PPGIO PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU* EM INSTRUMENTAÇÃO E ÓPTICA APLICADA

MODELAGEM ESTOCÁSTICA DE ALGORITMOS DE FILTRAGEM ADAPTATIVA COM RESTRIÇÃO DE NORMA ℓ_0

Lucas Paiva Rodrigues da Silva

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

Orientador(a): Diego Barreto Haddad
Coorientador(a): Fábio Augusto de Alcântara Andrade

Rio de Janeiro,

Junho 2025

MODELAGEM ESTOCÁSTICA DE ALGORITMOS DE FILTRAGEM ADAPTATIVA COM RESTRIÇÃO DE NORMA ℓ_0

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

Lucas Paiva Rodrigues da Silva

Banca Examinadora:

gov.br Documento assinado digitalmente
DIEGO BARRETO HADDAD
Data: 30/06/2025 12:30:33-0300
Verifique em https://validar.iti.gov.br

Presidente, Professor D.Sc. Diego Barreto Haddad (CEFET/RJ) (Orientador(a))

Professor Associado Ph.D. Fábio Augusto de Alcântara Andrade (USN) (Coorientador(a))

gov.br Documento assinado digitalmente
FELIPE DA ROCHA HENRIQUES
Data: 03/07/2025 15:03:58-0300
Verifique em https://validar.iti.gov.br

Professor D.Sc. Felipe da Rocha Henriques (CEFET/RJ)

gov.br Documento assinado digitalmente
GILSON ANTONIO GIRALDI
Data: 03/07/2025 16:17:18-0300
Verifique em https://validar.iti.gov.br

Professor D.Sc. Gilson Antonio Giraldi (LNCC)

gov.br Documento assinado digitalmente
MARIANE REMBOLD PETRAGLIA
Data: 04/07/2025 12:11:51-0300
Verifique em https://validar.iti.gov.br

Professora Ph.D. Mariane Rembold Petraglia (UFRJ)

gov.br Documento assinado digitalmente
LEONARDO CORREIA RESENDE
Data: 04/07/2025 13:44:01-0300
Verifique em https://validar.iti.gov.br

Professor D.Sc. Leonardo Correia Resende (IFRJ)

Rio de Janeiro,

Junho 2025

Ficha catalográfica elaborada pela Biblioteca Central do CEFET/RJ

AS586 Silva, Lucas Paiva Rodrigues da
Modelagem estocástica de algoritmos de filtragem adaptativa com restrição de norma$_{lo}$ / Lucas Paiva Rodrigues da Silva. — 2025.
85f. : il. color. , enc.

Tese (Doutorado) Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, 2025.
Bibliografia : f. 77-85
Orientador: Diego Barreto Haddad
Coorientador: Fábio Augusto Alcântara Andrade

1. Processamento de sinais. 2. Filtros adaptativos. 3. Algoritmos. 4. Processo estocástico. I. Haddad, Diego Barreto. (Orient.). II. Andrade, Fábio Augusto Alcântara. (Coorient.). III. Título.

CDD 621.3822

Elaborada pela bibliotecária Tania Mello – CRB/7 nº 5507/04

DEDICATÓRIA

À minha avó, Lucimar, que me ensinou que o conhecimento vai além dos livros. Seu amor e sabedoria ainda me guiam, mesmo que seus olhos não possam mais ver esta vitória.

AGRADECIMENTOS

Gostaria de expressar minha profunda gratidão ao meu orientador, Diego Haddad, pela paciência, dedicação e apoio incondicional ao longo desses anos. Sua capacidade de incentivo e atenção constante foram fundamentais para a conclusão deste trabalho. Sem sua expertise e generosidade, esta tese não seria possível. Ele é, sem dúvida, o professor que mais influenciou minha formação acadêmica, e espero levar sua amizade para além desta jornada.

Aos professores e colegas do PPGIO, pelas discussões enriquecedoras e pelo conhecimento compartilhado, assim como a todos os docentes que contribuíram para minha trajetória.

Aos meus amigos e familiares, pelo apoio, compreensão e presença nos momentos mais desafiadores.

Às pessoas que, direta ou indiretamente, colaboraram para a realização desta tese. Cada contribuição, por menor que pareça, foi peça essencial nesse processo.

Por fim, um agradecimento especial aos meus avós, que me deram amor incondicional, suporte e foram meus primeiros educadores.

O presente trabalho foi desenvolvido com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

RESUMO

Modelagem Estocástica de Algoritmos de Filtragem Adaptativa com Restrição de Norma ℓ_0

Um dos principais desafios na utilização de algoritmos de filtragem adaptativa é a emulação (ou identificação) eficiente de um sistema em um ambiente ruidoso. Isso pode ser facilitado em aplicações onde a resposta do sistema ao impulso é esparsa, o que permite acelerar a taxa de convergência se estratégias adequadas forem utilizadas. Neste contexto, métodos que impõem restrições de norma são amplamente utilizados. Esta tese propõe um novo modelo capaz de prever o desempenho do algoritmo ℓ_0-*Least Mean Square (LMS)* no caso de plantas não estacionárias sujeitas a uma perturbação markoviana de primeira ordem. O modelo teórico oferece uma equação analítica que prevê o desvio médio quadrático assintótico em termos da variância da perturbação markoviana. Também é proposto um novo algoritmo que utiliza uma aproximação linear por partes da função de penalização de norma. Este algoritmo tem como casos particulares diversos outros algoritmos previamente propostos. Por fim, é proposto um modelo teórico para este algoritmo, o qual obtém fórmulas analíticas tanto para o seu desempenho assintótico (com um erro absoluto inferior a 0,2 dB) quanto para o valor ótimo do parâmetro que regula a intensidade do termo de regularização de norma.

Palavras-chave: Filtragem Adaptativa Consciente da Esparsidade, Algoritmo ℓ_0-LMS, Análise de Rastreamento, Modelos Estocásticos.

ABSTRACT

Stochastic Modeling of Adaptive Filtering Algorithms with ℓ_0 Norm Constraint

One of the main challenges in using adaptive filtering algorithms is the efficient emulation (or identification) of a system in a noisy environment. This task can be facilitated in applications where the system's impulse response is sparse, allowing for faster convergence rates if appropriate strategies are employed. In this context, norm-constrained methods are widely used. This thesis proposes a new model capable of predicting the performance of the ℓ_0-*LMS* algorithm in the case of nonstationary plants subject to a first-order Markovian disturbance. The theoretical model provides an analytical expression that predicts the asymptotic mean-square deviation in terms of the variance of the Markovian perturbation. A new algorithm is also proposed, which employs a piecewise linear approximation of the norm penalty function. This algorithm generalizes several previously proposed algorithms as special cases. Finally, a theoretical model is developed for this algorithm, yielding analytical formulas for both its asymptotic performance (with an absolute error below 0.2 dB) and for the optimal value of the parameter that controls the strength of the norm regularization term.

Keywords:

Sparsity-Aware Adaptive Filtering, ℓ_0-LMS Algorithm, Tracking Analysis, Stochastic Models.

LISTA DE ILUSTRAÇÕES

Figura 1 – Diagrama em blocos da estrutura de um algoritmo adaptativo supervisionado de identificação de sistemas. 15

Figura 2 – Função $g_i(k)$ em termos de $w_i(k)$. Na curva em azul tem-se que α = 10 e em vermelho α = 5. 22

Figura 3 – MSD em regime permanente do Cenário 1. Em vermelho a curva teórica e em azul a empírica. 47

Figura 4 – MSD em regime permanente do Cenário 2. Em vermelho a curva teórica e em azul a empírica. 48

Figura 5 – MSD em regime permanente do Cenário 3. Em vermelho a curva teórica e em azul a empírica. 48

Figura 6 – MSD em regime permanente do Cenário 4. Em vermelho a curva teórica e em azul a empírica. 49

Figura 7 – MSD em regime permanente do Cenário 5. Em vermelho a curva teórica e em azul a empírica. 49

Figura 8 – Exemplo de uma função linear por partes $g[w_i(k)]$ com $P = 3$. 54

Figura 9 – Comparação do comportamento do Mean Square Deviation (MSD) dos algoritmos LMS e Reweighted Zero Attracting Least Mean Squares (RZA-LMS) e aproximações (com o termo de penalização sendo linear por partes) do algoritmo RZA-LMS com diferentes valores de P (com $P \in \{100, 1000, 10000\}$). As configurações do experimentos são as mesmas da Figura 1 de [31]. 55

Figura 10 – Comparação entre o termo de penalização de norma do algoritmo RZA-LMS [31] (linha vermelha tracejada, com $\epsilon = 10$, ver Equação (112)) e a aproximação linear por partes adotada nas simulações desta seção (linha azul contínua). O valor de ϵ foi escolhido de forma a se alinhar com as simulações apresentadas em [31]. 69

Figura 11 – MSD em regime permanente (em dB), empírico (azul) e teórico (vermelho), em função de β. O MSD empírico foi obtido a partir de $K = 100$ execuções independentes de Monte Carlo. Em cada execução, foi calculada a média das últimas 1000 iterações. 69

Figura 12 – MSD em regime permanente (em dB), empírico (azul) e teórico (vermelho), em função de σ_x^2. O MSD empírico foi obtido a partir de $K = 1000$ execuções independentes de Monte Carlo. Em cada execução, foi calculada a média das últimas 1000 iterações. 70

Figura 13 – MSD(em dB) em regime permanente empírico (em azul) e teórico (em vermelho), como uma função de σ_ν^2. O MSD empírico foi obtido por meio de $K = 100$ ensaios independentes de Monte Carlo. Em cada ensaio de Monte Carlo, a média das últimas 1000 iterações foi computada. 71

Figura 14 – MSD em regime permanente (em dB), empírico (círculos azuis) e teórico (curva vermelha), como uma função de κ. O valor teórico ótimo de κ computado pela Equação (193) é indicado por uma linha verde vertical tracejada. O MSD empírico foi obtido por meio de $K = 100$ ensaios independentes de Monte Carlo. Em cada ensaio de Monte Carlo, a média das últimas 10000 iterações foi computada. 71

Figura 15 – MSD em regime permanente (em dB), empírico (círculos azuis) e teórico (curva vermelha), como uma função de κ. O valor teórico ótimo de κ computado pela Equação (193) é indicado por uma linha verde vertical tracejada. O MSD empírico foi obtido por meio de $K = 100$ ensaios independentes de Monte Carlo. Em cada ensaio de Monte Carlo, a média das últimas 10000 iterações foi computada. 72

Figura 16 – MSD em regime permanente (em dB), empírico (círculos azuis) e teórico (curva vermelha), como uma função de κ. O valor teórico ótimo de κ computado pela Equação (193) é indicado por uma linha verde vertical tracejada. O MSD empírico foi obtido por meio de $K = 100$ ensaios independentes de Monte Carlo. Em cada ensaio de Monte Carlo, a média das últimas 10000 iterações foi computada. 72

Figura 17 – MSD em regime permanente (em dB), empírico (círculos azuis) e teórico (curva vermelha), como uma função de κ. O valor teórico ótimo de κ computado pela Equação (193) é indicado por uma linha verde vertical tracejada. O MSD empírico foi obtido por meio de $K = 100$ ensaios independentes de Monte Carlo. Em cada ensaio de Monte Carlo, a média das últimas 10000 iterações foi computada. 73

LISTA DE TABELAS

Tabela 1 – Algumas referências que têm contribuições no âmbito da família de algoritmos adaptativos proporcionais. 19

Tabela 2 – Referências de relevo que citam o trabalho que propõe o algoritmo ℓ_0-LMS. 23

Tabela 3 – Principais parâmetros dos cenários e seus respectivos valores. 46

Tabela 4 – Plantas dos cenários. 46

Tabela 5 – Exemplos de algoritmos que se ajustam ao paradigma proposto neste artigo. 54

LISTA DE ABREVIATURAS E SIGLAS

APA	Affine Projection Algorithm
GPLR-LMS	Generalized Piecewise Linear Regularization Least Mean Squares
IIR	Infinite Impulse Response
ILMS	Improved ℓ_0 LMS
IMPNLMS	Improved Proportionate Normalized Least Mean Squares
ITU	International Telecommunication Union
LMS	Least Mean Square
MSD	Mean Square Deviation
NLMS	Normalized Least Mean Squares
OZ-LMS	Optmized Zero Attracting LMS
PNLMS	Proportionate Normalized Least Mean Squares
PRLS	Proportionate Recursive Least Squares
RLS	Recursive Least Squares
RZA-LMS	Reweighted Zero Attracting Least Mean Squares
ZA-LMS	Zero Attracting Least Mean Squares

SUMÁRIO

1	**Introdução**	**10**
1.1	Objetivos deste trabalho	11
1.2	Estrutura deste trabalho	12
1.3	Publicações associadas	12
1.4	Notação Matemática	13
2	**Fundamentação Teórica**	**14**
2.1	O algoritmo LMS	16
2.2	Algoritmos Conscientes da Esparsidade	17
2.2.1	Algoritmos Proporcionais	17
2.2.2	Algoritmos com Restrição de Norma	19
2.3	Considerações Finais	23
3	**Análise Assintótica de Rastreamento do Algoritmo ℓ_0-LMS**	**24**
3.1	Introdução	24
3.2	Análise de Primeira Ordem	25
3.3	Análise de Segunda Ordem	30
3.4	Resultados	44
3.5	Cenários	45
3.6	Comentários Finais	50
4	**Modelagem Estocástica do Algoritmo de Penalização de Norma Generalizada**	**51**
4.1	Algoritmo GPLR-LMS	53
4.2	Análise de Regime Permanente do Algoritmo GPLR-LMS	56
4.3	Modelo estocástico simplificado	64
4.4	Resultados	67
4.5	Conclusões	74

5 Conclusões e Trabalhos Futuros **75**

Referências Bibliográficas **77**

1- Introdução

A estimativa de um sistema ótimo é um típico problema de engenharia, cujos sinais encontram-se comumente sujeitos a ruído de medição imprevisível [1]. Tal estimativa é normalmente denominada “identificação de sistemas”, o que compreende a determinação de um modelo matemático que descreva o comportamento de um sistema baseado nos seus respectivos sinais de entrada e saída [2]. Algoritmos de filtragem adaptativa, foco deste trabalho, têm um papel fundamental na implementação de estimadores em tempo real de plantas variantes no tempo, sendo extensivamente utilizados em aplicações tais como identificação de sistemas, cancelamento de eco, equalização de canais, cancelamento de interferência inter-simbólica e controle ativo de ruído [3]–[5]. Em anos recentes, os algoritmos adaptativos têm experimentado uma expansão em aplicações como telefonia sem fio, sistemas de audioconferência e sistemas de suporte à audição [6].

As diversas soluções de filtragem adaptativa compartilham uma estrutura em comum, na qual uma função custo é otimizada de modo iterativo [7]. A escolha da função custo adotada na derivação do algoritmo tem grande impacto na eficiência dos algoritmos de filtragem adaptativa resultantes [3]. Na verdade, a seleção da mesma é um dos maiores desafios em filtragem adaptativa [8]. O algoritmo LMS é um dos mais simples e populares algoritmos de filtragem adaptativa. Ele utiliza uma função custo quadrática, apresentando as desejáveis propriedades de convergir para a solução de Wiener caso o passo de aprendizagem encontre-se numa determinada faixa, ter uma estrutura simples e demandar baixo custo computacional [9], [10]. A estas características se soma o fato de que funções custo quadráticas apresentam tratabilidade matemática, convexidade e desempenho ótimo quando os sinais e o ruído são gaussianos [8].

Em diversas aplicações (como em sistemas acústicos, sísmicos, químicos e de comunicações [11], [12]), as respostas ao impulso dos sistemas a identificar são sabidamente esparsas. Tal esparsidade significa que poucos coeficientes destas respostas apresentam magnitude elevada, enquanto que os demais têm magnitude desprezível (ou mesmo nula). Atualmente, há duas abordagens maduras de exploração da esparsidade para obter melhora no desempenho dos algoritmos tradicionais: os métodos proporcionais e os métodos com restrição de norma [13]. No escopo desta última abordagem, o algo-

ritmo ℓ_0-LMS é um dos mais populares [14]. As abordagens "conscientes da esparsidade" têm gerado um contínuo interesse de pesquisa na última década [15]. Afinal, o projeto de algoritmos de filtragem adaptativa para filtros longos com baixa latência e alta taxa de convergência para aplicações como cancelamento de eco é algo tão importante quanto desafiador [16].

A adoção destes algoritmos na prática é normalmente mediada por diretrizes e garantias oferecidas por modelos estocásticos (como, por exemplo, a indicação da região de estabilidade ou a predição de desempenho assintótico), os quais suscitam grande interesse na literatura especializada. Assim, modelos estocásticos são úteis no estabelecimento de relações de causa e efeito, bem como oferecem uma compreensão mais profunda do comportamento de um algoritmo de filtragem adaptativa [17].

1.1- Objetivos deste trabalho

Diferentes modelos teóricos têm sido mobilizados para analisar o algoritmo ℓ_0-LMS [18], [19]. Um deles, proposto em [19], é particularmente expressivo, porque é capaz de gerar expressões analíticas para o desempenho assintótico, cujos termos estabelecem relações *interpretáveis* de causa e efeito relacionados ao comportamento do algoritmo. Por essa razão, esse modelo estocástico será o foco desta tese, a qual pretende estendê-lo para novas configurações nas quais opera o algoritmo ℓ_0-LMS (notadamente, em contextos não estacionários), assim como para uma versão generalizada do algoritmo.

Assim, podemos elencar os seguintes objetivos principais:

1. Adaptar o modelo estocástico proposto por [19] para modelar a identificação de uma planta esparsa que varia no tempo, consoante um processo markoviano de primeira ordem;

2. Deduzir uma equação analítica que descreva o desempenho assintótico do algoritmo ℓ_0-LMS operando numa planta não estacionária;

3. Propor uma versão generalizada de algoritmos de filtragem adaptativa com penalização de norma, na qual a função de penalização de norma é linear por partes;

4. Deduzir uma equação analítica que descreva o desvio quadrático médio assintótico da versão generalizada do algoritmo ℓ_0-LMS (que denominamos GPRL-LMS);

5. Deduzir, por meio de algumas simplificações, uma equação analítica que permita determinar o valor ótimo do parâmetro que regula a intensidade da regularização por norma da versão generalizada proposta do algoritmo proposto GPRL-LMS.

1.2- Estrutura deste trabalho

Este trabalho é estruturado do seguinte modo: o Capítulo 2 apresenta a fundamentação teórica dos algoritmos adaptativos conscientes da esparsidade. O Capítulo 3 propõe um modelo estocástico inédito capaz de prever o desempenho assintótico do algoritmo ℓ_0-LMS, o qual é comparado com resultados de simulações. Já o Capítulo 4 propõe uma versão generalizada de algoritmos de filtragem adaptativa com penalização por norma, propõe um novo modelo estocástico capaz de prever seu desempenho assintótico, e encontra fórmulas fechadas tanto para o seu desempenho assintótico quanto para o parâmetro de intensidade que regula o termo de penalização de norma. Por fim, o Capítulo 5 encerra as conclusões desta tese, assim como delineia futuras direções de pesquisa.

1.3- Publicações associadas

Ao longo desta pesquisa, o seguintes artigos foram *publicado* em revistas internacionais:

- **Silva, Lucas Paiva** et al., *Tracking Analysis of the ℓ_0-LMS Algorithm*, *Circuits, Systems, and Signal Processing*, vol. 43, p. 7867-7893, Aug. 2024.

O fator de impacto (em 2023) é de 1,8 e Qualis CAPES (referente ao Quadriênio 2017-2020) encontra-se no estrato A2.

- **Silva, Lucas Paiva** et al., *A Novel Stochastic Model for the Steady-state Performance of Norm-penalized Adaptive Algorithms*. *Digital Signal Processing*, vol.166, 2025.

Este artigo possui Qualis A1 (referente ao Quadriênio 2017-2020) e fator de impacto 3,0 (em 2024).

1.4- Notação Matemática

Ao longo deste trabalho, matrizes são denotadas por letras maiúsculas em negrito (por exemplo, $\boldsymbol{A}$). Já vetores, sempre do tipo coluna, são denotados por letras minúsculas em negrito (como $\boldsymbol{w}$). Para escalares, reservamos letras minúsculas sem negrito (por exemplo, d). O operador $(\cdot)^T$ indica transposição matricial, enquanto que $\mathbb{E}[\cdot]$ denota o operador de valor esperado (ou esperança estatística). Ademais, Tr$[\boldsymbol{A}]$ denota o traço da matriz quadrada $\boldsymbol{A}$ e $\nabla_{\boldsymbol{w}} f$ denota o gradiente de f com relação ao vetor $\boldsymbol{w}$. Por fim, $\|\boldsymbol{w}\|_q$ denota a norma ℓ_q do vetor $\boldsymbol{w}$ (ou seja: $\|\boldsymbol{w}\|_q \triangleq (\sum_k |w_k|^q)^{1/q}$).

2- Fundamentação Teórica

A filtragem adaptativa desempenha um papel central em diversas aplicações de processamento de sinais digitais, sobretudo em ambientes onde o sistema a ser modelado apresenta comportamento dinâmico ou estrutura esparsa [20]. Esse é o caso, por exemplo, de sistemas de comunicação em redes digitais, sensores inteligentes, dispositivos e aplicações em aprendizado de máquina embarcado, como redes neurais convolucionais otimizadas para execução em tempo real [21]–[24].

Nesse contexto, o sistema a ser identificado — ou planta — costuma variar ao longo do tempo, seja por alterações nas condições do ambiente, seja pela própria evolução temporal dos dados de entrada [25]. Para lidar com isso, os métodos utilizados devem ter complexidade computacional reduzida, apresentar taxa de convergência alta e pouca sensibilidade a efeitos de ponto fixo [26].

Diante desses desafios, uma estratégia amplamente empregada é o uso de algoritmos adaptativos voltados à estimativa da resposta do sistema ao longo do tempo. Esta tese se dedica a esse tipo de abordagem, explorando técnicas supervisionadas para ajustar os parâmetros do modelo com base em dados de entrada e saída. O desenvolvimento de soluções eficazes envolve etapas clássicas da identificação de sistemas, como o planejamento experimental, a seleção de uma estrutura do modelo, a estimativa de parâmetros e a validação do algoritmo adotado [27].

Considere o problema de rastrear o comportamento de uma planta digital inserida em uma rede de comunicação, seja $d(k)$ a saída desejada do sistema. Assumindo linearidade, invariância no tempo e uma resposta ao impulso finita e desconhecida cujos coeficientes encontram-se armazenados no vetor $\boldsymbol{w}^{\star}(k) \in \mathbb{R}^N$, podemos modelar sistema supracitado por meio da relação

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

onde

$$\boldsymbol{w}^{\star}(k) = \begin{bmatrix} w_0^{\star}(k) & w_1^{\star}(k) & w_2^{\star}(k) & \dots & w_N^{\star}(k) \end{bmatrix}^T, \tag{2}$$

contém N coeficientes desconhecidos a identificar,

$$\boldsymbol{x}(k) = \begin{bmatrix} x(k) & x(k-1) & x(k-2) & \dots & x(k-N+1) \end{bmatrix}^T \tag{3}$$

contém N amostras de entrada consecutivas e $\nu(k) \in \mathbb{R}$ denota o ruído de medição. O modelo (1) permite-nos assumir que a solução estimada por uma estrutura transversal será adequada num regime de pequenos desvios, uma hipótese tradicional em problemas estatísticos de processamento de sinais [28].

A identificação de sistemas (1) é costumeiramente empreendida por meio de um filtro adaptativo de uma estrutura transversal $\boldsymbol{w}(k) \in \mathbb{R}^N$ cuja saída $y(k) \in \mathbb{R}$ na k-ésima iteração

$$y(k) = \boldsymbol{w}^T(k)\boldsymbol{x}(k) \tag{4}$$

deve estimar $d(k)$ em (1) com a maior acurácia possível.

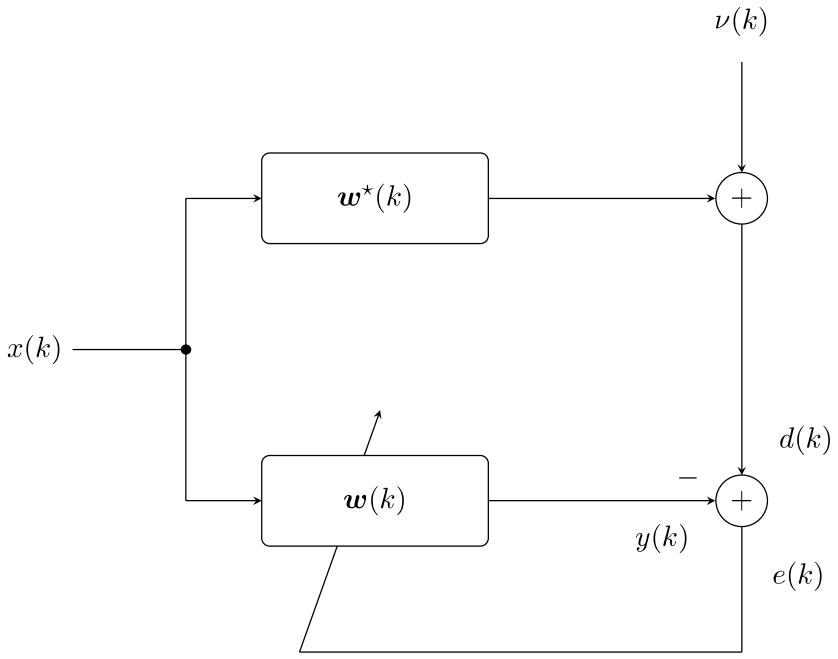


Figura 1 – Diagrama em blocos da estrutura de um algoritmo adaptativo supervisionado de identificação de sistemas.

Uma solução de baixo custo para este problema reside no apelo a algoritmos iterativos que adaptam continuamente $\boldsymbol{w}(k)$, ou seja:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \Delta\boldsymbol{w}(k), \tag{5}$$

onde o termo $\Delta \boldsymbol{w}(k)$ depende da discrepância atual entre o sinal de referência $d(k)$ e a saída do filtro $y(k)$. Denotamos este sinal de discrepância como o erro $e(k) \in \mathbb{R}$, cujo cômputo é

$$e(k) = d(k) - y(k), \tag{6}$$

o que caracteriza os algoritmos tratados neste trabalho como supervisionados, já que $d(k)$ é um sinal conhecido que atua continuamente na supervisão do treinamento do filtro. Esta configuração está inserida na modalidade de métodos de identificação conhecidos como de erro de predição mínima, os quais são guiados pelos sinais exteriores do comportamento do sistema [29]. A Figura 1 apresenta o diagrama em blocos da estrutura adaptativa que será doravante considerada.

2.1- O algoritmo LMS

O algoritmo LMS foi o pioneiro dos algoritmos de filtragem adaptativa. Ele conjuga a minimização do quadrado do erro instantâneo (ou seja, a função custo) com o método de otimização conhecido como *gradiente estocástico*. Assim, a atualização (5) é particularizada para:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \beta \nabla_{\boldsymbol{w}(k)} \left[\frac{1}{2} e^2(k) \right], \tag{7}$$

onde $\beta \in \mathbb{R}_+$ é um parâmetro ajustável, responsável por características tais como taxa de convergência e desempenho assintótico. Como o gradiente sinaliza a direção de subida mais íngreme de uma função, o sinal de negativo é adotado porque pretende-se caminhar na direção de minimização desta função. O algoritmo LMS pode então ser explicitamente escrito como

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \beta e(k) \nabla_{\boldsymbol{w}(k)} \left[e(k) \right] = \boldsymbol{w}(k) - \beta e(k) \nabla_{\boldsymbol{w}(k)} \left[d(k) - \boldsymbol{w}^T(k) \boldsymbol{x}(k) \right]$$

$$\Rightarrow \boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \boldsymbol{x}(k) e(k). \tag{8}$$

Quando o algoritmo é estável, seu desempenho progressivamente melhora nas primeiras iterações (caracterizando um regime conhecido como transiente), apresentando um desempenho constante após o transiente. Esta região de desempenho constante é

chamada de regime permanente (ou regime assintótico).

