# FILTRAGEM ADAPTATIVA ROBUSTA COM GRADIENTE CONJUGADO E MODELAGEM ESTOCÁSTICA

José Victor Gonçalez de Souza

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.

Orientador(a): Diego Barreto Haddad
Coorientador(a): Milena Faria Pinto

Rio de Janeiro,

Julho 2025

FILTRAGEM ADAPTATIVA ROBUSTA COM GRADIENTE CONJUGADO E MODELAGEM ESTOCÁSTICA

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.

José Victor Gonçalez de Souza

Banca Examinadora:

---

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

---

Professora D.Sc. Milena Faria Pinto (CEFET/RJ) (Coorientador(a))

---

Professor D.Sc. Fabio Augusto de Alcantara Andrade
(University of South-Eastern Norway - USN)

---

Professor D.Sc. Leonardo Correia Resende
(Instituto Federal do Rio de Janeiro - IFRJ)

---

Professor D.Sc. Mario Antonio Ribeiro Dantas
(Universidade Federal de Juiz de Fora - UFJF)

---

Professora PhD. Mariane Rembold Petraglia
(Universidade Federal do Rio de Janeiro - UFRJ)

Rio de Janeiro,
Julho 2025

Ficha catalográfica elaborada pela Biblioteca Central do CEFET/RJ

S729 Souza, Jose Victor Gonçalez de
Filtragem adaptativa robusta com gradiente conjugado e modelagem estocástica / Jose Victor Gonçalez de Souza. — 2025.
159f. : il. color. , enc.

Tese (Doutorado) Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, 2025.
Bibliografia : f. 141-159
Orientador: Diego Barreto Haddad
Coorientadora: Milena Faria Pinto

1. Processamento de sinais. 2. Controle de ruído. 3. Processo estocástico. 4. Algoritmos. I. Haddad, Diego Barreto. (Orient.). II. Pinto, Milena Faria (Coorient.) III. Título.

CDD 621.3822

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

# AGRADECIMENTOS

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

# RESUMO

**Filtragem Adaptativa Robusta com Gradiente Conjugado e Modelagem Estocástica**

A atenuação de ruídos e vibrações em baixas frequências representa um desafio técnico significativo, especialmente em ambientes nos quais métodos passivos se mostram ineficazes ou inviáveis. Neste contexto, esta tese apresenta contribuições relevantes na área de filtragem adaptativa robusta, com foco em cancelamento ativo de ruído (ANC) e identificação de sistemas. O trabalho propõe algoritmos originais robustos a ruídos não gaussianos com ênfase em distribuições super-gaussianas e ruído impulsivo, e também com maior eficiência em termos de taxa de convergência e custo computacional. Entre as contribuições, destaca-se a derivação de um algoritmo de supressão de ruído baseado no método do gradiente conjugado (CG), tradicionalmente utilizado em otimização, mas aqui aplicado de forma inédita à filtragem adaptativa. Além disso, é proposto o algoritmo SM-sign-NLMS, que apresenta baixo custo computacional, robustez estatística e requer poucos hiperparâmetros. A tese também desenvolve modelos estocásticos de desempenho para o algoritmo SM-sign-NLMS, tanto em regime permanente quanto em fase de transiente, além de propor uma versão com passo de aprendizagem variável (VSS-SM-sign-NLMS), capaz de equilibrar desempenho assintótico e velocidade de convergência. Os resultados obtidos mostraram, ainda, alta acurácia entre previsões analíticas e resultados empíricos, mesmo em cenários com plantas não estacionárias e filtros de comprimento deficiente, comprovando a eficácia e aplicabilidade das abordagens propostas.

Palavras-chave: Filtragem adaptativa; cancelamento de ruído; ruído não gaussiano; gradiente conjugado; modelagem estocástica; passo de aprendizagem variável; SM-sign-NLMS.

# ABSTRACT

**Robust Adaptive Filtering with Conjugate Gradient and Stochastic Modeling**

The attenuation of noise and vibrations at low frequencies represents a significant technical challenge, especially in environments where passive methods prove ineffective or unfeasible. In this context, this thesis presents relevant contributions in the area of robust adaptive filtering, with a focus on active noise cancellation (ANC) and system identification. The work proposes original algorithms that are robust to non-Gaussian noise, with an emphasis on super-Gaussian distributions and impulsive noise, and also with greater efficiency in terms of convergence rate and computational cost. Among the contributions, we highlight the derivation of a noise suppression algorithm based on the conjugate gradient (CG) method, which is traditionally used in optimization but is applied here in a novel way to adaptive filtering. Furthermore, we propose the SM-sign-NLMS algorithm, which offers low computational cost, statistical robustness, and requires a minimal number of hyperparameters. The thesis also develops stochastic performance models for the SM-sign-NLMS algorithm, both in steady-state and transient phases, and proposes a version with a variable learning step (VSS-SM-sign-NLMS), capable of balancing asymptotic performance and convergence speed. The results also demonstrated high accuracy between analytical predictions and empirical results, even in scenarios with non-stationary plants and poor filter lengths, demonstrating the effectiveness and applicability of the proposed approaches.

Keywords: Adaptive filtering; noise cancellation; non-Gaussian noise; conjugate gradient; stochastic modeling; variable learning step; SM-sign-NLMS.

# LISTA DE FIGURAS


Figura 1 – Os dois focos desta tese (cancelamento ativo de ruído e identificação de sistemas), vistos como subáreas do tema “Filtragem Adaptativa”. Cabe ressaltar que o problema de “Controle ativo de ruído” apresenta uma grande interseção com a área de controle adaptativo. 12

Figura 2 – Estrutura da tese e associação dos capítulos com os artigos respectivos. 18

Figura 3 – Esquemático de um sistema *feedforward* de banda larga. Adaptado de [12]. 26

Figura 4 – Diagrama simplificado de um método de controle ativo de ruído do tipo *feedforward*, numa aplicação de cancelamento de ruído oriundo de uma máquina rotativa (o que incluiu turbinas). Tipicamente, um alto-falante emite o sinal que cancela o ruído. 28

Figura 5 – Esquema simplificado de um método de controle ativo de ruído do tipo *feedback*. 28

Figura 6 – Diagrama em blocos do algoritmo FX-LMS. 31

Figura 7 – Comparativo entre os algoritmos LMS, LMLS e IMPLMLS-RC para $\rho = 0{,}45$, $L = 2$, $L = 5$ e $L = 10$. 50

Figura 8 – Comparativoe entre os algoritmos LMS, LMLS e IMPLMLS-RC para $L = 2$, $\rho = 0{,}1$, $\rho = 0{,}45$ e $\rho = 0{,}7$. 51

Figura 9 – Comparativo entre os algoritmos LAD, LLAD e IMPLLAD-RC para $\rho = 0{,}45$, $L = 2$, $L = 5$ e $L = 10$. 52

Figura 10 – Comparativo entre os algoritmos LAD, LLAD e IMPLLAD-RC para $L = 2$, $\rho = 0{,}1$, $\rho = 0{,}45$ and $\rho = 0{,}7$. 52

Figura 11 – MSD assintótico (em dB) para o experimento de ANC. 53

Figura 12 – Evolução do MSD (em dB) para o experimento de ANC para o FX-LMS (com $\beta = 0.000798$), FX-LMLS (com $\beta = 0.0028$) e FX-IMPLMLS-RC (com $\beta = 0.006$). 54

Figura 13 – Comportamento médio dos coeficientes do filtro adaptativo (em preto), com $K = 500$ e $\sigma_\nu^2 = 10^{-6}$. Linha vermelha tracejada: estimativa teórica utilizando a Equação (109). (a) $\hat{\boldsymbol{s}} = \boldsymbol{s}$. (b) $\hat{\boldsymbol{s}} \neq \boldsymbol{s}$ 65

Figura 14 – Evolução de MSE e ANR ao longo das iterações, com entrada estacionária, ruído de medição branco, $K = 100$, $\beta_{\text{FX-LMS}} = \beta_{\text{MFX-LMS}} = 2.5 \times 10^{-2}$ e $\sigma_\nu^2 = 10^{-4}$ 66

Figura 15 – Evolução do MSE (em dB) ao longo das iterações, com entrada de ruído branco e fala, $K = 100$, $\beta_{\text{FX-LMS}} = \beta_{\text{MFX-LMS}} = 3 \times 10^{-2}$ e $\sigma_\nu^2 = 10^{-4}$. As curvas do FX-LMS e MFX-LMS são indistinguíveis 67

Figura 16 – Diagrama de blocos da estrutura de um algoritmo de filtragem adaptativa aplicado a uma tarefa de identificação de sistemas. 70

Figura 17 – Evolução do MSD (em dB) dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 1, para $\tau = 2$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,059$, $\beta_{\text{SM-sign}} = 0,0828$. 78

Figura 18 – Evolução do MSD (em dB) dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 1, para $\tau = 5$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,0575$, $\beta_{\text{SM-sign}} = 0,0862$. 78

Figura 19 – Evolução da probabilidade de atualização dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 1, para $\tau = 2$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,059$, $\beta_{\text{SM-sign}} = 0,0828$. 79

Figura 20 – Evolução da probabilidade de atualização dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 1, para $\tau = 5$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,0575$, $\beta_{\text{SM-sign}} = 0,0862$. 79

Figura 21 – Evolução do MSD (em dB) dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 2, para $\tau = 2$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,1761$, $\beta_{\text{SM-sign}} = 0,4542$. 80

Figura 22 – Evolução do MSD (em dB) dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 2, para $\tau = 5$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,1753$, $\beta_{\text{SM-sign}} = 0,4563$. 80

Figura 23 – Evolução da probabilidade de atualização dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 2, para $\tau = 2$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,1761$, $\beta_{\text{SM-sign}} = 0,4542$. 81

Figura 24 – Evolução da probabilidade de atualização dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 2, para $\tau = 5$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,1753$, $\beta_{\text{SM-sign}} = 0,4563$. 81

Figura 25 – Evolução do MSE (em dB) dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 3, para $\tau = 5$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,03$, $\beta_{\text{SM-sign}} = 0,01$. 82

Figura 26 – Evolução do MSD (em dB) dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 3, para $\tau = 5$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,03$, $\beta_{\text{SM-sign}} = 0,01$. 83

Figura 27 – Evolução da probabilidade de atualização dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 3, para $\tau = 5$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,03$, $\beta_{\text{SM-sign}} = 0,01$. 83

Figura 28 – Regime permanente teórico e simulado para o Cenário 1. Linha azul sólida: curva simulada; linha vermelha tracejada: predição teórica. 91

Figura 29 – Regime permanente teórico e simulado para o Cenário 2. Linha azul sólida: curva simulada; linha vermelha tracejada: predição teórica. 92

Figura 30 – Regime permanente teórico e simulado para o Cenário 3 ($\tau = 2$). Linha azul sólida: curva simulada; linha vermelha tracejada: predição teórica. 92

