA-SS-WECM tells us a bit different story:

$$\begin{bmatrix} 6.556 \times 10^{-8} & 1.051 \times 10^{-9} & 1.86 \times 10^{-9} \\ 1.051 \times 10^{-9} & 6.97 \times 10^{-8} & -1.786 \times 10^{-9} \\ 1.86 \times 10^{-9} & -1.786 \times 10^{-9} & 7.873 \times 10^{-8} \end{bmatrix} \tag{139}$$

Note that the EEA-SS-WECM predicts that $\mathbb{E}\left[\tilde{w}_i^2(k)\right] \neq \mathbb{E}\left[\tilde{w}_j^2(k)\right]$, when $i \neq j$, which implies that the supposed symmetry is indeed a byproduct engendered by the independence assumption. Further, the EEA predicts that $\mathbb{E}\left[\tilde{w}_2^2(k)\right] > \mathbb{E}\left[\tilde{w}_1^2(k)\right] > \mathbb{E}\left[\tilde{w}_0^2(k)\right]$, and that $\mathbb{E}\left[\tilde{w}_0(k)\tilde{w}_2(k)\right]$ (*i.e.*, $1.86{\times}10^{-9}$) is much larger than the value predicted by the traditional method (4.091×10^{-10}), indicating that distinct deviations are much more correlated than presumed by traditional analyses. This fact means that the stochastic coupling between the adaptive weights is actually much more intricate than standard analyses portray.

The EMP-SS-WECM (*i.e.*, the WECM obtained through actual executions of the LMS) in this scenario is given by

$$\begin{bmatrix} 6.358 \times 10^{-8} & 2.066 \times 10^{-9} & 1.296 \times 10^{-9} \\ 2.066 \times 10^{-9} & 6.76 \times 10^{-8} & -8.467 \times 10^{-10} \\ 1.296 \times 10^{-9} & -8.467 \times 10^{-10} & 7.342 \times 10^{-8} \end{bmatrix} \tag{140}$$

In order to facilitate some comparisons, Table 6 presents the results in a distinct manner. Note that the simulations confirm the EEA prediction that $\mathbb{E}\left[\tilde{w}_2^2(k)\right] > \mathbb{E}\left[\tilde{w}_1^2(k)\right] >$

$\mathbb{E}\left[\tilde{w}_0^2(k)\right]$. The empirical moments $\mathbb{E}\left[\tilde{w}_0^2(k)\right]$, $\mathbb{E}\left[\tilde{w}_1^2(k)\right]$ and $\mathbb{E}\left[\tilde{w}_2^2(k)\right]$ are 9,1%, 15% and 26% larger (respectively) than the traditional estimates, whereas the EEA estimates exceeds the empirical correlations by 3,1%, 3,1% and 7,2% (respectively). Thus, it is possible to conclude that the EEA displays a much higher level of adherence to empirical outcomes. Please observe that the total absence of symmetry in the theoretical estimate is a feature of the proposed analysis, which has no precedent in the literature regarding LMS.

It is much more challenging to obtain strong empirical evidence in the case of cross-correlations between distinct deviations since such correlations are much smaller than the ones located in the main diagonal of the WECM. But, even with a finite number of Monte Carlo trials, the experimental correlation $\mathbb{E}\left[\tilde{w}_0(k)\tilde{w}_2(k)\right]$ (1.296×10^{-9}) *is more than three times that predicted by the traditional method*, an effect that is captured by the EEA analysis (whose estimate, $1,89 \times 10^{-9}$, exceeds the empirical estimate by approximately 40%). Figure 25 depicts the MSD evolution for the three considered curves, indicating that EEA is the most suitable analysis for the asymptotic operation.

$\mathbb{E}\left[\tilde{w}_i(k)\tilde{w}_j(k)\right]$	IA	EEA	EMP
$\mathbb{E}\left[\tilde{w}_0^2(k)\right]$	5.826×10^{-8}	6.556×10^{-8}	6.358×10^{-8}
$\mathbb{E}\left[\tilde{w}_1^2(k)\right]$	5.867×10^{-8}	6.97×10^{-8}	6.76×10^{-8}
$\mathbb{E}\left[\tilde{w}_2^2(k)\right]$	5.826×10^{-8}	7.873×10^{-8}	7.342×10^{-8}
$\mathbb{E}\left[\tilde{w}_0(k)\tilde{w}_2(k)\right]$	4.091×10^{-10}	1.86×10^{-9}	1.296×10^{-9}

Table 6 – Results for Scenario 1.

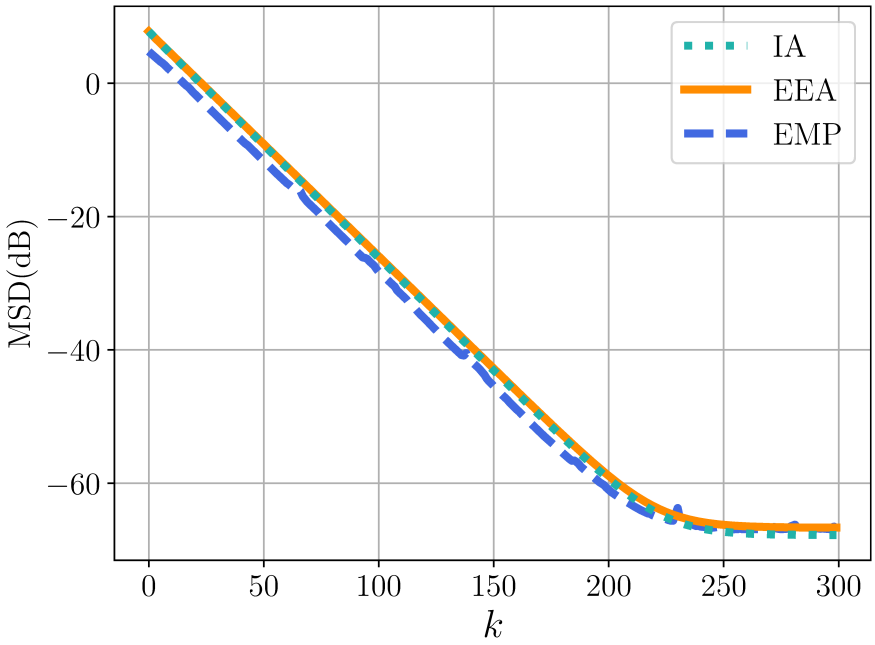


Figure 25 – MSD (dB) for Scenario 1.

6.2.2 Scenario 2

The major contrast between Scenario 2 and Scenario 1 is that the former, as opposed to the latter, includes colored measurement noise. In Scenario 2, the IA-SS-WECM is given by:

$$\begin{bmatrix} 1.054 \times 10^{-7} & -8.204 \times 10^{-9} & 6.984 \times 10^{-10} \\ -8.204 \times 10^{-9} & 1.061 \times 10^{-7} & -8.204 \times 10^{-9} \\ 6.984 \times 10^{-10} & -8.204 \times 10^{-9} & 1.054 \times 10^{-7} \end{bmatrix}, \tag{141}$$

where once again a symmetry is observed, given that $\mathbb{E}\left[\tilde{w}_0^2(k)\right] = \mathbb{E}\left[\tilde{w}_2^2(k)\right] = 1.054 \times 10^{-7}$.