Uma métrica de aferição de desempenho que será bastante útil ao longo desta tese é o *MSD*

$$\text{MSD (em dB)} = 10\log_{10}\left\{\mathbb{E}\left[\|\boldsymbol{w}^{\star}(k) - \boldsymbol{w}(k)\|^2\right]\right\}, \tag{9}$$

onde o valor esperado é tipicamente computado empiricamente (ou seja, por meio de simulações computacionais), com a média advinda de diversos ensaios independentes de Monte Carlo.

2.2- Algoritmos Conscientes da Esparsidade

Duas grandes famílias de algoritmos de filtragem adaptativa pretendem acelerar a taxa de convergência na identificação de sistemas esparsos: *i)* a família de algoritmos proporcionais (inaugurada por [30]), caracterizada por uma adaptação dos coeficientes que leva em conta sua magnitude; *ii)* e a família de restrição por norma, a qual propicia uma regularização promotora de esparsidade na função custo adotada pelo algoritmo [14], [31]. A seguir, alguns conceitos mobilizados por estas duas famílias serão apresentados.

2.2.1 Algoritmos Proporcionais

Em 2000, Duttweiler [30] propôs uma ideia bastante frutífera para a evolução dos algoritmos conscientes da esparsidade: determinar a direção de atualização utilizando a estimativa atual do filtro. Assim, o algoritmo *Proportionate Normalized Least Mean Squares (PNLMS)* atribui passos de aprendizagem distintos para diferentes coeficientes adaptativos, com coeficientes de maior magnitude recebendo passos de aprendizagem maiores [32].

A maioria dos algoritmos proporcionais pode ser colocada como solucionadores

do seguinte problema de otimização com restrições:

$$\min_{\boldsymbol{w}(k+1)} \frac{1}{2}\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|^2_{\boldsymbol{\Lambda}^{-1}(k)} \text{ sujeito a } e_{\mathsf{p}}(k) = (1-\beta)e(k), \tag{10}$$

onde $\boldsymbol{\Lambda} \in \mathbb{R}^{N\times N}$ é uma matriz diagonal cuja diagonal principal é responsável por alterar o passo de aprendizagem para cada coeficiente adaptativo, $\|\boldsymbol{w}\|^2_{\boldsymbol{A}} \triangleq \boldsymbol{w}^T\boldsymbol{A}\boldsymbol{w}(k)$ denotando a norma ℓ_2 do vetor $\boldsymbol{w}$ ponderada pela matriz $\boldsymbol{A}$ e com o erro *a posteriori* $e_{\mathsf{p}}(k)$ sendo definido como o erro recalculado com as informações da k-ésima iteração após a atualização:

$$e_{\mathsf{p}}(k) \triangleq d(k) - \boldsymbol{w}^T(k+1)\boldsymbol{x}(k). \tag{11}$$

Utilizando a técnica dos multiplicadores de Lagrange, o problema de otimização (10) pode ser transladado para um problema equivalente sem restrições:

$$\min_{\boldsymbol{w}(k+1)} \frac{1}{2}\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|^2_{\boldsymbol{\Lambda}^{-1}(k)} + \lambda\left[e_{\mathsf{p}}(k) - (1-\beta)e(k)\right], \tag{12}$$

onde $\lambda \in \mathbb{R}$ denota o multiplicador de Lagrange. Calculando-se o gradiente de (12) com relação a $\boldsymbol{w}(k+1)$ e igualando-se ao vetor nulo, obtemos:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \lambda\boldsymbol{\Lambda}(k)\boldsymbol{x}(k). \tag{13}$$

A inserção de (13) na restrição linear de (10) nos permite obter λ:

$$\lambda = \frac{\beta e(k)}{\|\boldsymbol{x}(k)\|^2_{\boldsymbol{\Lambda}(k)}} = \frac{\beta e(k)}{\boldsymbol{x}^T(k)\boldsymbol{\Lambda}(k)\boldsymbol{x}(k)}, \tag{14}$$

resultado que, quando inserido em (13), nos fornece a equação de atualização dos algoritmos proporcionais:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \frac{\beta e(k)\boldsymbol{\Lambda}(k)\boldsymbol{x}(k)}{\boldsymbol{x}^T(k)\boldsymbol{\Lambda}(k)\boldsymbol{x}(k)}, \tag{15}$$

com os diferentes algoritmos desta família geralmente se diferenciando na escolha dos coeficientes da matriz diagonal $\boldsymbol{\Lambda}(k)$. Por exemplo, o algoritmo *Improved Proportionate Normalized Least Mean Squares (IMPNLMS)* adota a seguinte fórmula para o i-ésimo

elemento da diagonal principal de $\mathbf{\Lambda}(k)$ [33]:

$$\lambda_i(k) = \frac{1-\epsilon_1}{2N} + (1+\epsilon_1)|\frac{|w_i(k)|}{2\|\boldsymbol{w}(k)\|_1 + \epsilon_2}, \tag{16}$$

onde $\epsilon_1 \in [-1, 1]$ é um parâmetro ajustável que regula a penalização por esparsidade e ϵ_2 (também ajustável) evita divisões por zero. A Tabela 1 apresenta algumas contribuições na literatura efetuadas com relação aos algoritmos proporcionais, sem pretensão de exaurir a imensa riqueza teórica que esta família de técnicas apresenta. Nesta tese, estamos interessados numa outra família de algoritmos, que será abordada na próxima seção. [1]

Tabela 1 – Algumas referências que têm contribuições no âmbito da família de algoritmos adaptativos proporcionais.

Referência	**Contribuição**
[35]	Algoritmo proporcional em subbandas e com *Set-membership*
[36]	Combinação adaptativa de filtros proporcionais
[37]	Algoritmos proporcionais do tipo sign-*Affine Projection Algorithm (APA)*
[38]	Simplificação do algoritmo APA proporcional
[39]	Versão eficiente do APA proporcional
[40]	Filtro proporcional minimizador de medidas de diversidade
[41]	Algoritmo adaptativo proporcional baseado na norma ℓ_0
[42]	Algoritmo proporcional do tipo APA robusto a interferência impulsiva
[43]	Algoritmo proporcional para sinais complexos
[44]	Algoritmo proporcional para sistemas esparsos em blocos
[45]	Algoritmo proporcional para controle ativo de ruído

2.2.2 Algoritmos com Restrição de Norma

A ideia básica dos algoritmos com restrição com norma reside na introdução explícita de uma penalidade que favorece a esparsidade na função custo do algoritmo. Em termos gerais, temos uma função custo dada por

$$\mathcal{F}_{\mathsf{RL}}\left[\boldsymbol{w}(k)\right] = \frac{1}{2}e^2(k) + \gamma H[\boldsymbol{w}(k)], \tag{17}$$

[1] Um estudo comparativo entre as famílias já foi feito em [34], não sendo este portanto, o objetivo deste trabalho.

onde “RL” denota “regularização por norma” e $H[\boldsymbol{w}(k)]$ é uma função que penaliza soluções pouco esparsas, a qual se soma à função custo quadrática utilizada pelo LMS (veja a Equação (7)).

Pertencente a esta família de técnicas, o algoritmo *Zero Attracting Least Mean Squares (ZA-LMS)*, proposto originalmente em [31], utiliza a norma ℓ_1 como penalidade para o vetor de coeficientes:

$$\mathcal{F}_{\text{ZA}}\left[\boldsymbol{w}(k)\right] = \frac{1}{2}e^2(k) + \gamma\|\boldsymbol{w}(k)\|_1. \tag{18}$$

Utilizando o gradiente estocástico para otimizar (18), obtemos a equação de atualização do ZA-LMS:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \beta\nabla_{\boldsymbol{w}(k)}\left[\frac{1}{2}e^2(k) + \gamma\|\boldsymbol{w}(k)\|_1\right] = \boldsymbol{w}(k) + \beta\boldsymbol{x}(k)e(k) - \kappa\text{sign}[\boldsymbol{w}(k)], \tag{19}$$

onde[2]

$$\kappa \triangleq \beta\gamma \tag{20}$$

e o i-ésimo coeficiente $p_i(k)$ do vetor $\boldsymbol{p}(k) = \text{sign}[\boldsymbol{w}(k)]$ armazena o sinal do i-ésimo coeficiente adaptativo no instante k:

$$p_i(k) = \text{sign}[w_i(k)] = \begin{cases} 1, & \text{se } w_i(k) > 0 \\ 0, & \text{se } w_i(k) = 0 \\ -1, & \text{se } w_i(k) < 0 \end{cases}. \tag{21}$$

Assim, a equação de atualização do algoritmo ZA-LMS

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta\boldsymbol{x}(k)e(k) - \kappa\text{sign}[\boldsymbol{w}(k)], \tag{22}$$

tem uma interpretação intuitiva, já que o termo adicional $-\kappa\text{sign}[\boldsymbol{w}(k)]$ sempre impõe uma atração para zero. Tal atração, embora possa ser vantajosa em termos de taxa de convergência quando o sistema a identificar é esparso, dá origem a um estimador enviesado [31], já que atrai para zero mesmo os coeficientes de alta magnitude.

Uma forma de contornar este viés assintótico reside no apelo a uma função custo

[2] A reescrita do produto $\beta\gamma$ por κ é um procedimento comum nos algoritmos com restrição de norma, o qual também será utilizado nas derivações dos próximos algoritmos.

reponderada, adotada pelo algoritmo *RZA-LMS* [31]:

$$\mathcal{F}_{\mathsf{RZA}}\left[\boldsymbol{w}(k)\right] = \frac{1}{2}e^2(k) + \gamma \sum_{i=0}^{N-1} \log\left(1 + |w_i(k)|/\epsilon'\right), \tag{23}$$

onde ϵ é um parâmetro ajustável. A penalidade $\sum_{i=0}^{N-1} \log\left(1 + |w_i(k)|/\epsilon'\right)$ se comporta de modo mais próximo ao da norma[3] ℓ_0 do que a norma ℓ_1 [46].

A otimização via gradiente estocástico de (23) nos fornece a equação de atualização do algoritmo RZA-LMS:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \boldsymbol{x}(k)e(k) - \kappa\overline{\boldsymbol{p}}(k), \tag{24}$$

onde o i-ésimo elemento $\overline{p}_i(k)$ de $\overline{\boldsymbol{p}}(k)$ é descrito por

$$\overline{p}_i(k) = \frac{\mathsf{sign}[w_i(k)]}{1 + \epsilon|w_i(k)|}, \tag{25}$$

onde $\epsilon \triangleq 1/\epsilon'$. O vetor adicional $\overline{\boldsymbol{p}}(k)$ adotado pelo algoritmo RZA-LMS apresenta uma redução de sua magnitude quando a magnitude do coeficiente adaptativo é alta, o que mitiga o viés dos coeficientes de alta magnitude, mantendo o termo de atração para zero dos termos de baixa magnitude praticamente intacto.

Já o algoritmo ℓ_0-LMS (foco desta tese) pretende minimizar uma aproximação da seguinte função custo [14]:

$$\mathcal{F}_{\mathsf{ZA}}\left[\boldsymbol{w}(k)\right] \approx \frac{1}{2}e^2(k) + \gamma\|\boldsymbol{w}(k)\|_0, \tag{26}$$

onde a aproximação decorre do fato de a minimização da norma ℓ_0 ser um problema NP-difícil. Assim, utiliza-se uma aproximação da norma ℓ_0 sob a qual, aplicando-se a otimização via gradiente estocástico e utilizando-se uma expansão de Taylor, dá origem à equação de atualização do algoritmo ℓ_0-LMS:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta e(k)\boldsymbol{x}(k) + \kappa\boldsymbol{g}[\boldsymbol{w}(k)], \tag{27}$$

[3] A norma ℓ_0 de um vetor afere o número de seus elementos não nulos.

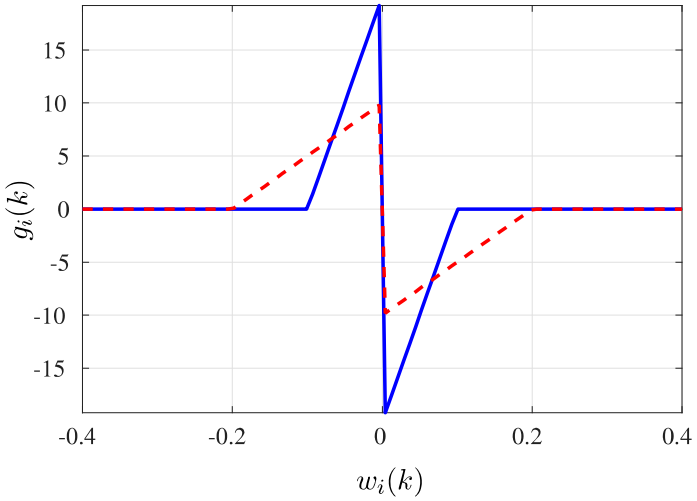


Figura 2 – Função $g_i(k)$ em termos de $w_i(k)$. Na curva em azul tem-se que α = 10 e em vermelho α = 5.

sendo o i-ésimo elemento $g_i(k) \in \mathbb{R}$ do vetor $\boldsymbol{g}[\boldsymbol{w}(k)] \in \mathbb{R}^N$ igual a

$$g_i(k) = \begin{cases} 2\alpha^2 w_i(k) - 2\alpha \text{sign}[w_i(k)], & |w_i(k)| \leq 1/\alpha \\ 0, & \text{caso contrário} \end{cases}, \tag{28}$$

onde $\alpha \in \mathbb{R}_+$ é um parâmetro ajustável, o qual controla o intervalo $[-1/\alpha, 1/\alpha]$ no qual existe a atração para zero. O fato de o algoritmo ℓ_0-LMS somente atrair para zero os coeficientes que se encontram no intervalo $[-1/\alpha, 1/\alpha]$ elimina o problema do viés assintótico na estimação dos coeficientes de alta magnitude. Uma representação disto pode ser notada na Figura 2, na qual observa-se para valores maiores de α (curva azul), menor é o intervalo de atração dos coeficientes. Essa é uma razão pelas quais o algoritmo ℓ_0-LMS é extremamente popular e gerou vasto interesse na comunidade de pesquisadores, tendo sido alvo de inúmeras publicações[4]. Cabe ressaltar também que [47] efetuou uma comparação das abordagens proporcional e com penalização de norma, revelando que esta última apresenta um compromisso melhor entre taxa de convergência e desempenho assintótico.

A Tabela 2 expõe algumas referências relevantes no atual estado da arte acerca do emprego da norma ℓ_0 em filtragem adaptativa.

[4]No IEEEXplore, o artigo conta com 375 citações. Consulta realizada em 22 de abril de 2025.

Tabela 2 – Referências de relevo que citam o trabalho que propõe o algoritmo ℓ_0-LMS.

Referência	Contribuição
[48]	Algoritmo ℓ_0-*Proportionate Recursive Least Squares (PRLS)* utilizado em comunicações acústicas submersas
[49]	Compressão de redes neurais convolucionais através da regularização do ℓ_0
[50]	Algoritmo ℓ_0-LMS com tamanho de passo múltiplo variável
[18]	Análise de transiente do algoritmo ℓ_0-LMS
[51]	Extração cega de componentes esparsos baseado na minimização da norma ℓ_0
[52]	Limiar suavizado da norma ℓ_0 para recuperação esparsa acelerada
[53]	Filtros Volterra adaptativos de norma ℓ_0
[54]	Desconvolução de aproximação rápida da norma ℓ_0 usando estrutura Wavelet
[55]	Método de super-resolução ustilizando filtragem adaptativa e norma ℓ_0 duplamente suavizada

2.3- Considerações Finais

Neste capítulo, alguns tópicos fundamentais para a pesquisa a ser empreendida foram mobilizados, como identificação de sistemas, filtragem adaptativa e algoritmos adaptativos conscientes da esparsidade. Esta tese tem por objetivo desenvolver modelos estocásticos inéditos para predição do desempenho do algoritmo ℓ_0-LMS no contexto de plantas não estacionárias e prever o desempenho de uma versão generalizada do algoritmo ℓ_0-LMS. A metodologia escolhida para análise é a proposta em [19], porque fornece soluções analíticas para o desempenho assintótico do algoritmo. No próximo capítulo, esta técnica será utilizada de maneira inédita para prever o MSD assintótico do algoritmo ℓ_0-LMS num cenário de identificação de sistemas no qual a planta a emular é variante no tempo.

3- Análise Assintótica de Rastreamento do Algoritmo ℓ_0-LMS

3.1- Introdução

Neste capítulo, será introduzido um modelo estocástico inédito capaz de prever o desempenho assintótico do algoritmo ℓ_0-LMS num cenário de não estacionariedade. Como na configuração sob análise o vetor de coeficientes ótimo se altera ao longo do tempo, podemos dizer que estamos analisando a *capacidade de rastreabilidade* do algoritmo ℓ_0-LMS. Este tipo de análise é importante na oferta de garantias de desempenho ou de estabilidade, porque algoritmos de filtragem adaptativa que apresentam um bom comportamento em ambientes estacionários não necessariamente apresentam uma boa habilidade de rastreamento em ambientes com plantas variantes no tempo [56]. Segundo [57], o rastreamento é o fator chave dos algoritmos adaptativos de todas as modalidades, pois a não estacionariedade reconhecidamente traz uma espécie de "atraso" no processo adaptativo de emulação do vetor desconhecido [58], característica sob a qual valores muito pequenos de β (passo de adaptação do algoritmo ou fator de aprendizagem) podem implicar redução do desempenho assintótico, já que um filtro com menor adaptabilidade tem menor capacidade de rastreamento. Em termos gerais, o desempenho assintótico advindo do gradiente estocástico se deteriora com o *aumento* de β, enquanto que o desempenho assintótico induzido pela variação temporal da planta desconhecida se deteriora com a *redução* de β [59], de modo que não raro existe um passo de aprendizagem que otimiza o desempenho assintótico na configuração não estacionária. Este tipo de fenômeno não existe na configuração estacionária, sob a qual o desempenho assintótico progressivamente melhora com a redução de β (com a desvantagem de a taxa de convergência se tornar progressivamente menor), já que apenas o ruído advindo do gradiente estocástico encontra-se presente quando não há variação no filtro ótimo a estimar. Como cenários não estacionários são comuns na prática, algoritmos com fatores de passo deterministicamente decrescentes ao longo do tempo (que otimizariam o desempenho num contexto estacionário) têm uso limitado na prática [60].

Em resumo, o propósito da análise de rastreamento de algoritmos de filtragem adaptativa está na caracterização de sua habilidade de rastreamento em ambientes não estacionários [61]. Iremos assumir que o vetor de coeficientes ótimo varia no tempo de acordo com a equação de passo aleatório, que sob algumas hipóteses estatísticas equivale a um modelo markoviano de primeira ordem, descrita em [61]:

$$\boldsymbol{w}^{\star}(k+1) = \boldsymbol{w}^{\star}(k) + \boldsymbol{q}(k), \tag{29}$$

onde $\boldsymbol{q}(k)$ é o vetor de ruído branco aleatório.

Tal modelo é encontrado em algumas aplicações realísticas, tais como sistemas de transmissão, processamento de sinais acústicos e cancelamento de eco acústico [62]–[65], não sendo portanto adotado apenas por mera conveniência analítica.

3.2- Análise de Primeira Ordem

Como dito no Capítulo 1, iremos aderir a técnica de modelagem estocástica preconizada por [19]. Devido à grande sofisticação do comportamento de aprendizado dos algoritmos de filtragem adaptativa, todos os modelos utilizam hipóteses simplificadoras (mesmo para o LMS, algoritmo mais simples), e nossa análise da capacidade de rastreamento do algoritmo ℓ_0-LMS não será diferente.

A análise de primeira ordem almeja identificar o comportamento médio dos coeficientes adaptativos. Seja a equação de atualização do algoritmo ℓ_0-LMS, aqui reproduzida por conveniência:[1]

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta e(k)\boldsymbol{x}(k) + \kappa \boldsymbol{g}[\boldsymbol{w}(k)], \tag{30}$$

a qual pretende manter próximo de $\boldsymbol{w}^{\star}(k)$ a estimativa $\boldsymbol{w}(k)$. Tal proximidade pode ser matematicamente capturada no vetor de desvios $\tilde{\boldsymbol{w}}(k)$ na k-ésima iteração:

$$\tilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}^{\star}(k) - \boldsymbol{w}(k). \tag{31}$$

Por meio de manipulações matemáticas, a equação de atualização (30) pode ser

[1] Note que κ é o parâmetro que controla a intensidade do termo de regularização, promovendo a esparsidade.

reescrita em termos do vetor de desvios:

$$\tilde{\boldsymbol{w}}(k+1) = \tilde{\boldsymbol{w}}(k) - \beta e(k)\boldsymbol{x}(k) - \kappa \boldsymbol{g}[\boldsymbol{w}(k)] + \boldsymbol{q}(k), \tag{32}$$

onde o modelo de variação temporal do filtro ótimo (29) foi utilizado.

Utilizando o modelo de regressão linear ruidosa (1) para o sinal desejado, podemos também expressar o sinal de erro em termos do vetor de desvio:

$$e(k) = \tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k) + \nu(k), \tag{33}$$

de modo que (32) pode ser reescrita como:

$$\tilde{\boldsymbol{w}}(k+1) = \left[\boldsymbol{I} - \beta \boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] \tilde{\boldsymbol{w}}(k) - \beta\nu(k)\boldsymbol{x}(k) - \kappa \boldsymbol{g}[\boldsymbol{w}(k)] + \boldsymbol{q}(k), \tag{34}$$

uma equação a diferenças estocástica variante no tempo cujo comportamento é bastante complexo de ser analisado [60]. Entretanto, ela nos permite estabelecer a seguinte recursão em termos de valores *médios* do seguinte modo:

$$\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] = \mathbb{E}\left\{\left[\boldsymbol{I} - \beta \boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] \tilde{\boldsymbol{w}}(k)\right\} - \beta\mathbb{E}\left[\nu(k)\boldsymbol{x}(k)\right] - \kappa\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\right\} + \mathbb{E}\left[\boldsymbol{q}(k)\right], \tag{35}$$

com as constantes β e κ podendo "sair" do operador de valor esperado devido à sua linearidade.

O resultado (35) é bastante complexo para o cômputo teórico da evolução do vetor médio $\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right]$, podendo ser simplificado mediante o concurso de hipóteses estatísticas simplificadoras. A primeira delas é relativa ao vetor aleatório de perturbação $\boldsymbol{q}(k)$:

$\mathbb{H}_{\boldsymbol{q}}$ (Hipótese referente à perturbação $\boldsymbol{q}(k)$): o vetor $\boldsymbol{q}(k) \in \mathbb{R}^N$ é de média zero, independente e identicamente distribuído, com matriz de autocorrelação

$$\boldsymbol{R}_{\boldsymbol{q}}(k) = \mathbb{E}\left[\boldsymbol{q}(k)\boldsymbol{q}^T(k)\right] = \sigma_q^2 \boldsymbol{I}_N, \tag{36}$$

onde $\boldsymbol{I}_N$ denota a matriz identidade de dimensões $N \times N$ e σ_q^2 denota a variância do ruído.

Comentário: $\mathbb{H}_{\boldsymbol{q}}$ é extensamente utilizada na análise de rastreamento de algoritmos de filtragem adaptativa [59], [66]–[68], sendo válida em diversas aplicações realísticas [62]–[65].

Outra hipótese que utilizaremos é relacionada ao ruído de medição $\nu(k)$ presente no sinal de referência $d(k)$ (segundo (1)):

$\mathbb{H}_\nu$ (Hipótese referente ao ruído $\nu(k)$): o sinal $\nu(k)$ é de média zero, é estatisticamente independente das demais variáveis aleatórias, com suas amostras sendo estatisticamente independentes e indenticamente distribuídas (i.i.d.).

Comentário: Embora $\mathbb{H}_\nu$ não seja válida quando fenômenos físicos não lineares (tais como saturação e intermodulação [66]) encontram-se presentes, trata-se de uma hipótese no mais das vezes fisicamente plausível e ubíqua na literatura referente a algoritmos de filtragem adaptativa [18], [69], sendo adotada mesmo em simulações de identificação de algoritmos adaptativos que *não* utilizam a estrutura transversal (como exemplo, [70] utiliza um sinal de ruído que adere a $\mathbb{H}_\nu$ com uma estrutura de identificação *Infinite Impulse Response (IIR)*). Cabe notar que a referência [71] efetua uma análise chamada de esperança exata sem utilizar $\mathbb{H}_\nu$, a qual entretanto só é factível para filtros de comprimento pequeno. Por fim, [72] ressalta que $\mathbb{H}_\nu$ assume a correção do modelo linear, assim como o fato de que o comprimento da solução ótima de Wiener para a estrutura transversal coincide com a ordem da planta desconhecida.

A adoção de $\mathbb{H}_{\boldsymbol{q}}$ e de $\mathbb{H}_\nu$ permite-nos simplificar (35), dando origem a

$$\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] = \mathbb{E}\left\{\left[\boldsymbol{I} - \beta \boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\tilde{\boldsymbol{w}}(k)\right\} - \kappa\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\right\}, \tag{37}$$

equação cujo termo $\mathbb{E}\left\{\left[\boldsymbol{I} - \beta \boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\tilde{\boldsymbol{w}}(k)\right\}$ exige o cômputo de momentos conjuntos entre o sinal de entrada $x(k)$ e os coeficientes adaptativos. Tal cômputo pode ser bastante facilitado com a introdução de uma nova hipótese:

$\mathbb{H}_{\mathbf{I}}$ (Hipótese da independência). O vetor $\boldsymbol{x}(k)$ é estatisticamente independente de $\boldsymbol{w}(k)$ (e, por conseguinte, de $\tilde{\boldsymbol{w}}(k)$).

Comentários: A hipótese da independência atraiu muita atenção a partir de [58], sendo quase onipresente nas análises teóricas de algoritmos de filtragem adaptativa. Importa ressaltar que tal hipótese não é, ao contrário das demais, fisicamente plausível. Ela costuma ser adotada por tratabilidade matemática, já que em estruturas transversais (como as analisadas nesta tese), vetores de entrada $\boldsymbol{x}(k)$ em instantes sucessivos compartilham quase a totalidade das amostras em comum (ainda que deslocadas), caracterizando o que [73] denomina de "coerência determinística". Assim, o vetor $\boldsymbol{x}(k)$ contém diversas amostras que foram utilizadas nas iterações anteriores para atualizar o

filtro adaptativo, de modo que é inexato assumir independência estatística entre $\boldsymbol{x}(k)$ e $\boldsymbol{w}(k)$. Por essa razão, em [60] tal hipótese é chamada de *heurística da independência*. Ela geralmente produz predições acuradas para valores baixos de β, o que pode ser justificado pelo fato de que a estimação empreendida pelo algoritmo de filtragem adaptativa atua como um filtro passa-baixas em relação ao sinal de entrada [72].

Assim, acrescentando à aplicação de $\mathbb{H}_{\boldsymbol{q}}$ e $\mathbb{H}_{\nu}$ a aproximação $\mathbb{H}_{\mathbf{I}}$, a recursão (37) se reduz a

$$\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] = \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right] - \beta \boldsymbol{R}_{\boldsymbol{x}} \tilde{\boldsymbol{w}}(k) - \kappa \mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\right\}, \tag{38}$$

onde

$$\boldsymbol{R}_{\boldsymbol{x}} \triangleq \mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] \tag{39}$$

é a matriz de autocorrelação do vetor de entrada $\boldsymbol{x}(k)$.