Figura 31 – Regime permanente teórico e simulado para o Cenário 3 ($\tau = 5$). Linha azul sólida: curva simulada; linha vermelha tracejada: predição teórica. 93

Figura 32 – Regime permanente teórico e simulado para o Cenário 4 ($\tau = 2$). Linha azul sólida: curva simulada; linha vermelha tracejada: predição teórica. 93

Figura 33 – Regime permanente teórico e simulado para o Cenário 4 ($\tau = 5$). Linha azul sólida: curva simulada; linha vermelha tracejada: predição teórica. 94

Figura 34 – Evolução do MSE (em dB) teórico (em vermelho tracejado) e empírico (em azul sólido), referente ao Cenário 1. 102

Figura 35 – Evolução do MSE (em dB) teórico (em vermelho tracejado) e empírico (em azul sólido), referente ao Cenário 2. 103

Figura 36 – Evolução do MSE (em dB) teórico (em vermelho tracejado) e empírico (em azul sólido), referente ao Cenário 3. 103
Figura 37 – Evolução do MSE (em dB) teórico (em vermelho tracejado) e empírico (em azul sólido), referente ao Cenário 4. 104
Figura 38 – Evolução do MSE (em dB) para o Cenário 1. Em vermelho: algoritmo SM-sign-NLMS com $\beta = 0,1$. Em verde: algoritmo SM-sign-NLMS com $\beta = 0,5$. Em azul: algoritmo SM-sign-NLMS com $\beta = 1$. Em preto: algoritmo proposto VSS-SM-sign-NLMS. 113
Figura 39 – Evolução do MSE (em dB) para o Cenário 2. Em vermelho: algoritmo SM-sign-NLMS com $\beta = 0,1$. Em verde: algoritmo SM-sign-NLMS com $\beta = 0,5$. Em azul: algoritmo SM-sign-NLMS com $\beta = 1$. Em preto: algoritmo proposto VSS-SM-sign-NLMS. 114
Figura 40 – Evolução do MSE (em dB) para o Cenário 3. Em vermelho: algoritmo SM-sign-NLMS com $\beta = 0,1$. Em verde: algoritmo SM-sign-NLMS com $\beta = 0,5$. Em azul: algoritmo SM-sign-NLMS com $\beta = 1$. Em preto: algoritmo proposto VSS-SM-sign-NLMS. 114
Figura 41 – Evolução do MSE (em dB) para o Cenário 4. Em vermelho: algoritmo SM-sign-NLMS com $\beta = 0,1$. Em verde: algoritmo SM-sign-NLMS com $\beta = 0,5$. Em azul: algoritmo SM-sign-NLMS com $\beta = 1$. Em preto: algoritmo proposto VSS-SM-sign-NLMS. 115
Figura 42 – Evolução do MSE (em dB) para o Cenário 5. Em vermelho: algoritmo SM-sign-NLMS com $\beta = 0,1$. Em verde: algoritmo SM-sign-NLMS com $\beta = 0,5$. Em azul: algoritmo SM-sign-NLMS com $\beta = 1$. Em preto: algoritmo proposto VSS-SM-sign-NLMS. 115
Figura 43 – Evolução do valor médio do passo de aprendizagem $\beta(k)$ para o algoritmo VSS-SM-sign-NLMS, no Cenário 1. 117
Figura 44 – Evolução do valor médio do passo de aprendizagem $\beta(k)$ para o algoritmo VSS-SM-sign-NLMS, no Cenário 2. 118
Figura 45 – Evolução do valor médio do passo de aprendizagem $\beta(k)$ para o algoritmo VSS-SM-sign-NLMS, no Cenário 3. 118
Figura 46 – Evolução do valor médio do passo de aprendizagem $\beta(k)$ para o algoritmo VSS-SM-sign-NLMS, no Cenário 4. 119

Figura 47 – Evolução do valor médio do passo de aprendizagem $\beta(k)$ para o algoritmo VSS-SM-sign-NLMS, no Cenário 5. 119

Figura 48 – MSE (em dB) em regime permanente em termos de $\beta$, para o Cenário 1. Em azul sólido, curva obtida por simulações. Em vermelho tracejado, previsão teórica engendrada pela resolução numérica da Equação (195). 129

Figura 49 – MSE (em dB) em regime permanente em termos de $\beta$, para o Cenário 2. Em azul sólido, curva obtida por simulações. Em vermelho tracejado, previsão teórica engendrada pela resolução numérica da Equação (195). 129

Figura 50 – MSE (em dB) em regime permanente em termos de $\beta$, para o Cenário 3. Em azul sólido, curva obtida por simulações. Em vermelho tracejado, previsão teórica engendrada pela resolução numérica da Equação (195). 130

Figura 51 – MSE (em dB) em regime permanente em termos de $\beta$, para o Cenário 4. Em azul sólido, curva obtida por simulações. Em vermelho tracejado, previsão teórica engendrada pela resolução numérica da Equação (195). 130

Figura 52 – Evolução da média dos coeficientes adaptativos. As retas tracejadas horizontais sinalizam o coeficiente correspondente na resposta ao impulso da planta desconhecida. As curvas representam a evolução da média de $w_1(k)$ (em vermelho), $w_4(k)$ (em verde) e $w_8(k)$ (em azul). As curvas foram computadas por meio de $10^4$ ensaios independentes de Monte Carlo. 131

Figura 53 – Evolução do MSE ao longo das iterações, para o Cenário 1. 132

Figura 54 – Evolução do MSE ao longo das iterações, para o Cenário 2. 133

Figura 55 – Evolução do MSE ao longo das iterações, para o Cenário 3. 133

Figura 56 – Evolução do MSE ao longo das iterações, para o Cenário 4. 134

Figura 57 – Evolução do MSE ao longo das iterações, para o Cenário 5. 134

Figura 58 – Evolução do MSE ao longo das iterações, para o Cenário 6. 135

# LISTA DE TABELAS


Tabela 1 – Exemplificação de alguns algoritmos de ANC da família FX. Versão adaptada e estendida de [10]. 30

Tabela 2 – Parâmetros utilizados nas simulações da Figura 7. 50

Tabela 3 – Parâmetros utilizados nas simulações da Figura 8. 50

Tabela 4 – Iteração em que a convergência foi atingida nas simulações das Figuras 7 e 8. 50

Tabela 5 – Parâmetros utilizados nas simulações da Figura 9. 51

Tabela 6 – Parâmetros utilizados nas simulações da Figura 10. 52

Tabela 7 – Iteração em que a convergência foi atingida nas simulações das Figuras 9 e 10 53

Tabela 8 – Parâmetros utilizados nas simulações das Figuras 11 e 12. 54

Tabela 9 – Características principais dos cenários analisados. O #Modelo reflete o número da função de transferência estimada por medições, constante no relatório da ITU [122]. 76

Tabela 10 – Características dos 4 cenários. Nos Cenários 3 e 4, o número de iterações $N_{\text{it}}$ foi maior para garantir o regime permanente, já que o número de coeficientes necessários para identificar a planta desconhecida é maior. #Modelo reflete o número da função de transferência estimada por medições, constante no relatório da ITU [122]. 90

Tabela 11 – Máxima diferença (em dB) entre as curvas teórica e simulada nos diferentes cenários. 90

Tabela 12 – Especificação dos cenários adotados para avaliação da predição teórica de transiente. Todas as plantas tiveram seus coeficientes multiplicados por 100. A coluna #Modelo reflete o número da função de transferência estimada por medições, constante no relatório da ITU [122]. 102

Tabela 13 – Comparação da estimativa da taxa de convergência com relação ao valor simulado. $N_{\text{iter}}^{\Delta}$: número de iterações necessárias (pelo modelo teórico e por simulações) para que haja uma queda de $\Delta_{\text{MSE}}$ dB em relação ao MSE inicial. $R_{\text{iter}}^{\Delta}$ (em dB/iteração): valor estimado da razão entre $\Delta$ (queda do MSE, em dB, a partir do MSE inicial) e $N_{\text{iter}}^{\Delta}$. Observe que $R_{\text{iter}}^{\Delta}$ relaciona-se com a taxa de convergência do algoritmo. 105

Tabela 14 – Especificações dos cenários testados. A coluna #Modelo reflete o número da função de transferência estimada por medições, constante no relatório da ITU [122]. 113

Tabela 15 – Características do filtro de colorimento do sinal de entrada para os Cenários 1-5. 113

Tabela 16 – MSE em regime permanente (em dB) do algoritmo SM-sign-NLMS com $\beta$ constante e do algoritmo proposto VSS-SM-sign-NLMS. A estimativa do MSE assintótico foi efetuada por meio da média das últimas 1000 iterações. 116

Tabela 17 – Especificação dos Cenários analisados para avaliação da capacidade de rastreamento do algoritmo SM-sign-NLMS. 128

Tabela 18 – Especificações dos 6 cenários de simulação adotados para avaliação da predição teórica de desempenho assintótico no caso de comprimento deficiente. $N_{\text{total}} \triangleq N + L$ denota o comprimento da planta ótima a identificar. $N_{\text{it}}$ denota o número de iterações executadas. A coluna #Modelo reflete o número da planta estimada por medições, constante no relatório da ITU [122]. 132

Tabela 19 – Comparação entre o MSE em regime permanente obtido por meio de simulações ($MSE_{EMP}$) e o obtido pela resolução das Equações (213) e (214) ($MSE_{TEOR}$). O $MSE_{EMP}$ foi obtido pela média das últimas 1000 iterações da curva de MSE computada com 1000 ensaios independentes de Monte Carlo. 135

# LISTA DE ALGORITMOS


Algoritmo 1 – Algoritmo CG não adaptativo 58

Algoritmo 2 – Algoritmo adaptativo CG proposto para cancelamento de ruído 61

Algoritmo 3 – Algoritmo VSS-SM-sign-NLMS proposto. 112

# LISTA DE ABREVIATURAS E SIGLAS

| | |
|---|---|
| ANC | *Active Noise Cancelling* |
| ANR | *Average noise reduction* |
| CG | *Conjugate gradient* |
| CG-FX | *Conjugate Gradient Filtered-X* |
| FIR | *Finite Impulse Response* |
| FX-LMS | *Filtered-X Least Mean Square Algorithm* |
| HR | Hipótese Acerca Do Ruído |
| IMPLLAD-RC | *Improved $\mu$-law least logarithmic absolute difference with reuse of coefficients* |
| IMPLMLS-RC | *Improved $\mu$-law least mean logarithmic square with reuse of coefficients* |
| LAD | *Least absolute difference* |
| LLAD | *Least logarithmic absolute difference* |
| LMF | *Least mean fourth* |
| LMLS | *Least mean logarithmic square* |
| LMS | *Least Mean Square* |
| MFX-LMS | *Modified Filtered-X LMS* |
| MSD | *Mean squared deviation* |
| MSE | *Mean squared error* |
| RLS | *Recursive Least Square* |
| SNR | *Signal-to-noise ratio* |
| VSS | *Variable Step Size* |