In the same scenario, the EEA-SS-WECM is

$$\begin{bmatrix} 1.234 \times 10^{-7} & -2.98 \times 10^{-8} & 1.803 \times 10^{-9} \\ -2.98 \times 10^{-8} & 1.11 \times 10^{-7} & -3.003 \times 10^{-8} \\ 1.803 \times 10^{-9} & -3.003 \times 10^{-8} & 1.257 \times 10^{-7} \end{bmatrix}, \tag{142}$$

which implies that the statistical correlations between distinct deviations is much more intense than the one predicted by IA. Such a finding can be confirmed by the respective EMP-SS-WECM, which is

$$\begin{bmatrix} 1.23 \times 10^{-7} & -3.036 \times 10^{-8} & 1.061 \times 10^{-9} \\ -3.036 \times 10^{-8} & 1.085 \times 10^{-7} & -2.848 \times 10^{-8} \\ 1.061 \times 10^{-9} & -2.848 \times 10^{-8} & 1.214 \times 10^{-7} \end{bmatrix}. \tag{143}$$

Table 7 suggests that the EEA-SS-WECM fits the data better than IA-SS-WECM. For instance, EEA predicts that $\mathbb{E}[\tilde{w}_0(k)\tilde{w}_1(k)]$ is -2.98×10^{-8}, which is 3.63 times larger than the value predicted by traditional analysis (-8.204×10^{-9}). The simulated value of $E[\tilde{w}_0(k)\tilde{w}_1(k)]$ is -3.036×10^{-8}, just 1.88% above the value predicted by the advanced analysis. In addition, EEA predicts that $\mathbb{E}\left[\tilde{w}_0^2(k)\right]$ is 17.08% above the value estimated by IA. The simulations uphold this finding, revealing a value for $\mathbb{E}\left[\tilde{w}_0^2(k)\right]$ that is 16.70% higher than the prediction made by traditional analysis.

$\mathbb{E}\left[\tilde{w}_i(k)\tilde{w}_j(k)\right]$	IA	EEA	EMP
$\mathbb{E}\left[\tilde{w}_0^2(k)\right]$	1.054×10^{-7}	1.234×10^{-7}	1.23×10^{-7}
$\mathbb{E}\left[\tilde{w}_1^2(k)\right]$	1.061×10^{-7}	1.11×10^{-7}	1.085×10^{-7}
$\mathbb{E}\left[\tilde{w}_2^2(k)\right]$	1.054×10^{-7}	1.257×10^{-7}	1.214×10^{-7}
$\mathbb{E}\left[\tilde{w}_0(k)\tilde{w}_1(k)\right]$	-8.204×10^{-9}	-2.98×10^{-8}	-3.036×10^{-8}
$\mathbb{E}\left[\tilde{w}_0(k)\tilde{w}_2(k)\right]$	6.984×10^{-10}	1.803×10^{-9}	1.061×10^{-9}
$\mathbb{E}\left[\tilde{w}_1(k)\tilde{w}_2(k)\right]$	-8.204×10^{-9}	-3.003×10^{-8}	1.214×10^{-7}

Table 7 – Results for Scenario 2.

6.2.3 Scenario 3

Scenario 3 uses a filter of size $N = 4$, larger than the size adopted in Scenarios 1 and 2. In this scenario, the IA-SS-WECM is given by:

$$\begin{bmatrix} 3.483 \times 10^{-8} & -1.605 \times 10^{-9} & 2.033 \times 10^{-10} & 6.838 \times 10^{-11} \\ -1.605 \times 10^{-9} & 3.503 \times 10^{-8} & -1.537 \times 10^{-9} & 2.033 \times 10^{-10} \\ 2.033 \times 10^{-10} & -1.537 \times 10^{-9} & 3.503 \times 10^{-8} & -1.605 \times 10^{-9} \\ 6.838 \times 10^{-11} & 2.033 \times 10^{-10} & -1.605 \times 10^{-9} & 3.483 \times 10^{-8} \end{bmatrix}, \tag{144}$$

where the expect symmetry also happens, since $\mathbb{E}\left[\tilde{w}_0^2(k)\right] = \mathbb{E}\left[\tilde{w}_3^2(k)\right]$, $\mathbb{E}\left[\tilde{w}_1^2(k)\right] = \mathbb{E}\left[\tilde{w}_2^2(k)\right]$ and $\mathbb{E}\left[\tilde{w}_0(k)\tilde{w}_1(k)\right] = \mathbb{E}\left[\tilde{w}_2(k)\tilde{w}_3(k)\right]$, for example. Put differently, classical analysis constrains the freedom of the WECM to a greater extent than the constraint imposed by the symmetry of the WECM.

In the same scenario, the computation of EEA-SS-WECM leads to

$$\begin{bmatrix} 3.499 \times 10^{-8} & -2.159 \times 10^{-10} & 2.002 \times 10^{-10} & 8.654 \times 10^{-11} \\ -2.159 \times 10^{-10} & 3.534 \times 10^{-8} & -6.934 \times 10^{-11} & 4.258 \times 10^{-10} \\ 2.002 \times 10^{-10} & -6.934 \times 10^{-11} & 3.539 \times 10^{-8} & -4.636 \times 10^{-10} \\ 8.654 \times 10^{-11} & 4.258 \times 10^{-10} & -4.636 \times 10^{-10} & 3.537 \times 10^{-8} \end{bmatrix}, \tag{145}$$

whereas the EMP-SS-WECM is

$$\begin{bmatrix} 3.472 \times 10^{-8} & -3.137 \times 10^{-10} & 1.299 \times 10^{-10} & 8.762 \times 10^{-11} \\ -3.137 \times 10^{-10} & 3.491 \times 10^{-8} & -1.35 \times 10^{-10} & 3.428 \times 10^{-10} \\ 1.299 \times 10^{-10} & -1.35 \times 10^{-10} & 3.496 \times 10^{-8} & -2.054 \times 10^{-10} \\ 8.762 \times 10^{-11} & 3.428 \times 10^{-10} & -2.054 \times 10^{-10} & 3.44 \times 10^{-8} \end{bmatrix}. \tag{146}$$

Table 8 demonstrates two scenarios in which the IA-SS-WECM's estimation of the joint moment misses the empirical value by an order of magnitude. Conversely, EEA yields estimates that are much closer to the simulated values, always within the same order of magnitude.