Para facilitar o encontro de fórmulas analíticas, iremos classificar os coeficientes $w_i^\star(0) \in \mathbb{R}$ da planta ótima no instante inicial $\boldsymbol{w}^\star(0) \in \mathbb{R}^N$ em três grupos:

$$\mathcal{C}_L = \{i | 1/\alpha \leq |w_i^\star(0)|\} \text{ (coeficientes de alta magnitude)}, \tag{40}$$

$$\mathcal{C}_S = \{i | 0 < |w_i^\star(0)| < 1/\alpha\} \text{ (coeficientes de baixa magnitude)}, \tag{41}$$

$$\mathcal{C}_0 = \{i | w_i^\star(0) = 0\} \text{ (coeficientes nulos)}, \tag{42}$$

Doravante, Q será o número de coeficientes não nulos (isto é, $Q = |\mathcal{C}_L \cup \mathcal{C}_S|$). As esperanças estatísticas com a não linearidade $g[\cdot]$ serão simplificadas mediante recurso à seguinte hipótese, introduzida em [19]:

$\mathbb{H}_g$. (Hipótese referente à função de atração $g[\cdot]$). Para todo $i \in \mathcal{C}_L \cup \mathcal{C}_S$, o coeficiente $w_i(k)$ tem o mesmo sinal que seu valor ótimo correspondente (isto é, $w_i^\star(k)$). Ademais, assume-se que ele está fora do termo de atração $g[\cdot]$ quando $i \in \mathcal{C}_L$, estando sob atração caso contrário.

Comentário: $\mathbb{H}_g$ foi uma inovação da metodologia de [19], a qual estamos aplicando à configuração não estacionária. Ela pode ser justificada caso assumamos valores baixos de β e o regime permanente, porque neste caso os coeficientes adaptativos oscilam pouco em torno de seus valores médios. Como o termo de atração é restrito ao intervalo $[-1/\alpha, 1/\alpha]$ (vide (28)), tal hipótese não é muito forte. Ademais, cabe ressaltar que a análise neste capítulo é restrita ao regime permanente, situação na qual o vetor adaptativo encontra-se próximo de seus valores ótimos. Por fim, num contexto de

perturbação markoviana (veja a Equação (29)), tal hipótese se manteria válida caso a variância da perturbação σ_q^2 fosse suficientemente pequena. Caso esta variação seja elevada, possivelmente o filtro adaptativo (que atua num aprendizado amostra a amostra) dificilmente conseguiria ser útil.

A aplicação de $\mathbb{H}_g$ em (38), assumindo que o algoritmo converge em regime permanente, dá origem ao resultado seguinte:

$$\lim_{k\to\infty} \mathbb{E}\left[\tilde{w}_i(k)\right] = \begin{cases} -\frac{\kappa}{\beta\sigma_x^2}\mathbb{E}\left\{g\left[w_i(k)\right]\right\}, & i \in \mathcal{C}_S \\ 0, & \text{caso contrário} \end{cases}, \tag{43}$$

o que significa que o algoritmo não introduz um viés na estimativa assintótica dos coeficientes nulos, tampouco nos de alta magnitude (desvantagem do algoritmo ZA-LMS abordado no Capítulo 2), embora introduza um viés na estimativa dos coeficientes pertencentes a $\mathcal{C}_S$. Este viés é contrabalançado por uma menor variabilidade na estimativa, o que explica parcialmente o melhor desempenho do algoritmo ℓ_0-LMS caso comparado a uma versão agnóstica com relação à esparsidade. Cabe ressaltar também que este viés somente ocorre para os coeficientes $w_i(k)$ com $i \in \mathcal{C}_S$, sendo os demais coeficientes estimados de modo não polarizado.

O resultado (43) é equivalente à Equação (14) do artigo [19]. A diferença de sinal deriva do fato de definirmos o desvio de modo distinto. Logo, podemos concluir que, sob a perturbação markoviana (29), o comportamento de primeira ordem do algoritmo é idêntico ao do caso estacionário abordado em [19]. O mesmo entretanto não ocorre na análise de segunda ordem, alvo da próxima seção.

3.3- Análise de Segunda Ordem

Quando abordamos estatísticas de segunda ordem dos coeficientes adaptativos, a análise é bem mais complexa. Ao multiplicarmos (118) por sua transposta, obtemos

$$\begin{aligned}\tilde{\boldsymbol{w}}(k+1)\tilde{\boldsymbol{w}}^T(k+1) &= \left[\boldsymbol{I}-\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\left[\boldsymbol{I}-\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\\ &\quad -\beta\nu(k)\left[\boldsymbol{I}-\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\tilde{\boldsymbol{w}}(k)\boldsymbol{x}^T(k)\\ &\quad -\kappa\left[\boldsymbol{I}-\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\tilde{\boldsymbol{w}}(k)\boldsymbol{g}^T\left[\boldsymbol{w}(k)\right]\\ &\quad -\beta\nu(k)\boldsymbol{x}(k)\tilde{\boldsymbol{w}}^T(k)\left[\boldsymbol{I}-\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\\ &\quad +\beta^2\nu^2(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\\ &\quad +\beta\kappa\nu(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\\ &\quad -\kappa\boldsymbol{g}\left[\boldsymbol{w}(k)\right]\tilde{\boldsymbol{w}}^T(k)\left[\boldsymbol{I}-\beta\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\\ &\quad +\beta\kappa\nu(k)\boldsymbol{g}[\boldsymbol{w}(k)]\boldsymbol{x}^T(k)\\ &\quad +\kappa^2\boldsymbol{g}\left[\boldsymbol{w}(k)\right]\boldsymbol{g}^T\left[\boldsymbol{w}(k)\right]\\ &\quad +\mathcal{O}[\boldsymbol{q}(k)]+\boldsymbol{q}(k)\boldsymbol{q}^T(k), \end{aligned} \tag{44}$$

onde $\mathcal{O}[\boldsymbol{q}(k)]$ agrega termos de primeira ordem da perturbação markoviana $\boldsymbol{q}(k)$.

Aplicando o operador de valor esperado em conjunção com $\mathbb{H}_{\boldsymbol{q}}$ e $\mathbb{H}_{\nu}$, nós obtemos:

$$\begin{aligned}\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\tilde{\boldsymbol{w}}^T(k+1)\right] &= \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right]-\beta\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\\ &\quad -\beta\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right]\\ &\quad +\beta^2\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right]\\ &\quad -\kappa\mathbb{E}\left\{\tilde{\boldsymbol{w}}(k)\boldsymbol{g}^T\left[\boldsymbol{w}(k)\right]\right\}+\kappa\beta\mathbb{E}\left\{\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\boldsymbol{g}^T\left[\boldsymbol{w}(k)\right]\right\}\\ &\quad +\beta^2\sigma_\nu^2\boldsymbol{R}_x\\ &\quad -\kappa\mathbb{E}\left\{\boldsymbol{g}\left[\boldsymbol{w}(k)\right]\tilde{\boldsymbol{w}}^T(k)\right\}+\beta\kappa\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right\}\\ &\quad +\kappa^2\mathbb{E}\left\{\boldsymbol{g}\left[\boldsymbol{w}(k)\right]\boldsymbol{g}^T\left[\boldsymbol{w}(k)\right]\right\}+\sigma_q^2\boldsymbol{I}_N \end{aligned} \tag{45}$$

Para reduzir a complexidade das equações resultantes, iremos adotar a seguinte hipótese:

$\mathbb{H}_{\text{BG}}$. (Hipótese referente à brancura do sinal de entrada). O sinal de entrada é branco e gaussiano.

Comentários: $\mathbb{H}_{\text{BG}}$ é uma hipótese restritiva, porque embora diversos sinais de entrada sejam gaussianos, na prática não costumam ser brancos. Ela foi adotada para reduzir a complexidade da análise e facilitar a extração de *insights* acerca dos fatores qualitativos que influenciam o desempenho do algoritmo ℓ_0-LMS. Embora tal hipótese não seja ubíqua na literatura, é bastante comum [74]–[77]. Tal hipótese implica que a matriz de autocorrelação do vetor de entrada $\boldsymbol{x}(k)$ é proporcional à matriz identidade, com uma constante de proporcionalidade igual a σ_x^2, a variância do sinal de entrada.

Utilizando $\mathbb{H}_{\text{BG}}$ em (45) nos leva à seguinte equação:

$$\begin{aligned}
\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k+1) \;=\; & \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) - \beta\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] \\
& +\beta^2\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] \\
& -\beta\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right] - \kappa\mathbb{E}\left\{\tilde{\boldsymbol{w}}(k)\boldsymbol{g}^T\left[\boldsymbol{w}(k)\right]\right\} \\
& +\kappa\beta\mathbb{E}\left\{\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\boldsymbol{g}^T[\boldsymbol{w}(k)]\right\} + \beta^2\sigma_\nu^2\sigma_x^2\boldsymbol{I} \\
& -\kappa\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\tilde{\boldsymbol{w}}^T(k)\right\} + \beta\kappa\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right\} \\
& +\kappa^2\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\boldsymbol{g}^T\left[\boldsymbol{w}(k)\right]\right\} + \sigma_q^2\boldsymbol{I}, \qquad (46)
\end{aligned}$$

onde

$$\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) \triangleq \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right], \qquad (47)$$

é a matriz de autocorrelação dos desvios.

Já que $\mathbb{H}_{\text{BG}}$ foi assumida, podemos empregar o teorema da fatoração de momentos gaussianos [78], o qual pode ser expresso como:

$$\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\boldsymbol{A}\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] = 2\boldsymbol{\Lambda}\boldsymbol{A}\boldsymbol{\Lambda} + \boldsymbol{\Lambda}\text{Tr}\left\{\boldsymbol{\Lambda}\boldsymbol{A}\right\}, \qquad (48)$$

onde assume-se a decomposição espectral[2] $\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$.

A aplicação de (48) permite fazer a seguinte decomposição:

$$\begin{aligned}
\beta^2\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] &= 2\beta^2\sigma_x^4\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) + \beta^2\sigma_x^4\text{Tr}\left\{\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right\}\boldsymbol{I} \\
&= 2\beta^2\sigma_x^4\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) + \beta^2\sigma_x^4 D(k)\boldsymbol{I}. \qquad (49)
\end{aligned}$$

[2]Observe que a decomposição espectral apresentada não assume a brancura do vetor de entrada, mas apenas a sua gaussianidade, sendo portanto mais geral do que o necessário neste momento.

A aplicação de (49) em (46) dá origem a

$$\begin{aligned}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k+1) &= \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) - \beta\sigma_x^2\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) + 2\beta^2\sigma_x^4\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) + \beta^2\sigma_x^4 D(k)\boldsymbol{I} \\ &\quad -\beta\sigma_x^2\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) - \kappa\mathbb{E}\left\{\tilde{\boldsymbol{w}}(k)\boldsymbol{g}^T\left[\boldsymbol{w}(k)\right]\right\} + \beta\kappa\sigma_x^2\mathbb{E}\left\{\tilde{\boldsymbol{w}}(k)\boldsymbol{g}^T[\boldsymbol{w}(k)]\right\} \\ &\quad +\beta^2\sigma_\nu^2\sigma_x^2\boldsymbol{I} - \kappa\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\tilde{\boldsymbol{w}}^T(k)\right\} + \beta\kappa\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\tilde{\boldsymbol{w}}^T(k)\right\} \\ &\quad +\kappa^2\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\boldsymbol{g}^T[\boldsymbol{w}(k)]\right\} + \sigma_q^2\boldsymbol{I}_N, \end{aligned} \tag{50}$$

onde $D(k)$ denota o MSD na k-ésima iteração:

$$D(k) \triangleq \mathbb{E}\left[\|\boldsymbol{w}(k) - \boldsymbol{w}^\star(k)\|^2\right] = \sum_{i=0}^{N-1}\mathbb{E}\left\{\tilde{w}_i^2(k)\right\}. \tag{51}$$

A recursão (50) pode ser equivalentemente descrita como

$$\begin{aligned}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k+1) &= (1 - 2\beta\sigma_x^2 + 2\beta^2\sigma_x^4)\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) + \beta^2\sigma_x^4 D(k)\boldsymbol{I} + \beta^2\sigma_\nu^2\sigma_x^2\boldsymbol{I} \\ &\quad -\kappa\left(1 - \beta\sigma_x^2\right)\mathbb{E}\left\{\tilde{\boldsymbol{w}}(k)\boldsymbol{g}^T\left[\boldsymbol{w}(k)\right]\right\} - \kappa(1 - \beta\sigma_x^2)\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\tilde{\boldsymbol{w}}^T(k)\right\} \\ &\quad +\kappa^2\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\boldsymbol{g}^T[\boldsymbol{w}(k)]\right\} + \sigma_q^2\boldsymbol{I}, \end{aligned} \tag{52}$$

ou

$$\begin{aligned}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k+1) &= (1 - 2\beta\sigma_x^2\Delta_0)\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) + \beta^2\sigma_x^4 D(k)\boldsymbol{I} + \beta^2\sigma_\nu^2\sigma_x^2\boldsymbol{I} \\ &\quad -\kappa\Delta_0\mathbb{E}\left\{\tilde{\boldsymbol{w}}(k)\boldsymbol{g}^T\left[\boldsymbol{w}(k)\right]\right\} - \kappa\Delta_0\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\tilde{\boldsymbol{w}}^T(k)\right\} \\ &\quad +\kappa^2\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\boldsymbol{g}^T[\boldsymbol{w}(k)]\right\} + \sigma_q^2\boldsymbol{I}, \end{aligned} \tag{53}$$

onde

$$\Delta_0 \triangleq 1 - \beta\sigma_x^2. \tag{54}$$

Aplicando o operador de traço em ambos os lados de (53) para obter uma recursão escalar, temos:

$$\begin{aligned}D(k+1) &= (1 - 2\beta\sigma_x^2\Delta_0)D(k) + \beta^2\sigma_x^4 N D(k) + N\beta^2\sigma_\nu^2\sigma_x^2 \\ &\quad -2\kappa\Delta_0\mathbb{E}\left\{\tilde{\boldsymbol{w}}^T(k)\boldsymbol{g}[\boldsymbol{w}(k)]\right\} + \kappa^2\mathbb{E}\left\{\|\boldsymbol{g}[\boldsymbol{w}(k)]\|^2\right\} + N\sigma_q^2, \end{aligned} \tag{55}$$

ou

$$\begin{aligned} D(k+1) &= (1-2\beta\sigma_x^2\Delta_0+\beta^2\sigma_x^4N)D(k)+N\beta^2\sigma_\nu^2\sigma_x^2+N\sigma_q^2 \\ &\quad -2\kappa\Delta_0\mathbb{E}\left\{\tilde{\boldsymbol{w}}^T(k)\boldsymbol{g}[\boldsymbol{w}(k)]\right\}+\kappa^2\mathbb{E}\left\{\|\boldsymbol{g}[\boldsymbol{w}(k)]\|^2\right\}. \end{aligned} \tag{56}$$

Seja a simplificação seguinte:

$$\begin{aligned} 1-2\beta\sigma_x^2+2\beta^2\sigma_x^4+\beta^2\sigma_x^4N &= 1-2\beta\sigma_x^2+\beta^2(N+2)\sigma_x^4 \\ &= 1-\beta\sigma_x^2\left(2-\beta(N+2)\sigma_x^2\right) \\ &= 1-\beta\sigma_x^2\Delta_N, \end{aligned} \tag{57}$$

onde

$$\Delta_N \triangleq \left(2-\beta(N+2)\sigma_x^2\right). \tag{58}$$

Aplicando-se (57) em (56), obtemos:

$$\begin{aligned} D(k+1) &= (1-\beta\sigma_x^2\Delta_N)D(k)+N\beta^2\sigma_\nu^2\sigma_x^2+N\sigma_q^2 \\ &\quad -2\kappa\Delta_0\mathbb{E}\left\{\tilde{\boldsymbol{w}}^T(k)\boldsymbol{g}[\boldsymbol{w}(k)]\right\}+\kappa^2\mathbb{E}\left\{\|\boldsymbol{g}[\boldsymbol{w}(k)]\|^2\right\}, \end{aligned} \tag{59}$$

um resultado importante que será utilizado futuramente. Entretanto, como a função $g[\cdot]$ tem impacto distinto em coeficientes diferentes (devido a $\mathbb{H}_g$), retornemos ao i-ésimo elemento da diagonal principal de $\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)$ em (53):

$$\begin{aligned} \mathbb{E}\left[\tilde{w}_i^2(k+1)\right] &= (1-2\beta\sigma_x^2\Delta_0)\mathbb{E}\left[\tilde{w}_i^2(k)\right]+\beta^2\sigma_x^4D(k)+\beta^2\sigma_\nu^2\sigma_x^2 \\ &\quad -2\kappa\Delta_0\mathbb{E}\left\{\tilde{w}_i(k)g[w_i(k)]\right\}+\kappa^2\mathbb{E}\left\{g^2\left[w_i(k)\right]\right\}+\sigma_q^2. \end{aligned} \tag{60}$$

Neste capítulo, estamos interessados no comportamento assintótico do algoritmo ℓ_0-LMS, região na qual $\mathbb{H}_g$ torna-se razoável. Assim, assumindo que o algoritmo converge e que estamos em regime assintótico, podemos escrever:

$$\mathbb{E}\left[\tilde{w}_i(\infty)\right] \triangleq \lim_{k\to\infty}\mathbb{E}\left[\tilde{w}_i^2(k+1)\right] = \lim_{k\to\infty}\mathbb{E}\left[\tilde{w}_i^2(k)\right], \tag{61}$$

resultado que simplifica (60) para:

$$\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] = \frac{\beta^2\sigma_x^4 D(\infty) + \beta^2\sigma_\nu^2\sigma_x^2 - 2\kappa\Delta_0\mathbb{E}\left\{\tilde{w}_i(\infty)g[w_i(\infty)]\right\} + \kappa^2\mathbb{E}\left\{g^2[w_i(\infty)]\right\} + \sigma_q^2}{2\beta\sigma_x^2\Delta_0}, \tag{62}$$

a qual é generalização da Equação (46) de [19] para a configuração não estacionária sob consideração.

A Equação (62) pode ser particularizada para cada tipo de coeficiente (a depender da classe a qual ele pertence: $\mathcal{C}_0$, $\mathcal{C}_S$, ou $\mathcal{C}_L$), porque os termos $\mathbb{E}\left\{\tilde{w}_i(\infty)g[w_i(\infty)]\right\}$ e $\mathbb{E}\left\{g^2[w_i(\infty)]\right\}$ diferem bastante para cada um destes conjuntos (de acordo com $\mathbb{H}_g$).

Assim, iremos explicitar os valores que estes termos assumem em cada uma das classes de coeficientes, utilizando $\mathbb{H}_g$:

- Para $i \in \mathcal{C}_L$:

$$\mathbb{E}\left\{\tilde{w}_i(\infty)g[w_i(\infty)]\right\} = 0, \tag{63}$$

$$\mathbb{E}\left\{g^2[w_i(\infty)]\right\} = 0, \tag{64}$$

  já que sob $\mathbb{H}_g$ o termo de atração é sempre inativo para os coeficientes de alta magnitude.

- Para $i \in \mathcal{C}_S$:

$$g[w_i(\infty)] = g\left[w_i^\star(\infty) + \tilde{w}_i(\infty)\right] \tag{65}$$

  para $\tilde{w}_i(\infty)$ pequeno (pois assumimos $i \in \mathcal{C}_S$):

$$g[w_i(\infty)] \approx g\left[w_i^\star(\infty)\right] + 2\alpha^2\tilde{w}_i(\infty) \tag{66}$$

  Daí:

$$\begin{aligned}\mathbb{E}\left\{\tilde{w}_i(\infty)g[w_i(\infty)]\right\} &= \mathbb{E}\left\{\tilde{w}_i(\infty)\left\{g[w_i^\star(\infty)] + 2\alpha^2\tilde{w}_i(\infty)\right\}\right\} \\ &\approx g[w_i^\star(0)]\mathbb{E}\left[\tilde{w}_i(\infty)\right] + 2\alpha^2\mathbb{E}\left[\tilde{w}_i^2(\infty)\right],\end{aligned} \tag{67}$$

onde assumimos as seguintes aproximações[3]:

$$\begin{aligned}\mathbb{E}\left[\tilde{w}_i(\infty)g[w_i^\star(\infty)]\right] &\approx \mathbb{E}\left[g[w_i^\star(\infty)]\right]\mathbb{E}\left[\tilde{w}_i(\infty)\right] \approx g[\mathbb{E}\left[w_i^\star(\infty)\right]]\,\mathbb{E}\left[\tilde{w}_i(\infty)\right] \\ &= g[w_i^\star(0)]\mathbb{E}\left[\tilde{w}_i(\infty)\right], \end{aligned} \tag{68}$$

as quais são justificáveis quando σ_q^2 é pequeno.

Comentário: as aproximações (68) são bastante fortes, limitando o escopo desta análise a configurações nas quais σ_q^2 é baixo. Cabe notar entretanto ser extremamente difícil na prática um filtro adaptativo acompanhar uma função de transferência variante no tempo com σ_q^2 alto.

Utilizando as aproximações (68), obtemos:

$$\begin{aligned}\mathbb{E}\left\{g^2\left[w_i(\infty)\right]\right\} &= \mathbb{E}\left\{g^2\left[w_i^\star(\infty)\right] + 4\alpha^2\tilde{w}_i(\infty)g\left[w_i^\star(\infty)\right] + 4\alpha^4\tilde{w}_i^2(\infty)\right\} \\ &\approx g^2\left[w_i^\star(\infty)\right] + 4\alpha^2 g\left[w_i^\star(\infty)\right]\mathbb{E}\left[\tilde{w}_i(\infty)\right] + 4\alpha^4\mathbb{E}\left[\tilde{w}_i^2(\infty)\right], \end{aligned} \tag{69}$$

que é generalização da equação (49) de [19] para o caso não estacionário.

- Para $i \in \mathcal{C}_0$:

$$\mathbb{E}\left[\tilde{w}_i(k)\right] = \overbrace{\mathbb{E}\left[w_i^\star(\infty)\right]}^{=0} - \mathbb{E}\left[w_i(\infty)\right] = -\mathbb{E}[w_i(\infty)] \tag{70}$$

Como $w_i(\infty) \approx -\tilde{w}_i(\infty)$ (assumindo σ_q^2 pequeno), temos que, para $i \in \mathcal{C}_0$, vale:

$$\mathbb{E}\left[w_i^2(\infty)\right] = \mathbb{E}\left[\tilde{w}_i^2(\infty)\right]. \tag{71}$$

Assim, ainda considerando $i \in \mathcal{C}_0$, temos:

$$\begin{aligned}\mathbb{E}\left\{\tilde{w}_i(\infty)g\left[w_i(\infty)\right]\right\} &= -\mathbb{E}\left\{w_i(\infty)g[w_i(\infty)]\right\} \\ &= -\mathbb{E}\left\{2\alpha^2 w_i^2(\infty) - 2\alpha\,\mathsf{sign}\left[w_i(\infty)\right]w_i(\infty)\right\} \\ &= -2\alpha^2\mathbb{E}\left[w_i^2(\infty)\right] - 2\alpha\mathbb{E}\left\{\mathsf{sign}\left[w_i(\infty)\right]w_i(\infty)\right\}. \end{aligned} \tag{72}$$

Para podermos simplificar (72), utilizamos a seguinte hipótese:

$\mathbb{H}_\mathsf{G}$. Os coeficientes adaptativos $w_i(k)$ apresentam distribuição gaussiana.

[3] Estas aproximações não foram necessárias em [19], porque esta referência assumiu uma configuração estacionária.

Comentário: $\mathbb{H}_{\mathsf{G}}$ não é uma hipótese muito utilizada, sendo necessária na análise de alguns algoritmos com funções não lineares na equação de atualização (caso do ℓ_0-LMS). É utilizada geralmente para simplificar a análise (veja, por exemplo, [18], [72]). Não é considerada uma hipótese muito forte, porque pelo teorema central do limite, em regime assintótico os coeficientes adaptativos são o resultado de diversas somas com componentes estocásticos.

Para uma variável aleatória gaussiana, temos, por meio do teorema de Price [79]:

$$\mathbb{E}\left[x\mathsf{sign}(x)\right] = \sqrt{\frac{2}{\pi\sigma_x^2}}\mathbb{E}\left[x^2\right]. \tag{73}$$

Assumindo $\mathbb{H}_{\mathsf{G}}$, podemos aplicar (73) em (72), obtendo:

$$\begin{aligned}
\mathbb{E}\left\{\tilde{w}_i(\infty)g\left[w_i(\infty)\right]\right\} &= -2\alpha^2\mathbb{E}\left\{w_i^2(\infty)\right\} + 2\alpha\sqrt{\frac{2}{\pi\mathbb{E}\left\{w_i^2(\infty)\right\}}}\mathbb{E}\left[w_i^2(\infty)\right] \\
&= -2\alpha^2\mathbb{E}\left\{w_i^2(\infty)\right\} + \frac{4\alpha}{\sqrt{2\pi}}\sqrt{\mathbb{E}\left[w_i^2(\infty)\right]} \\
&= -2\alpha^2\mathbb{E}\left\{\tilde{w}_i^2(\infty)\right\} + \frac{4\alpha}{\sqrt{2\pi}}\sqrt{\mathbb{E}\left[\tilde{w}_i^2(\infty)\right]},
\end{aligned} \tag{74}$$

que é a Equação (50) de [19].

Usando o mesmo raciocínio, temos:

$$\begin{aligned}
\mathbb{E}\left\{g^2\left[w_i(\infty)\right]\right\} &= 4\alpha^4\mathbb{E}\left[w_i^2(\infty)\right] - 8\alpha^3\left\{w_i(\infty)\mathsf{sign}\left[w_i(\infty)\right]\right\} \\
&\quad +4\alpha^2\overbrace{\mathsf{sign}^2[w_i(\infty)]}^{=1} \\
&= 4\alpha^4\mathbb{E}\left[w_i^2(\infty)\right] - 8\alpha^3\sqrt{\frac{2}{\pi\mathbb{E}\left[w_i^2(\infty)\right]}}\mathbb{E}[w_i^2(\infty)] + 4\alpha^2 \\
&= 4\alpha^4\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] - \frac{16\alpha^3}{\sqrt{\pi}}\sqrt{\mathbb{E}\left[\tilde{w}_i^2(\infty)\right]} + 4\alpha^2 \\
&= 4\alpha^4\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] - \frac{16\alpha^3}{\sqrt{2\pi}}\sqrt{\mathbb{E}\left[\tilde{w}_i^2(\infty)\right]} + 4\alpha^2,
\end{aligned} \tag{75}$$

que é a equação (51) de [19].

Com estes resultados, podemos retornar à Equação (62), fazendo as análises para os diferentes conjuntos $\mathcal{C}_S$, $\mathcal{C}_L$ e $\mathcal{C}_0$:

- $i \in \mathcal{C}_L$: Partindo de (62):

$$2\beta\sigma_x^2\Delta_0\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] - \beta^2\sigma_x^4 D(\infty) - \beta^2\sigma_\nu^2\sigma_x^2 - \sigma_q^2 = 0. \tag{76}$$

- $i \in \mathcal{C}_S$

  Partindo de (62):

$$\begin{aligned}
2\beta\sigma_x^2\Delta_0\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] &= \beta^2\sigma_x^4 D(\infty) + \beta^2\sigma_\nu^2\sigma_x^2 - 2\kappa\Delta_0\mathbb{E}\left\{\tilde{w}_i(\infty)g[w_i(\infty)]\right\} \\
&\quad +\kappa^2\mathbb{E}\left\{g^2[w_i(\infty)]\right\} + \sigma_q^2 \\
&= \beta^2\sigma_x^4 D(\infty) + \beta^2\sigma_\nu^2\sigma_x^2 + \sigma_q^2 \\
&\quad -2\kappa\Delta_0\left\{g[w_i^\star(0)]\mathbb{E}\left[\tilde{w}_i(k)\right] + 2\alpha^2\mathbb{E}\left[\tilde{w}_i^2(k)\right]\right\} \\
&\quad +\kappa^2\left\{g^2[w_i^\star(0)] + 4\alpha^2 g[w_i^\star]\mathbb{E}\left[\tilde{w}_i(k)\right] + 4\alpha^4\mathbb{E}\left[\tilde{w}_i^2(k)\right]\right\} \\
&= \beta^2\sigma_x^4 D(\infty) + \beta^2\sigma_\nu^2\sigma_x^2 + \sigma_q^2 \\
&\quad -2\kappa\Delta_0 g[w_i^\star(0)]\mathbb{E}\left[\tilde{w}_i(k)\right] - 4\alpha^2\kappa\Delta_0\mathbb{E}\left[\tilde{w}_i^2(k)\right] \\
&\quad +\kappa^2 g^2[w_i^\star] + 4\alpha^2\kappa^2 g[w_i^\star]\mathbb{E}\left[\tilde{w}_i(k)\right] + 4\alpha^4\kappa^2\mathbb{E}\left[\tilde{w}_i^2(k)\right].
\end{aligned} \tag{77}$$

Para simplificar este resultado, iremos lançar mão da seguinte hipótese:

$\mathbb{H}_\kappa$ (Hipótese referente ao parâmetro κ). O parâmetro κ é suficientemente pequeno, de modo que potências de κ com expoente igual a 2 (ou superior) podem ser desprezadas. Adicionalmente, $2\alpha^2\kappa \ll \beta\sigma_x^2$.