# SUMÁRIO


**Introdução** **11**

**1 Fundamentação Teórica** **19**

1.1 Identificação de Sistemas em Tempo Real: o Algoritmo LMS 19

1.2 Controle Ativo de Ruído (ANC) 25

1.2.1 Algoritmo FX-LMS Aplicado ao Cancelamento Ativo de Ruído 29

1.2.2 Solução Ótima do Algoritmo FX-LMS 32

1.3 Comentários Finais 34

**2 Extensões dos Algoritmos LMS e FX-LMS** **35**

2.1 Algoritmos LAD, LMP LMF 36

2.2 Algoritmos LMLS e LLAD 38

2.3 Revisão sobre Algoritmos Proporcionais 42

2.4 Reúso de Coeficientes 43

2.5 Algoritmos propostos: FX-IMPLMLS-RC e FX-IMPLLAD-RC 45

2.6 Relação de conservação de energia 47

2.7 Resultados de simulações 49

2.7.1 Identificação de sistema 49

2.7.2 Controle Ativo de Ruído 52

2.8 Comentários Conclusivos 55

**3 Cancelamento Ativo de Ruído com Gradiente Conjugado** **56**

3.1 Identificação adaptativa de sistemas por meio de gradiente conjugado 57

3.2 Proposição de um método de cancelamento automático de ruído baseado em gradiente conjugado 59

3.3 Análise Estocástica 61

3.4 Resultados de simulação 63

3.5 Considerações Parciais 67

**4 Algoritmo SM-Sign-NLMS 68**
4.1 Introdução 68
4.2 Algoritmos Adaptativos do tipo Set-Membership 69
4.3 Algoritmo Proposto SM-Sign-NLMS 72
4.3.1 Complexidade computacional 75
4.4 Simulações 75
4.4.1 Cenário 1 77
4.4.2 Cenário 2 77
4.4.3 Cenário 3 79
4.5 Comentários Finais 83

**5 Análise de Regime Permanente do Algoritmo SM-Sign-NLMS 85**
5.1 Análise de regime permanente 86
5.2 Resultados 90
5.3 Comentários finais 94

**6 Análise de Transiente do Algoritmo SM-Sign-NLMS 96**
6.1 Análise de Transiente do algoritmo SM-sign-NLMS 97
6.2 Resultados 101
6.3 Comentários finais 106

**7 Algoritmo SM-Sign-NLMS com Passo de Aprendizagem Variável 107**
7.1 Algoritmos com passos de aprendizagem variantes no tempo 108
7.2 Algoritmo proposto: VSS-SM-sign-NLMS 110
7.3 Resultados 112
7.4 Comentários Finais 117

**8 Análise de Rastreamento e Comprimento Deficiente do Algoritmo SM-Sign-NLMS 121**
8.1 Introdução 121
8.2 Rastreamento 122
8.3 Análise de Comprimento Deficiente 124
8.3.1 Comportamento de primeira ordem 124
8.3.2 Análise de Segunda Ordem 126
8.4 Resultados 128

8.4.1 Rastreamento 128
8.4.2 Comprimento Deficiente 131
8.5 Comentários finais 136

**9 Conclusões 138**

**Referências 140**

# Introdução

Os avanços na disponibilidade de dados e na tecnologia levaram a uma “explosão de dados” disponíveis, a qual deu origem a implementações de modelos do tipo *data driven* em múltiplas áreas [1], [2]. Nas últimas décadas, algoritmos de filtragem adaptativa vêm sendo empregados nas mais diferentes aplicações com dados obtidos dos mais diversos sensores, em tarefas como em controle/cancelamento adaptativo de ruído, cancelamento de eco acústico, equalização de canal, identificação de sistemas, controle ativo de ruído, conformação de feixe, entre outros [3]–[7]. Isso se deve ao fato de não necessitarem de estatísticas do sinal de entrada e também da possibilidade de serem empregados em sistemas cujos parâmetros de operação variam no tempo [8], [9]. Inserida neste contexto, esta tese tem como foco dois grandes campos de aplicação da filtragem adaptativa: cancelamento automático de ruído e identificação de sistemas, conforme ilustrado pela Figura 1.

Cabe ressaltar que frequentemente uma etapa necessária para o cancelamento automático de ruído consiste na modelagem do sistema secundário (que é, em muitas aplicações, variante no tempo [10]), a qual exige o concurso de técnicas de identificação de sistemas em tempo real, de modo que tais campos encontram-se de fato estreitamente relacionados. Estas técnicas em tempo real costumeiramente mobilizam estratégias de otimização estocástica, as quais tornam cada iteração do algoritmo de aprendizagem independente da quantidade de informações por meio da seleção (aleatória ou não) de um subconjunto dos dados disponíveis [1]. Parte significativa destas técnicas pode ser enquadrada em técnicas baseadas em métodos de primeira ordem e métodos de segunda ordem. Enquanto os métodos de segunda ordem podem oferecer taxas de convergência quadrática (isto é, com melhora acelerada a cada iteração), os métodos de primeira ordem, mais simples, geralmente apresentam apenas convergência linear, o que significa um progresso mais lento em direção à solução ideal. No entanto, esses métodos lineares possuem a vantagem de demandar menor custo computacional e serem mais robustos em aplicações com dados ruidosos ou incompletos [1]. A otimização estocástica é um processo que se baseia na minimização ou na maximização do valor do valor esperado (ou de uma função matemática) quando um ou mais parâmetros de entrada dependem

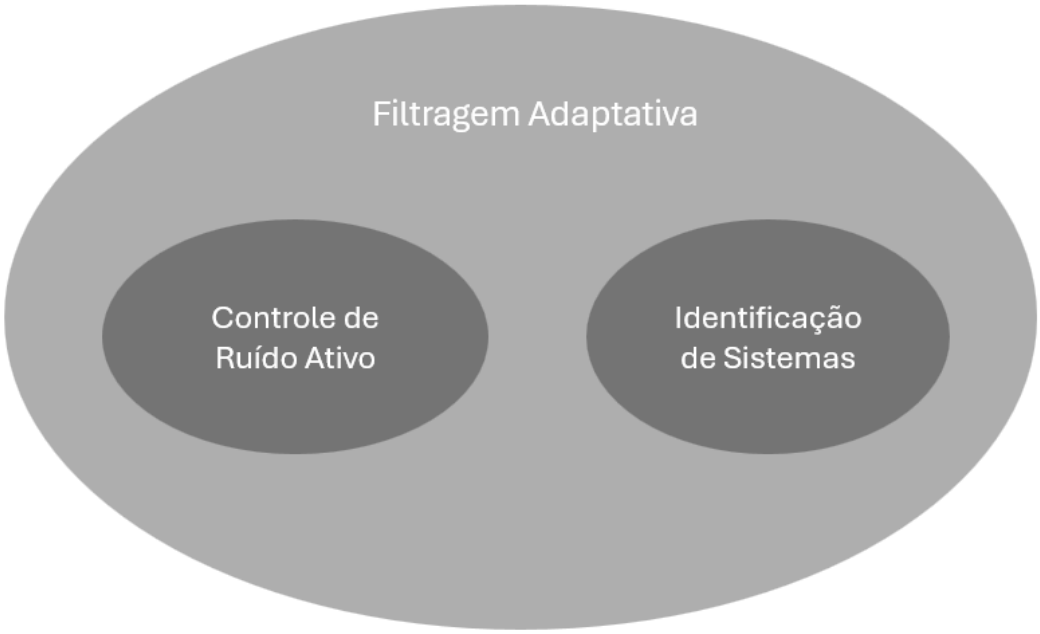



Figura 1 – Os dois focos desta tese (cancelamento ativo de ruído e identificação de sistemas), vistos como subáreas do tema “Filtragem Adaptativa”. Cabe ressaltar que o problema de “Controle ativo de ruído” apresenta uma grande interseção com a área de controle adaptativo.

de variáveis aleatórias [11]. Em termos mais gerais, a busca de uma solução aproximada para uma determinada função ótima num espaço de dimensão finita é um dos problemas mais comuns em matemática aplicada [11].

A supressão de ruídos e vibrações[1], em particular em baixas frequências, consiste num grande desafio tecnológico. O método tradicional de cancelamento de ruídos e vibrações de altas frequências, que utiliza técnicas passivas como abafadores acústicos e amortecimento mecânico, é economicamente inviável ou tecnicamente impraticável em baixas frequências (abaixo de $600\,\mathrm{Hz}$) [12], pois os comprimentos de onda mais longos requerem estruturas grandes e caras, como silenciadores e barreiras, para serem suprimidos adequadamente [13], [14]. Ademais, um grande número de problemas de ruído acústico são dominados por contribuições de baixa frequência [15]. Justamente nesta faixa de frequência os métodos ativos atuam melhor, pois neste caso os comprimentos de onda são grandes, comparados com a região na qual cumpre cancelar o ruído [16]. Cabe ressaltar que alguns autores ressaltam a complementaridade dos métodos de redução de ruído ativos e passivos [17].

[1] Ao longo desta tese, não distinguiremos a supressão de ruído da supressão de vibrações, já que matematicamente ambos os tipos de cancelamento são equivalentes.

Os métodos ativos de controle (ou *cancelamento*) de ruído (doravante também denominados *Active Noise Cancelling* (ANC), do inglês *active noise cancelling*) se valem do princípio da interferência destrutiva das ondas mecânicas, tendo sido originalmente propostos por Lueg [18]. A patente [18] já delineava a estratégia essencial das técnicas de ANC, através da qual os sinais utilizados provocam a atenuação desejada por efeitos de interferência destrutiva, de modo puramente mecânico. Tal interferência destrutiva se dá entre uma fonte sonora gerada pela fonte primária (ou original) e a fonte secundária, controlável e judiciosamente sintetizada para obter o efeito almejado [16]. Tal possibilidade é facilitada pelo fato de que a propagação sonora é um fenômeno bastante linear, uma condição necessária para que o princípio da superposição seja válido [16].

Já a identificação de sistemas pode ser compreendida como um processo de inferência estatística no qual, a partir de medições de entrada e saída de um determinado sistema, pretende-se identificar as operações que este sistema efetua no sinal de entrada, de modo a gerar seu respectivo sinal de saída. Em outras palavras, a identificação de sistemas almeja construir modelos matemáticos de sistemas dinâmicos a partir de dados observados destes sistemas. Sendo tais sistemas abundantes, as técnicas de identificação de sistemas apresentam um amplo espectro de aplicações [19].

Uma importante característica das técnicas tratadas nesta tese reside no fato de que focamos em técnicas de atualização *em tempo real*. Se por um lado essa característica permite a essas técnicas, no mais das vezes, rastrear modificações nos sistemas subjacentes e reduzir seu custo computacional, por outro tende a gerar perturbações nas soluções, advindas do fato de que apenas informações localizadas temporalmente são adotadas, o que introduz incertezas no processo de estimação. Para lidar com os desafios inerentes a tais soluções *real time*, estratégias baseadas em otimização estocástica serão mobilizadas. Tais soluções constituem um *framework* bastante disseminado que permite a incorporação explícita de incerteza e de falta de informação.