$\mathbb{E}\left[\tilde{w}_i(k)\tilde{w}_j(k)\right]$	IA	EEA	EMP
$\mathbb{E}\left[\tilde{w}_0(k)\tilde{w}_1(k)\right]$	-1.605×10^{-9}	-2.159×10^{-10}	-3.137×10^{-10}
$\mathbb{E}\left[\tilde{w}_2(k)\tilde{w}_3(k)\right]$	-1.605×10^{-9}	-4.636×10^{-10}	-2.054×10^{-10}

Table 8 – Some results for Scenario 3.

6.2.4 Scenario 4

Scenario 4 features a larger filter size ($N = 5$) compared to the previous scenarios. This is the first scenario in which the input signal is white. In this case, as observed, classical analysis tends to generate estimates of very low magnitude for the cross-correla-

tion of deviations (i.e., $\mathbb{E}\left[\tilde{w}_i(k)\tilde{w}_j(k)\right]$ for $i \neq j$). In this scenario, the IA-SS-WECM is given by:

$$\begin{bmatrix} 5.092 \times 10^{-8} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} \\ 1.232 \times 10^{-13} & 5.092 \times 10^{-8} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} \\ 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 5.092 \times 10^{-8} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} \\ 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 5.092 \times 10^{-8} & 1.232 \times 10^{-13} \\ 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 1.232 \times 10^{-13} & 5.092 \times 10^{-8} \end{bmatrix}, \quad (147)$$

which makes a strong prediction: that all cross-correlations are equal to 1.232×10^{-13}.

On the other hand, EEA-SS-WECM assumes that such cross-correlations have significantly higher (and heterogeneous) magnitudes, depicting a much richer learning behavior:

$$\begin{bmatrix} 5.102 \times 10^{-8} & 2.96 \times 10^{-11} & -4.416 \times 10^{-11} & 2.254 \times 10^{-11} & -6.17 \times 10^{-11} \\ 2.96 \times 10^{-11} & 5.064 \times 10^{-8} & 3.466 \times 10^{-11} & 2.08 \times 10^{-11} & 2.147 \times 10^{-11} \\ -4.416 \times 10^{-11} & 3.466 \times 10^{-11} & 5.034 \times 10^{-8} & 3.149 \times 10^{-11} & 4.696 \times 10^{-11} \\ 2.254 \times 10^{-11} & 2.08 \times 10^{-11} & 3.149 \times 10^{-11} & 5.005 \times 10^{-8} & 2.244 \times 10^{-11} \\ -6.17 \times 10^{-11} & 2.147 \times 10^{-11} & 4.696 \times 10^{-11} & 2.244 \times 10^{-11} & 4.968 \times 10^{-8} \end{bmatrix} \quad (148)$$

The EMP-SS-WECM computed in this scenario is as follows:

$$\begin{bmatrix} 5.101 \times 10^{-8} & 4.916 \times 10^{-12} & -3.739 \times 10^{-11} & 7.144 \times 10^{-12} & -3.448 \times 10^{-11} \\ 4.916 \times 10^{-12} & 5.063 \times 10^{-8} & 3.611 \times 10^{-11} & 3.229 \times 10^{-11} & 3.522 \times 10^{-11} \\ -3.739 \times 10^{-11} & 3.611 \times 10^{-11} & 5.035 \times 10^{-8} & 4.145 \times 10^{-11} & 5.355 \times 10^{-11} \\ 7.144 \times 10^{-12} & 3.229 \times 10^{-11} & 4.145 \times 10^{-11} & 5.005 \times 10^{-8} & 2.622 \times 10^{-11} \\ -3.448 \times 10^{-11} & 3.522 \times 10^{-11} & 5.355 \times 10^{-11} & 2.622 \times 10^{-11} & 4.970 \times 10^{-8} \end{bmatrix}. \quad (149)$$

The EMP-SS-WECM corroborates the findings of the exact analysis, validating that the stochastic coupling among the various deviations is far more intricate than implied by conventional analysis. It is worth noting that in some cases, classical analysis (and only classical analysis) even gets the sign of the joint moment wrong, such as in the estimates of $\mathbb{E}\left[\tilde{w}_0(k)\tilde{w}_2(k)\right]$ and $\mathbb{E}\left[\tilde{w}_0(k)\tilde{w}_4(k)\right]$.

6.2.5 Scenario 5

In Scenario 5, an even larger filter size is observed (*i.e.*, $N = 6$). In this scenario, the IA-SS-WECM is given by:

$$\begin{bmatrix} 3.525 \times 10^{-8} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} \\ 9.562 \times 10^{-20} & 3.525 \times 10^{-8} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} \\ 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 3.525 \times 10^{-8} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} \\ 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 3.525 \times 10^{-8} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} \\ 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 3.525 \times 10^{-8} & 9.562 \times 10^{-20} \\ 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 9.562 \times 10^{-20} & 3.525 \times 10^{-8} \end{bmatrix}, \tag{150}$$

whereas the EEA-SS-WECM is given by

$$\begin{bmatrix} 3.528 \times 10^{-8} & 1.659 \times 10^{-16} & -3.721 \times 10^{-11} & 1.255 \times 10^{-16} & -4.317 \times 10^{-11} & 1.48 \times 10^{-16} \\ 1.659 \times 10^{-16} & 3.514 \times 10^{-8} & 1.889 \times 10^{-16} & -1.981 \times 10^{-11} & 1.806 \times 10^{-16} & -2.631 \times 10^{-11} \\ -3.721 \times 10^{-11} & 1.889 \times 10^{-16} & 3.501 \times 10^{-8} & 2.341 \times 10^{-16} & -6.932 \times 10^{-12} & 1.179 \times 10^{-16} \\ 1.255 \times 10^{-16} & -1.981 \times 10^{-11} & 2.341 \times 10^{-16} & 3.49 \times 10^{-8} & 1.719 \times 10^{-16} & 4.568 \times 10^{-12} \\ -4.317 \times 10^{-11} & 1.806 \times 10^{-16} & -6.932 \times 10^{-12} & 1.719 \times 10^{-16} & 3.478 \times 10^{-8} & 1.382 \times 10^{-16} \\ 1.48 \times 10^{-16} & -2.631 \times 10^{-11} & 1.179 \times 10^{-16} & 4.568 \times 10^{-12} & 1.382 \times 10^{-16} & 3.463 \times 10^{-8} \end{bmatrix}. \tag{151}$$