Comentário: Esta hipótese é justificada em [19] com o argumento de valores altos de κ podem introduzir viés significativo na estimativa dos coeficientes, deteriorando o MSD assintótico. Por essa razão, κ tem que assumir valores pequenos, na prática.

Assumindo $\mathbb{H}_\kappa$, vários termos da Equação (77) podem ser desprezados, gerando a seguinte versão simplificada:

$$\begin{aligned}
2\beta\sigma_x^2\Delta_0\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] &= \beta^2\sigma_x^4 D(\infty) + \beta^2\sigma_\nu^2\sigma_x^2 + \sigma_q^2 \\
&\quad -2\kappa\Delta_0 g[w_i^\star]\mathbb{E}\left[\tilde{w}_i(k)\right] + \kappa^2 g^2[w_i^\star].
\end{aligned} \tag{78}$$

Como estamos assumindo $i \in \mathcal{C}_S$, temos:

$$\lim_{k\to\infty} \mathbb{E}\left[\tilde{w}_i(k)\right] = -\frac{\kappa}{\beta\sigma_x^2}\mathbb{E}\left\{g[w_i(\infty)]\right\} \approx -\frac{\kappa}{\beta\sigma_x^2}\left\{g[w_i^\star(0)] + 2\alpha^2\mathbb{E}\left[\tilde{w}_i(\infty)\right]\right\}. \quad (79)$$

Logo:

$$\begin{aligned} 2\beta\sigma_x^2\Delta_0\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] &= \beta^2\sigma_x^4 D(\infty) + \beta^2\sigma_\nu^2\sigma_x^2 + \sigma_q^2 \\ &\quad + \frac{2\kappa^2\Delta_0}{\beta\sigma_x^2} g^2[w_i^\star] + \kappa^2 g^2[w_i^\star], \end{aligned} \quad (80)$$

onde utilizamos a aproximação $4\kappa^2\alpha^2\Delta_0 g[w_i^\star(\infty)]\mathbb{E}\left[\tilde{w}_i(\infty)\right] \approx 0$, baseados em $\mathbb{H}_\kappa$.

Daí, encontramos:

$$2\beta\sigma_x^2\Delta_0\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] - \beta^2\sigma_x^4 D(\infty) - \beta^2\sigma_\nu^2\sigma_x^2 - \kappa^2\left(1 + \frac{2\Delta_0}{\beta\sigma_x^2}\right) g^2[w_i^\star(0)] - \sigma_q^2 = 0. \quad (81)$$

Como $\Delta_0 = 1 - \beta\sigma_x^2$, temos:

$$2\beta\sigma_x^2\Delta_0\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] - \beta^2\sigma_x^4 D(\infty) - \beta^2\sigma_\nu^2\sigma_x^2 - \kappa^2\left(1 + \frac{2}{\beta\sigma_x^2} - 2\right) g^2[w_i^\star(0)] - \sigma_q^2 = 0. \quad (82)$$

Logo:

$$2\beta\sigma_x^2\Delta_0\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] - \beta^2\sigma_x^4 D(\infty) - \beta^2\sigma_\nu^2\sigma_x^2 - \kappa^2\left(\frac{2}{\beta\sigma_x^2} - 1\right) g^2[w_i^\star] - \sigma_q^2 = 0. \quad (83)$$

que é a Equação (53) de [19], agora generalizada para o caso não estacionário.

- $i \in \mathcal{C}_0$

Já no caso dos coeficientes nulos, partindo de (62), obtemos:

$$\begin{aligned} 2\beta\sigma_x^2\Delta_0\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] &= \beta^2\sigma_x^4 D(\infty) + \beta^2\sigma_\nu^2\sigma_x^2 - 2\kappa\Delta_0\mathbb{E}\left\{\tilde{w}_i(\infty)g[w_i(\infty)]\right\} \\ &\quad + \kappa^2\mathbb{E}\left\{g^2\left[w_i(\infty)\right]\right\} + \sigma_q^2. \end{aligned} \quad (84)$$

Já que todos os desvios quadráticos $\mathbb{E}\left[\tilde{w}_i^2(k)\right]$ são iguais (devido a $\mathbb{H}_{\text{BG}}$), podemos definir

$$\omega \triangleq \lim_{k\to\infty} \sqrt{\mathbb{E}\left[\tilde{w}_i^2(k)\right]}, \quad (85)$$

identidade que nos leva à seguinte equação quadrática em ω:

$$\left[2\beta\sigma_x^2\Delta_0 - 4\kappa\Delta_0\alpha^2 - 4\kappa^2\alpha^4 + \frac{16\alpha^3\kappa^2}{\sqrt{2\pi}}\right]\omega^2$$
$$+8\frac{\alpha\kappa\Delta_0}{\sqrt{2\pi}}\omega - \beta^2\sigma_x^4 D(\infty) - \beta^2\sigma_\nu^2\sigma_x^2 - 4\alpha^2\kappa^2 - \sigma_q^2 = 0. \tag{86}$$

Como $2\alpha^2\kappa \ll \beta\sigma_x^2$ (já que adotamos $\mathbb{H}_\kappa$), temos que

$$2\left(\beta\sigma_x^2\Delta_0 - 2\kappa\alpha^2\Delta_0 - 2(\alpha^2\kappa)^2 + \frac{8\alpha\kappa(\alpha^2\kappa)}{\sqrt{2\pi}}\right) \approx 2\beta\sigma_x^2\Delta_0. \tag{87}$$

Logo, obtemos o seguinte resultado:

$$2\beta\sigma_x^2\Delta_0\omega^2 + \frac{8\alpha\kappa\Delta_0}{\sqrt{2\pi}}\omega - \beta^2\sigma_x^4 D(\infty) - \beta^2\sigma_\nu^2\sigma_x^2 - 4\alpha^2\kappa^2 - \sigma_q^2 = 0, \tag{88}$$

o qual generaliza a Equação (54) de [19] para o caso não estacionário.

Fazendo o somatório em $\mathcal{C}_L \cup \mathcal{C}_S$, temos:

$$2\beta\sigma_x^2\Delta_0 \sum_{i\in\mathcal{C}_L\cup\mathcal{C}_S} \mathbb{E}\left[\tilde{w}_i^2(\infty)\right] - \beta^2\sigma_x^4 Q D(\infty) - \beta^2\sigma_\nu^2\sigma_x^2 Q$$
$$-\kappa^2\left(\frac{2}{\beta\sigma_x^2} - 1\right)G^\star - Q\sigma_q^2 = 0, \tag{89}$$

onde

$$Q \triangleq |\mathcal{C}_S \cup \mathcal{S}_L| \tag{90}$$

é o número de coeficientes não nulos de $\boldsymbol{w}^\star(0)$ e

$$G^\star \triangleq \sum_{i\in\mathcal{C}_S} g^2[w_i^\star(0)]. \tag{91}$$

Por meio de $\mathbb{H}_{\text{BG}}$, temos que

$$\sum_{i\in\mathcal{C}_L\cup\mathcal{C}_S} \mathbb{E}\left[\tilde{w}_i^2(\infty)\right] = D(\infty) - (N-Q)\omega^2, \tag{92}$$

de modo que a Equação (89) se torna

$$\begin{aligned} D(\infty) &= \frac{2(N-Q)\Delta_0}{2\Delta_0-\beta Q\sigma_x^2}\omega^2+\frac{\beta Q\sigma_\nu^2}{2\Delta_0-\beta Q\sigma_x^2}+\frac{\kappa^2\left(\frac{2}{\beta\sigma_x^2}-1\right)G^\star}{2\beta\sigma_x^2\Delta_0-\beta^2\sigma_x^4 Q} \\ &\quad +\frac{Q\sigma_q^2}{2\beta\sigma_x^2\Delta_0-\beta^2\sigma_x^4 Q}, \end{aligned} \tag{93}$$

ou

$$D(\infty) = \frac{2\Delta_0(N-Q)}{\Delta_Q}\omega^2+\frac{\beta Q\sigma_\nu^2}{\Delta_Q}+\frac{\kappa^2\Delta_0' G^\star}{\beta^2\sigma_x^4\Delta_Q}+\frac{Q\sigma_q^2}{\beta\sigma_x^2\Delta_Q}. \tag{94}$$

Sabendo que

$$\begin{aligned} 2\beta\sigma_x^2\Delta_0-\beta^2\sigma_x^4 Q &= \beta\sigma_x^2(2\Delta_0-2\beta\sigma_x^2 Q)=\beta\sigma_x^2(2(1-\beta\sigma_x^2)-\beta\sigma_x^2 Q) \\ &= \beta\sigma_x^2(2-2\beta\sigma_x^2-\beta\sigma_x^2 Q)=\beta\sigma_x^2\underbrace{(2-(Q+2)\beta\sigma_x^2)}_{=\Delta_Q}, \end{aligned}$$

temos

$$2\beta\sigma_x^2\Delta_0\omega^2+\frac{8\alpha\kappa\Delta_0}{\sqrt{2\pi}}\omega-\beta^2\sigma_x^4 D(\infty)-\beta^2\sigma_\nu^2\sigma_x^2-4\alpha^2\kappa^2-\sigma_q^2=0, \tag{95}$$

ou, utilizando (94):

$$\begin{aligned} 2\beta\sigma_x^2\Delta_0\omega^2+\frac{8\alpha\kappa\Delta_0}{\sqrt{2\pi}}\omega-\beta^2\sigma_x^4\left[\frac{2\Delta_0(N-Q)}{\Delta_Q}\omega^2+\frac{\beta Q\sigma_\nu^2}{\Delta_Q}+\frac{\kappa^2\Delta_0' G^\star}{\beta^2\sigma_x^4\Delta_Q}+\frac{Q\sigma_q^2}{\beta\sigma_x^2\Delta_Q}\right] \\ -\beta^2\sigma_\nu^2\sigma_x^2-4\alpha^2\kappa^2-\sigma_q^2=0, \end{aligned}$$

que nos leva a

$$\begin{aligned} 2\beta\sigma_x^2\Delta_0\Delta_N\omega^2+\frac{8\alpha\kappa\Delta_0\Delta_Q}{\sqrt{2\pi}}\omega-2\beta^2\sigma_\nu^2\sigma_x^2\Delta_0-4\alpha^2\kappa^2\Delta_Q \\ -\kappa^2\Delta_0' G^\star-\beta\sigma_x^2 Q\sigma_q^2-\sigma_q^2\Delta_Q=0. \end{aligned} \tag{96}$$

Resolvendo esta equação, temos:

$$
\begin{aligned}
\omega &= -\frac{8\alpha\kappa\Delta_0\Delta_Q}{\sqrt{2\pi}}\frac{1}{4\beta\sigma_x^2\Delta_0\Delta_N} \\
&\quad +\frac{1}{4\beta\sigma_x^2\Delta_0\Delta_N}\sqrt{\begin{array}{l}\frac{32\alpha^2\Delta_0^2\Delta_Q^2}{\pi}\kappa^2+16\beta^3\sigma_x^4\Delta_0^2\Delta_N\sigma_\nu^2+32\beta\sigma_x^2\Delta_0\Delta_N\alpha^2\Delta_Q\kappa^2 \\ +8\beta\sigma_x^2\Delta_0\Delta_N\Delta_0'G^\star\kappa^2+8\beta^2\sigma_x^4\Delta_0\Delta_N Q\sigma_q^2+8\beta\sigma_x^2\Delta_0\Delta_N\Delta_Q\sigma_q^2\end{array}} \\
&= -\frac{\sqrt{2}\alpha\Delta_Q}{\beta\sigma_x^2\Delta_N\sqrt{\pi}}\kappa \\
&\quad +\frac{1}{4\beta\sigma_x^2\Delta_0\Delta_N}\sqrt{\begin{array}{l}\frac{32\alpha^2\Delta_0^2\Delta_Q^2}{\pi}\kappa^2+16\beta^3\sigma_x^4\Delta_0^2\Delta_N\sigma_\nu^2+32\beta\sigma_x^2\Delta_0\Delta_N\alpha^2\Delta_Q\kappa^2 \\ +8\beta\sigma_x^2\Delta_0\Delta_N\Delta_0'G^\star\kappa^2+8\beta^2\sigma_x^4\Delta_0\Delta_N Q\sigma_q^2+8\beta\sigma_x^2\Delta_0\Delta_N\Delta_Q\sigma_q^2.\end{array}}
\end{aligned}
\tag{97}
$$

Utilizando (94), obtemos:

$$
\begin{aligned}
D(\infty) &= \frac{2\Delta_0(N-Q)\beta\sigma_\nu^2}{\Delta_Q\Delta_N}+\frac{\beta Q\sigma_\nu^2}{\Delta_Q} \\
&\quad +\frac{2\Delta_0(N-Q)Q\sigma_q^2}{2\Delta_0\Delta_N}+\frac{2\Delta_0(N-Q)\Delta_Q\sigma_q^2}{2\beta\sigma_x^2\Delta_Q\Delta_0\Delta_N}+\frac{Q\sigma_q^2}{2\beta\sigma_x^2\Delta_0-\beta^2\sigma_x^4 Q} \\
&\quad +\left[\frac{8\Delta_0(N-Q)\alpha^2\Delta_Q}{\pi\beta^2\sigma_x^4\Delta_N^2}+\frac{4\alpha^2(N-Q)}{\beta\sigma_x^2\Delta_N}+\frac{(N-Q)\Delta_0'G^\star}{\beta\sigma_x^2\Delta_Q\Delta_N}+\frac{\Delta_0'G^\star}{\beta^2\sigma_x^4\Delta_Q}\right]\kappa^2 \\
&\quad -\frac{4\alpha(N-Q)}{\beta^2\sigma_x^4\Delta_N^2}\kappa\sqrt{\frac{\Delta_0}{\pi}\left(\frac{4\alpha^2\Delta_0\Delta_Q^2}{\pi}+4\beta\sigma_x^2\Delta_N\alpha^2\Delta_Q+\beta\sigma_x^2\Delta_N\Delta_0'G^\star\right)} \\
&\quad \times\sqrt{\kappa^2+\frac{\frac{\Delta_0}{\pi}\left(2\beta^3\sigma_x^4\Delta_0\Delta_N\sigma_\nu^2+\beta^2\sigma_x^4\Delta_N Q\sigma_q^2+\beta\sigma_x^2\Delta_N\Delta_Q\sigma_q^2\right)}{\frac{\Delta_0}{\pi}\left(\beta\sigma_x^2\Delta_N\Delta_0'G^\star+4\alpha^2\Delta_Q\left(\beta\sigma_x^2\Delta_N+\frac{\Delta_0\Delta_Q}{\pi}\right)\right)}},
\end{aligned}
\tag{98}
$$

o que é uma equação bastante extensa. Com o propósito de simplificá-la, seja

$$
\theta_0 \triangleq \left(\beta\sigma_x^2\Delta_N\Delta_0'G^\star+4\alpha^2\Delta_Q\left(\beta\sigma_x^2\Delta_N+\frac{\Delta_0\Delta_Q}{\pi}\right)\right). \tag{99}
$$

Aplicando (99) em (98) temos:

$$
\begin{aligned}
D(\infty) &= \frac{2\Delta_0(N-Q)\beta\sigma_\nu^2+\beta Q\sigma_\nu^2\Delta_N}{\Delta_N\Delta_Q} \\
&+\frac{(N-Q)Q\sigma_q^2}{\Delta_N}+\frac{(N-Q)\sigma_q^2}{\beta\sigma_x^2\Delta_N}+\frac{Q\sigma_q^2}{2\beta\sigma_x^2(1-\beta\sigma_x^2)-\beta^2\sigma_x^4Q} \\
&+\left\{\frac{\beta\sigma_x^2\frac{(N-Q)}{\Delta_Q}\Delta_0'G^\star+\Delta_0'G^\star\frac{\Delta_N}{\Delta_Q}}{\beta^2\sigma_x^4\Delta_N}+\frac{4(N-Q)\alpha^2}{\beta^2\sigma_x^4\Delta_N}\left[\beta\sigma_x^2+\frac{2\Delta_0\Delta_Q}{\pi\Delta_N}\right]\right\}\kappa^2 \\
&-\frac{4\alpha(N-Q)}{\beta^2\sigma_x^4\Delta_N^2}\sqrt{\frac{\Delta_0}{\pi}\left[\beta\sigma_x^2\Delta_0'\Delta_NG^\star+4\alpha^2\Delta_Q\left(\beta\sigma_x^2\Delta_N+\frac{\Delta_0\Delta_Q}{\pi}\right)\right]}\kappa \\
&\times\sqrt{\kappa^2+\frac{2\beta^3\sigma_x^4\Delta_0\Delta_N\sigma_\nu^2+\beta\sigma_x^2\Delta_N\sigma_q^2(\beta\sigma_x^2Q+\Delta_Q)}{\theta_0}}, \qquad (100)
\end{aligned}
$$

equação que pode ser simplificada com a constante:

$$
\theta_2 \triangleq \frac{4\alpha(N-Q)}{\beta^2\sigma_x^4\Delta_N^2}\sqrt{\frac{\Delta_0}{\pi}\left[\beta\sigma_x^2\Delta_0'\Delta_NG^\star+4\alpha^2\Delta_Q\left(\beta\sigma_x^2\Delta_N+\frac{\Delta_0\Delta_Q}{\pi}\right)\right]}, \qquad (101)
$$

de modo que (100) se torna:

$$
\begin{aligned}
D(\infty) &= \frac{(2N-2\beta N\sigma_x^2-2Q+2\beta Q\sigma_x^2+2Q-NQ\beta\sigma_x^2-2Q\beta\sigma_x^2)\beta\sigma_\nu^2}{\Delta_N\Delta_Q} \\
&+\frac{(N-Q)Q\sigma_q^2}{\Delta_N}+\frac{(N-Q)\sigma_q^2}{\beta\sigma_x^2\Delta_N}+\frac{Q\sigma_q^2}{2\beta\sigma_x^2(1-\beta\sigma_x^2)-\beta^2\sigma_x^4Q} \\
&+\left\{\frac{\Delta_0'G^\star}{\beta^2\sigma_x^4\Delta_N}\left[\frac{\beta\sigma_x^2N-\beta\sigma_x^2Q+2-N\beta\sigma_x^2-2\beta\sigma_x^2}{\Delta_Q}\right]+\frac{4(N-Q)\alpha^2}{\beta^2\sigma_x^4\Delta_N}\left[\beta\sigma_x^2+\frac{2\Delta_0\Delta_Q}{\pi\Delta_N}\right]\right\}\kappa^2 \\
&-\theta_2\kappa\sqrt{\kappa^2+\frac{2\beta^3\sigma_x^4\Delta_0\Delta_N\sigma_\nu^2+2\beta\sigma_x^2\Delta_N\Delta_0\sigma_q^2}{\theta_0}}. \qquad (102)
\end{aligned}
$$

Por meio das definições

$$
\theta_3' \triangleq \frac{2\beta^3\sigma_x^4\Delta_0\Delta_N\sigma_\nu^2+2\beta\sigma_x^2\Delta_N\Delta_0\sigma_q^2}{\theta_0}, \qquad (103)
$$

e

$$
\theta_1 \triangleq \frac{\Delta_0'G^\star+4(N-Q)\alpha^2\left(\beta\sigma_x^2+\frac{2\Delta_0\Delta_Q}{\pi\Delta_N}\right)}{\beta^2\sigma_x^4\Delta_N}, \qquad (104)
$$

a Equação (102) pode ser simplificada para:

$$\begin{aligned} D(\infty) &= \frac{N(2-(Q+2)\beta\sigma_x^2)}{\Delta_N\Delta_Q}\beta\sigma_\nu^2 \\ &\quad + \frac{(N-Q)Q\sigma_q^2}{\Delta_N} + \frac{(N-Q)\sigma_q^2}{\beta\sigma_x^2\Delta_N} + \frac{Q\sigma_q^2}{2\beta\sigma_x^2(1-\beta\sigma_x^2)-\beta^2\sigma_x^4Q} \\ &\quad + \theta_1\kappa^2 - \theta_2\kappa\sqrt{\kappa^2+\theta_3'}, \end{aligned} \tag{105}$$

a qual, após algumas manipulações, nos fornece:

$$\begin{aligned} D(\infty) &= \frac{\beta\sigma_\nu^2 N}{\Delta_N} + \theta_1\kappa^2 + \frac{(N-Q)Q\sigma_q^2}{\Delta_N} + \frac{(N-Q)\sigma_q^2}{\beta\sigma_x^2\Delta_N} + \frac{Q\sigma_q^2}{2\beta\sigma_x^2(1-\beta\sigma_x^2)-\beta^2\sigma_x^4Q} \\ &\quad - \theta_2\kappa\sqrt{\kappa^2+\theta_3'}, \end{aligned} \tag{106}$$

que é o principal resultado deste capítulo: uma expressão analítica que prevê o MSD assintótico do algoritmo ℓ_0-LMS, quando a planta ótima a identificar é variante no tempo de acordo com (29) e a hipótese $\mathbb{H}_q$ é válida. Podemos dividir a expressão obtida em quatro partes:

1. Termo $\frac{\beta\sigma_\nu^2 N}{\Delta_N}$

   Este termo coincide exatamente com o MSD assintótico do algoritmo LMS [80]. Ou seja, se a soma dos três termos restantes for negativa, o ℓ_0-LMS apresentaria (em teoria) um MSD inferior ao do algoritmo LMS clássico (em outras palavras, seria vantajoso adotar o algoritmo ℓ_0-LMS).

2. Termo $\theta_1\kappa^2$

   Este termo não depende de σ_q^2, sendo idêntico ao encontrado na análise do algoritmo ℓ_0-LMS na configuração estacionária.

3. Termo $\frac{(N-Q)Q\sigma_q^2}{\Delta_N} + \frac{(N-Q)\sigma_q^2}{\beta\sigma_x^2\Delta_N} + \frac{Q\sigma_q^2}{2\beta\sigma_x^2(1-\beta\sigma_x^2)-\beta^2\sigma_x^4Q}$

   Este termo - inédito - é uma expressão analítica que reflete um impacto direto (rigorosamente aditivo) que a variância da perturbação markoviana apresenta no MSD em regime permanente. Obviamente, fazendo $\sigma_q^2 = 0$, retornamos ao caso da análise da configuração estacionária. Cabe ressaltar que caso tenhamos todos os coeficientes a identificar nulos[4] (isto é, $Q = 0$), apenas a fração intermediária seria

[4] Cabe ressaltar que os coeficientes nulos a que nos referimos são os associados ao valor inicial dos coeficientes da planta desconhecida, já que a perturbação markoviana irá necessariamente alterar esta propriedade.

não nula, gerando uma contribuição igual a $\frac{N\sigma_q^2}{\beta\sigma_x^2\Delta_N}$ no MSD assintótico. Neste caso, tal contribuição seria proporcional ao tamanho do filtro adaptativo. Caso todos os coeficientes a identificar sejam não nulos (isto é, $Q = N$), apenas a última fração seria não nula, de modo que o MSD assintótico seria $\frac{N\sigma_q^2}{2\beta\sigma_x^2(1-\beta\sigma_x^2)-\beta^2\sigma_x^4 Q}$, também proporcional ao tamanho do filtro.

4. Termo $-\theta_2\kappa\sqrt{\kappa^2+\theta_3'}$

   Este termo depende tanto de κ quanto de σ_ν^2, porque θ_3' depende de σ_ν^2 (vide (103)). A constante de proporcionalidade θ_2 é a mesma encontrada na análise do ℓ_0-LMS num contexto estacionário. Já θ_3' se degenera na constante θ_3 de [19] quando $\sigma_q^2 = 0$, conforme esperado. Este termo demonstra uma contribuição não linear e não trivial que a variância σ_q^2 provoca no MSD em regime assintótico. Cabe ressaltar que este termo modela um impacto sofisticado no MSD assintótico oriundo de um acoplamento não linear entre o fator de atração (que depende de κ) e a variância da perturbação markoviana.

A qualidade das predições engendradas pela Equação (106) deve ser avaliada por meio de comparações com resultados de experimentos computacionais. Nestes experimentos, o algoritmo é executado num ambiente cujas estatísticas são conhecidas, de modo a possibilitar o cálculo de todos os termos da fórmula. A próxima seção se dedica a efetuar tais comparações.

3.4- Resultados

Para aquilatar a acurácia do modelo proposto, diversos cenários são simulados, para testar a robustez da equação encontrada (Equação (106)) perante diferentes configurações. Ao longo de todos os cenários alguns fatores foram mantidos constantes, sendo eles: número de iterações $= 160000$, $N = 30$ e $\alpha = 5$. A estimativa do desempenho assintótico foi efetuada por meio da média das últimas 10000 iterações. Testes preliminares sinalizaram comportamentos similares aos que serão apresentados para valores maiores de N.

Com exceção do Cenário 3, que utilizou os 12 primeiros coeficientes da função de

transferência do caso 2 ($N = 96$) da base de dados de [81][5], os coeficientes das plantas nos demais cenários são amostrados de uma distribuição Gaussiana com média zero e variância unitária, sendo também empregado um sinal de entrada branco, o que coincide com a hipótese $\mathbb{H}_{\text{BG}}$.

A base de dados da ITU foi escolhida por fornecer perfis de resposta ao impulso que representam com fidelidade canais reais de comunicação, tanto em ambientes móveis quanto fixos. Esses perfis são derivados de medições experimentais e padronizados por meio de recomendações técnicas amplamente reconhecidas pela comunidade científica e industrial.

Além de garantir reprodutibilidade dos experimentos e facilitar a comparação com outros trabalhos da literatura, essa escolha está diretamente alinhada com o contexto desta tese. A identificação e o rastreamento de plantas refletem de forma realista os desafios enfrentados em aplicações práticas, como cancelamento de eco, equalização adaptativa e redes digitais com variações de canal.

Dessa forma, o uso da base ITU não apenas fortalece o rigor metodológico da validação, como assegura que os experimentos propostos sejam realizados em cenários condizentes com o tipo de aplicação para o qual o algoritmo foi projetado.

Para os experimentos desta tese, foram realizados testes prévios com diferentes valores de α, e observou-se que valores muito altos limitavam a ação da penalização, enquanto valores muito baixos afetavam negativamente a capacidade de rastreamento do algoritmo, eliminando coeficientes potencialmente relevantes.

Dessa forma, foi adotado um valor intermediário, compatível com os encontrados na literatura, que proporciona um bom equilíbrio entre indução de esparsidade e desempenho adaptativo. Isso assegura que apenas coeficientes de baixa magnitude sejam penalizados, preservando os componentes significativos da planta.

3.5- Cenários

A Tabela 3 expõe os parâmetros dos diferentes cenários, assim como a 4 apresenta suas respectivas plantas.

[5] A base de dados foi obtida no site da *International Telecommunication Union (ITU)*, uma agência da ONU especializada em tecnologias de informação e comunicação.