À incerteza derivada da ausência de informação completa acerca do ambiente sob o qual o sistema de ANC opera, soma-se à incerteza oriunda do processo de medição. Em ambientes digitais, o processo de medição impõe uma quantização das grandezas medidas, a qual é responsável por corromper a qualidade dos dados adquiridos, estando estes sujeitos ao chamado *ruído de quantização*. Outro ruído com efeitos bastante deletérios no desempenho das técnicas aqui contempladas é o *ruído impulsivo*, o qual ocorre a fenômenos tais como grande número de determinada espécie de camarões

em canais hidroacústicos [20], *double-talk* em sistemas de ANC [21] e fenômenos atmosféricos em sistemas de telecomunicações [22]. Sinais de ruído, nas suas diversas modalidades, reconhecidamente acarretam prejuízo ao desempenho das técnicas enfocadas, particularmente quando estas são implementadas em tempo real. O fato de tais sinais potencialmente tornarem inúteis alguns dados de entrada (ou mesmo prejudiciais à otimização pretendida) pode ser utilizado para reduzir o custo computacional de alguns algoritmos, por meio de estratégias de *censura de dados*, tal como a estratégia *Set-membership* [23], a qual evita atualizações desnecessárias do estimador adaptativo quando os dados são pouco informativos.

**Objetivos da tese**

Esta tese efetua contribuições em algoritmos de cancelamento automático de ruído e de identificação de sistemas. São efetuadas contribuições nas dimensões de robustez a ruído impulsivo, aumento da taxa de convergência e de redução de custo computacional, bem como na obtenção de algoritmos inéditos e de aplicações inéditas de técnicas de otimização a problemas de cancelamento ativo de ruído. Em termos concisos, as contribuições dessa tese podem ser elencadas da seguinte maneira:

- ⋆ Propor novos algoritmos de identificação de sistemas e de cancelamento automático de ruído robustos a ruídos não gaussianos e com maior taxa de convergência quando o vetor ótimo da solução é esparso;
- ⋆ Utilizar técnica de gradiente conjugado para proposição de um algoritmo inédito para cancelamento automático de ruído, o qual apresenta maior taxa de convergência com algoritmos no estado da arte, à custa de um aumento de complexidade computacional;
- ⋆ Propor e obter um algoritmo de filtragem adaptativa inédito (intitulado SM-Sign-NLMS) capaz de apresentar menor custo computacional, robustez a ruído impulsivo e com poucos hiperparâmetros a ajustar;
- ⋆ Propor um modelo estocástico para previsão analítica do desempenho assintótico do algoritmo SM-Sign-NLMS;

- ⋆ Propor um modelo estocástico para previsão analítica da evolução do aprendizado no transiente do algoritmo SM-Sign-NLMS;
- ⋆ Propor uma versão com passo de aprendizagem variável do algoritmo SM-Sign-NLMS, baseada no modelo teórico de transiente, para contornar o compromisso entre taxa de convergência e desempenho assintótico e comparar o desempenho real da versão proposta com a versão tradicional do algoritmo;
- ⋆ Desenvolver modelos estocásticos para predição do desempenho do algoritmo SM-Sign-NLMS quando o sistema a identificar é não estacionário e quando o filtro apresenta comprimento deficiente;
- ⋆ Comparar as predições analíticas com resultados reais, obtidos por meio de simulações com ensaios independentes de Monte Carlo.

**Notação matemática**

Ao longo deste trabalho, vetores são denotados por letras minúsculas em negrito. Ademais, assume-se que os vetores são do tipo coluna, a menos que se explicite o contrário. Escalares são denotados por letras minúsculas não negritadas, enquanto matrizes encontram-se em letras maiúsculas e em negrito. $\nabla_{\boldsymbol{w}} f$ denota o gradiente da função escalar $f$ em termos do vetor $\boldsymbol{w}$, enquanto que $\frac{\partial f}{\partial x}$ denota a derivada parcial de $f$ com relação a $x$. Já a função sign$(x)$ fornece o sinal de seu argumento $x$ conforme:

$$\text{sign}[x] = \begin{cases} -1, & \text{se } x < 0 \\ 0, & \text{se } x = 0 \\ 1, & \text{se } x > 0 \end{cases} . \tag{1}$$

**Produções associadas a este trabalho**

Esta tese resultou, direta ou indiretamente, na publicação das seguintes produções em revistas internacionais:

1. **de Souza, José V. G.**; Haddad, Diego B.; Henriques, Felipe da R.; Petraglia, Mariane R. "*Novel Proportionate Adaptive Filters with Coefficient Vector Reusing*", *Circuits, Systems and Signal Processing*, v. 39, n. 5, p. 2473-2488, Set., 2019.

    - Esse artigo propôs o desenvolvimento do algoritmo de filtragem adaptativa baseado em reúso de coeficientes, dando origem, portanto, ao Capítulo 2 dessa tese. A revista, no ano de 2024, apresenta um Qualis A2 e um fator de impacto de 2,3.

2. **de Souza, José V. G.** et al., "*Conjugate gradient-based active noise cancellation*", *Electronics Letters*, pp. e13234, v. 60, n. 12, jun., 2024.

    - Esse artigo propôs a aplicação do algoritmo de otimização conhecido como gradiente conjugado para o contexto de cancelamento ativo de ruídos, dando origem ao Capítulo 3 desta tese. A revista, no ano de 2024, apresenta um Qualis A3 e um fator de impacto de 3,1.

3. **de Souza, José Victor G.** et al., "*Stochastic Modelling of the Set-Membership-sign-NLMS Algorithm*", *IEEE Acess*, vol. 12, pp. 32739-32752, Mar., 2024.

    - Este artigo contém as principais contribuições desta tese, sendo derivado dos resultados constantes ao longo dos Capítulos 4-8. A revista, no ano de 2024, apresenta Qualis A1 e um fator de impacto de 3,9.

**Estrutura da tese**

O restante desta tese é dividido da seguinte maneira: o Capítulo 1 conceitualiza os algoritmos de filtragem adaptativa e trabalhos correlatos a essa tese. Ademais, é

apresentado o algoritmo clássico *Filtered-X Least Mean Square Algorithm* (FX-LMS), fornecendo o tecido conceitual básico para o restante do texto. A adoção de filtros digitais adaptativos para sistemas de cancelamento ativo de ruído remonta a Burgess [24], onde um algoritmo de controle em malha fechada baseado no *Least Mean Square* (LMS) é proposto para a atenuação de ruído, seja em banda estreita, seja em banda larga. Esta relação deve ser destacada porque, como veremos, há muita sinergia entre as áreas de filtragem adaptativa (em especial à aplicada à identificação de sistemas, como tipicamente o LMS é usado) e a área de cancelamento ativo de ruído.

O Capítulo 2 apresenta novos algoritmos de filtragem adaptativa e de cancelamento de ruído obtidos ao longo da pesquisa, os quais são mais robustos a uma baixa razão sinal-ruído (SNR) e a ruído aditivo impulsivo. Tais algoritmos conjugam a técnica de reúso de coeficientes com a adoção de funções custo mais adequadas para a ocorrência de ruído de distribuição supergaussiana.

O Capítulo 3 adapta a técnica de otimização conhecida como gradiente conjugado para o problema de cancelamento de ruído. É então proposto um algoritmo inédito, o qual apresenta maior taxa de convergência do que as alternativas tradicionais. O desempenho superior é também observado com sinais reais (sinais de voz).

O Capítulo 4 propõe um algoritmo adaptativo inédito, chamado de SM-sign-NLMS, para o problema de identificação de sistemas. O algoritmo alia características desejáveis tais como robustez a ruído impulsivo, baixo custo computacional e poucos parâmetros ajustáveis.

O Capítulo 5 propõe um modelo estocástico para predição do desempenho assintótico do algoritmo proposto SM-sign-NLMS. O modelo utiliza princípios de conservação de energia e dá origem a uma equação de ponto fixo, cuja resolução numérica dá origem à estimativa teórica de desempenho assintótico.

O Capítulo 6 efetua uma modelagem estocástica inédita do algoritmo SM-sign-NLMS, a qual emprega um modelo estatístico para os sucessivos vetores de entrada que alimentam o filtro adaptativo. Aproximações estocásticas, tais como a heurística de Price [25] também são mobilizadas com o propósito de encontrar equações concisas que descrevam com acurácia a dinâmica de aprendizado do algoritmo.

O Capítulo 7 utiliza o modelo de transiente explicitado no Capítulo 6 para propor uma versão do algoritmo SM-sign-NLMS com passos de aprendizagem variantes no tempo, de modo a conjugar uma alta taxa de convergência a um bom desempenho




Figura 2 – Estrutura da tese e associação dos capítulos com os artigos respectivos.

assintótico.

Por sua vez, o Capítulo 8 apresenta modelos estocásticos que preveem o desempenho assintótico (tanto em termos de média dos coeficientes quanto em termos de erro quadrático médio) do algoritmo SM-sign-NLMS em configurações nas quais a planta a identificar é não estacionária (e sujeita a perturbações markovianas de primeira ordem) e nas quais o filtro opera com comprimento deficiente (isto é, com menos coeficientes do que o número de coeficientes da resposta ao impulso da planta a identificar).

Por fim, o Capítulo 9 encerra a tese, contendo as suas conclusões, bem como as perspectivas de trabalhos futuros consideradas promissoras. A Figura 2 apresenta os capítulos, bem como os associa aos artigos correspondentes.

# 1- Fundamentação Teórica

Como vimos na Introdução, encontram-se no escopo desta tese dois tipos de tarefas: identificação de sistemas e controle ativo de ruído. Este capítulo pretende contextualizar tais tarefas, para facilitar a compreensão do leitor, bem como introduzir conceitos importantes, os quais serão mobilizados nos capítulos seguintes.

Os métodos que serão contemplados podem ser interpretados como *metodologias de otimização*. Diversos desafios encontrados em problemas práticos, tais como uma taxa de convergência deficiente, a emergência de ruído impulsivo ou necessidade de redução de complexidade computacional, instigam a criatividade do projetista para a proposição de novas metodologias de otimização que consigam contornar limitações das técnicas existentes. As contribuições desta tese devem ser compreendidas à luz de tais considerações.

## 1.1- Identificação de Sistemas em Tempo Real: o Algoritmo LMS

O trabalho de Burgess [24], mencionado anteriormente devido a sua relevância histórica, adaptou o algoritmo LMS, introduzido originalmente num contexto de filtragem adaptativa, para prover uma solução adaptativa em malha fechada do tipo *feedforward* para cancelamento ativo de ruído. Esta relação histórica, como vimos, reflete a conexão íntima entre as áreas de filtragem adaptativa e controle adaptativo, de modo que é possível enriquecer o espectro das soluções para técnicas de ANC por meio de técnicas que se revelaram bem-sucedidas em filtragem adaptativa.