Comparing the EEA-SS-WECM with IA-SS-WECM, it can be observed that the exact analysis predicts that the off-diagonal elements of the WECM have a much higher magnitude than predicted by classical analysis. This finding is confirmed by the computed EMP-SS-WECM:

$$\begin{bmatrix} 3.527 \times 10^{-8} & 9.484 \times 10^{-13} & -5.0262 \times 10^{-11} & 2.528 \times 10^{-13} & -5.334 \times 10^{-11} & 8.317 \times 10^{-12} \\ 9.484 \times 10^{-13} & 3.514 \times 10^{-8} & -1.258 \times 10^{-11} & -2.08 \times 10^{-11} & 1.51 \times 10^{-12} & -2.954 \times 10^{-11} \\ -5.02 \times 10^{-11} & -1.258 \times 10^{-11} & 3.503 \times 10^{-8} & 1.853 \times 10^{-12} & -2.397 \times 10^{-11} & 1.212 \times 10^{-11} \\ 2.528 \times 10^{-13} & -2.08 \times 10^{-11} & 1.853 \times 10^{-12} & 3.493 \times 10^{-8} & -2.811 \times 10^{-11} & 3.357 \times 10^{-11} \\ -5.33 \times 10^{-11} & 1.51 \times 10^{-12} & -2.397 \times 10^{-11} & -2.811 \times 10^{-11} & 3.478 \times 10^{-8} & -1.688 \times 10^{-12} \\ 8.317 \times 10^{-12} & -2.954 \times 10^{-11} & 1.212 \times 10^{-11} & 3.357 \times 10^{-11} & -1.688 \times 10^{-12} & 3.463 \times 10^{-8} \end{bmatrix}. \tag{152}$$

It is worth mentioning, however, in this case that even the exact analysis occasionally misjudges the order of magnitude of the cross-correlations of the deviations. Nonetheless, in all cases, the error of the exact analysis is lower than that of the classical analysis. It can be noticed that with the increase in N, there appears to be a tendency for a reduction in the magnitude of correlations between different deviations. Thus, even with

the adoption of a large number of independent Monte Carlo trials (*i.e.*, 10^6 trials), it is challenging to compute a reasonable approximation of such small numbers.

6.2.6 Scenario 6

The key difference in Scenario 6 compared to the previous scenarios lies in the optimal values of the adaptive coefficients (all of which are now set to 10). The purpose of this scenario is to determine whether there is a notable difference in the asymptotic behavior when the final values of the filters change. Intuitively, after convergence, the variance and deviations' correlations are not highly sensitive to the values around which the adaptive estimator oscillates (as LMS is an unbiased algorithm). The objective of the simulations in this scenario is to validate or refine such intuition.

In this scenario, the IA-SS-WECM is given by:

$$\begin{bmatrix} 3.525 \times 10^{-8} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} \\ 9.562 \times 10^{-18} & 3.525 \times 10^{-8} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} \\ 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 3.525 \times 10^{-8} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} \\ 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 3.525 \times 10^{-8} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} \\ 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 3.525 \times 10^{-8} & 9.562 \times 10^{-18} \\ 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 9.562 \times 10^{-18} & 3.525 \times 10^{-8} \end{bmatrix}, \tag{153}$$

whereas the EEA-SS-WECM gives us:

$$\begin{bmatrix} 3.528 \times 10^{-8} & 1.659 \times 10^{-14} & -3.719 \times 10^{-11} & 1.255 \times 10^{-14} & -4.316 \times 10^{-11} & 1.48 \times 10^{-14} \\ 1.659 \times 10^{-14} & 3.514 \times 10^{-8} & 1.889 \times 10^{-14} & -1.979 \times 10^{-11} & 1.806 \times 10^{-14} - & 2.63 \times 10^{-11} \\ -3.719 \times 10^{-11} & 1.889 \times 10^{-14} & 3.501 \times 10^{-8} & 2.341 \times 10^{-14} & -6.915 \times 10^{-12} & 1.179 \times 10^{-14} \\ 1.255 \times 10^{-14} & -1.979 \times 10^{-11} & 2.341 \times 10^{-14} & 3.49 \times 10^{-8} & 1.719 \times 10^{-14} & 4.579 \times 10^{-12} \\ -4.316 \times 10^{-11} & 1.806 \times 10^{-14} & -6.915 \times 10^{-12} & 1.719 \times 10^{-14} & 3.478 \times 10^{-8} & 1.382 \times 10^{-14} \\ 1.48 \times 10^{-14} & -2.63 \times 10^{-11} & 1.179 \times 10^{-14} & 4.579 \times 10^{-12} & 1.382 \times 10^{-14} & 3.463 \times 10^{-8} \end{bmatrix} \tag{154}$$

The EMP-SS-WECM in this scenario is given by

$$\begin{bmatrix} 3.531 \times 10^{-8} & -1.489 \times 10^{-11} & -3.799 \times 10^{-11} & -1.79 \times 10^{-11} & -4.126 \times 10^{-11} & 3.896 \times 10^{-13} \\ -1.489 \times 10^{-11} & 3.514 \times 10^{-8} & -1.876 \times 10^{-13} & -1.877 \times 10^{-11} & 1.069 \times 10^{-12} & -4.442 \times 10^{-11} \\ -3.799 \times 10^{-11} & -1.876 \times 10^{-13} & 3.502 \times 10^{-8} & 3.381 \times 10^{-12} & -1.722 \times 10^{-11} & -2.047 \times 10^{-11} \\ -1.79 \times 10^{-11} & -1.877 \times 10^{-11} & 3.381 \times 10^{-12} & 3.489 \times 10^{-8} & 1.175 \times 10^{-12} & 2.373 \times 10^{-11} \\ -4.126 \times 10^{-11} & 1.069 \times 10^{-12} & -1.722 \times 10^{-11} & 1.175 \times 10^{-12} & 3.476 \times 10^{-8} & -1.029 \times 10^{-12} \\ 3.896 \times 10^{-13} & -4.442 \times 10^{-11} & -2.047 \times 10^{-11} & 2.373 \times 10^{-11} & -1.029 \times 10^{-12} & 3.464 \times 10^{-8} \end{bmatrix}, \tag{155}$$

from which we can extract very similar conclusions to those of Scenario 5. The results of this scenario support the intuition that the statistical properties of the adaptive estimator in steady-state are reasonably unaffected by the optimal value of the coefficients.

6.3- Concluding Remarks

In this chapter, the exact analysis was employed for a different purpose than in the other chapters: to predict asymptotic characteristics of the deviation autocorrelation matrix that classical analysis is unable to accurately predict. The WECM provides more statistical information than metrics such as MSE and MSD, which can be useful in situations where the adaptive filter is adopted to predict a narrowband component or for some statistical test. In addition, it was observed that exact analysis outlines a much richer asymptotic behavior of the adaptive estimator than traditional analyses assume. Further, one concludes that several symmetries found in WECM prediction through classical analysis are actually artifacts produced by the assumption of independence. It was also observed that the stochastic coupling between the adaptive weights can be quite different from that predicted by these analyses. The experiments confirmed the indications of the exact analysis that the cross-correlations between these deviations seem to decrease as the filter size increases. Lastly, such couplings (or, more precisely, the joint moments) do not appear to be sensitive to the optimal value of these coefficients, which aligns with what is expected from traditional stochastic models.

This chapter concludes the chapters of original contributions in the thesis. The next chapter serves as the final conclusion of the thesis, encompassing the final remarks and providing an overview of potential future work.