Com a finalidade de verificar o desempenho do modelo são apresentadas duas curvas: a teórica obtida pelo modelo estocástico proposto (vermelha) e a empírica, obtida mediante a média das últimas iterações (azul). A métrica de avaliação utilizada foi o MSD, já que a análise teórica apresenta uma fórmula fechada para o MSD. No eixo das abscissas o fator β foi escolhido, pois além de sua importância em relação a convergência do algoritmo, também é importante observar se a análise teórica prevê corretamente o valor de β que otimiza o desempenho assintótico.

Tabela 3 – Principais parâmetros dos cenários e seus respectivos valores.

Cenário	$\log_{10}\beta$	κ	σ_ν^2	σ_q^2
1	$\{-4; -3,75; ...; -1,5\}$	10^{-9}	10^{-6}	5×10^{-11}
2	$\{-4; -3,75; ...; -1,5\}$	10^{-9}	10^{-4}	5×10^{-11}
3	$\{-4; -3,75; ...; -1,5\}$	10^{-9}	10^{-6}	5×10^{-10}
4	$\{-4; -3,75; ...; -1,5\}$	5×10^{-9}	10^{-6}	5×10^{-11}
5	$\{-4; -3,75; ...; -1,2\}$	10^{-9}	10^{-6}	5×10^{-11}

Tabela 4 – Plantas dos cenários.

Cenário	Função de Transferência
1	$-0,3069 - 0,9273z^{-1} + 0,1397z^{-2}$
2	$0,1530 + 1,4434z^{-1} + 0,5546z^{-2}$
3	Referência [81]
4	$-0,0208 - 1,5990z^{-1} + 0,5236z^{-2}$
5	$-0,5571 - 0,0294z^{-1} + 0,4945z^{-2}$

A Fig. 3 exibe os resultados para o Cenário 1, o qual foi utilizado como base de comparação com o demais, de modo que as alterações seguintes pudessem ser observadas através de um contraponto. Na Fig. 4 o valor de σ_ν^2 foi modificado com o intuito de averiguar a atuação do algoritmo em uma configuração com um maior ruído de medição. Assim como na Fig. 5, a perturbação markoviana σ_q^2 foi aumentada e a planta da ITU foi utilizada. Esta planta trata-se de uma aplicação real, na qual foram medidas as respostas ao impulso de redes digitais com o objetivo de projetar canceladores de eco e avaliar o desempenho deles por meio de testes laboratoriais. Deste modo, pode-se

observar um caso prático no qual a acurácia do modelo é independente da função de transferência adotada. Na Fig. 6 o parâmetro κ foi testado com um valor maior. Percebe-se que houve perda de aderência com relação à curva empírica, justamente pelo fato do modelo proposto adotar $\mathbb{H}_\kappa$, o que faz com que o modelo tenha uma resposta melhor para valores pequenos de κ. Já na Fig. 7 onde a faixa de valores de β foi estendida, é possível notar que para valores altos o modelo perde desempenho em regime permanente.

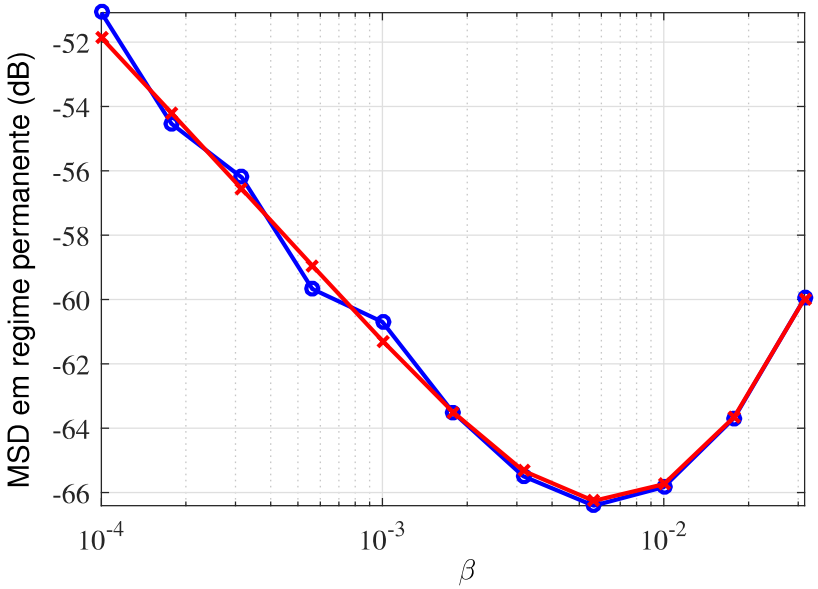


Figura 3 – MSD em regime permanente do Cenário 1. Em vermelho a curva teórica e em azul a empírica.

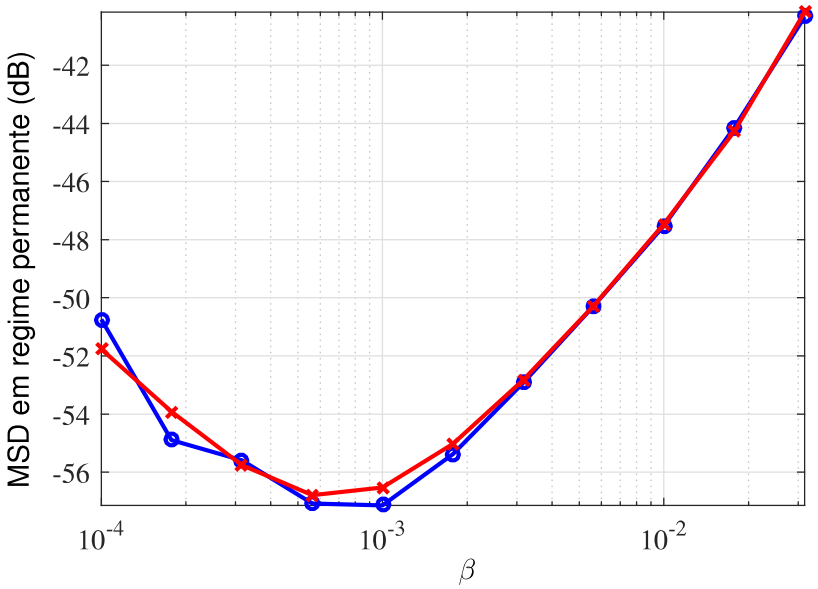


Figura 4 – MSD em regime permanente do Cenário 2. Em vermelho a curva teórica e em azul a empírica.

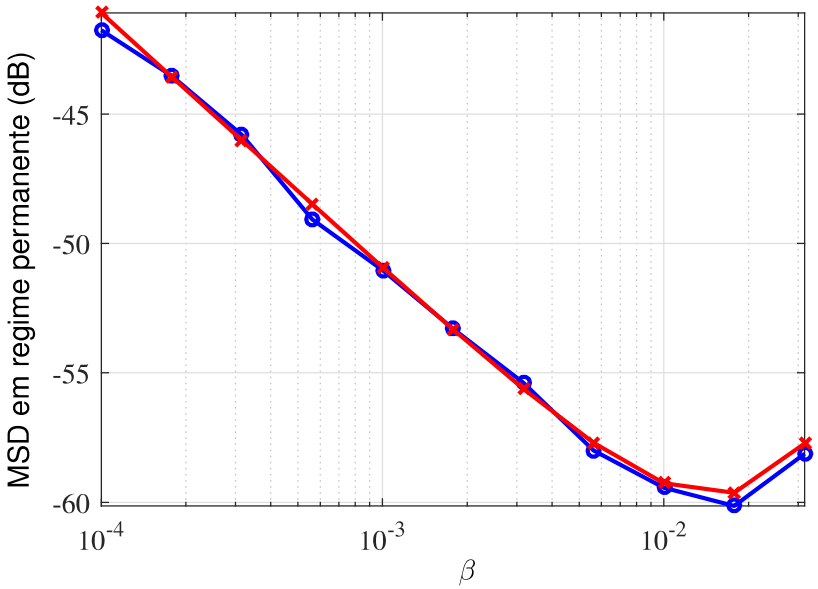


Figura 5 – MSD em regime permanente do Cenário 3. Em vermelho a curva teórica e em azul a empírica.

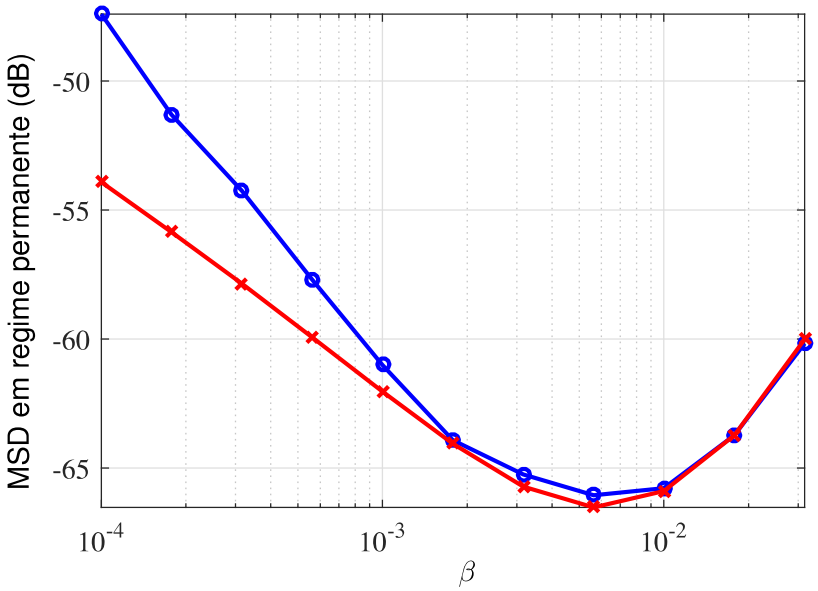


Figura 6 – MSD em regime permanente do Cenário 4. Em vermelho a curva teórica e em azul a empírica.

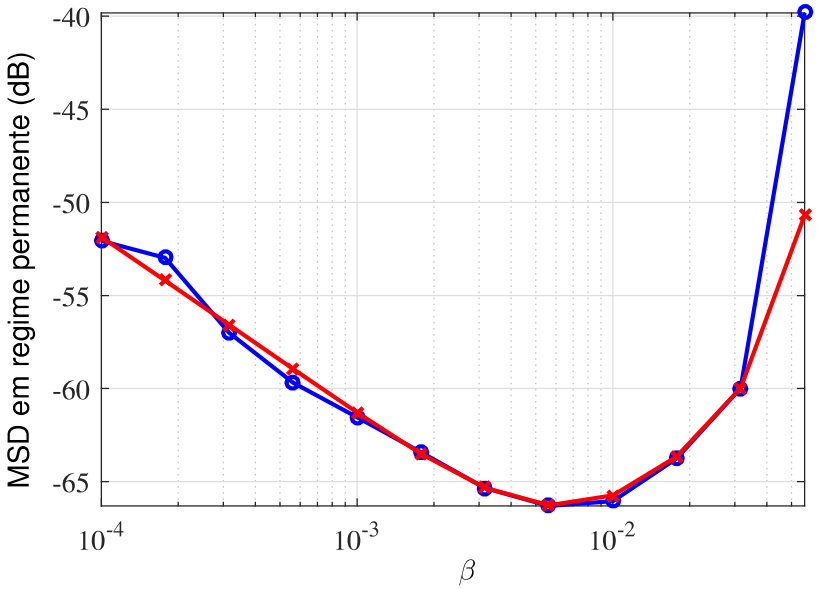


Figura 7 – MSD em regime permanente do Cenário 5. Em vermelho a curva teórica e em azul a empírica.

Em todos os cenários, observa-se que o desempenho assintótico em termos de β apresenta uma concavidade para cima. Esse comportamento já foi bastante discutido na literatura concernente, refletindo o fato de que valores de β muito pequenos prejudicam a capacidade de rastreamento, enquanto que valores de β muito altos penalizam o desem-

penho assintótico ao induzirem oscilações de alta magnitude no estimador adaptativo. A análise teórica proposta permite obter predições precisas para o desempenho assintótico ao longo de uma ampla faixa de valores de β, assim como consegue prever acuradamente o valor deste parâmetro que otimiza o MSD em regime permanente.

3.6- Comentários Finais

Este capítulo propõe um modelo estocástico inédito, baseado na metodologia de [19], capaz de prever o desempenho assintótico do algoritmo ℓ_0-LMS na identificação de uma planta sujeita a perturbações markovianas de primeira ordem. O modelo foi comparado com resultados de simulações. Dentro de algumas restrições, ele foi capaz de prever com acurácia o desempenho assintótico do algoritmo, e ainda sinalizar com razoável precisão qual seria o valor ótimo de β para otimização do MSD em regime permanente. O próximo capítulo propõe uma generalização do algoritmo ℓ_0-LMS (a qual contempla outros algoritmos de penalização de norma, como casos particulares), propõe uma equação analítica capaz de prever seu desempenho em regime permanente e também, após algumas aproximações, propõe uma fórmula fechada capaz de computar o valor ótimo de κ que otimiza o desempenho do algoritmo em regime permanente.

4- Modelagem Estocástica do Algoritmo de Penalização de Norma Generalizada

Como vimos, algoritmos de filtragem adaptativa convencionais, tais como o LMS, o *Normalized Least Mean Squares (NLMS)*, o APA e o *Recursive Least Squares (RLS)* não levam em conta a esparsidade da planta a identificar, negligenciando a estrutura inerente ao problema de identificação de sistemas. Tal esparsidade pode ser utilizada para melhorar a taxa de convergência destes algoritmos, ou então o seu desempenho assintótico [82].

O Capítulo 2 explica que algoritmos regularizados por normas de esparsidade apresentam menor custo computacional, podendo obter um melhor compromisso entre taxa de convergência e desempenho assintótico, tornando-os uma opção viável, na prática.

Uma penalização baseada na norma ℓ_1 apresenta problemas relativos ao viés nas estimativas, fazendo a adoção de uma penalidade não convexa, tal como a norma ℓ_0 ou a norma ℓ_q (para $0 < q < 1$), frequentemente uma alternativa interessante [83], [84]. Como a minimização de uma função custo com uma penalidade de norma ℓ_0 é um problema NP-difícil, a norma ℓ_0 é tipicamente aproximada usando técnicas tais como uma função baseada em soma de logaritmos, na correntropia (a qual quantifica a similaridade local entre duas variáveis aleatórias) ou uma aproximação baseada numa função exponencial [85]. Uma abordagem comum - adotada pelo algoritmo ℓ_0-LMS - aproxima a norma ℓ_0 por meio de funções diferenciáveis em quase todas as partes, permitindo que o problema de otimização associado (não convexo) seja tratado com métodos de gradiente estocástico [82].

Neste capítulo, propomos uma penalização por norma linear por partes capaz de aproximar, com uma precisão arbitrária, diversos termos de penalização de norma previamente introduzidos na literatura. Idealmente, o desenvolvimento de um novo algoritmo deve vir acompanhado por um modelo analítico capaz de predizer o seu desempenho. Entretanto, como vimos no Capítulo 3, a modelagem do desempenho de algoritmos de filtragem adaptativa é uma tarefa bastante desafiadora, especialmente porque o seu pro-

cesso de adaptação é governado por equações a diferenças estocásticas e não lineares, as quais dependem do sinal de referência e de dados de entrada estocásticos de um modo não linear e variante no tempo. Como resultado, a análise de desempenho de algoritmos de filtragem adaptativa é complexa, e análises exatas são raras (veja, por exemplo, [66], [69]) e tipicamente limitadas a casos específicos. Dada esta contextualização, as contribuições deste capítulo podem ser sucintamente descritas como:

1. Proposta de um algoritmo adaptativo generalizado de regularização de norma que utiliza um termo de penalização linear por partes, o qual pode aproximar alguns algoritmos previamente propostos tão acuradamente quanto desejado;

2. Executar uma análise estocástica do desvio quadrático médio do algoritmo proposto;

3. Obter uma fórmula fechada que descreve o desempenho assintótico do algoritmo proposto;

4. Após a aplicação de heurísticas de simplificação para a expressão obtida, obter uma fórmula fechada que determina o valor ótimo do parâmetro κ, o qual, como vimos, regula a intensidade do termo de regularização.

A relevância para a obtenção de uma fórmula capaz de especificar a intensidade do termo de regularização advém do fato de que este termo, na prática, é um elemento crítico para que filtros adaptativos regularizados por norma obtenham um bom desempenho [34]. Adicionalmente, cumpre enfatizar que, como o algoritmo proposto inclui diversos algoritmos previamente publicados como casos particulares, os resultados teóricos obtidos neste capítulo também se aplicam a estes algoritmos.

Este capítulo é organizado do seguinte modo: a Seção 4.1 introduz o algoritmo proposto *Generalized Piecewise Linear Regularization Least Mean Squares (GPLR-LMS)*. Um modelo estocástico é desenvolvido na Seção 4.2, provendo uma fórmula fechada que descreve o desvio quadrático médio do algoritmo GPLR-LMS. A Seção 4.3 apresenta simplificações no modelo estocástico capazes de permitir a obtenção de uma fórmula fechada para a determinação da magnitude do termo de penalização. Os resultados do modelo estocástico proposto, comparados com curvas de desempenho do algoritmo obtidas por meio de simulações computacionais, são discutidos na Seção 4.4. Por fim, a Seção 4.5 contém as conclusões respectivas do capítulo.

4.1- Algoritmo GPLR-LMS

Como vimos no Capítulo 2, a equação de atualização do algoritmo ℓ_0-LMS pode ser descrita como

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta e(k)\boldsymbol{x}(k) + \kappa \boldsymbol{g}[\boldsymbol{w}(k)], \tag{107}$$

onde diferentes escolhas da função $\boldsymbol{g}[\boldsymbol{w}(k)]$ podem dar origens a algoritmos distintos. Neste capítulo, a função $\boldsymbol{g}[\boldsymbol{w}(k)]$ é considerada uma função linear por partes ímpar, com P segmentos de reta. Assim, assuma limiares (ajustáveis pelo projetista) $t_i \in \mathbb{R}_+$ (para $i \in \{1, 2, \ldots, P\}$) e as definições seguintes para os conjuntos[1] $\mathcal{C}_i$ associado com os índices do vetor $\boldsymbol{w}(k)$:

$$\begin{aligned}
\mathcal{C}_0 &= \{i \,|\, w_i(k) = 0\}, \\
\mathcal{C}_1 &= \{i \,|\, 0 < w_i(k) \leq t_1\}, \\
\mathcal{C}_{-1} &= \{i \,|\, -t_1 \leq w_i(k) < 0\}, \\
\mathcal{C}_2 &= \{i \,|\, t_1 < w_i(k) \leq t_2\}, \\
\mathcal{C}_{-2} &= \{i \,|\, -t_2 \leq w_i(k) < -t_1\}, \\
&\vdots \\
\mathcal{C}_P &= \{i \,|\, t_{P-1} < w_i(k) \leq t_P\}, \\
\mathcal{C}_{-P} &= \{i \,|\, -t_P \leq w_i(k) < -t_{P-1}\}, \\
\mathcal{C}_L &= \{i \,|\, |w_i(k)| > t_P\},
\end{aligned}$$

onde $\mathcal{C}_L$ contém os coeficientes de alta magnitude, para os quais $g[w_i(k)]$ (para $i \in \mathcal{C}_L$) assume o valor zero[2]. Adicionalmente, note que $t_P \to \infty \Rightarrow |\mathcal{C}_L| = 0$. Assumindo que $i \in \mathcal{C}_j$ (para $j \in \{-P, -P+1, \ldots, -1, 0, 1, \ldots, P-1, P\}$), o i-ésimo elemento de $\boldsymbol{g}[\boldsymbol{w}(k)]$ pode ser escrito como a seguinte equação linear (com relação ao j-ésimo segmento de reta):

$$g_j[w_i(k)] = a_j w_i(k) + b_j, \text{ para } i \in \mathcal{C}_j. \tag{108}$$

Um exemplo de uma função linear por partes que respeita a condição (108) é

[1] Por simplicidade, o índice k, quando referenciado aos conjuntos $\mathcal{C}_i$, será omitido.

[2] Essa hipótese é razoável num contexto de penalização de norma, sendo comum em algoritmos adaptativos conscientes de esparsidade.

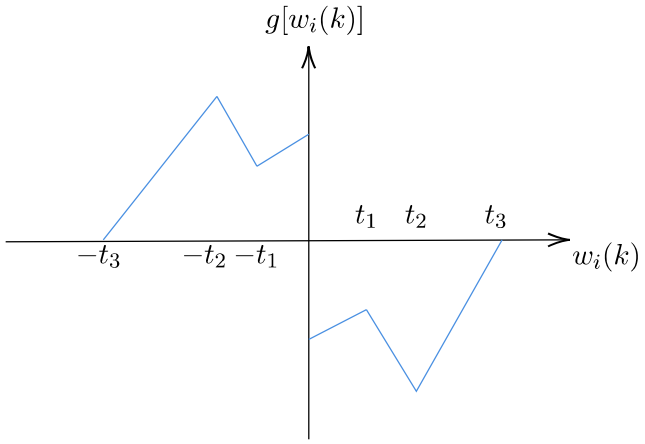


Figura 8 – Exemplo de uma função linear por partes $g[w_i(k)]$ com $P = 3$.

Tabela 5 – Exemplos de algoritmos que se ajustam ao paradigma proposto neste artigo.

Algoritmo	Escolha de parâmetros
ℓ_0-LMS [14]	$a_1 = b_1^2$, $t_1 = \frac{1}{b_1}$, $P = 1$
ZA-LMS [31]	$a_1 = 0$, $b_1 = -1$, $t_1 \to \infty$, $P = 1$
Optmized Zero Attracting LMS (OZ-LMS) [86]	$b_1 = 0$, $a_1 < 0$, $P = 1$
Improved ℓ_0 *LMS (ILMS)* [87]	$t_1 = \frac{\sqrt{a_2}}{a_2 - a_1}$, $b_1 = 0$, $b_2 = -a_2$, $t_2 = \frac{1}{\sqrt{a_2}}$, $P = 2$

apresentado na Figura 8.

Como assume-se que o termo de penalização é ímpar, as seguintes simetrias são consideradas válidas:

$$a_{-j} = a_j, \tag{109}$$

$$b_{-j} = -b_j. \tag{110}$$

Note também que, como os conjuntos $\mathcal{C}_{-P}, \mathcal{C}_{-P+1}, \ldots, \mathcal{C}_P$ e $\mathcal{C}_L$ são disjuntos, vale a seguinte restrição:

$$|C_0| + |C_1 \cup C_2 \cup \ldots \cup C_P| + |C_{-1} \cup C_{-2} \cup \ldots \cup C_{-P}| + |\mathcal{C}_L| = N. \tag{111}$$

Notavelmente, o paradigma proposto generaliza alguns algoritmos previamente propostos, como demonstrado pelos exemplos da Tabela 5. É importante destacar que a Tabela 5 não exaure todos os algoritmos que podem ser emulados com qualquer grau de acurácia por meio do paradigma proposto. Por exemplo, seja o algoritmo RZA-LMS,

introduzido em [31]. Sua função de penalidade pode ser descrita como:

$$g_{\text{RZA-LMS}}[w_i(k)] = -\frac{\text{sign}[w_i(k)]}{1+\epsilon|w_i(k)|}, \tag{112}$$

onde

$$\text{sign}(x) \triangleq \begin{cases} 1, & \text{para } x > 0 \\ 0, & \text{para } x = 0 \\ -1, & \text{para } x < 0 \end{cases}. \tag{113}$$

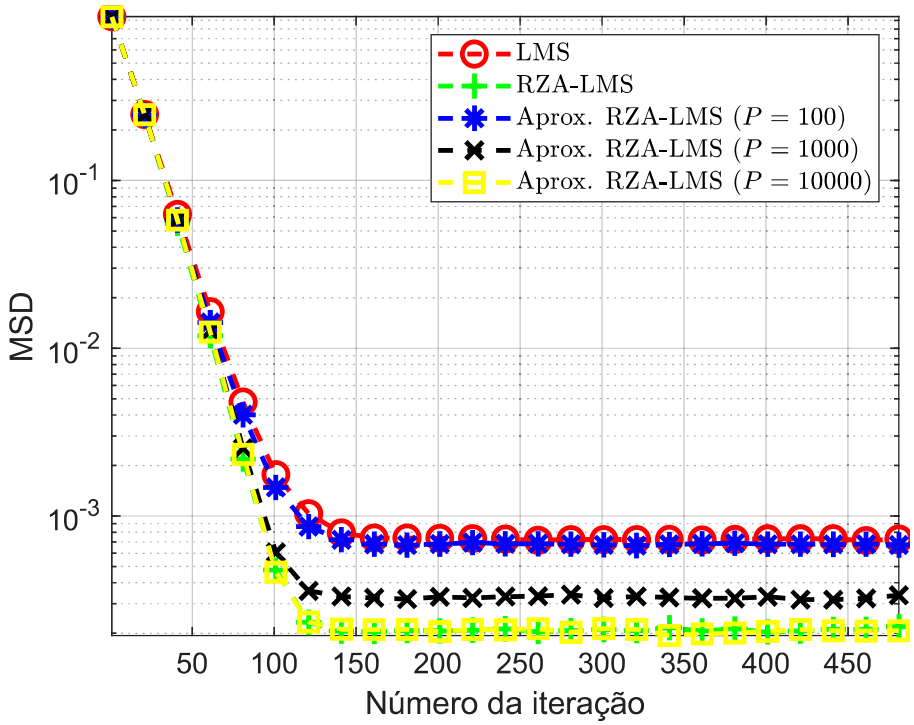


Figura 9 – Comparação do comportamento do MSD dos algoritmos LMS e RZA-LMS e aproximações (com o termo de penalização sendo linear por partes) do algoritmo RZA-LMS com diferentes valores de P (com $P \in \{100, 1000, 10000\}$). As configurações do experimentos são as mesmas da Figura 1 de [31].

Embora a penalidade imposta pelo algoritmo RZA-LMS (veja a Equação (112)) não seja linear por partes, ela é ímpar e pode ser aproximada por uma função linear por partes. Para demonstrar isso, um experimento de [31] é replicado, com $\beta = 0.05$, $\kappa = 5 \times 10^{-4}$ e $\epsilon = 10$. Tanto o sinal de entrada quanto o ruído observado são brancos e seguem distribuições gaussianas, com variâncias $\sigma_x^2 = 1$ e $\sigma_\nu^2 = 10^{-3}$, respectivamente. O vetor ótimo (e desconhecido) de coeficientes a identificar consiste de 16 coeficientes nulos, à exceção do quinto, o qual apresenta uma magnitude unitária. Considere uma

aproximação linear por partes com P segmentos de igual comprimento que interpolam a função $g_{\text{RZA-LMS}}[w_i(k)]$, no intervalo (0,10). A Figura 9 apresenta a evolução do MSD para os algoritmos LMS, RZA-LMS e a aproximação linear por partes da função de penalização do algoritmo RZA-LMS, usando P segmentos de reta. Note que às medida que P aumenta, as versões aproximadas crescentemente emulam o algoritmo RZA-LMS, e com $P = 10000$ o desempenho da versão aproximada é, para efeitos práticos, equivalente ao do algoritmo RZA-LMS. Isto significa que os resultados teóricos deste capítulo também se aplicam, no limite, ao algoritmo RZA-LMS, cuja função de penalização de norma não pode ser considerada, em termos rigorosos, linear por partes.