A primeira motivação do algoritmo LMS foi modulada por desafios na área de reconhecimento de padrões, na qual, por vezes, o importante é adaptar uma estrutura puramente linear, conhecida como *Adaline* (do inglês *adaptive linear*). Tais combinadores lineares adaptativos têm sido aplicados de maneira bem-sucedida na modelagem paramétrica de sistemas desconhecidos, em predição linear, em sistemas adaptativos de antena, em sistemas de equalização de canal para comunicações de alta velocidade, em

cancelamento de eco, em estimação de frequência instantânea, em receptores de sinais em banda estreita imersos em ruído, em sistemas de controle adaptativo, e em muitas outras aplicações [26].

Nesta tese, assume-se que as estruturas lineares envolvidas são do tipo *tapped-delay line*, nas quais sucessivas amostras de um sinal de entrada são concatenadas para compor o vetor de entrada $\boldsymbol{x}(k) \in \mathbb{R}^N$, ou seja:

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

onde $x(k) \in \mathbb{R}$ denota a $k$-ésima amostra do sinal de entrada $x(k) \in \mathbb{R}$. Observe que o vetor de entrada, definido na Equação (2), efetua uma operação similar à de uma janela deslizante ao longo das amostras do sinal de entrada, incorporando ordenadamente $N$ amostras (no mais das vezes, as mais recentes) do sinal de entrada.

A saída da estrutura considerada é computada por meio de uma *soma ponderada* das amostras presentes em $\boldsymbol{x}(k)$, o que dá origem à saída (ou estimativa) $y(k) \in \mathbb{R}$ na $k$-ésima iteração:

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

onde $w_i(k) \in \mathbb{R}$ denota o $i$-ésimo peso (ou coeficiente) da aludida ponderação. A natureza variante no tempo dos pesos $\{w_i(k)\}_{i=0}^{N-1}$ (já que um cenário de *adaptação* é assumido) é incorporada na notação por meio do índice temporal $k$. Por conveniência matemática, define-se normalmente um vetor de pesos $\boldsymbol{w}(k) \in \mathbb{R}^N$, cuja estrutura é similar à de $\boldsymbol{x}(k)$ (vide (2)):

$$\boldsymbol{w}(k) = \begin{bmatrix} w_0(k) & w_1(k) & \dots & w_{N-1}(k) \end{bmatrix}^T, \tag{4}$$

onde normalmente considera-se que tais pesos sejam modificados por meio de um algoritmo específico de aprendizado, o qual, quando supervisionado, se vale da discrepância entre a saída da estrutura e um dado sinal desejado $d(k)$ (ou seja, o erro) como informação útil para uma atualização contínua.

A conjugação de (4) e de (2) nos permite reescrever $y(k)$ (vide (3)) de um modo mais conciso, por meio de álgebra linear:

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

uma formulação que permite uma derivação mais compacta dos algoritmos adaptativos.

Num contexto supervisionado, assume-se que o filtro tem acesso a um sinal desejado (por vezes também chamado de referência, embora tal termo engendre ambiguidades quando as técnicas de ANC forem introduzidas). Este sinal desejado, doravante denotado por $d(k) \in \mathbb{R}$, pode ser adquirido de formas bem diferentes, a depender da aplicação. Assim, num problema de cancelamento de eco acústico em telefonia viva-voz, $d(k)$ normalmente é o sinal capturado pelo microfone. Já num problema de equalização de canais em comunicação digital, $d(k)$ pode consistir de amostras de uma sequência de treinamento, já pré-acordada, entre o transmissor e o receptor. Ainda, considerando-se um problema de reconhecimento de padrões, tal valor de referência pode consistir numa codificação numérica que sinalize a classe de interesse do vetor de exemplo que no momento alimenta a entrada do filtro.

O fato de estarem disponíveis, na $k$-ésima iteração, a saída do filtro $y(k)$ e o respectivo sinal desejado $d(k)$ viabiliza a aprendizagem dos pesos constantes na estrutura linear por meio do cômputo da discrepância entre tais sinais. Tal discrepância costumeiramente é matematicamente calculada por meio de um sinal de *erro* $e(k) \in \mathbb{R}$, o qual sinaliza (em termos aproximados estatisticamente) o quão boa é a solução atual, codificada no conjunto dos pesos adaptativos $\boldsymbol{w}(k)$:

$$e(k) = d(k) - y(k) = d(k) - \boldsymbol{w}^T(k)\boldsymbol{x}(k). \tag{6}$$

O sinal de erro contém informação relevante para o aprendizado da estrutura linear de interesse. Espera-se magnitudes baixas deste sinal sejam estatisticamente correlacionadas com situações de aprendizado efetivo. Como o erro, que alimenta o processo de adaptação, depende da estrutura linear (já que $y(k)$ é diretamente dependente de $\boldsymbol{w}(k)$), observa-se a ocorrência de retroalimentação (ou *feedback*) no processo de aprendizagem, fenômeno que descabe confundir conceitualmente com a modalidade *feedback* de soluções ANC, vista na introdução desta tese. É exatamente este *feedback* que confere flexibilidade ao processo adaptativo, viabilizando o aprendizado pretendido. Todavia, esta retroalimentação é responsável pelo fato de a saída $y(k)$ (para todo $k$) não poder ser predita a partir de $x(k)$ por meio do princípio da superposição, o que acaba por caracterizar o sistema entrada-saída (do filtro adaptativo) como não linear e variante no tempo [27]. Como na prática os sinais envolvidos são estocásticos e não há linearidade no sistema como um todo (a despeito de, num determinado instante, a estrutura linear

conter coeficientes fixos), temos que a análise matemática de algoritmos adaptativos é complexa e desafiadora [26].

Fixados tais conceitos básicos, cabe descrever como a informação contida no sinal de erro pode ser aproveitada para ensejar uma judiciosa modificação dos pesos constantes em $\boldsymbol{w}(k)$. Neste ponto, cumpre distinguir algoritmos que utilizam globalmente os dados disponíveis dos algoritmos que se adaptam em tempo real, ou seja, a cada vez que uma nova amostra do sinal de entrada é adquirida. Ao longo desta tese, como visto no capítulo introdutório, assumiremos esta última modalidade, a qual caracteriza algoritmos de aprendizado *on-line*. Estes algoritmos são de grande interesse, já que conjugam a um baixo custo computacional a possibilidade de lidar com sistemas variantes no tempo (enquanto os métodos que atuam em lotes com grande quantidade de dados assumem que os sistemas subjacentes são fixos).

Existem duas grandes famílias de técnicas de derivação de algoritmos de filtragem adaptativa: as baseadas em gradiente estocástico, motivadas pela aproximação iterativa de uma função custo global (ou seja, que atente para todos os dados disponíveis) por uma função custo "instantânea" e as baseadas no princípio da distorção mínima, as quais descrevem um problema de otimização determinístico puramente local, cuja solução (seja exata, seja aproximada) resulta no algoritmo de interesse.

Normalmente, a atualização dos coeficientes adaptativos é orientada pela escolha de uma função custo. No caso de algoritmos derivados via gradiente estocástico, por razões de conveniência, maturidade teórica e tratabilidade matemática, historicamente o erro quadrático instantâneo tem sido a função custo mais popular:

$$\mathcal{F}_2[e(k)] = \frac{1}{2}\left[e(k)\right]^2. \tag{7}$$

Cabe notar que há autores que também argumentam que, como o ouvido humano responde principalmente ao valor quadrático médio da pressão acústica nele incidente, o valor quadrático do sinal de erro também é respaldado por considerações físicas [16].

A otimização via gradiente estocástico de (7) apresenta o seguinte formato:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \beta\nabla_{\boldsymbol{w}(k)}\left\{\mathcal{F}_2[e(k)]\right\}, \tag{8}$$

onde $\beta \in \mathbb{R}_+$ denota o fator de aprendizagem (também conhecido como passo de

aprendizagem) e $\nabla_{\boldsymbol{w}(k)}[\mathcal{F}_2[e(k)]]$ significa o gradiente de $\mathcal{F}_2[e(k)]$ com relação a $\boldsymbol{w}(k)$:

$$\nabla_{\boldsymbol{w}(k)}[\mathcal{F}_2[e(k)]] = \begin{bmatrix} \frac{\partial \mathcal{F}_2[e'(k)]}{\partial w_0(k)} \\ \frac{\partial \mathcal{F}_2[e'(k)]}{\partial w_1(k)} \\ \vdots \\ \frac{\partial \mathcal{F}_2[e'(k)]}{\partial w_{N-1}(k)} \end{bmatrix}, \tag{9}$$

o qual em geral é não nulo, já que $e(k)$ depende de uma combinação linear estocástica dos pesos constantes em $\boldsymbol{w}(k)$.

O gradiente estocástico (vide Eq. (8)) estabelece um aprendizado recursivo do vetor de parâmetros $\boldsymbol{w}(k)$, atualizando-o na direção mais íngreme de descida. A direção mais íngreme *de subida* da função custo em função de $\boldsymbol{w}(k)$ é fornecida pelo gradiente. O sinal de negativo é adotado porque estamos tratando de minimizar a função custo (e por isso, temos que "descer" na hipersuperfície). O fator de aprendizagem $\beta$ é de fundamental importância, porque de sua escolha dependem características de aprendizagem cruciais, tais como taxa de convergência, desempenho assintótico, capacidade de rastreamento de sistemas variantes no tempo e mesmo estabilidade.

Utilizando a função custo definida na Equação (7) no cálculo de atualização via gradiente estocástico, encontramos:

$$\boldsymbol{w}(k+1) \quad = \quad \boldsymbol{w}(k) + \beta e(k) \underbrace{\nabla_{\boldsymbol{w}(k)} \left[\boldsymbol{w}^T(k)\boldsymbol{x}(k)\right]}_{=\boldsymbol{x}(k)}, \tag{10}$$

de modo que a equação de atualização do algoritmo LMS pode ser descrita como:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta e(k)\boldsymbol{x}(k), \tag{11}$$

a qual requer $2N$ adições e $2N+1$ multiplicações por iteração. Duas métricas populares para aferição da qualidade do aprendizado do algoritmo LMS são o *mean squared error* (MSE), do português, erro quadrático médio, e o *mean squared deviation* (MSD), do português, desvio quadrático médio, a saber:

$$\text{MSE}(k) \quad \triangleq \quad \mathbb{E}\left[e^2(k)\right], \tag{12}$$

$$\text{MSD}(k) \quad \triangleq \quad \mathbb{E}\left[\|\boldsymbol{w}(k) - \boldsymbol{w}^\star\|^2\right], \tag{13}$$

onde $\boldsymbol{w}^\star$ denota a solução ótima, a qual coincide com a solução de Wiener para funções custo quadráticas [28]. Num contexto de estatísticas conjuntas estacionárias, a evolução destas métricas pode ser dividida em dois estágios: o transiente, durante o qual há um decréscimo contínuo do MSE e do MSD e o regime permanente (ou assintótico), no qual a métrica se estabiliza num determinado patamar.