7- Conclusions

Adaptive filtering algorithms implement widely employed online estimators that exploit statistical relationships from recorded and processed data. When successful, they furnish the information of interest after a learning process that typically occurs in a recursive manner. Although their flexibility guarantees their attractiveness in a plethora of applications, their feedback mechanism makes them nonlinear processors, even when their filtering structure, assuming fixed weights, respects the superposition principle.

Unfortunately, most analyses of adaptive filtering algorithms suffer from two important restrictions: *i*) model uncertainty, since features such as non-linearity of the plant the adaptive filter aims to estimate are commonly disregarded, and *ii*) accuracy losses when the step size assume large values. The latter issue is caused by the usage of the ubiquitous independence assumption, which states that excitation data is statistically independent from the adaptive weights. This thesis extended the exact expectation analysis method (which avoids the independence assumption). The advanced models take into account important non-idealities that occur in practice, such as non-linearity in the system that the adaptive filter intends to emulate and nonstationary environments (which was addressed in Chapter 4). Some interesting phenomena regarding a worsening of the learning in the first phase of the learning are inferred from the model. Simulations are carried out and reveal that the resulting theoretical analysis predicts the behavior of the least mean squares algorithm in the square sense in an accurate manner. Furthermore, it provides a reliable operation region where the algorithm self adaptation is stable and may offer a deterministic sequence of step size values that maximizes actual performance of the LMS.

The EEA was extended to address a non-stationary plant in Chapter 5. The model results in a linear state space model that predicts the learning behavior of the LMS when the ideal plant varies according to a first-order Markovian model. The model closely follows the experimental curves, even when the step size is not small or when the unknown system has a highly nonstationary degree. The asymptotic operation was decoupled in a closed-form manner as the a sum of tracking-related terms and steady-state quantities related to the stationary setting. A deterministic optimization that derives

a sequence of distint step sizes (one for each adaptive tap) was advanced, whose superiority was confirmed by simulations.

Chapter 6 utilizes the EEA aproach to derive the asymptotic WECM of the LMS algorithm. The resulting predictions depicts a much more subtle scenario of stochastic coupling among deviations than the one offered by traditional analyses. Symmetries in the joint moments $\mathbb{E}\left[\tilde{w}_i(k)\tilde{w}_j(k)\right]$ are demonstrated to be mere artifacts produced by the independence assumption. Further, distinct deviations can be much more correlated than assumed by standard models, which sometimes even get the sign of the joint moment wrong. All these findings were corroborated by computational experiments.

Some directions in this topic are regarded as especially promising. One of them is to investigate whether it is possible to construct an intermediary analysis method, which assumes a weak formulation of IA in order to require a smaller computational burden than the one demanded by the EEA. Other alternative is to address higher-order statistics (such as the fourth-order), in order to predict, for example, whether the variance of the MSE grows without limit, even when the MSE does not diverge.

References

[1] LI, B. Forecasting financial time series using linear predictive filters. Imperial College London, 2013.

[2] NAIR, B. B. et al. Application of hybrid adaptive filters for stock market prediction. In: IEEE. *2010 International Conference on Communication and Computational Intelligence (INCOCCI)*. [S.l.], 2010. p. 443–447.

[3] SIMON, H. Adaptive filter theory. *Prentice Hall*, v. 2, p. 478–481, 2002.

[4] ZAKNICH, A. *Principles of adaptive filters and self-learning systems*. London: Springer London, 2005. ISBN 978-1-84628-121-1.

[5] KEPPENNE, C. L.; GHIL, M. Adaptive filtering and prediction of the southern oscillation index. *Journal of Geophysical Research: Atmospheres*, Wiley Online Library, v. 97, n. D18, p. 20449–20454, 1992.

[6] THAKOR, N. V.; ZHU, Y.-S. Applications of adaptive filtering to ecg analysis: noise cancellation and arrhythmia detection. *IEEE transactions on biomedical engineering*, IEEE, v. 38, n. 8, p. 785–794, 1991.

[7] CHOI, Y.-S. Subband adaptive filtering with-norm constraint for sparse system identification. *Mathematical Problems in Engineering*, Hindawi, v. 2013, 2013.

[8] COX, H.; ZESKIND, R.; OWEN, M. Robust adaptive beamforming. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, IEEE, v. 35, n. 10, p. 1365–1376, 1987.

[9] LI, J.; STOICA, P. An adaptive filtering approach to spectral estimation and sar imaging. *IEEE Transactions on Signal Processing*, IEEE, v. 44, n. 6, p. 1469–1484, 1996.

[10] PÄRLIN, K.; RIIHONEN, T.; TURUNEN, M. Sweep jamming mitigation using adaptive filtering for detecting frequency agile systems. In: IEEE. *2019 International Conference on Military Communications and Information Systems (ICMCIS)*. [S.l.], 2019. p. 1–6.

[11] MARTINEK, R. et al. A robust approach for acoustic noise suppression in speech using anfis. *Journal of electrical engineering*, v. 66, n. 6, p. 301–310, 2015.

[12] RUPP, M. Asymptotic equivalent analysis of the LMS algorithm under linearly filtered processes. *EURASIP Journal on Advances in Signal Processing*, SpringerOpen, v. 2016, n. 1, p. 1–16, 2016.

[13] RUPP, M.; BUTTERWECK, H.-J. Overcoming the independence assumption in LMS filtering. In: *The Thirty-Seventh Asilomar Conference on Signals, Systems Computers, 2003*. [S.l.: s.n.], 2003. v. 1, p. 607–611 Vol.1.

[14] Widrow, B. et al. Stationary and nonstationary learning characteristics of the LMS adaptive filter. *Proceedings of the IEEE*, v. 64, n. 8, p. 1151–1162, Aug. 1976.

[15] LI, Y.; LIU, K. R. Static and dynamic convergence behavior of adaptive blind equalizers. *IEEE Transactions on Signal Processing*, v. 44, n. 11, p. 2736–2745, 1996.

[16] Solo, V. LMS: Past, present and future. In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2019. p. 7740–7744.

[17] EWEDA, E. A new approach for analyzing the limiting behavior of the normalized LMS algorithm under weak assumptions. *Signal Processing*, v. 89, n. 11, p. 2143 – 2151, 2009. ISSN 0165-1684.

[18] HAYKIN, S.; WIDROW, B. *Least-Mean-Square Adaptive Filters*. [S.l.]: Wiley Interscience, 2003.

[19] DINIZ, P. S. *Adaptive filtering*. [S.l.]: Springer, 1997.

[20] DOUGLAS, S.; MENG, T.-Y. Exact expectation analysis of the LMS adaptive filter without the independence assumption. In: IEEE. *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*. [S.l.], 1992. v. 4, p. 61–64.

[21] DOUGLAS, S. C. Exact expectation analysis of the LMS adaptive filter for correlated gaussian input data. In: *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*. [S.l.: s.n.], 1993. v. 3, p. 519–522 vol.3.