4.2- Análise de Regime Permanente do Algoritmo GPLR-LMS

O fato de a função de penalização adotada pelo algoritmo GPLR-LMS apresentar vários segmentos de reta requer uma atualização das definições dos conjuntos $\mathcal{C}_{\text{L}}$, $\mathcal{C}_{\text{S}}$ e $\mathcal{C}_0$ definidos no Capítulo 3. Assim, neste capítulo consideraremos que os coeficientes ótimos necessariamente pertencem a um dos três conjuntos distintos:

$$\mathcal{C}_{\text{L}} \triangleq \{i||w_i^\star| \geq t_P\} \text{ (coeficientes de alta magnitude)}, \tag{114}$$

$$\mathcal{C}_{\text{S}} \triangleq \{i|0 < |w_i^\star| < t_P\} \text{ (coeficientes de baixa magnitude)}, \tag{115}$$

$$\mathcal{C}_0 \triangleq \{i|w_i^\star = 0\} \text{ (coeficientes nulos)}, \tag{116}$$

onde $\mathcal{C}_{\text{S}} \triangleq \mathcal{C}_{-P} \cup \mathcal{C}_{-P+1} \cup \ldots \cup \mathcal{C}_{-1} \cup \mathcal{C}_1 \cup \ldots \mathcal{C}_P$. Note que $|\mathcal{C}_L \cup \mathcal{C}_S \cup \mathcal{C}_0| = N$. As cardinalidades destes conjuntos serão denotadas como:

$$Q_{\text{L}} \triangleq |\mathcal{C}_{\text{L}}|, Q_{\text{S}} \triangleq |\mathcal{C}_{\text{S}}|, Q_{\text{S}} \triangleq |\mathcal{C}_{\text{S}}|. \tag{117}$$

Como ponto de partida para uma análise do comportamento médio do algoritmo, seja a seguinte recursão do vetor de desvios[3]:

$$\tilde{\boldsymbol{w}}(k+1) = \left[\boldsymbol{I}_N - \beta \boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] \tilde{\boldsymbol{w}}(k) - \beta\nu(k)\boldsymbol{x}(k) - \kappa \boldsymbol{g}[\boldsymbol{w}(k)]. \tag{118}$$

[3]Observe que esta equação é muito similar à Equação (38), que consta no Capítulo 3.

Todas as hipóteses utilizadas no Capítulo 3 serão mobilizadas para uma análise média do comportamento de primeira e segunda ordens dos coeficientes do vetor de desvio $\tilde{\boldsymbol{w}}(k)$. No entanto, o fato de a função $g[w_i(k)]$ ser linear por partes, admitindo diversos segmentos de reta, demanda uma pequena alteração em $\mathbb{H}_{\mathsf{g}}$, de modo que neste capítulo será adotada a hipótese $\mathbb{H}'_{\mathsf{g}}$, especificada a seguir:

$\mathbb{H}'_{\mathsf{g}}$ (Hipótese referente à função de atração $g[\cdot]$). O coeficiente adaptativo $w_i(k)$ compartilha o mesmo segmento de reta (isto é, a função $g_j[w_i(k)]$ - vide Equação (108)) que o coeficiente ótimo desconhecido correspondente. Ademais, assume-se que ele está fora do termo de atração $g[\cdot]$ quando $i \in \mathcal{C}_L$, estando sob atração caso contrário.

Comentário: Embora $\mathbb{H}'_{\mathsf{g}}$ seja uma hipótese muito forte durante o transiente do algoritmo, é razoável esperar que após a convergência ela seja uma aproximação válida. Cabe ressaltar que valores elevados de β podem implicar aumento da variância dos coeficientes adaptativos, o que pode prejudicar a validade de $\mathbb{H}'_{\mathsf{g}}$. Na prática, como veremos, tal impacto negativo não foi observado nas simulações.

Aplicando-se o operador de valor esperado em (118) e aduzindo a aplicação de $\mathbb{H}_{\mathsf{BG}}$, $\mathbb{H}_{\mathsf{I}}$, $\mathbb{H}_{\nu}$ e $\mathbb{H}'_{\mathsf{g}}$ nos leva a:

$$\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] = \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right] - \beta\sigma_x^2\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right] - \kappa\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\right\}, \tag{119}$$

onde σ_x^2, como no Capítulo 3, denota a variância do sinal de entrada $x(k)$.

Assumindo que o algoritmo converge, iremos a seguir calcular $\lim_{k\to\infty}\mathbb{E}\left[\tilde{w}_i(k)\right]$ para os conjuntos $\mathcal{C}_0$, $\mathcal{C}_{\mathsf{S}}$ and $\mathcal{C}_{\mathsf{L}}$, assumindo que o algoritmo opera em regime permanente (operação assintótica).

- Para $i \in \mathcal{C}_0$:

  Notando que $g(x)$ tem o sinal contrário de x no intervalo $(-t_1, t_1)$ e usando $\mathbb{H}'_g$, chegamos a:

  $$\lim_{k\to\infty}\mathbb{E}\left[\tilde{w}_i(k)\right] = 0, \text{ para } i \in \mathcal{C}_0. \tag{120}$$

- Para $i \in \mathcal{C}_{\mathsf{S}}$:

  Usando $\mathbb{H}'_{\mathsf{g}}$:

  $$\lim_{k\to\infty}\beta\sigma_x^2\mathbb{E}\left[\tilde{w}_i(k)\right] = -\kappa g_j[w_i(k)] = -\kappa\left[a_j w_i(k) + b_j\right] \tag{121}$$

  $$= -\kappa a_j w_i^\star - \kappa b_j + \kappa a_j\mathbb{E}\left[\tilde{w}_i(k)\right]. \tag{122}$$

$$\Rightarrow \mathbb{E}\left[\tilde{w}_i(k)\right] = \frac{-\kappa}{\beta\sigma_x^2 - \kappa a_j} g_j[w_i^\star], \text{ para } i \in \mathcal{C}_{\mathsf{S}}. \tag{123}$$

- Para $i \in \mathcal{C}_L$: Neste caso, $\mathbb{H}_\nu$ leva a

$$\lim_{k\to\infty} \mathbb{E}\left[\tilde{w}_i(k)\right] = 0, \text{ para } i \in \mathcal{C}_L. \tag{124}$$

Comentário: Note que a estimativa do algoritmo adaptativo com penalização de norma é enviesada no caso de coeficientes pertencentes a $\mathcal{C}_{\mathsf{S}}$ (veja a Equação (123)). Entretanto, este viés leva a uma redução benéfica na variância do estimador quando o vetor desconhecido a ser estimado é suficientemente esparso.

A primeira contribuição deste capítulo, a qual depende das hipóteses $\mathbb{H}_{\mathsf{BG}}$, $\mathbb{H}_\nu$, $\mathbb{H}_{\mathsf{I}}$, $\mathbb{H}_{\mathsf{G}}$ e $\mathbb{H}'_{\mathsf{g}}$, está descrita no Teorema 1 a seguir. Este teorema provê uma fórmula fechada para o desvio quadrático médio assintótico, obtida a partir das hipóteses previamente explicitadas.

Teorema 1. Considere o desvio quadrático médio assintótico $D(\infty)$ como:

$$D(\infty) \triangleq \lim_{k\to\infty} D(k) = \lim_{k\to\infty} \mathbb{E}\left[\|\tilde{\boldsymbol{w}}(k)\|^2\right]. \tag{125}$$

Esta métrica assintótica pode ser computada usando a seguinte fórmula fechada:

$$D(\infty) = -\frac{\zeta_2}{\zeta_1} + \frac{\zeta_4 - \sqrt{\zeta_4^2 - 4\zeta_1\zeta_2\zeta_4 + 4\zeta_1^2\zeta_3}}{2\zeta_1^2}, \tag{126}$$

cujas variáveis ζ_1, ζ_2, ζ_3 e ζ_4 serão definidas na prova do Teorema 1, apresentada a seguir.

Demonstração: Após multiplicar ambos os lados da Equação (118) por sua versão transposta e aplicando o operador de valor esperado, em conjunto com $\mathbb{H}_{\mathsf{BG}}$, $\mathbb{H}_\nu$ e $\mathbb{H}_{\mathsf{I}}$, obtém-se:

$$\begin{aligned}
\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k+1) \;=\; & \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) - \beta\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] + \beta^2\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] \\
& -\beta\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right] - \kappa\mathbb{E}\left\{\tilde{\boldsymbol{w}}(k)\boldsymbol{g}^T\left[\boldsymbol{w}(k)\right]\right\} \\
& +\kappa\beta\mathbb{E}\left\{\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\boldsymbol{g}^T[\boldsymbol{w}(k)]\right\} + \beta^2\sigma_\nu^2\sigma_x^2\boldsymbol{I} \\
& -\kappa\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\tilde{\boldsymbol{w}}^T(k)\right\} + \beta\kappa\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right\} \\
& +\kappa^2\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\boldsymbol{g}^T\left[\boldsymbol{w}(k)\right]\right\},
\end{aligned} \tag{127}$$

onde

$$\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) \triangleq \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right] \tag{128}$$

denota a matriz de correlação dos desvios.

A aplicação de $\mathbb{H}_{\mathsf{G}}$, em conjunção com o Teorema da fatoração de momentos gaussianos [78] implica a seguinte identidade:

$$\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\boldsymbol{A}\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] = 2\boldsymbol{\Lambda}\boldsymbol{A}\boldsymbol{\Lambda} + \boldsymbol{\Lambda}\mathsf{Tr}\left\{\boldsymbol{\Lambda}\boldsymbol{A}\right\}, \tag{129}$$

onde $\boldsymbol{\Lambda}$ é a matriz diagonal que contém na sua diagonal principal os autovalores da matriz de autocorrelação da matriz $\boldsymbol{R}_{\boldsymbol{x}} \triangleq \mathbb{E}[\boldsymbol{x}(k)\boldsymbol{x}^T(k)]$.

A aplicação da Equação (129), em conjunção com $\mathbb{H}_{\mathsf{BG}}$, permite a seguinte decomposição:

$$\beta^2\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] = 2\beta^2\sigma_x^4\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) + \beta^2\sigma_x^4 D(k)\boldsymbol{I}_N. \tag{130}$$

A combinação de (130) com (127) engendra

$$\begin{aligned}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k+1) &= (1-2\beta\sigma_x^2\Delta_0)\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) + \beta^2\sigma_x^4 D(k)\boldsymbol{I} + \beta^2\sigma_\nu^2\sigma_x^2\boldsymbol{I} \\ &\quad -\kappa\Delta_0\mathbb{E}\left\{\tilde{\boldsymbol{w}}(k)\boldsymbol{g}^T\left[\boldsymbol{w}(k)\right]\right\} - \kappa\Delta_0\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\tilde{\boldsymbol{w}}^T(k)\right\} \\ &\quad +\kappa^2\mathbb{E}\left\{\boldsymbol{g}[\boldsymbol{w}(k)]\boldsymbol{g}^T[\boldsymbol{w}(k)]\right\}.\end{aligned} \tag{131}$$

Quando o operador de "traço" é aplicado em ambos os lados de (131), chega-se a:

$$\begin{aligned}D(k+1) &= (1-\beta\sigma_x^2\Delta_N)D(k) + N\beta^2\sigma_\nu^2\sigma_x^2 \\ &\quad -2\kappa\Delta_0\mathbb{E}\left\{\tilde{\boldsymbol{w}}^T(k)\boldsymbol{g}[\boldsymbol{w}(k)]\right\} + \kappa^2\mathbb{E}\left\{\|\boldsymbol{g}[\boldsymbol{w}(k)]\|^2\right\}.\end{aligned} \tag{132}$$

Expressando a identidade matricial (131) em termos do i-ésimo elemento da diagonal principal de $\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)$ leva a:

$$\begin{aligned}\mathbb{E}\left[\tilde{w}_i^2(k+1)\right] &= (1-2\beta\sigma_x^2\Delta_0)\mathbb{E}\left[\tilde{w}_i^2(k)\right] + \beta^2\sigma_x^4 D(k) + \beta^2\sigma_\nu^2\sigma_x^2 \\ &\quad -2\kappa\Delta_0\mathbb{E}\left\{\tilde{w}_i(k)g_j[w_i(k)]\right\} + \kappa^2\mathbb{E}\left\{g_j^2\left[w_i(k)\right]\right\},\end{aligned} \tag{133}$$

a qual, em operação assintótica, se torna na seguinte identidade:

$$\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] = \frac{\beta^2\sigma_x^4 D(\infty) + \beta^2\sigma_\nu^2\sigma_x^2 - 2\kappa\Delta_0\mathbb{E}\left\{\tilde{w}_i(\infty)g_j[w_i(\infty)]\right\} + \kappa^2\mathbb{E}\left\{g_j^2[w_i(\infty)]\right\}}{2\beta\sigma_x^2\Delta_0}. \tag{134}$$

Comentário: Observe que a Equação (62) é similar à (134), sendo as diferenças o fato de haver uma variância da perturbação markoviana em (62) e haver uma função de penalização mais sofisticada em (134), a qual implicará manipulações algébricas bem mais desafiadoras.

Utilizando as hipóteses $\mathbb{H}_{\mathsf{BG}}$, $\mathbb{H}_\nu$, $\mathbb{H}_{\mathsf{I}}$, $\mathbb{H}_{\mathsf{G}}$ e $\mathbb{H}'_{\mathsf{g}}$, a fórmula (134) é calculada para os conjuntos $\mathcal{C}_0$, $\mathcal{C}_{\mathsf{S}}$ e $\mathcal{C}_{\mathsf{L}}$ a seguir.

- Para $i \in \mathcal{C}_{\mathsf{L}}$:

$$\mathbb{E}\left\{\tilde{w}_i(k)g_j[w_i(k)]\right\} = 0, \tag{135}$$

- Para $i \in \mathcal{C}_{\mathsf{S}}$:

$$\begin{aligned} g_j[w_i(k)] &= a_j w_i(k) + b_j = a_j w_i^\star - a_j\tilde{w}_i(k) + b_j, & (136)\\ &= g_j[w_i^\star] - a_j\tilde{w}_i(k). & (137)\end{aligned}$$

$$\begin{aligned} \mathbb{E}\left\{\tilde{w}_i(k)g_j[w_i(k)]\right\} &= \mathbb{E}\left[\tilde{w}_i(k)\right]g_j\left[w_i^\star\right] - a_j\mathbb{E}\left[\tilde{w}_i^2(k)\right], & (138)\\ \mathbb{E}\left\{g_j^2[w_i(k)]\right\} &= g_j^2\left[w_i^\star\right] - 2a_j g_j\left[w_i^\star\right]\mathbb{E}\left[\tilde{w}_i(k)\right] + a_j^2\mathbb{E}\left[\tilde{w}_i^2(k)\right]. & (139)\end{aligned}$$

- Para $i \in \mathcal{C}_0$:

$$\begin{aligned} \mathbb{E}\left[\tilde{w}_i(k)\right] &= 0, & (140)\\ \mathbb{E}\left[\tilde{w}_i^2(k)\right] &= \mathbb{E}\left[w_i^2(k)\right]. & (141)\end{aligned}$$

Quando $i \in \mathcal{C}_0$, vale[4] $g_1(x) = g_{-1}(x) = a_1 x + b_1\mathsf{sign}(x)$. Logo, usando $\mathbb{H}_{\mathsf{G}}$ e o Teorema de Price [79] dá lugar a:

$$\begin{aligned} \mathbb{E}\left\{\tilde{w}_i(k)g_1[w_i(k)]\right\} &= \mathbb{E}\left\{\tilde{w}_i(k)\left[a_1 w_1^\star - a_1\tilde{w}_i(k) + b_1\mathsf{sign}[w_i(k)]\right]\right\} & (142)\\ &= -a_1\mathbb{E}\left[\tilde{w}_i^2(k)\right] - b_1\sqrt{\frac{2}{\pi}}\sqrt{\mathbb{E}\left[\tilde{w}_i^2(k)\right]}. & (143)\end{aligned}$$

[4]Sob o ponto de vista da expressão analítica, embora os domínios das funções $g_1(x)$ e $g_{-1}(x)$ sejam claramente distintos.

Finalmente, $\mathbb{E}\left\{g_1^2\left[w_i(k)\right]\right\}$, para $i \in \mathcal{C}_0$, pode ser formulado como:

$$\begin{aligned}\mathbb{E}\left\{g_1^2\left[w_i(k)\right]\right\} &= \mathbb{E}\left\{\left\{a_1 w_i(k) + b_1 \text{sign}\left[w_i(k)\right]\right\}^2\right\} \\ &= a_1^2\mathbb{E}\left[\tilde{w}_i^2(k)\right] + b_1^2 + 2a_1 b_1 \frac{\sqrt{2}}{\sqrt{\pi}}\sqrt{\mathbb{E}\left[\tilde{w}_i^2(k)\right]}. \end{aligned} \quad (144)$$

Comentário: escolhendo $a_1 = 2\alpha^2$ e $b_1 = -2\alpha$, a Equação (143) (respectivamente (144)) é equivalente à Equação (50) (respectivamente, (51)) em [19].

Tendo estes resultados, podemos revisitar a Equação (134), pode efetuar a análise para os conjuntos $\mathcal{C}_\text{S}$, $\mathcal{C}_\text{L}$ e $\mathcal{C}_0$.

- $i \in \mathcal{C}_L$:

$$2\beta\sigma_x^2\Delta_0\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] - \beta^2\sigma_x^4 D(\infty) - \beta^2\sigma_\nu^2\sigma_x^2 = 0. \quad (145)$$

Logo, somando todos os coeficientes $\mathcal{C}_\text{L}$, obtém-se:

$$2\beta\sigma_x^2\Delta_0\sum_{i\in\mathcal{C}_L}\mathbb{E}[\tilde{w}_i^2(\infty)] - \beta^2 Q_\text{L}\sigma_x^4 D(\infty) - \beta^2 Q_\text{L}\sigma_\nu^2\sigma_x^2 = 0. \quad (146)$$

- $i \in \mathcal{C}_S$:

$$\begin{aligned}(2\beta\sigma_x^2\Delta_0 - 2\kappa_j\Delta_0 - a_j^2\kappa^2)\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] &= \beta^2\sigma_x^4 D(\infty) + \beta^2\sigma_\nu^2\sigma_x^2 + \kappa^2 g_j^2[w_i^\star] \quad (147)\\ &\quad -(2\kappa\Delta_0 g_j[w_i^\star] + 2a_j\kappa^2 g_j[w_i^\star])\mathbb{E}\left[\tilde{w}_i(\infty)\right].\end{aligned}$$

Aplicando-se (123), obtém-se:

$$\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] = \frac{\beta^2\sigma_x^4 D(\infty) + \beta^2\sigma_\nu^2\sigma_x^2 + \left(\kappa^2 + \frac{2\kappa^2\Delta_0 + 2a_j\kappa^3}{\beta\sigma_x^2 - \kappa a_j}\right) g_j^2[w_i^\star]}{2\beta\sigma_x^2\Delta_0 - 2\kappa a_j\Delta_0 - a_j^2\kappa^2}, \quad (148)$$

ou

$$\mathbb{E}\left[\tilde{w}_i^2(\infty)\right] = \eta_{1,i}^\text{s} D(\infty) + \eta_{2,i}^\text{s} + \frac{\beta^2\sigma_\nu^2\sigma_x^2}{\eta_{3,i}^\text{s}}, \text{ para } i \in \mathcal{C}_\text{S}, \quad (149)$$

onde

$$\eta_{1,i}^\text{s} \triangleq \frac{\beta^2\sigma_x^4}{\eta_{3,i}^\text{s}} \quad (150)$$

$$\eta_{2,i}^\text{s} \triangleq \frac{\left(\kappa^2 + \frac{2\kappa^2\Delta_0 + 2a_j\kappa^3}{\beta\sigma_x^2 - \kappa a_j}\right) g_j^2[w_i^\star]}{\eta_{3,i}^\text{s}} \quad (151)$$

$$\eta_{3,i}^\text{s} \triangleq 2\beta\sigma_x^2\Delta_0 - 2\kappa a_j\Delta_0 - a_j^2\kappa^2, \quad (152)$$

- Para $i \in \mathcal{C}_0$:

  Considere a seguinte definição:

$$\omega \triangleq \sqrt{\mathbb{E}\left[\tilde{w}_i^2(k)\right]}, \text{ para } i \in \mathcal{C}_0. \tag{153}$$

  Combinando as Equações (134), (143) e (144) implica que:

$$\overline{\eta}_1 \omega^2 + \overline{\eta}_2 \omega + \overline{\eta}_3 - \beta^2 \sigma_x^4 D(\infty) = 0, \tag{154}$$

  onde

$$\overline{\eta}_1 \triangleq 2\beta\sigma_x^2\Delta_0 - 2\kappa\Delta_0 a_1 - \kappa^2 a_1^2, \tag{155}$$

$$\overline{\eta}_2 \triangleq -2\kappa\Delta_0 b_1\sqrt{\frac{2}{\pi}} - 2a_1 b_1 \kappa^2\sqrt{\frac{2}{\pi}}, \tag{156}$$

$$\overline{\eta}_3 \triangleq -\beta^2\sigma_\nu^2\sigma_x^2 - \kappa^2 b_1^2. \tag{157}$$

Comentário: Note que para $i \in \mathcal{C}_0$ o termo $\mathbb{E}\left[\tilde{w}_i^2(k)\right]$ se degenera em ω, não dependendo de i, pois os coeficientes desse conjunto se encontram todos na mesma faixa de atração.

A partir da Equação (154), podemos determinar uma fórmula analítica para ω:

$$\omega = \frac{-\overline{\eta}_2 + \sqrt{\overline{\eta}_2^2 - 4\overline{\eta}_1\left[\overline{\eta}_3 - \beta^2\sigma_x^4 D(\infty)\right]}}{2\overline{\eta}_1}, \tag{158}$$

o qual revela um acoplamento sofisticado entre o MSD assintótico $D(\infty)$ e ω. Cabe ressaltar que a fórmula revela que o MSD não depende apenas de coeficientes adaptativos de alta magnitude mas também daqueles penalizados pela regularização de norma, cujas variâncias são influenciadas pelos coeficientes lineares e angulares dos segmentos de reta associados à função de penalização definida pelo projetista.

A combinação dos resultados prévios nos permite obter uma fórmula fechada para o MSD assintótico, a qual pode ser descrita como:

$$D(\infty) = \sum_{i\in\mathcal{C}_0} \mathbb{E}\left[\tilde{w}_i^2(\infty)\right] + \sum_{i\in\mathcal{C}_S} \mathbb{E}\left[\tilde{w}_i^2(\infty)\right] + \sum_{i\in\mathcal{C}_L} \mathbb{E}\left[\tilde{w}_i^2(\infty)\right], \tag{159}$$

ou, após a aplicação das Equações (146), (149) e (158):

$$\begin{aligned} D(\infty) &= \frac{\beta Q_{\mathsf{L}}\sigma_x^2 D(\infty) + \beta Q_{\mathsf{L}}\sigma_\nu^2}{2\Delta_0} + Q_0\left(\frac{-\overline{\eta}_2 + \sqrt{\overline{\eta}_2^2 - 4\overline{\eta}_1\left[\overline{\eta}_3 - \beta^2\sigma_x^4 D(\infty)\right]}}{2\overline{\eta}_1}\right)^2 \\ &\quad + D(\infty)\sum_{i\in\mathcal{C}_{\mathsf{S}}}\eta_{1,i}^{\mathsf{S}} + \sum_{i\in\mathcal{C}_{\mathsf{S}}}\left(\eta_{2,i} + \frac{\beta^2\sigma_\nu^2\sigma_x^2}{\eta_{3,i}^{\mathsf{S}}}\right), \end{aligned} \tag{160}$$

implicando a seguinte igualdade:

$$\begin{aligned} &\frac{2\overline{\eta}_1^2}{Q_0\overline{\eta}_2}\left(1 - \frac{\beta Q_{\mathsf{L}}\sigma_x^2}{2\Delta_0} - \frac{Q_0\beta^2\sigma_x^4}{\overline{\eta}_1} - \sum_{i\in\mathcal{C}_S}\eta_{1,i}^s\right)D(\infty) \\ &-\frac{\beta Q_{\mathsf{L}}\sigma_\nu^2\overline{\eta}_1^2}{\Delta_0 Q_0\overline{\eta}_2} - \eta_2 + \frac{2\overline{\eta}_1\overline{\eta}_3}{\overline{\eta}_2} - \frac{2\overline{\eta}_1^2}{Q_0\overline{\eta}_2}\sum_{i\in\mathcal{C}_S}\left(\eta_{2,i}^s + \frac{\beta^2\sigma_\nu^2\sigma_x^2}{\eta_{3,i}^s}\right) \\ &= -\sqrt{\overline{\eta}_2^2 - 4\overline{\eta}_1\left[\overline{\eta}_3 - \beta^2\sigma_x^4 D(\infty)\right]}. \end{aligned} \tag{161}$$

Para facilitar as manipulações algébricas seguintes, considere as definições a seguir:

$$\zeta_1 \triangleq \frac{2\overline{\eta}_1^2}{Q_0\overline{\eta}_2}\left(1 - \frac{\beta Q_{\mathsf{L}}\sigma_x^2}{2\Delta_0} - \frac{Q_0\beta^2\sigma_x^4}{\overline{\eta}_1} - \sum_{i\in\mathcal{C}_S}\eta_{1,i}^s\right), \tag{162}$$

$$\zeta_2 \triangleq -\frac{\beta Q_{\mathsf{L}}\sigma_\nu^2\overline{\eta}_1^2}{\Delta_0 Q_0\overline{\eta}_2} - \eta_2 + \frac{2\overline{\eta}_1\overline{\eta}_3}{\overline{\eta}_2} - \frac{2\overline{\eta}_1^2}{Q_0\overline{\eta}_2}\sum_{i\in\mathcal{C}_S}\left(\eta_{2,i}^s + \frac{\beta^2\sigma_\nu^2\sigma_x^2}{\eta_{3,i}^s}\right), \tag{163}$$

$$\zeta_3 \triangleq \overline{\eta}_2^2 - 4\overline{\eta}_1\overline{\eta}_3, \tag{164}$$

$$\zeta_4 \triangleq 4\beta^2\overline{\eta}_1\sigma_x^4. \tag{165}$$

Logo:

$$\zeta_1 D(\infty) + \zeta_2 = -\sqrt{\zeta_3 + \zeta_4 D(\infty)}, \tag{166}$$

cuja solução analítica é dada por (126), completando a prova do Teorema 1. ■

Comentário: A Equação (126) revela a sofisticação dos fatores que concorrem para o desempenho do algoritmo GPLR-LMS. Desafortunadamente, a fórmula analítica obtida é demasiadamente complexa, o que dificulta a extração de percepções simples acerca de impactos de escolhas de parâmetros por parte do projetista. Por outro lado, especialmente quando são adotados filtros com comprimentos muito longos, a fórmula (126) requer muito menos custo computacional para prever o desempenho assintótico do

algoritmo do que simulações computacionais baseadas em ensaios independentes de Monte Carlo. Na próxima seção, algumas aproximações serão introduzidas para facilitar a obtenção de uma fórmula fechada para o valor de κ que otimiza o desempenho assintótico do algoritmo.

4.3- Modelo estocástico simplificado

Esta seção introduz uma versão simplificada do modelo estocástico apresentado na Seção 4.2. O modelo simplificado é mobilizado para obter uma fórmula fechada para o valor ótimo de κ que garante desempenho assintótico do algoritmo GPLR-LMS. As simplificações serão primariamente baseadas na adoção adicional de $\mathbb{H}_\kappa$, discutida no Capítulo 3.