Uma derivação alternativa do LMS mobiliza a técnica matemática conhecida como multiplicadores de Lagrange, a qual tem se demonstrado bastante fértil para a proposição de novos algoritmos de filtragem adaptativa [21], [29]–[33].

Tal derivação alternativa apela a um problema de otimização com restrição linear:

$$\min_{\boldsymbol{w}(k+1)} \frac{1}{2}\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|^2 \text{ sujeito a } e_{\mathrm{p}}(k) = (1-\beta\|\boldsymbol{x}(k)\|^2)e(k), \tag{14}$$

onde $e_{\mathrm{p}}(k) \in \mathbb{R}$ denota o erro *a posteriori*, o qual recalcula o erro com o par de dados $\{\boldsymbol{x}(k), d(k)\}$ *após* tais dados terem sido usados na atualização:

$$e_{\mathrm{p}}(k) \triangleq d(k) - \boldsymbol{w}^T(k+1)\boldsymbol{x}(k), \tag{15}$$

fórmula a qual deve ser cotejada com a fórmula do erro:

$$e(k) = d(k) - \boldsymbol{w}^T(k)\boldsymbol{x}(k). \tag{16}$$

O termo $\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|^2$ minimizado em (14) implica que a nova estimativa $\boldsymbol{w}(k+1)$ não deve se posicionar muito longe da estimativa atual $\boldsymbol{w}(k)$. Este é o fundamento do chamado *princípio da mínima distorção*, que é um princípio conservador, no sentido de que entende, num contexto de aprendizado em tempo real, que as informações atuais, contidas no par de dados $\{\boldsymbol{x}(k), d(k)\}$, contém uma quantidade limitada de informação e podem não ser confiáveis (devido a ruído impulsivo, por exemplo), devendo ser dada maior atenção à informação contida o vetor $\boldsymbol{w}(k)$, já que este enfeixa toda a informação pregressa que alimentou o filtro.

A aplicação da técnica dos multiplicadores de Lagrange nos permite reescrever (14) como um problema equivalente e *sem restrições*:

$$\min_{\boldsymbol{w}(k+1)} \left\{ \frac{1}{2}\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|^2 + \lambda\left[e_{\mathrm{p}}(k) - (1-\beta\|\boldsymbol{x}(k)\|^2)e(k)\right] \right\}, \tag{17}$$

cuja solução pode ser obtida zerando-se o gradiente com relação a $\boldsymbol{w}(k+1)$, dando

origem a:

$$\boldsymbol{w}(k+1) - \boldsymbol{w}(k) - \lambda \boldsymbol{x}(k) = \mathbf{0}, \tag{18}$$

ou, equivalentemente:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \lambda \boldsymbol{x}(k), \tag{19}$$

onde $\lambda$ pode ser encontrado impondo-se a restrição afim de (14), o que dá origem a

$$\lambda = \beta e(k), \tag{20}$$

o que faz que (19) coincida com (11). Assim, temos um paradigma alternativo para a derivação do LMS que dá origem à mesma equação de atualização. Existem diversas proposições de algoritmos de filtragem adaptativa que são obtidas apenas por meio de gradiente estocástico, enquanto que outras são classicamente derivadas mediante apelo ao princípio da distorção mínima. Por vezes, reposicionar a derivação de um algoritmo de um destes métodos para o outro pode permitir a inserção inéditas de estratégias que são típicas do novo paradigma. Embora tal reposicionamento seja por vezes simples, em outras vezes exige bastante engenhosidade. Quando obtido, ele permite a proposição de algoritmos inéditos, que por vezes superam o estado da arte. No Capítulo 2, ao reposicionarmos algoritmos costumeiramente derivados por meio de gradiente estocástico como algoritmos que resolvem um problema fundamentado no princípio da distorção mínima, podemos incorporar aos algoritmos iniciais diversas estratégias típicas de algoritmos obtidos por meio deste princípio.

## 1.2- Controle Ativo de Ruído (ANC)

Os sistemas de controle ativo de ruído em malha fechada são baseados em dois possíveis métodos: *feedforward* e *feedback*. No controle *feedforward*, uma entrada de ruído de referência coerente é detectada antes de se propagar pelo alto-falante de cancelamento. Já no controle de *feedback*, o controlador de ruído ativo tenta cancelar o ruído sem uma entrada de referência [12].

A Figura 3 ilustra um esquemático de um sistema *feedforward* simples de banda

larga. Note que uma quantidade considerável de ruído de banda larga é produzida no duto. O sinal de referência $x(k)$ é detectado por um microfone de entrada próximo à fonte de ruído antes de passar por um alto-falante. O cancelador de ruído utiliza de um sinal de entrada de referência para gerar um sinal $y(k)$ de igual amplitude, mas com uma defasagem de $180°$. Este sinal anti-ruído é usado para fazer com que o alto-falante produza um som de cancelamento que atenue o ruído primário no duto.

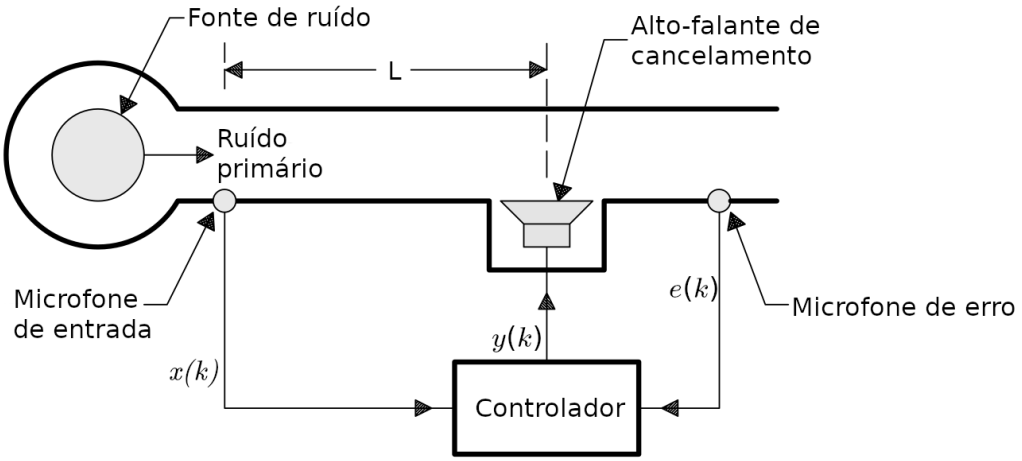



Figura 3 – Esquemático de um sistema *feedforward* de banda larga. Adaptado de [12].

O princípio básico da abordagem *feedforward* de banda larga consiste no fato de que o atraso de propagação entre o sensor de ruído (ou seja, microfone de entrada) e a fonte de controle ativo (ou seja, alto-falante) possibilita que o ruído original seja reintroduzido em oposição de fase. O espaçamento entre o microfone e o alto-falante deve satisfazer os princípios de causalidade e alta coerência, o que significa que a referência deve ser medida com antecedência suficiente para que o sinal anti-ruído possa ser gerado no momento em que o sinal de ruído atinge o alto-falante. Além disso, o sinal de ruído no alto-falante deve ser muito semelhante ao ruído medido no microfone de entrada de entrada, ou seja, o canal acústico não pode alterar significativamente o ruído. O cancelador de ruído usa o sinal de entrada para gerar um sinal $y(k)$ de igual amplitude e defasagem de $180°$ em relação a $x(k)$. Este ruído é enviado para um alto-falante e usado para cancelar o ruído indesejado [12].

Uma primeira distinção entre vertentes de mecanismos de ANC secciona-os entre os de malha aberta e de malha fechada. Em geral, as soluções em malha aberta despertam pouco interesse na comunidade científica, por razões que foram elucidadas em trabalhos anteriores da literatura, como a necessidade de que os componentes do sistema

de controle apresentam atraso com fase zero ao longo de toda a faixa de frequência de operação e de que os parâmetros não controláveis do sistema (como temperatura) sejam mantidos praticamente constantes [24]. Por essas razões, essa tese se concentrará em sistemas de malha fechada para propósitos de ANC. Como vimos, dentre as técnicas de ANC de malha fechada, duas diferentes modalidades são disseminadas na literatura: a direta (*feedforward*) e a com retroalimentação (*feedback*).

Consideremos por ora a modalidade *feedforward*, num exemplo prático: a supressão de ruído numa cabine de avião, com um sinal oriundo da turbina utilizado como referência. A partir deste sinal de referência, cabe a um cancelador eletrônico (mais especificamente, microprocessado) extrair um sinal de supressão, a partir de um filtro cuja entrada é o sinal de referência. Cumpre então à saída deste filtro alimentar a fonte secundária - no caso, um alto-falante, cuja função de transferência pode ser incorporada à planta eletroacústica do secundário. Neste exemplo, o sinal de referência pode ser inclusive uma senoide simples, adquirida a partir de um tacômetro situado próximo a um dos motores [16].

Após a saída do secundário, existe uma interferência física entre as formas de onda do secundário e do primário, cujo resultado é um sinal cuja energia se deseja reduzir. Ou seja, este sinal é um distúrbio, cuja medição dá origem a um sinal de erro que, medido por um microfone, é utilizado pelo controlador para ajustar os coeficientes do filtro, conforme ilustrado pela Figura 4. Uma adoção precoce de tal esquema (ajustado manualmente) foi usada, ainda em 1956, por William Conover, na General Electric Co., para ajuste do ruído produzido por transformadores de distribuição de potência [16]. Como exemplo de sistema MIMO da modalidade *feedforward* de cancelamento ativo de ruído, podemos elencar o sistema feito pela Ultra Electronics, de Cambridge, para aeronaves Saab 2000 de 50 assentos, num sistema que adotava 37 alto-falantes e 72 microfones [16].

O artigo de Burgess [24] aplica de forma pioneira conceitos da área de processamento de sinais para o problema de cancelamento ativo de ruído, constatação que serve para ilustrar o fato de que esta classe de problemas é objeto de atenção tanto da área de controle quanto da área de processamento de sinais. Mesmo utilizando uma tecnologia relativamente primitiva, Burgess [24] reporta o cancelamento de tons puros para um patamar pelo menos 10 dB abaixo do espectro do ruído para uma *Signal-to-noise ratio* (SNR) de até 0 dB.