[22] DOUGLAS, S. C.; PAN, W. Exact expectation analysis of the LMS adaptive filter. *IEEE Transactions on Signal Processing*, v. 43, n. 12, p. 2863–2871, Dec. 1995.

[23] FAIFER, M. et al. Definition of simplified frequency-domain Volterra models with quasi-sinusoidal input. *IEEE Transactions on Circuits and Systems I: Regular Papers*, v. 65, n. 5, p. 1652–1663, May. 2018.

[24] SAABE, W. et al. Volterra-based modeling of traveling wave tube amplifier for system level simulation. In: *2017 Eighteenth International Vacuum Electronics Conference (IVEC)*. [S.l.: s.n.], 2017. p. 1–2.

[25] SERUNJOGI, S. M.; SANDULEANU, M. A.; RASRAS, M. S. Volterra series based linearity analysis of a phase-modulated microwave photonic link. *Journal of Lightwave Technology*, v. 36, n. 9, p. 1537–1551, May 2018.

[26] VALLIARAMPATH, J. T.; SINHA, S. Designing linear pas at millimeter-wave frequencies using Volterra series analysis. *Canadian Journal of Electrical and Computer Engineering*, v. 38, n. 3, p. 232–237, Summer 2015.

[27] ASSIS, L. S. de et al. Efficient volterra systems identification using hierarchical genetic algorithms. *Applied Soft Computing*, Elsevier BV, v. 85, p. 105745, dez. 2019. Disponível em: <https://doi.org/10.1016/j.asoc.2019.105745>.

[28] LOPES, C. G.; BERMUDEZ, J. C. M. Evaluation and design of variable step size adaptive algorithms. In: *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*. [S.l.: s.n.], 2001. v. 6, p. 3845–3848 vol.6. ISSN 1520-6149.

[29] EWEDA, E.; BERSHAD, N. J.; BERMUDEZ, J. C. M. Stochastic analysis of the recursive least squares algorithm for cyclostationary colored inputs. *IEEE Transactions on Signal Processing*, v. 68, p. 676–686, 2020.

[30] EWEDA, E.; BERSHAD, N. J.; BERMUDEZ, J. C. M. Stochastic analysis of the lms and nlms algorithms for cyclostationary white gaussian and non-gaussian inputs. *IEEE Transactions on Signal Processing*, v. 66, n. 18, p. 4753–4765, 2018.

[31] BERSHAD, N. J.; EWEDA, E.; BERMUDEZ, J. C. M. Stochastic analysis of an adaptive line enhancer/canceler with a cyclostationary input. *IEEE Transactions on Signal Processing*, v. 64, n. 1, p. 104–119, 2016.

[32] BERMUDEZ, J.; BERSHAD, N. Nonlinear quantization effects an the lms algorithm-analytical models for the mse transient and convergence behavior. In: *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*. [S.l.: s.n.], 1994. v. 1, p. 627–631 vol.1.

[33] AL-SHABILI, A.; WERUAGA, L.; JIMAA, S. Modal analysis of the ℓ_0-LMS and ℓ_0-NLMS sparse adaptive algorithms. In: *2016 IEEE 59th International Midwest Symposium on Circuits and Systems (MWSCAS)*. [S.l.: s.n.], 2016. p. 1–4.

[34] BERSHAD, N. On weight update saturation nonlinearities in lms adaptation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 38, n. 4, p. 623–630, 1990.

[35] SU, G. et al. Performance analysis of ℓ_0-norm constraint least mean square algorithm. *IEEE Transactions on Signal Processing*, v. 60, n. 5, p. 2223–2235, 2012.

[36] BERMUDEZ, J.; BERSHAD, N. A nonlinear analytical model for the quantized LMS algorithm-the arbitrary step size case. *IEEE Transactions on Signal Processing*, v. 44, n. 5, p. 1175–1183, 1996.

[37] BERSHAD, N. J.; BERMUDEZ, J. C. M.; TOURNERET, J.-Y. An affine combination of two LMS adaptive filters—transient mean-square analysis. *IEEE Transactions on Signal Processing*, v. 56, n. 5, p. 1853–1864, 2008.

[38] DOUGLAS, S. C.; PAN, W. Exact expectation analysis of the LMS adaptive filter. *IEEE Transactions on Signal Processing*, IEEE, v. 43, n. 12, p. 2863–2871, 1995.

[39] WIDROW, B. et al. Stationary and nonstationary learning characteristics of the LMS adaptive filter. *Aspects of Signal Processing*, Springer Netherlands, p. 355–393, 1977.

[40] BERSHAD, N. Analysis of the normalized LMS algorithm with gaussian inputs. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, IEEE, v. 34, n. 4, p. 793–806, 1986.

[41] KIM, S. E.; LEE, J. W.; SONG, W. J. Steady-state analysis of the NLMS algorithm with reusing coefficient vector and a method for improving its performance. In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2011. p. 4120–4123.

[42] ZHANG, S.; ZHANG, J. Transient analysis of zero attracting NLMS algorithm without gaussian inputs assumption. *Signal Processing*, v. 97, p. 100 – 109, 2014. ISSN 0165-1684. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0165168413004131>.

[43] YU YI; ZHAO, H. L. L. Steady-state mean-square analysis of the deficient-length NLMS algorithm with reusing coefficient vector for white input. *Control Conference (CCC), 2016 35th Chinese. IEEE*, p. 4787–4791, 2016.

[44] LARA, P.; TARRATACA, L. D.; HADDAD, D. B. Exact expectation analysis of the deficient-length LMS algorithm. *Signal Processing*, v. 162, p. 54 – 64, 2019.

[45] HADDAD, D. B.; PETRAGLIA, M. R. Transient and steady-state MSE analysis of the IMPNLMS algorithm. *Digital Signal Processing*, v. 33, p. 50 – 59, Oct. 2014.

[46] OLINTO, K. da S.; HADDAD, D. B.; PETRAGLIA, M. R. Transient analysis of ℓ_0-LMS and ℓ_0-NLMS algorithms. *Signal Processing*, v. 127, p. 217 – 226, 2016.

[47] OLINTO, K. da S.; HADDAD, D. B.; PETRAGLIA, M. R. Transient analysis of ℓ_0-LMS and ℓ_0-NLMS algorithms. *Signal Processing*, v. 127, p. 217 – 226, Oct. 2016.

[48] PETRAGLIA, M. R.; HADDAD, D. B.; MARQUES, E. L. Normalized subband adaptive filtering algorithm with reduced computational complexity. *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 62, n. 12, p. 1164–1168, 2015.

[49] ASSIS, L. S. de et al. Efficient Volterra systems identification using hierarchical genetic algorithms. *Applied Soft Computing*, v. 85, p. 105745, Dec. 2019.