Logo, consideremos as seguintes definições simplificadas:

$$\overline{\eta}'_1 \triangleq 2\beta\sigma_x^2\Delta_0, \tag{167}$$

$$\overline{\eta}'_2 \triangleq \chi_1\kappa, \tag{168}$$

$$\chi_1 \triangleq -2\Delta_0 b_1\sqrt{\frac{2}{\pi}}, \tag{169}$$

$$\chi_2 \triangleq 2\beta\sigma_x^2\Delta_0, \tag{170}$$

$$\overline{\eta}'_3 \triangleq -\beta^2\sigma_\nu^2\sigma_x^2, \tag{171}$$

$$\chi_3 \triangleq \frac{\chi_2}{\beta^2\sigma_x^2}, \tag{172}$$

$$\hat{\chi}_j \triangleq \frac{\overline{\chi}_j}{\beta^2\sigma_x^2}, \tag{173}$$

$$\eta_{1,i}^{\mathsf{s}'} \triangleq \frac{1}{\chi_3 + \hat{\chi}_j\kappa^2}, \tag{174}$$

$$\breve{\chi}_j \triangleq \frac{\beta\sigma_x^2\chi_2}{2\Delta_0 g_j^2\left[w_i^\star\right]}, \tag{175}$$

$$\dot{\chi}_j \triangleq \frac{\beta\sigma_x^2\overline{\chi}_j}{2\Delta_0 g_j^2\left[w_i^\star\right]}, \tag{176}$$

$$\eta_{2,i}^{\mathsf{s}'} \approx \frac{\kappa^2}{\breve{\chi}_j + \dot{\chi}_j\kappa^2}, \tag{177}$$

$$\overline{\chi}_j \triangleq -a_j^2, \tag{178}$$

$$\eta_{3,i}^{s'} \triangleq \chi_2 + \overline{\chi}_j \kappa^2, \tag{179}$$

$$\chi_4 \triangleq \frac{Q_0 \chi_1}{8\beta^2 \sigma_x^4 \Delta_0^2 \left(1 - \frac{Q_0 \beta^2 \sigma_x^4}{2\beta \sigma_x^2 \Delta_0}\right)}, \tag{180}$$

$$\Psi_1 \triangleq \frac{Q_0 \chi_1 \chi_3}{8\beta^2 \sigma_x^4 \Delta_0^2}, \tag{181}$$

$$\dot{\Psi}_j \triangleq \frac{Q_0 \chi_1 \hat{\chi}_j}{8\beta^2 \sigma_x^4 \Delta_0^2}, \tag{182}$$

$$\zeta_1' \triangleq \frac{1}{\chi_4 \kappa} - \sum_{i \in \mathcal{C}_S} \frac{1}{\Psi_1 \kappa + \dot{\Psi}_j \kappa^3}. \tag{183}$$

Para efeitos de simplificação adicional do termo ζ_1', a seguinte aproximação será efetuada:

$$\Psi_1 \kappa + \dot{\Psi}_j \kappa^3 \approx \Psi_1 \kappa. \tag{184}$$

A aplicação de (184) em (183) resulta em:

$$\zeta_1' \approx \frac{1}{\chi_4 \kappa} - \sum_{i \in \mathcal{C}_S} \frac{1}{\Psi_1 \kappa} = \frac{1}{\chi_4 \kappa} - \frac{Q_S}{\Psi_1 \kappa} = \frac{\Psi_1 - Q_S \chi_4}{\chi_4 \Psi_1 \kappa} = \frac{1}{\Omega_0 \kappa}, \tag{185}$$

com

$$\Omega_0 \triangleq \frac{\chi_4 \Psi_1}{\Psi_1 - Q_{\mathsf{S}} \chi_4}. \tag{186}$$

Com relação a ζ_2, as aproximações seguintes serão adotadas:

$$\begin{aligned} \zeta_2' &\approx -\overline{\eta}_2 + \frac{2\overline{\eta}_1 \overline{\eta}_3}{\overline{\eta}_2} - \frac{2\overline{\eta}_1^2}{Q_0 \overline{\eta}_2} \sum_{i \in \mathcal{C}_S} \left(\eta_{2,i}^s + \frac{\beta^2 \sigma_\nu^2 \sigma_x^2}{\eta_{3,i}^s} \right) \\ &= -\chi_1 \kappa + \frac{4\beta \sigma_x^2 \Delta_0 (-\beta^2 \sigma_\nu^2 \sigma_x^2)}{\chi_1} \frac{1}{\kappa} - \frac{2(2\beta \sigma_x^2 \Delta_0)^2}{Q_0 \chi_1} \frac{1}{\kappa} \sum_{i \in \mathcal{C}_S} \frac{\kappa^2}{\breve{\chi}_j + \dot{\chi}_j \kappa^2} \\ &\quad - \frac{2(2\beta \sigma_x^2 \Delta_0)^2}{Q_0 \chi_1} \frac{1}{\kappa} \sum_{i \in \mathcal{C}_S} \frac{\beta^2 \sigma_\nu^2 \sigma_x^2}{\chi_2 + \overline{\chi}_j \kappa^2}. \end{aligned} \tag{187}$$

Usando as aproximações: $\breve{\chi}_j + \dot{\chi}_j \kappa^2 \approx \tilde{\chi}_j$ e $\chi_2 + \overline{\chi}_j \kappa^2 \approx \chi_2$, as seguintes expressões são

obtidas:

$$\begin{aligned}
\zeta_2' &\approx -\chi_1\kappa - \frac{4\beta^3\sigma_x^4\Delta_0\sigma_\nu^2}{\chi_1}\frac{1}{\kappa} - \frac{8\beta^2\sigma_x^4\Delta_0^2}{Q_0\chi_1}\frac{1}{\kappa}\sum_{i\in\mathcal{C}_S}\left(\frac{1}{\breve{\chi}_j}\kappa^2 + \frac{\beta^2\sigma_\nu^2\sigma_x^2}{\chi_2}\right) \\
&= \underbrace{\left(-\chi_1 - \frac{8\beta^2\sigma_x^4\Delta_0^2}{Q_0\chi_1}\sum_{i\in\mathcal{C}_S}\frac{1}{\breve{\chi}_j}\right)}_{\triangleq\Omega_1}\kappa + \underbrace{\left(-\frac{4\beta^3\sigma_x^4\Delta_0\sigma_\nu^2}{\chi_1} - \frac{8\beta^4\sigma_x^6\Delta_0^2\sigma_\nu^2}{Q_0\chi_1\chi_2}Q_S\right)}_{\triangleq\Omega_2}\frac{1}{\kappa} \\
&= \Omega_1\kappa + \Omega_2\frac{1}{\kappa}. \qquad (188)
\end{aligned}$$

$$\zeta_3' \triangleq \overline{\eta}_2'^2 - 4\overline{\eta}_1'\overline{\eta}_3' = \chi_1^2\kappa^2 + \Omega_3, \qquad (189)$$

com

$$\Omega_3 \triangleq 8\beta^3\sigma_x^4\Delta_0\sigma_\nu^2, \qquad (190)$$

$$\zeta_4' \triangleq 4\beta^2\overline{\eta}_1'\sigma_x^4. \qquad (191)$$

Comentário: Cumpre notar que ζ_4' não depende de κ, podendo ser considerada uma constante quando efetuamos uma otimização em relação a κ.

A conjugação das simplificações pregressas, permite expressar o MSD assintótico como

$$D(\infty) \approx -\frac{\zeta_2'}{\zeta_1'} + \frac{\zeta_4' - 2\sqrt{\zeta_1'^2\zeta_3' - \zeta_1'\zeta_2'\zeta_4'}}{2\zeta_1'^2}. \qquad (192)$$

O resultado (192) é a base para o Teorema 2, a seguir explicitado, que é a segunda contribuição (em termos de modelagem estocástica) deste capítulo:

Teorema 2. A identidade (192) implica que o valor de κ que otimiza o comportamento assintótico do algoritmo pode ser computado por meio da seguinte fórmula fechada:

$$\kappa_{\text{ótimo}} = \frac{\sqrt{2}}{2}\sqrt{\frac{\Omega_6(\Omega_4(\Omega_4 + \sqrt{\Omega_4^2 - \Omega_0^4\Omega_5}) + \Omega_0^4\Omega_5 - 2\Omega_4^2)}{\Omega_5(\Omega_4^2 - \Omega_0^4\Omega_5)}}, \qquad (193)$$

cujos termos Ω_4, Ω_5 e Ω_6 serão definidos na demonstração do Teorema, a qual é efetuada a seguir.

Demonstração: Quando expressamos $D(\infty)$ em termos de κ, o resultado é:

$$D(\infty) = \underbrace{\left(\frac{\zeta_4\Omega_0^2}{2} - \Omega_0\Omega_1\right)}_{\triangleq\Omega_4}\kappa^2 - \Omega_0\Omega_2 - \Omega_0^2\kappa^2\sqrt{\underbrace{\left(\frac{\chi_1^2}{\Omega_0^2} - \frac{\Omega_1\zeta_4}{\Omega_0}\right)}_{\triangleq\Omega_5} + \underbrace{\left(\frac{\Omega_3}{\Omega_0^2} - \frac{\Omega_2\zeta_4}{\Omega_0}\right)}_{\triangleq\Omega_6}\frac{1}{\kappa^2}}. \qquad (194)$$

Logo:

$$D(\infty) = \Omega_4\kappa^2 - \Omega_0\Omega_2 - \Omega_0^2\kappa^2\sqrt{\Omega_5 + \Omega_6\frac{1}{\kappa^2}}. \tag{195}$$

Para determinar o valor ótimo de κ, cumpre-nos calcular a derivada parcial de (195) com relação a κ, dando origem a:

$$\frac{\partial D(\infty)}{\partial\kappa} = 2\Omega_4\kappa - 2\Omega_0^2\kappa\sqrt{\Omega_5 + \Omega_6/\kappa^2} + \frac{\Omega_0^2\Omega_6}{\kappa\sqrt{\Omega_5 + \Omega_6/\kappa^2}}. \tag{196}$$

Impondo a condição de que $\frac{\partial D(\infty)}{\partial\kappa}$ seja nulo de modo a encontrar o MSD mínimo, obtém-se a identidade (193), completando a prova do Teorema 2. ■

Comentários: O principal resultado desta seção é a Equação (193), a qual é bastante genérica porque, como vimos, diversos algoritmos são casos especiais do algoritmo GPLR-LMS proposto (veja a Tabela 5). Esta generalidade, entretanto, tem um custo: a fórmula obtida é algebricamente complexa, embora tenha o potencial de permitir a obtenção de políticas que especifiquem valores de κ variantes no tempo, desde que determinadas aproximações sejam feitas, as quais garantam que grandezas tipicamente observáveis durante a execução do algoritmo (tais como o sinal de erro ou os próprios valores atuais dos coeficientes adaptativos) sejam utilizadas para modular os valores de κ. Tal política de variação pode ser benéfica em configurações nas quais a esparsidade da planta a identificar é variante no tempo.

4.4- Resultados

As principais contribuições deste capítulo são: *i)* a proposição de um algoritmo bastante flexível GPLR-LMS, o qual generaliza diversos algoritmos prévios na literatura; *ii)* a obtenção de uma fórmula analítica que prediz o desempenho assintótico do algoritmo; *iii)* a obtenção de uma fórmula analítica que permite computar o valor ótimo de κ ($\kappa_{\text{ótimo}}$) que otimiza o desempenho assintótico do algoritmo GPLR-LMS. As fórmulas encontradas - Equações (126) e (193) - demandaram algumas hipóteses estatísticas, de modo que é necessário cotejá-las com resultados de experimentos computacionais. Em consonância com a literatura científica em filtragem adaptativa, os resultados destes experimentos

serão chamados de resultados empíricos, a despeito de terem sido obtidos num ambiente puramente computacional.

Para validar os modelos analíticos propostos, será apresentado um conjunto de simulações comparando as previsões teóricas com os resultados empíricos. Essas simulações têm como objetivo avaliar o impacto de parâmetros-chave, incluindo a variância do ruído branco Gaussiano aditivo σ_ν^2, a taxa de aprendizado β e a variância do sinal de entrada σ_x^2, também assumido como seguindo uma distribuição gaussiana branca. Ao variar sistematicamente esses parâmetros, avalia-se a precisão das expressões teóricas do MSD e sua capacidade de descrever o comportamento do algoritmo sob diferentes condições de operação. Ao longo desta seção, utilizamos a seleção de parâmetros $a_1 = 0{,}9091$, $a_2 = 0{,}0178$, $a_3 = 0{,}0039$, $b_1 = -1$, $b_2 = -0{,}1087$, $b_3 = -0{,}039$, $t_1 = 1$, $t_2 = 5$, $t_3 = 10$ e $P = 3$.

Esses valores adotados para os parâmetros a_j, b_j, t_j e P, que definem a função penalizante linear por partes utilizada nesta seção, foram extraídos com base na curva da penalização do algoritmo RZA-LMS proposto em [31] (ver Figura 10). Estes valores foram definidos de forma a aproximar, com boa precisão gráfica e numérica, o comportamento da função do RZA-LMS para o valor de $\epsilon = 10$, o mesmo adotado nos experimentos do artigo original. Outros parâmetros escolhidos (e mantidos constantes) obedeceram aos mesmos critérios metodológicos explicados na seção 3.4.

Para avaliar a capacidade do modelo teórico em acompanhar o desempenho assintótico do algoritmo para diferentes valores de β, foi criado um cenário utilizando o segundo modelo de [81] ($N = 96$), com 100 zeros concatenados ao final. A variância do ruído aditivo foi $\sigma_\nu^2 = 10^{-6}$, a intensidade do fator de penalização de norma foi $\kappa = 10^{-7}$, o número de iterações foi fixado em 50.000, e a variância do sinal de entrada Gaussiano branco foi $\sigma_x^2 = 1$. O resultado pode ser observado na Figura 11. Nota-se uma concordância notavelmente alta entre os resultados empíricos e os valores previstos teoricamente (com erro máximo inferior a 0,1 dB), demonstrando a elevada precisão da fórmula analítica proposta (Equação (126)) para o MSD em regime permanente, com diferentes valores de β.

Para avaliar se o modelo teórico captura o impacto da variância do sinal de entrada σ_x^2, foram realizadas simulações utilizando o segundo modelo de [81] (com a concatenação de 100 zeros ao final), com $\sigma_\nu^2 = 10^{-5}$, $\kappa = 10^{-7}$, $\beta = 0{,}005$ e σ_x^2 variando no intervalo de 0,1 a 1,5. O número de iterações em cada execução do algoritmo foi

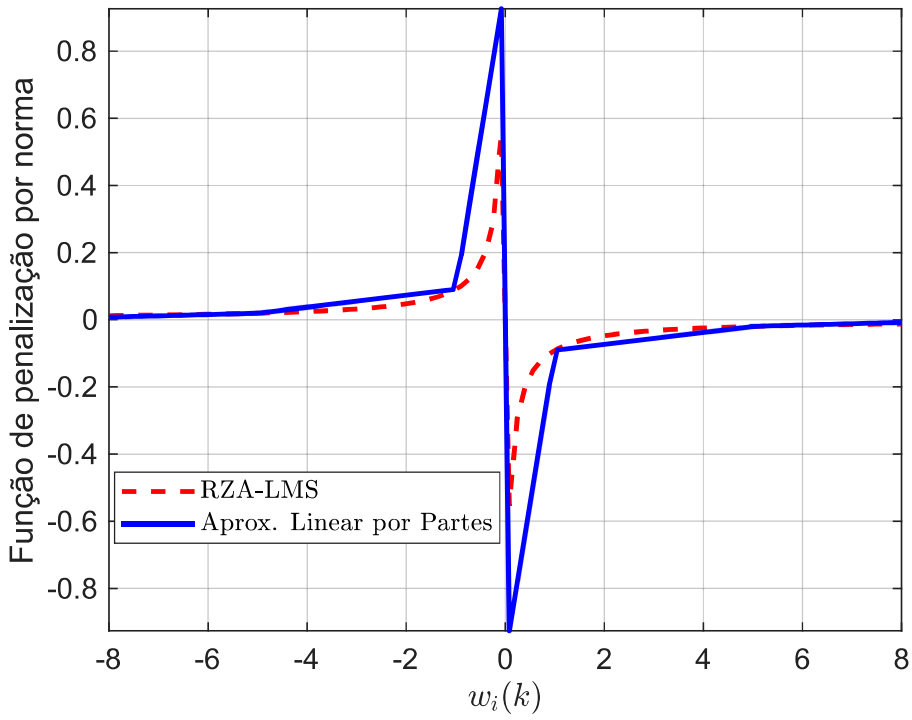


Figura 10 – Comparação entre o termo de penalização de norma do algoritmo RZA-LMS [31] (linha vermelha tracejada, com $\epsilon = 10$, ver Equação (112)) e a aproximação linear por partes adotada nas simulações desta seção (linha azul contínua). O valor de ϵ foi escolhido de forma a se alinhar com as simulações apresentadas em [31].

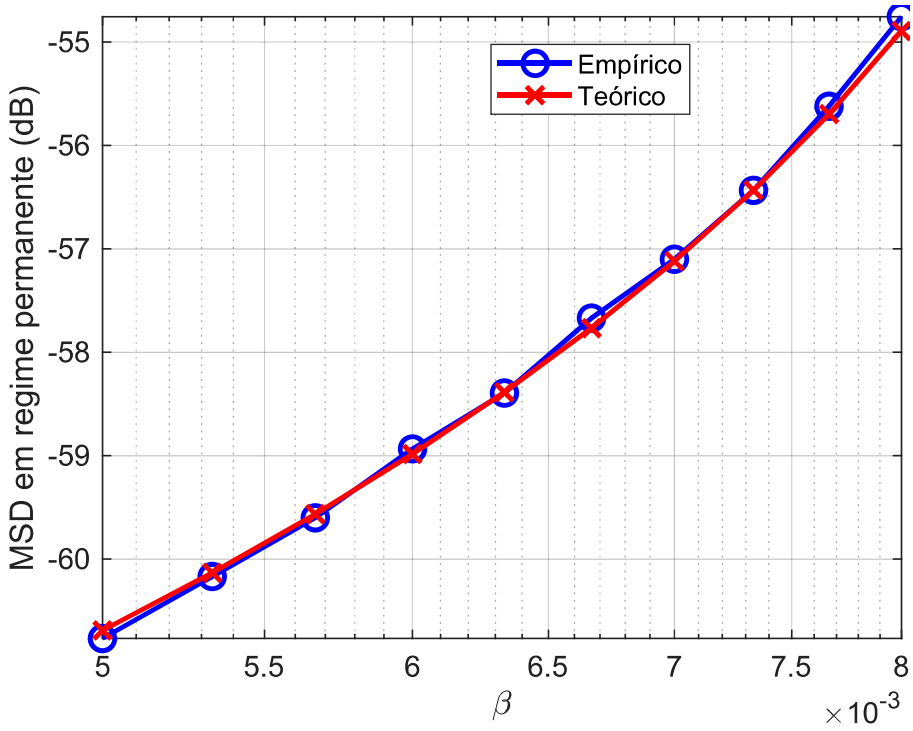


Figura 11 – MSD em regime permanente (em dB), empírico (azul) e teórico (vermelho), em função de β. O MSD empírico foi obtido a partir de $K = 100$ execuções independentes de Monte Carlo. Em cada execução, foi calculada a média das últimas 1000 iterações.

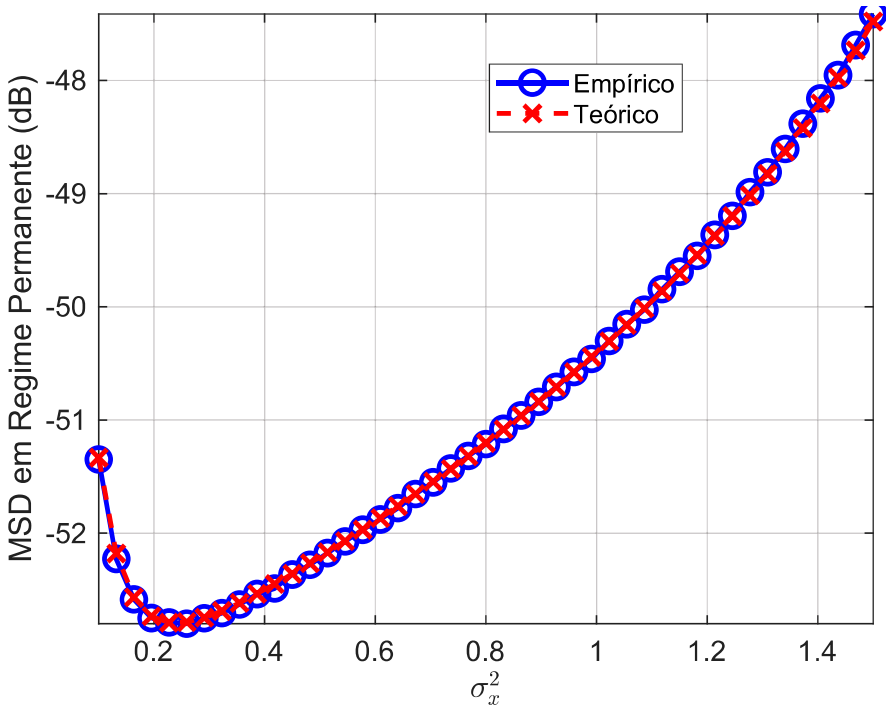


Figura 12 – MSD em regime permanente (em dB), empírico (azul) e teórico (vermelho), em função de σ_x^2. O MSD empírico foi obtido a partir de $K = 1000$ execuções independentes de Monte Carlo. Em cada execução, foi calculada a média das últimas 1000 iterações.

fixado em 50.000. A comparação entre os resultados teóricos (ver Equação (126)) e aqueles obtidos por meio de simulação pode ser observada na Figura 12. As simulações revelam um comportamento não monótono do MSD assintótico em relação a σ_x^2, o qual é adequadamente capturado pelo modelo analítico proposto.

A capacidade da fórmula analítica (126) de modelar o impacto da variação na variância do ruído σ_ν^2 é avaliada na Figura 13. Neste experimento, o sistema a ser identificado é o Modelo 3 de [81] ($N = 96$), seguido por 10 zeros. A variância do ruído varia de 10^{-5} a 10^{-2}. A variância do sinal de entrada gaussiano branco é unitária, $\kappa = 10^{-6}$ e $\beta = 0,01$. O número de iterações realizadas pelo algoritmo em cada execução é de 50.000. Observa-se uma forte concordância entre o modelo teórico e os resultados experimentais na Figura 13, com uma discrepância máxima observada de 0,17 dB.

Adicionalmente, é examinada a escolha ótima do peso de penalização κ, o qual equilibra a promoção de esparsidade e o desempenho de convergência no processo de adaptação. O valor ótimo de κ é obtido de forma analítica (por meio da Equação (193)) e validado por meio de resultados numéricos, demonstrando seu papel na minimização do desvio quadrático médio assintótico.

Nas Figuras 14–17, adotam-se os valores $\sigma_\nu^2 = 10^{-6}$, $\beta = 0{,}001$ e $\sigma_x^2 = 1$. Um

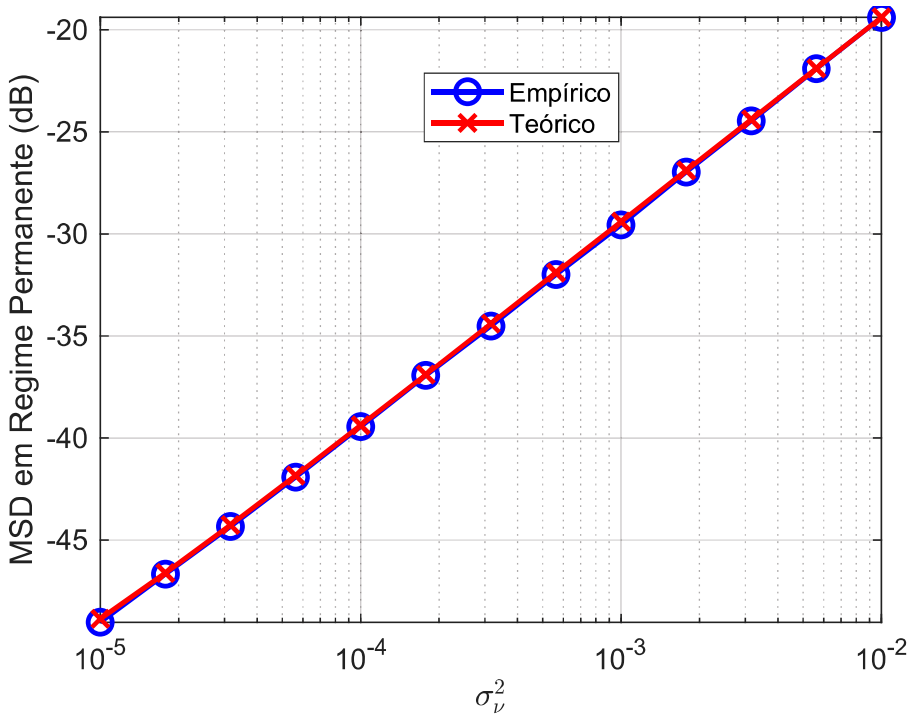


Figura 13 – MSD(em dB) em regime permanente empírico (em azul) e teórico (em vermelho), como uma função de σ_ν^2. O MSD empírico foi obtido por meio de $K = 100$ ensaios independentes de Monte Carlo. Em cada ensaio de Monte Carlo, a média das últimas 1000 iterações foi computada.

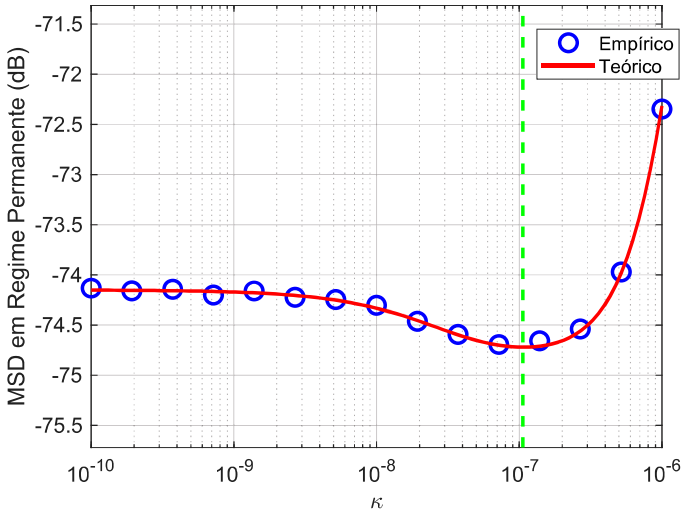


Figura 14 – MSD em regime permanente (em dB), empírico (círculos azuis) e teórico (curva vermelha), como uma função de κ. O valor teórico ótimo de κ computado pela Equação (193) é indicado por uma linha verde vertical tracejada. O MSD empírico foi obtido por meio de $K = 100$ ensaios independentes de Monte Carlo. Em cada ensaio de Monte Carlo, a média das últimas 10000 iterações foi computada.

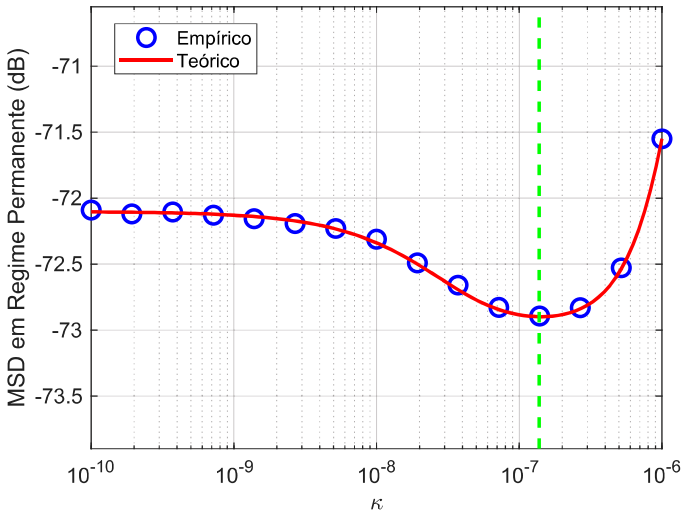


Figura 15 – MSD em regime permanente (em dB), empírico (círculos azuis) e teórico (curva vermelha), como uma função de κ. O valor teórico ótimo de κ computado pela Equação (193) é indicado por uma linha verde vertical tracejada. O MSD empírico foi obtido por meio de $K = 100$ ensaios independentes de Monte Carlo. Em cada ensaio de Monte Carlo, a média das últimas 10000 iterações foi computada.