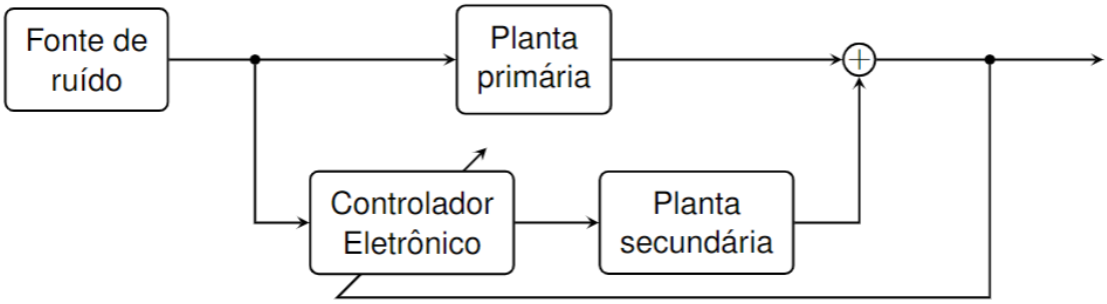



Figura 4 – Diagrama simplificado de um método de controle ativo de ruído do tipo *feedforward*, numa aplicação de cancelamento de ruído oriundo de uma máquina rotativa (o que incluiu turbinas). Tipicamente, um alto-falante emite o sinal que cancela o ruído.

Já a modalidade *feedback* em técnicas de ANC, assumindo-se a mesma aplicação, utilizaria apenas um microfone, um controlador eletrônico e um alto-falante, conforme descrito pela Figura 5. Note que a topologia desta modalidade é mais simples, já que a saída do microfone é modificada diretamente pelo algoritmo adaptativo (controlador eletrônico) para ser emitida pelo alto-falante. Observe que na modalidade com *feedback*, não é mais necessário um sinal de referência. Ademais, cabe projetar o controlador de sorte a prover *feedback* negativo (o que atenua a perturbação) ao longo da faixa de frequência de interesse. O projeto de tais esquemas não raro requer a confecção de diagramas de Bode e/ou de Nyquist. Cabe ressaltar que embora apresente um custo de implementação inferior, tal topologia tem diversas limitações, como problemas de instabilidade (similares aos apresentados por filtros de resposta infinita adaptativos), a incapacidade de supressão de ruído em todas as frequências simultaneamente e a restrição da largura de banda da atenuação do ruído [10], [16]. Por essa razão, esta tese enfoca técnicas de ANC pertencentes à modalidade *feedforward*.

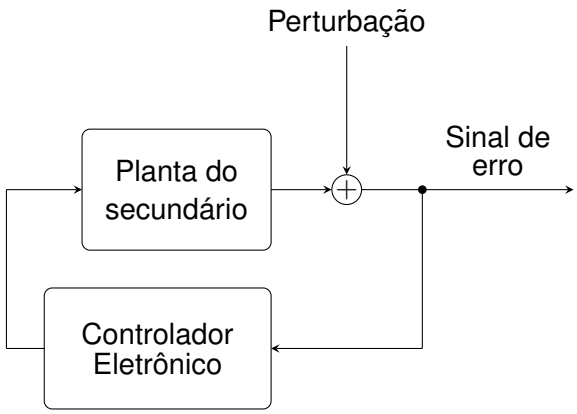



Figura 5 – Esquema simplificado de um método de controle ativo de ruído do tipo *feedback*.

### 1.2.1 Algoritmo FX-LMS Aplicado ao Cancelamento Ativo de Ruído

O algoritmo FX-LMS (do inglês *Filtered-X LMS*) é o algoritmo clássico de filtragem adaptativa adotado para efeitos de controle automático de ruído, que será generalizado no Capítulo 2, dando origem a algoritmos novos. Assumiremos por simplicidade que o sistema eletrônico ANC sob discussão compreende um microfone de referência, um alto-falante secundário, um microfone de erro e um controlador adaptativo [34]. Assume-se também um cancelador digital, implicando que os sinais egressos dos transdutores eletroacústicos são amostrados e processados em tempo real.

Devido ao seu grande leque de aplicações, as diferentes técnicas de ANC devem ser avaliadas sob prismas distintos, tal como as suas limitações fundamentais de desempenho, os limites práticos que restringem o seu desempenho, o balanço entre desempenho e complexidade, e ainda o projeto de uma arquitetura prática [34]. A Tabela 1 apresenta algum trabalhos empregando métodos da família FX-LMS ao cancelamento de ruído. Note que os algoritmos vêm sendo desenvolvidos para aplicações com condições específicas. Por exemplo, os algoritmos propostos por Huang, Xiao, Sun et al. [35] e Song e Park [36] possuem o passo de adaptação variável, porém, são aplicados na presença de ruído de banda estreita e banda larga, respectivamente. Mais recentemente, o algoritmo FX-LMS foi modificado a fim de acelerar a taxa de convergência quando o ruído a ser cancelado é periódico [37]. É possível notar que essa técnica ainda vem sendo pesquisada e utilizada para essa funcionalidade.

O algoritmo FX-LMS pode ser considerado um dos algoritmos chave, a partir do qual diversos outros algoritmos são desenvolvidos. Seu diagrama em blocos é apresentado na Fig. 6. Primeiramente, observe como se modela a vibração que se pretende remover como a soma de dois termos: $d(k)$ e $\nu(k)$, onde $d(k)$ é obtido pela convolução entre a função de transferência (desconhecida) $\boldsymbol{w}^\star$ e o sinal de referência $x(k)$. Ao contrário de um sistema de identificação, no qual cumpriria ao filtro adaptativo $\boldsymbol{w}(k)$ identificar $\boldsymbol{w}^\star$, na configuração em questão importa ressaltar que a saída do filtro $y(k)$ ainda sofre o efeito de uma função de transferência $\boldsymbol{S}$ (ou seja, o filtro do *secundário*), a qual modela o acoplamento acústico entre a saída do filtro e o ponto cuja vibração se almeja mitigar. Assumindo o caso de eliminação de vibração num ponto de um duto, a função de transferência $\boldsymbol{S}$ captura o fato de que o sinal do alto-falante (obtido a partir da

| Ano | Contribuições | Referência |
|---|---|---|
| 1987 | FX-LMS para controle ativo de sons e de vibrações | [38] |
| 1992 | FX-LMS no domínio da frequência para ANC | [39] |
| 1993 | FX-LMS mutuamente acoplado com estrutura entrelaçada para descorrelacionar a referência | [40] |
| 1993 | Algoritmo de sub-banda sem atraso para ANC | [41] |
| 2000 | FX-RLS (*recursive least square*) multi-canal e FX-FTF (*fast tranversal filter*) multi-canal | [42] |
| 2001 | Algoritmo de sub-banda sem atraso de baixo custo | [43] |
| 2003 | FX-LMS com pesos restringidos para ruídos de banda larga | [44] |
| 2005 | FX-AP (*affine projection*) multi-canal e análise do seu desempenho | [45] |
| 2006 | FX-LMS com sinal de referência saturado para controle de ruído impulsivo | [46] |
| 2007 | FX-AP com regularização ótima | [47] |
| 2007 | FX-LMS em blocos de estrutura reduzida baseado na transformada rápida de Fourier e na transformada rápida de Hartley | [48] |
| 2007 | FX-LMS em cascata para ANC não-linear | [49] |
| 2007 | FX-LMS, sem estimação do caminho secundário, baseado em análise geométrica e na propriedade SPR (*strictly positive real*) | [50] |
| 2013 | Algoritmo FX-LMS com passo de adaptação variável para ANC de banda estreita | [35] |
| 2015 | Algoritmo FX-AP de passo de adaptação variável ótimo | [36] |
| 2016 | Algoritmo FX-LMK (*least mean kurtosis*) | [51] |
| 2018 | FX-LMS modal para ANC global dentro de uma cavidade acústica | [52] |
| 2019 | Escolha ótima do fator de esquecimento do algoritmo *leaky* FX-LMS com saída limitada | [53] |
| 2020 | Emprego do FX-LMS com compensação de viés | [54] |
| 2021 | FX-LMS modificado para acelerar a taxa de convergência de um sistema ANC que lida com distúrbios periódicos. | [37] |
| 2022 | Modelo analítico que prevê o desempenho do algoritmo FXLMS distribuído na presença de descasamento de frequência | [55] |
| 2023 | Inserção de penalidade de convolução circular no algoritmo FXLMS estendido | [56] |
| 2023 | Sistema de ANC multicanal e direcional, baseado na direção de chegada | [57] |
| 2024 | Combinação de sub filtros de controle por meio de filtro bayesiano e rede neural convolucional | [58] |
| 2024 | Reutilização do erro do algoritmo FX-LMS para compensar a discrepância entre o ruído no primário e a saída estimada | [59] |

Tabela 1 – Exemplificação de alguns algoritmos de ANC da família FX. Versão adaptada e estendida de [10].

saída do filtro adaptativo $y(k)$) ainda passará por transformações antes de ser capturado pelo microfone situado no ponto das vibrações.

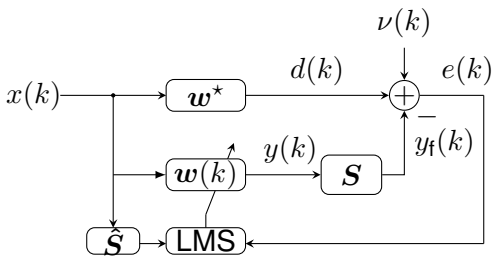



Figura 6 – Diagrama em blocos do algoritmo FX-LMS.

Na prática, assim como $\boldsymbol{w}^\star$, o filtro do secundário também deve ser estimado; normalmente, tal procedimento de identificação utiliza algum identificador de sistema cujo aprendizado se efetiva com a inserção de um sinal (possivelmente ruído branco) na saída do filtro adaptativo. Tal estimação não será objeto de atenção desta tese, embora modelemos por vezes o impacto que uma estimativa inacurada do secundário pode apresentar no desempenho do algoritmo.

Assumindo-se que o filtro do secundário é *Finite Impulse Response* (FIR) e apresenta $M$ coeficientes, a componente $y_{\mathrm{f}}(k)$ no ponto de vibração pode ser descrita como

$$y_{\mathrm{f}}(k) = \sum_{i=0}^{M-1} s_i y'(k-i), \tag{21}$$

onde $s_i$ denota o $i$-ésimo coeficiente do filtro secundário. O efeito do secundário deve ser compensado pelo algoritmo adaptativo, o que é feito por meio de um conjunto de estimativas dos parâmetros $\hat{s}_i$ (para $i \in \{0, 1, \ldots, \hat{M}-1\}$), o qual é usado para sintetizar o sinal de entrada do algoritmo FX-LMS:

$$\hat{\boldsymbol{x}}_{\mathrm{f}}(k) = \sum_{i=0}^{\hat{M}-1} \hat{s}_i \boldsymbol{x}(k-i), \tag{22}$$

onde $\hat{M}$ é uma estimativa de $M$. Suporemos doravante, a menos que se especifique o contrário, que $\hat{M} = M$, de modo que $\hat{M}$ e $M$ podem ser utilizados indistintamente. A *filtragem* do sinal de entrada ($x(k)$) é comumente descrita pela expressão em inglês "'Filtered-X", cuja sigla (FX) dá origem ao nome do algoritmo (FX-LMS).