[50] DOUGLAS, S. C.; PAN, W. Exact expectation analysis of the LMS adaptive filter. *IEEE Transactions on Signal Processing*, v. 43, n. 12, p. 2863–2871, Dec 1995.

[51] GOLUB, G. H.; LOAN, C. F. V. *Matrix Computations (3rd Ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996. ISBN 0-8018-5414-8.

[52] LARA, P. et al. Exact expectation evaluation and design of variable step-size adaptive algorithms. *IEEE Signal Processing Letters*, IEEE, v. 26, n. 1, p. 74–78, 2018.

[53] DOUGLAS, S. C.; MENG, T. H. Y. 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. [S.l.: s.n.], 1992. v. 4, p. 61–64 vol.4.

[54] Douglas, S. C. Exact expectation analysis of the LMS adaptive filter for correlated Gaussian input data. In: *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*. [S.l.: s.n.], 1993. v. 3, p. 519–522 vol.3.

[55] Lara, P. et al. Exact analysis of the least-mean-square algorithm with coloured measurement noise. *Electronics Letters*, v. 54, n. 24, p. 1401–1403, Nov. 2018.

[56] LARA, P.; HADDAD, D. B.; TARRATACA, L. Advances on the analysis of the LMS algorithm with a colored measurement noise. *Signal, Image and Video Processing*, v. 14, p. 529–536, Oct. 2019.

[57] Douglas, S. C.; Meng, T. H.-Y. 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*. [S.l.: s.n.], 1992. v. 4, p. 61–64.

[58] Lara, P. et al. Exact expectation analysis of the LMS adaptive identification of non-linear systems. *Electronics Letters*, v. 56, n. 1, p. 45–48, Jan. 2020.

[59] CARDOSO, J.-F. Blind signal separation: statistical principles. *Proceedings of the IEEE*, v. 86, n. 10, p. 2009–2025, Oct 1998.

[60] HASSIBI, B.; SAYED, A.; KAILATH, T. LMS is h/sup /spl infin// optimal. In: *Proceedings of 32nd IEEE Conference on Decision and Control*. [S.l.: s.n.], 1993. p. 74–79 vol.1.

[61] HASSIBI, B.; SAYED, A. H.; KAILATH, T. H^{∞} optimality of the LMS algorithm. *IEEE Transactions on Signal Processing*, v. 44, n. 2, p. 267–280, Feb. 1996.

[62] HAYKIN, S.; WIDROW, B. *Least-mean-square adaptive filters*. [S.l.]: John Wiley & Sons, 2003. v. 31.

[63] ZHOU, D.; DEBRUNNER, V. E. Novel adaptive nonlinear predistorters based on the direct learning algorithm. *IEEE Transactions on Signal Processing*, v. 55, n. 1, p. 120–133, 2007.

[64] Yong Hoon Lim et al. An adaptive nonlinear prefilter for compensation of distortion in nonlinear systems. *IEEE Transactions on Signal Processing*, v. 46, n. 6, p. 1726–1730, June 1998.

[65] MAZO, J. E. On the independence theory of equalizer convergence. *The Bell System Technical Journal*, v. 58, n. 5, p. 963–993, May 1979.

[66] ŚLIWIŃSKI, P. et al. Non-linear system modelling based on constrained volterra series estimates. *IET Control Theory & Applications*, Institution of Engineering and Technology (IET), v. 11, n. 15, p. 2623–2629, out. 2017. Disponível em: <https://doi.org/10.1049/iet-cta.2016.1360>.

[67] LIN, R.; NG, T.-Y. Identification of volterra kernels for improved predictions of nonlinear aeroelastic vibration responses and flutter. *Engineering Structures*, Elsevier BV, v. 171, p. 15–28, set. 2018.

[68] PRADO, R. A. d.; HENRIQUES, F. d. R.; HADDAD, D. B. Sparsity-aware distributed adaptive filtering algorithms for nonlinear system identification. In: *2018 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2018. p. 1–8.

[69] SILVA, F. B. da; MARTINS, W. A. Data-selective volterra adaptive filters. *Circuits, Systems, and Signal Processing*, Springer Science and Business Media LLC, v. 37, n. 10, p. 4651–4664, jan. 2018. Disponível em: <https://doi.org/10.1007/s00034-018-0765-0>.

[70] KAMALAKIS, T.; DEDE, G. Nonlinear degradation of a visible-light communication link: A volterra-series approach. *Optics Communications*, Elsevier BV, v. 417, p. 46–53, jun. 2018. Disponível em: <https://doi.org/10.1016/j.optcom.2018.02.022>.

[71] FENG, G. et al. Face recognition based on volterra kernels direct discriminant analysis and effective feature classification. *Information Sciences*, Elsevier BV, v. 441, p. 187–197, maio 2018. Disponível em: <https://doi.org/10.1016/j.ins.2018.02.028>.

[72] RUDKO, M.; WEINER, D. Volterra systems with random inputs: A formalized approach. *IEEE Transactions on Communications*, v. 26, n. 2, p. 217–227, 1978.

[73] YAO, L.; LIN, C.-C. Identification of nonlinear systems by the genetic programming-based Volterra filter. *IET Signal Processing*, Institution of Engineering and Technology

(IET), v. 3, n. 2, p. 93, 2009. Disponível em: <https://doi.org/10.1049/iet-spr:20070203>.

[74] MARCOS, S.; MACCHI, O. Tracking capability of the least mean square algorithm: Application to an asynchronous echo canceller. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 35, n. 11, p. 1570–1578, Nov 1987.

[75] Resende, L. C.; Haddad, D. B.; Petraglia, M. R. A variable step-size NLMS algorithm with adaptive coefficient vector reusing. In: *2018 IEEE International Conference on Electro/Information Technology (EIT)*. [S.l.: s.n.], 2018. p. 0181–0186.

[76] Hsia, C. et al. Low resolution method using adaptive LMS scheme for moving objects detection and tracking. In: *2010 International Symposium on Intelligent Signal Processing and Communication Systems*. [S.l.: s.n.], 2010. p. 1–4.

[77] Douglas, S. C.; Q. Zhu; Smith, K. F. A pipelined LMS adaptive FIR filter architecture without adaptation delay. *IEEE Transactions on Signal Processing*, v. 46, n. 3, p. 775–779, Mar. 1998.

[78] Rusu, A.; Ciochină, S.; Paleologu, C. On the step-size optimization of the LMS algorithm. In: *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*. [S.l.: s.n.], 2019. p. 168–173.

[79] Cattivelli, F. S.; Sayed, A. H. Analysis of spatial and incremental LMS processing for distributed estimation. *IEEE Transactions on Signal Processing*, v. 59, n. 4, p. 1465–1480, Apr. 2011.

[80] ALOUANE, M. T.-H.; JAIDANE-SAIDANE, M. A new nonstationary LMS algorithm for tracking markovian time varying systems. *Signal Processing*, v. 86, n. 1, p. 50 – 70, Jan. 2006. ISSN 0165-1684.