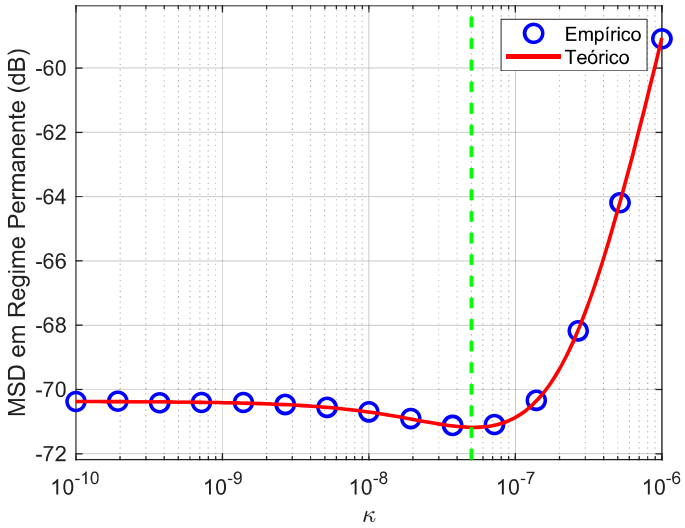


Figura 16 – MSD em regime permanente (em dB), empírico (círculos azuis) e teórico (curva vermelha), como uma função de κ. O valor teórico ótimo de κ computado pela Equação (193) é indicado por uma linha verde vertical tracejada. O MSD empírico foi obtido por meio de $K = 100$ ensaios independentes de Monte Carlo. Em cada ensaio de Monte Carlo, a média das últimas 10000 iterações foi computada.

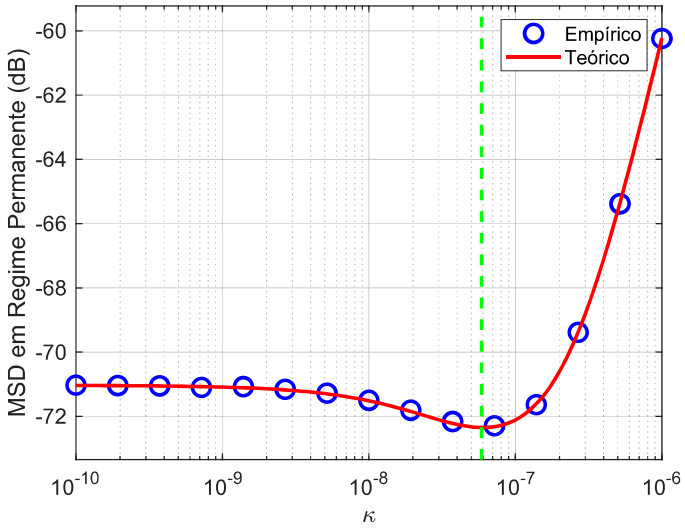


Figura 17 – MSD em regime permanente (em dB), empírico (círculos azuis) e teórico (curva vermelha), como uma função de κ. O valor teórico ótimo de κ computado pela Equação (193) é indicado por uma linha verde vertical tracejada. O MSD empírico foi obtido por meio de $K = 100$ ensaios independentes de Monte Carlo. Em cada ensaio de Monte Carlo, a média das últimas 10000 iterações foi computada.

total de 60.000 iterações foi realizado em cada execução do algoritmo. A Figura 14 apresenta o MSD assintótico quando o sistema é o primeiro modelo de [81] ($N = 64$), concatenado com 10 zeros. A Figura 15 foi gerada utilizando o segundo sistema de [81] ($N = 96$), concatenado com 20 zeros. A Figura 16 utilizou o quarto modelo de referência de [81] ($N = 128$), concatenado com 40 zeros. Por fim, a Figura 17 utilizou o quinto modelo de [81] ($N = 96$), concatenado com 50 zeros finais. O objetivo principal dessas figuras não é avaliar diretamente o ganho em dB nem a taxa de convergência, mas sim demonstrar como a fórmula analítica proposta (isto é, a Equação (193)) determina com precisão o valor ótimo de κ que otimiza o desempenho assintótico. Importa notar que os parâmetros da função de penalização linear por partes podem ser otimizados, dependendo da aplicação, para melhorar ainda mais o ganho assintótico, como é comum em algoritmos adaptativos com penalização por norma.

4.5- Conclusões

Este capítulo propõe um algoritmo adaptativo com penalização de norma generalizado, chamado de GPLR-LMS. Também desenvolve modelos estocásticos capazes de estimar o desempenho assintótico deste algoritmo. Também é introduzido um modelo estocástico simplificado para derivar uma expressão analítica para o valor ótimo do parâmetro κ, o qual assegura o melhor desempenho assintótico do algoritmo proposto. Por meio de uma série de aproximações e suposições razoáveis, conseguimos reduzir a complexidade do modelo e obter uma equação em forma fechada para $\kappa_{\text{ótimo}}$, facilitando sua aplicação, para o projetista.

Os resultados da análise destacam os compromissos inerentes à escolha de κ e evidenciam o delicado equilíbrio entre viés e desempenho assintótico. Embora as fórmulas obtidas forneçam uma aproximação útil, elas também revelam os desafios enfrentados ao se tentar modelar as sofisticadas capacidades de aprendizagem dos algoritmos adaptativos. Essas percepções são essenciais para o avanço da compreensão e da otimização de algoritmos de filtragem adaptativa em ambientes ruidosos e sistemas esparsos.

5- Conclusões e Trabalhos Futuros

A ubiquidade da esparsidade em diversos sistemas motiva o estudo de algoritmos adaptativos de identificação de sistemas que penalizem soluções pouco esparsas, para evitar uma perda de desempenho. Assim, a literatura científica sobre algoritmos adaptativos com penalização de norma é muito extensa. Esta tese enfocou em algoritmos de filtragem adaptativa com penalização de norma, os quais tendem a apresentar melhor compromisso entre desempenho assintótico e taxa de convergência.

O algoritmo ℓ_0-LMS é um dos mais eficazes na promoção de uma estimativa pouco enviesada de uma resposta ao impulso esparsa. A literatura, entretanto, não elucida as habilidades e características de seu comportamento quando a planta a identificar é variante no tempo. Nesta tese, tal lacuna foi preenchida assumindo que a planta a identificar encontra-se submetida a uma perturbação markoviana. Sob esta hipótese (e mais outras, tais como a de que o sinal de entrada é branco), utilizou-se um modelo estocástico capaz de oferecer uma fórmula fechada não matricial que prediz o desempenho assintótico (em termos de desvio quadrático médio) do algoritmo ℓ_0-LMS. A fórmula encontrada consiste de quatro termos. O primeiro deles consiste exatamente no MSD assintótico do algoritmo LMS. O segundo deles independe da não estacionariedade da planta. O terceiro termo reflete uma contribuição rigorosamente aditiva que a variância da perturbação markoviana apresenta no MSD assintótico. Já o último termo ($-\theta_2\kappa\sqrt{\kappa^2+\theta_3'}$) sinaliza uma contribuição sofisticada induzida pela variância da perturbação markoviana. Além disso, para a maioria dos cenários simulados, o erro do modelo teórico na previsão do desvio quadrático médio permanece abaixo de 1 dB, mesmo quando o passo de aprendizagem varia em várias ordens de magnitude.

O Capítulo 4 generaliza diversos algoritmos de filtragem adaptativa com penalização de norma propondo uma penalização de norma linear por partes. Mesmo quando algoritmos que não utilizam uma penalização linear por partes podem ser emulados, com uma acurácia pré-estabelecida, pelo paradigma proposto (como mostramos no caso do algoritmo RZA-LMS, vide Figura 10). O fato de o algoritmo proposto ser bastante flexível (permitindo a inclusão de um número qualquer de segmentos de reta no termo de penalização de norma) torna sua modelagem muito desafiadora, exigindo extensas

manipulações algébricas. Mesmo a escolha de hipóteses é uma arte, que deve levar em consideração a obtenção de fórmulas ao mesmo tempo acuradas e concisas. As hipóteses adotadas permitiram encontrar uma fórmula analítica escalar capaz de prever o MSD assintótico do algoritmo GPLR-LMS, dando origem ao Teorema 1 do Capítulo 4. Após algumas simplificações (bem como a adoção da hipótese $\mathbb{H}_\kappa$), conseguimos obter (no Teorema 2) uma fórmula analítica que permite computar o valor ótimo de κ que otimiza o desempenho assintótico do algoritmo. Todas as fórmulas foram extensamente cotejadas com resultados de experimentos computacionais, revelando uma boa aderência a eles, com um desvio máximo na faixa de 0,1 dB.

Entendemos que futuras direções promissoras de pesquisa residem na modelagem de versões normalizadas do algoritmo ℓ_0-LMS e no projeto otimizado de funções penalizadoras de norma mediante informações *a priori* eventualmente disponíveis sobre os sistemas nos quais o algoritmo GPLR-LMS irá operar. Isso pode permitir o ajuste automático dos segmentos da função penalizante durante a operação, levando a um comportamento mais eficiente em cenários variantes no tempo.

Além disso, reconhece-se como limitação desta contribuição a complexidade analítica das expressões finais, que, apesar de rigorosas, envolvem um número elevado de variáveis, aproximações e simplificações intermediárias. Como consequência, a extração de *insights* diretos sobre o comportamento do algoritmo pode se tornar restrita em certas situações.

Assim, uma possível extensão futura consiste na exploração de abordagens alternativas, como métodos de análise assintótica, simplificações guiadas por simulação ou mesmo técnicas baseadas em aprendizado de máquina interpretável, com o objetivo de tornar os modelos analíticos mais acessíveis e úteis como ferramenta de projeto e ajuste de algoritmos adaptativos.

Por fim, também se considera promissor o uso de respostas simuladas (como canais acústicos artificiais ou sistemas digitais controlados) para estudar o impacto de diferentes configurações da função $g[\cdot]$ e, assim, fundamentar o projeto dos segmentos lineares com base em desempenho teórico e experimental. Essas possibilidades podem beneficiar-se diretamente dos modelos analíticos desenvolvidos nesta tese.

Referências Bibliográficas

[1] S. Qiao, Y. Fan, G. Wang, D. Mu e Z. He, “Strong tracking square-root modified sliding-window variational adaptive Kalman filtering with unknown noise covariance matrices,” *Signal Processing*, v. 204, p. 108 837, 2023.

[2] J. He, G. Wang, X. Zhang, H. Wang e B. Peng, “Maximum total generalized correntropy adaptive filtering for parameter estimation,” *Signal Processing*, v. 203, p. 108 787, 2023.

[3] O. M. Abdelrhman e L. Sen, “Robust adaptive filtering algorithms based on the half-quadratic criterion,” *Signal Processing*, v. 202, p. 108 775, 2023.

[4] D. Liu e H. Zhao, “A family of variable step-size sparsity-aware SSAF algorithms with individual-weighting-factors under model-driven method,” *Journal of the Franklin Institute*, v. 359, n. 17, pp. 10 172–10 205, 2022.

[5] M. T. Khan e R. A. Shaik, “High-Performance VLSI Architecture of DLMS Adaptive Filter for Fast-Convergence and Low-MSE,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 69, n. 4, pp. 2106–2110, 2022.

[6] T. Yu, W. Li, R. C. de Lamare e Y. Yu, “M-estimate affine projection spline adaptive filtering algorithm: Analysis and implementation,” *Digital Signal Processing*, v. 123, p. 103 452, 2022.

[7] H. Buchner, K. Helwani, B. I. Ahmad e S. Godsill, “Efficient adaptive filtering in compressive domains for sparse systems and relation to transform-domain adaptive filtering,” em *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 3859–3863.

[8] K. Kumar, R. Pandey, S. S. Bora e N. V. George, “A Robust Family of Algorithms for Adaptive Filtering Based on the Arctangent Framework,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 69, n. 3, pp. 1967–1971, 2022.

[9] Y. Zong e J. Ni, “Cluster-sparsity-induced affine projection algorithm and its variable step-size version,” *Signal Processing*, v. 195, p. 108 490, 2022.

[10] T. Liang, Y. Li, X. Han, W. Xue e H. Tu, “Channel Estimation and Beamforming Using Constrained q-Rényi Kernel Functioned Adaptive Algorithm,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 69, n. 4, pp. 2356–2360, 2022.

[11] Z. Huang, Y. Yu, R. C. de Lamare, Y. Fan e K. Li, “Proportionate M-estimate adaptive filtering algorithms: Insights and improvements,” *Signal Processing*, v. 200, p. 108 671, 2022.

[12] F. Luis Perez, C. André Pitz e R. Seara, “A two-gain NLMS algorithm for sparse system identification,” *Signal Processing*, v. 200, p. 108 636, 2022.

[13] D. Liu e H. Zhao, “Sparsity-Aware Logarithmic Hyperbolic Cosine Normalized Subband Adaptive Filter Algorithm With Step-Size Optimization,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 69, n. 9, pp. 3964–3968, 2022.

[14] Y. Gu, J. Jin e S. Mei, “l_0 Norm Constraint LMS Algorithm for Sparse System Identification,” *IEEE Signal Processing Letters*, v. 16, n. 9, pp. 774–777, 2009.

[15] S. Radhika, F. Albu e A. Chandrasekar, “Proportionate Maximum Versoria Criterion-Based Adaptive Algorithm for Sparse System Identification,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 69, n. 3, pp. 1902–1906, 2022.

[16] Y. Sun, J. Chen e K. K. Parhi, “Multi-delay block frequency domain adaptive filters with sparse partial subblock update,” em *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, 2009, pp. 206–209.

[17] K. J. Bakri, E. V. Kuhn, M. V. Matsuo e R. Seara, “On the behavior of a combination of adaptive filters operating with the NLMS algorithm in a nonstationary environment,” *Signal Processing*, v. 196, p. 108 465, 2022.

[18] K. da S. Olinto, D. B. Haddad e M. R. Petraglia, “Transient analysis of l0-LMS and l0-NLMS algorithms,” *Signal Processing*, v. 127, pp. 217–226, 2016.

[19] G. Su, J. Jin, Y. Gu e J. Wang, “Performance Analysis of l_0 Norm Constraint Least Mean Square Algorithm,” *IEEE Transactions on Signal Processing*, v. 60, n. 5, pp. 2223–2235, 2012.

[20] G. Zhang, M. Yu, Y.-J. Liu, G. Zhao, D. Zhang e W. Zheng, “SparseDGCNN: Recognizing Emotion From Multichannel EEG Signals,” *IEEE Transactions on Affective Computing*, v. 14, n. 1, pp. 537–548, 2023.

[21] M. Quasdane, H. Ramchoun e T. Masrour, “Sparse smooth group L0L1/2 regularization method for convolutional neural networks,” *Knowledge-Based Systems*, v. 284, p. 111 327, 2024.

[22] F. Chu, G. Wang, J. Wang, C. P. Chen e X. Wang, “Learning broad learning system with controllable sparsity through L0 regularization,” *Applied Soft Computing*, v. 136, p. 110 068, 2023.

[23] Z. Wei, Q. Li, J. Wei e W. Bian, “Neural network for a class of sparse optimization with L0-regularization,” *Neural Networks*, v. 151, pp. 211–221, 2022.

[24] L. Xu, S. Zheng e J. Jia, “Unnatural L0 Sparse Representation for Natural Image Deblurring,” em *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1107–1114.

[25] H.-X. Wen, S.-Q. Yang, Y.-Q. Hong e H. Luo, “A Partial Update Adaptive Algorithm for Sparse System Identification,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 28, pp. 240–255, 2020.

[26] R. A. Dyba e P. P. He, “Identification of Multi-Reflective Echo Paths using Sub-Rate Adaptive Filtering in Presence of Moderate Aliasing,” em *15th International Conference on Digital Signal Processing*, 2007, pp. 51–54.

[27] K. Nishiyama, “An H/sub /spl infin// optimization and its fast algorithm for time-variant system identification,” *IEEE Transactions on Signal Processing*, v. 52, n. 5, pp. 1335–1342, 2004.

[28] S. Kalluri e G. Arce, “Adaptive weighted myriad filter algorithms for robust signal processing in /spl alpha/-stable noise environments,” *IEEE Transactions on Signal Processing*, v. 46, n. 2, pp. 322–334, 1998.

[29] I. Semushin, “The APA based time-variant system identification,” em *53rd IEEE Conference on Decision and Control*, 2014, pp. 4137–4141.

[30] D. Duttweiler, “Proportionate normalized least-mean-squares adaptation in echo cancelers,” *IEEE Transactions on Speech and Audio Processing*, v. 8, n. 5, pp. 508–518, 2000.

[31] Y. Chen, Y. Gu e A. O. Hero, “Sparse LMS for system identification,” em *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 3125–3128.

[32] H. Deng e M. Doroslovacki, “Improving convergence of the PNLMS algorithm for sparse impulse response identification,” *IEEE Signal Processing Letters*, v. 12, n. 3, pp. 181–184, 2005.

[33] J. Benesty e S. L. Gay, “An improved PNLMS algorithm,” em *IEEE International Conference on Acoustics Speech and Signal Processing*, IEEE, mai. de 2002.

[34] B. K. Das, L. A. Azpicueta-Ruiz, M. Chakraborty e J. Arenas-García, “A comparative study of two popular families of sparsity-aware adaptive filters,” em *2014 4th International Workshop on Cognitive Information Processing (CIP)*, 2014, pp. 1–6.

[35] Y. Dong e Z. Zheng, “Set-membership proportionate NSAF algorithm with individual activation factors,” em *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, 2017, pp. 1579–1583.

[36] J. Arenas-Garcia e A. R. Figueiras-Vidal, “Adaptive Combination of Proportionate Filters for Sparse Echo Cancellation,” *IEEE Transactions on Audio, Speech, and Language Processing*, v. 17, n. 6, pp. 1087–1098, 2009.

[37] F. Albu e H. Coanda, “A fast filtering proportionate affine projection sign algorithm,” em *2014 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, 2014, pp. 25–30.

[38] F. Albu, “Simplified proportionate affine projection algorithms,” em *2012 19th International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2012, pp. 368–371.

[39] C. Paleologu, S. Ciochina e J. Benesty, “An Efficient Proportionate Affine Projection Algorithm for Echo Cancellation,” *IEEE Signal Processing Letters*, v. 17, n. 2, pp. 165–168, 2010.

[40] C.-H. Lee, B. D. Rao e H. Garudadri, “Proportionate Adaptive Filters Based on Minimizing Diversity Measures for Promoting Sparsity,” em *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 769–773.

[41] C. Paleologu, J. Benesty e S. Ciochină, “An improved proportionate NLMS algorithm based on the l0 norm,” em *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 309–312.

[42] Z. Yang, Y. R. Zheng e S. L. Grant, “Proportionate affine projection sign algorithms for sparse system identification in impulsive interference,” em *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 4068–4071.

[43] K. T. Wagner e M. I. Doroslovački, “Complex proportionate-type normalized least mean square algorithms,” em *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 3285–3288.

[44] J. Liu e S. L. Grant, “Proportionate Adaptive Filtering for Block-Sparse System Identification,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 24, n. 4, pp. 623–630, 2016.

[45] J. Arenas-García, M. de Diego, L. A. Azpicueta-Ruiz, M. Ferrer e A. Gonzalez, “Combinations of proportionate adaptive filters in acoustics: An application to active noise control,” em *2011 19th European Signal Processing Conference*, 2011, pp. 1269–1273.

[46] E. J. Candès, M. B. Wakin e S. P. Boyd, “Enhancing Sparsity by Reweighted ℓ 1 Minimization,” *Journal of Fourier Analysis and Applications*, v. 14, n. 5-6, pp. 877–905, out. de 2008.

[47] B. K. Das, L. A. Azpicueta-Ruiz, M. Chakraborty e J. Arenas-García, “A Comparative Study of Two Popular Families of Sparsity-Aware Adaptive Filters,” em *2014 4th International Workshop on Cognitive Information Processing (CIP)*, IEEE, 2014, pp. 1–6.

[48] Y. Wang, Z. Qin, J. Tao, F. Tong e Y. Qiao, “Sparse Adaptive Channel Estimation based on l0-PRLS Algorithm for Underwater Acoustic Communications,” em *OCEANS 2022 - Chennai*, 2022, pp. 1–5.

[49] A. Formanek e D. Hadházi, “Compressing Convolutional Neural Networks by L0 Regularization,” em *2019 International Conference on Control, Artificial Intelligence, Robotics Optimization (ICCAIRO)*, 2019, pp. 155–162.

[50] Z. Youwen, X. Shuang, L. Lu e S. Dajun, “A variable multiple step-size LMS algorithm with l0-norm,” em *2016 IEEE/OES China Ocean Acoustics (COA)*, 2016, pp. 1–5.

[51] L. T. Duarte, R. Suyama, R. Attux, J. M. T. Romano e C. Jutten, “Blind extraction of sparse components based on 0-norm minimization,” em *2011 IEEE Statistical Signal Processing Workshop (SSP)*, 2011, pp. 617–620.

[52] H. Wang, Q. Guo, G. Zhang, G. Li e W. Xiang, “Thresholded Smoothed ℓ_0 Norm for Accelerated Sparse Recovery,” *IEEE Communications Letters*, v. 19, n. 6, pp. 953–956, 2015.

[53] H. Yazdanpanah, A. Carini e M. V. S. Lima, “L0-Norm Adaptive Volterra Filters,” em *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.

[54] T. D. Roberts e N. Kingsbury, “Fast approximate L0-norm deconvolution using structured wavelet domain priors,” em *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 1881–1885.

[55] Z. Liu, S. Shao, H. Liu e T. Su, “Robust Range Super-Resolution Imaging With Multi-criteria Joint Constraints via Double Smoothed l_0-Norm Under Limited Resources,” *IEEE Transactions on Aerospace and Electronic Systems*, v. 61, n. 2, pp. 5145–5165, 2025.

[56] M. T. M. Silva e V. H. Nascimento, “Convex Combination of Adaptive Filters with Different Tracking Capabilities,” em *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP ’07*, vol. 3, 2007, pp. III–925–III–928.

[57] L. Guo e L. Ljung, “Performance analysis of general tracking algorithms,” em *Proceedings of 1994 33rd IEEE Conference on Decision and Control*, vol. 3, 1994, 2851–2855 vol.3.

[58] B. Widrow, J. McCool, M. Larimore e C. Johnson, “Stationary and nonstationary learning characteristics of the LMS adaptive filter,” *Proceedings of the IEEE*, v. 64, n. 8, pp. 1151–1162, 1976.

[59] O. Macchi, “Optimization of adaptive identification for time-varying filters,” em *The 23rd IEEE Conference on Decision and Control*, 1984, pp. 330–334.

[60] V. Solo, “LMS: Past, Present and Future,” em *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 7740–7744.

[61] M. S. E. Abadi e F. Moradiani, “A unified approach to tracking performance analysis of the selective partial update adaptive filter algorithms in nonstationary environment,” *Digital Signal Processing*, v. 23, n. 3, pp. 817–830, 2013.

[62] M. T.-H. Alouane e M. Jaidane-Saidane, “A new nonstationary LMS algorithm for tracking Markovian time varying systems,” *Signal Processing*, v. 86, n. 1, pp. 50 –70, 2006.

[63] N. R. Yousef e A. H. Sayed, “Ability of adaptive filters to track carrier offsets and channel nonstationarities,” *IEEE Transactions on Signal Processing*, v. 50, n. 7, pp. 1533–1544, 2002.

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

[65] W. Fong, S. J. Godsill, A. Doucet e M. West, “Monte Carlo smoothing with application to audio signal enhancement,” *IEEE Transactions on Signal Processing*, v. 50, n. 2, pp. 438–449, 2002.

[66] T. T. P. Silva, P. Lara, F. Igreja, F. D. V. R. Oliveira, L. Tarrataca e D. B. Haddad, “An Exact Expectation Model for the LMS Tracking Abilities,” *IEEE Transactions on Signal Processing*, v. 68, pp. 5882–5893, 2020.

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

[68] W. Gardner, “Nonstationary learning characteristics of the LMS algorithm,” *IEEE Transactions on Circuits and Systems*, v. 34, n. 10, pp. 1199–1207, 1987.

[69] T. T. P. Silva, P. Lara, T. Gonzaga, D. Brandão, L. Tarrataca e D. B. Haddad, “Recent Advances in Overcoming the Independence Assumption in Adaptive Filtering Analyses,” em *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020, pp. 317–322.

[70] A. Carini, V. Mathews e G. Sicuranza, “Sufficient stability bounds for slowly varying direct-form recursive linear filters and their applications in adaptive IIR filters,” *IEEE Transactions on Signal Processing*, v. 47, n. 9, pp. 2561–2567, 1999.

[71] P. Lara, D. B. Haddad e L. Tarrataca, “Advances on the analysis of the LMS algorithm with a colored measurement noise,” *Signal, Image and Video Processing*, v. 14, n. 3, pp. 529–536, out. de 2019.

[72] K. T. Wagner e M. I. Doroslovacki, “Towards analytical convergence analysis of proportionate-type NLMS algorithms,” em *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 3825–3828.

[73] H. J. Butterweck, “An approach to LMS adaptive filtering without use of the independence assumption,” em *1996 8th European Signal Processing Conference (EUSIPCO 1996)*, 1996, pp. 1–4.

[74] Z. Li, L. Xing e B. Chen, “Adaptive filtering with quantized minimum error entropy criterion,” *Signal Processing*, v. 172, p. 107 534, jul. de 2020.

[75] N. J. Bershad, E. Eweda e J. C. Bermudez, “Stochastic analysis of soft limiters in the LMS algorithm for stationary white Gaussian inputs—A unified theory,” *Signal Processing*, v. 142, pp. 27–35, jan. de 2018.

[76] K. Shi e P. Shi, “Convergence analysis of sparse LMS algorithms with l1-norm penalty based on white input signal,” *Signal Processing*, v. 90, n. 12, pp. 3289–3293, dez. de 2010.

[77] N. J. Bershad e J. C. Bermudez, “Mean-square stability of the Normalized Least-Mean Fourth algorithm for white Gaussian inputs,” *Digital Signal Processing*, v. 21, n. 6, pp. 694–700, dez. de 2011.

[78] L. Horowitz e K. Senne, “Performance advantage of complex LMS for controlling narrow-band adaptive arrays,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 29, n. 3, pp. 722–736, 1981.

[79] R. Price, “A useful theorem for nonlinear devices having Gaussian inputs,” *IRE Transactions on Information Theory*, v. 4, n. 2, pp. 69–72, 1958.

[80] S. Haykin, *Adaptive filter theory*, 4th. Upper Saddle River, NJ: Prentice Hall, 2002.

[81] I.-T. S. G. 15, “Digital Network Echo Cancellers (Recommendation),” ITU-T, rel. técn. G.168, 2004.

[82] M. V. S. Lima, T. N. Ferreira, W. A. Martins e P. S. R. Diniz, “Sparsity-Aware Data-Selective Adaptive Filters,” *IEEE Transactions on Signal Processing*, v. 62, n. 17, pp. 4557–4572, 2014.

[83] F. Wen, L. Chu, P. Liu e R. C. Qiu, “A Survey on Nonconvex Regularization-Based Sparse and Low-Rank Recovery in Signal Processing, Statistics, and Machine Learning,” *IEEE Access*, v. 6, pp. 69 883–69 906, 2018.

[84] K. Kumar, M. L. N. S. Karthik e N. V. George, “A Novel Family of Sparsity-Aware Robust Adaptive Filters Based on a Logistic Distance Metric,” *IEEE Transactions on Signal Processing*, v. 70, pp. 6128–6141, 2022.

[85] M. Lee, T. Park e P. Park, “Variable Step-Size ℓ_0-Norm Constraint NLMS Algorithms Based on Novel Mean Square Deviation Analyses,” *IEEE Transactions on Signal Processing*, v. 70, pp. 5926–5939, 2022.

[86] L. Luo e W.-Z. Zhu, “An Optimized Zero-Attracting LMS Algorithm for the Identification of Sparse System,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 30, pp. 3060–3073, 2022.

[87] L. Luo e A. Xie, “Steady-state mean-square deviation analysis of improved ℓ_0-norm-constraint LMS algorithm for sparse system identification,” *Signal Processing*, v. 175, p. 107 658, 2020.