[81] Yousef, N. R.; Sayed, A. H. Ability of adaptive filters to track carrier offsets and channel nonstationarities. *IEEE Transactions on Signal Processing*, v. 50, n. 7, p. 1533–1544, Jul. 2002.

[82] van de Kerkhof, L. M.; Kitzen, W. J. W. Tracking of a time-varying acoustic impulse response by an adaptive filter. *IEEE Transactions on Signal Processing*, v. 40, n. 6, p. 1285–1294, Jun. 1992.

[83] Fong, W. et al. Monte Carlo smoothing with application to audio signal enhancement. *IEEE Transactions on Signal Processing*, v. 50, n. 2, p. 438–449, Feb. 2002.

[84] P. Lin; Rapajic, P.; Krusevac, Z. On the tracking performance of LMS and RLS algorithms in an adaptive MMSE CDMA receiver. In: *2005 Australian Communications Theory Workshop*. [S.l.: s.n.], 2005. p. 175–178.

[85] Hernandez, W.; Dominguez, M. E.; Sansigre, G. Analysis of the error signal of the LMS algorithm. *IEEE Signal Processing Letters*, v. 17, n. 3, p. 229–232, Mar. 2010.

[86] Silva, M. T. M.; Nascimento, V. H. Convex combination of adaptive filters with different tracking capabilities. In: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*. [S.l.: s.n.], 2007. v. 3, p. III–925–III–928.

[87] Butterweck, H. J. A wave theory of long adaptive filters. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, v. 48, n. 6, p. 739–747, Jun. 2001.

[88] BERMUDEZ, J. C.; BERSHAD, N. J.; EWEDA, E. Stochastic analysis of the LMS algorithm for cyclostationary colored Gaussian inputs. *Signal Processing*, v. 160, p. 127 – 136, 2019. ISSN 0165-1684.

[89] EWEDA, E.; BERSHAD, N. J. Stochastic analysis of a stable normalized least mean fourth algorithm for adaptive noise canceling with a white gaussian reference. *IEEE Transactions on Signal Processing*, v. 60, n. 12, p. 6235–6244, Dec 2012.

[90] ZHANG, S.; ZHANG, J.; SO, H. C. Mean square deviation analysis of LMS and NLMS algorithms with white reference inputs. *Signal Processing*, v. 131, p. 20 – 26, Feb. 2017.

[91] Eweda, E.; Bershad, N. J.; Bermudez, J. C. M. Stochastic analysis of the LMS and NLMS algorithms for cyclostationary white Gaussian and non-Gaussian inputs. *IEEE Transactions on Signal Processing*, v. 66, n. 18, p. 4753–4765, Sep. 2018.

[92] Hriljac, P. Learning curves for LMS and regular Gaussian processes. *IEEE Transactions on Automatic Control*, v. 47, n. 2, p. 284–289, Feb. 2002.

[93] Mayyas, K.; Aboulnasr, T. Leaky LMS algorithm: MSE analysis for gaussian data. *IEEE Transactions on Signal Processing*, v. 45, n. 4, p. 927–934, Apr. 1997.

[94] BENESTY, J. et al. A nonparametric VSS NLMS algorithm. *IEEE Signal Processing Letters*, v. 13, n. 10, p. 581–584, Oct. 2006.

[95] KWONG, R. H.; JOHNSTON, E. W. A variable step size LMS algorithm. *IEEE Transactions on Signal Processing*, v. 40, n. 7, p. 1633–1642, Jul 1992.

[96] Hwang, J.; Li, Y. Variable step-size LMS algorithm with a gradient-based weighted average. *IEEE Signal Processing Letters*, v. 16, n. 12, p. 1043–1046, Dec. 2009.

[97] Petraglia, M. R.; Haddad, D. B. New adaptive algorithms for identification of sparse impulse responses - analysis and comparisons. In: *2010 7th International Symposium on Wireless Communication Systems*. [S.l.: s.n.], 2010. p. 384–388.

[98] Ferreira, T. N. et al. Low-complexity proportionate algorithms with sparsity-promoting penalties. In: *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*. [S.l.: s.n.], 2016. p. 253–256.

[99] Benesty, J.; Paleologu, C.; Ciochină, S. Proportionate adaptive filters from a basis pursuit perspective. *IEEE Signal Processing Letters*, v. 17, n. 12, p. 985–988, Dec. 2010.

[100] Yukawa, M. Krylov-proportionate adaptive filtering techniques not limited to sparse systems. *IEEE Transactions on Signal Processing*, v. 57, n. 3, p. 927–943, Mar. 2009.

[101] Gelfand, S. B.; Y. Wei; Krogmeier, J. V. The stability of variable step-size LMS algorithms. *IEEE Transactions on Signal Processing*, v. 47, n. 12, p. 3277–3288, Dec. 1999.

[102] Macchi, O. Optimization of adaptive identification for time-varying filters. *IEEE Transactions on Automatic Control*, v. 31, n. 3, p. 283–287, Mar. 1986.

[103] OLLILA, E.; PALOMAR, D. P.; PASCAL, F. Shrinking the eigenvalues of m-estimators of covariance matrix. *IEEE Transactions on Signal Processing*, v. 69, p. 256–269, 2021.

[104] HUMPHREYS, D. A. et al. Principal component compression method for covariance matrices used for uncertainty propagation. *IEEE Transactions on Instrumentation and Measurement*, v. 64, n. 2, p. 356–365, 2015.

[105] BERSHAD, N.; QU, L. "on the probability density function of the lms adaptive filter weights". In: *ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing*. [S.l.: s.n.], 1987. v. 12, p. 109–112.

[106] MORRISON, R.; BAPTISTA, R.; BASOR, E. Diagonal nonlinear transformations preserve structure in covariance and precision matrices. *Journal of Multivariate Analysis*, v. 190, p. 104983, 2022. ISSN 0047-259X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0047259X22000252>.

[107] KWONG, R.; JOHNSTON, E. A variable step size lms algorithm. *IEEE Transactions on Signal Processing*, v. 40, n. 7, p. 1633–1642, 1992.

[108] FOLEY, J.; BOLAND, F. A note on the convergence analysis of lms adaptive filters with gaussian data. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 36, n. 7, p. 1087–1089, 1988.

[109] BUTTERWECK, H. An approach to lms adaptive filtering without use of the independence assumption. In: *1996 8th European Signal Processing Conference (EUSIPCO 1996)*. [S.l.: s.n.], 1996. p. 1–4.

[110] DOUGLAS, S.; MENG, T.-Y. Stochastic gradient adaptation under general error criteria. *IEEE Transactions on Signal Processing*, v. 42, n. 6, p. 1335–1351, 1994.

[111] MATHEWS, V.; XIE, Z. A stochastic gradient adaptive filter with gradient adaptive step size. *IEEE Transactions on Signal Processing*, v. 41, n. 6, p. 2075–2087, 1993.