A equação de atualização do algoritmo FX-LMS utiliza o mesmo formato da equação de atualização do LMS, porém com a substituição do sinal de entrada $\boldsymbol{x}(k)$ por

sua versão filtrada artificialmente $\boldsymbol{x}_{\mathrm{f}}(k)$, dando origem a

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta e(k)\hat{\boldsymbol{x}}_{\mathrm{f}}(k), \tag{23}$$

onde o sinal de erro $e(k)$ é descrito por

$$e(k) = d(k) - y_{\mathrm{f}}(k) + \nu(k). \tag{24}$$

Por fim, cabe ressaltar que a equação de atualização, Equação (23), requer um notal de $(M+1)N-1$ adições e $(M+2)N+1$ multiplicações por iteração.

### 1.2.2 Solução Ótima do Algoritmo FX-LMS

Assumindo sinais estocásticos estacionários e uma hipótese estatística fraca acerca do ruído de medição $\nu(k)$, podemos encontrar uma fórmula fechada para o valor ótimo em termos quadráticos (ou seja, a solução de Wiener) $\boldsymbol{w}^{\star}$ do vetor de coeficientes adaptativos $\boldsymbol{w}(k)$ [60]. A supracitada hipótese pode ser enunciada do seguinte modo:

Hipótese acerca do ruído (HR). O ruído aditivo $\nu(k)$ apresenta média zero, sendo estatisticamente independente das demais variáveis aleatórias.

A hipótese HR é ubíqua na literatura, embora possa ser contornada (veja [61], [62], por exemplo). Assumindo-se um filtro fixo $\boldsymbol{w}$, temos que

$$y_{\mathrm{f}}(k) = \sum_{i=0}^{M-1} s_i \boldsymbol{w}^T \boldsymbol{x}(k-i), \tag{25}$$

com o respectivo erro quadrático podendo ser obtido a partir de (24):

$$\begin{aligned} e^2(k) &= d^2(k) - 2\sum_{i=0}^{M-1} s_i \boldsymbol{w}^T \boldsymbol{x}(k-i)d(k) + \mathcal{O}[\nu(k)] \\ &\quad + \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} s_i s_j \boldsymbol{w}^T \boldsymbol{x}(k-i)\boldsymbol{x}^T(k-j)\boldsymbol{w} + \nu^2(k), \end{aligned} \tag{26}$$

onde $\mathcal{O}[\nu(k)]$ contém termos de primeira ordem do sinal de ruído $\nu(k)$.

Adotando-se o símbolo $\xi[\boldsymbol{w}]$ para o MSE, podemos obtê-lo pela aplicação do

operador de valor esperado $\mathbb{E}[\cdot]$ em (26). Após a adoção das hipóteses estatísticas descritas, $\xi[\boldsymbol{w}]$ pode ser formulado como

$$\xi[\boldsymbol{w}] = \mathbb{E}\left[d^2(k)\right] - 2\sum_{i=0}^{M-1} s_i \boldsymbol{p}_i^T \boldsymbol{w} + \mathbb{E}\left[\nu^2(k)\right] + \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} s_i s_j \boldsymbol{w}^T \boldsymbol{R}_{j-i} \boldsymbol{w}, \tag{27}$$

onde

$$\boldsymbol{p}_i \triangleq \mathbb{E}\left[d(k)\boldsymbol{x}(k-i)\right] \tag{28}$$

e

$$\boldsymbol{R}_{j-i} \triangleq \mathbb{E}\left[\boldsymbol{x}(k-i)\boldsymbol{x}^T(k-j)\right]. \tag{29}$$

Por meio das definições

$$\tilde{\boldsymbol{p}} \triangleq \sum_{i=0}^{M-1} s_i \boldsymbol{p}_i \tag{30}$$

e

$$\tilde{\boldsymbol{R}} \triangleq \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} s_i s_j \boldsymbol{R}_{j-i}, \tag{31}$$

podemos reescrever (27) de uma maneira mais sucinta:

$$\xi[\boldsymbol{w}] = \mathbb{E}\left[d^2(k)\right] - 2\tilde{\boldsymbol{p}}^T \boldsymbol{w} + \boldsymbol{w}^T \tilde{\boldsymbol{R}} \boldsymbol{w} + \sigma_\nu^2, \tag{32}$$

cujo ponto de mínimo (em termos de $\boldsymbol{w}$) pode ser obtido igualando-se ao vetor nulo o gradiente com relação a $\boldsymbol{w}$:

$$\nabla_{\boldsymbol{w}} \xi[\boldsymbol{w}] = -2\tilde{\boldsymbol{p}} + 2\tilde{\boldsymbol{R}}\boldsymbol{w} = \boldsymbol{0}, \tag{33}$$

o que nos leva a [63]

$$\boldsymbol{w}^\star = \left[\tilde{\boldsymbol{R}}\right]^{-1} \tilde{\boldsymbol{p}}, \tag{34}$$

igualdade que expressa a solução ótima de Wiener da configuração considerada na qual aplica-se o algoritmo FX-LMS. Observe que tal solução independe dos coeficientes estimados do secundário, pois estes apenas impactam a estimativa (adaptativa) do filtro $\boldsymbol{w}(k)$, e não do seu valor ótimo. Em termos práticos, a solução (34) pode ser implementada por meio de estimativas empíricas de $\tilde{\boldsymbol{R}}$ e de $\tilde{\boldsymbol{p}}$. Tal abordagem entretanto apresenta duas limitações severas: *i*) apresenta um alto custo computacional, devido à inversão da matriz $\tilde{\boldsymbol{R}}$ [64] e *ii*) exige que o sistema subjacente seja estacionário, uma hipótese inválida em diversas aplicações [26].

## 1.3- Comentários Finais

Este capítulo forneceu uma base conceitual necessária para a compreensão dos demais capítulos deste trabalho. Os algoritmos clássicos LMS (sob um enfoque de identificação de sistemas) e FX-LMS (que contempla o problema de ANC) foram apresentados, tal como as revisões de algumas derivações importantes. O próximo capítulo propõe variantes inéditas dos algoritmos LMS e FX-LMS, focadas em aumento da taxa de convergência quando a solução ótima é esparsa e com maior robustez perante ruídos de medição não gaussianos.

# 2- Extensões dos Algoritmos LMS e FX-LMS

Algoritmos de filtragem adaptativa tipicamente utilizam uma aproximação estocástica do erro quadrático médio para o processo de aprendizagem (vide o exemplo fornecido pela Equação (7)). Entretanto, diferentes expoentes do sinal de erro podem ser mobilizados na função custo, objetivando-se, por exemplo, melhor convergência ou desempenho assintótico [65], [66].

Nesse capítulo, os algoritmos de filtragem adaptativa *least mean logarithmic square* (LMLS) e *least logarithmic absolute difference* (LLAD) (originalmente propostos em [65]) são derivados de maneira inédita como a solução local de um problema de otimização. Esse novo paradigma permite a generalização desses algoritmos incorporando as técnicas de reúso de coeficientes e adaptação proporcional, obtendo-se os algoritmos propostos *improved $\mu$-law least mean logarithmic square with reuse of coefficients* (IMPLMLS-RC) e *improved $\mu$-law least logarithmic absolute difference with reuse of coefficients* (IMPLLAD-RC). Tais algoritmos são então estendidos para o contexto de cancelamento de ruído. As contribuições deste capítulo deram origem ao artigo [67].

Este capítulo é estruturado do seguinte modo: a Seção 2.1 apresenta os algoritmos LAD e LMF, enquanto a Seção 2.2 detalha os algoritmos LMLS e LLAD (que são, respectivamente, versões “logarítmicas” dos primeiros algoritmos). Já a Seção 2.3 apresenta a família proporcional de algoritmos adaptativos, enquanto a Seção 2.4 apresenta técnicas de reúso de coeficientes. Por sua vez, a Seção 2.5 combina todas estas técnicas, por meio de uma judiciosa transposição do problema de otimização resolvido por gradiente estocástico para um baseado no princípio da distorção mínima. O algoritmo resultante é então estendido para o contexto de ANC, e uma equação de fluxo de energia que o rege é derivada na Seção 2.6, a qual precede a Seção 2.7, que contém os resultados dos algoritmos propostos.

## 2.1- Algoritmos LAD, LMP LMF

Por razões didáticas, repetimos aqui a equação de atualização do LMS:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \boldsymbol{x}(k) e(k), \tag{35}$$

onde podemos observar que cada peso adaptativo é incrementado por uma pequena quantidade proporcional ao produto entre a sua entrada respectiva (isto é, as amostras $x(k)$) pelo sinal de erro. O cômputo deste produto é característico da maior parte dos algoritmos adaptativos. Muitos autores perceberam que o produto $x(k)e(k)$ corresponde a uma estimativa grosseira (ou estocástica) de uma operação de correlação, a qual pode inclusive ser alterada por um cômputo mais simples (em termos computacionais), tais como as escolhas sign$[x(k)]e(k)$, $x(k)$sign$[e(k)]$ ou mesmo sign$[x(k)]$sign$[e(k)]$. Algoritmos que utilizam o operador sign$(\cdot)$ apresentam a vantagem óbvia de redução de complexidade computacional, mas também demonstraram vantagens outras, tais como robustez a ruído impulsivo (ou, no caso de cancelamento de eco, uma menor sensibilidade à "*near-end speech*" [68]). A escolha mais tradicional consiste na "correlação não linear" $x(k)$sign$[e(k)]$, que deu origem ao algoritmo LAD (do inglês *least absolute difference*), cuja equação de atualização é

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \boldsymbol{x}(k) \text{sign}[e(k)], \tag{36}$$

onde a função sign$[\cdot]$ foi definida em (1). Observe que (36) revela que o algoritmo LAD (também conhecido como algoritmo sign-LMS) implementa um *clipping* (isto é, uma podagem ou saturação) do erro estimado [69]. Algoritmos do tipo LAD (que utilizam a função sign$[\cdot]$) têm atraído muita atenção em filtragem adaptativa. Uma das principais razões para esta popularidade ocorre em comunicações de dados em alta velocidade, nas quais um intervalo de um *baud* pode não ser suficiente longo para o cômputo de uma iteração do LMS; com a inclusão de uma mera operação de chaveamento (devido ao uso da função sign$(\cdot)$), o tempo para execução de uma iteração é reduzido [70]. Tal vantagem é cada vez mais importante, já que devido ao acréscimo sempre crescente nas taxas de dados, existe um interesse permanente em aumentar a velocidade de operação

dos filtros adaptativos [71]. A referência [72] trata da implementação de um equalizador do tipo *decision-feedback* em tecnologia CMOS de 0,13 $\mu$m. Ele utiliza o algoritmo LAD numa arquitetura do tipo *look ahead* exatamente para facilitar uma adaptação dependente da decisão atual que seja rápida o suficiente. Em termos gerais, pelo fato de reduzir (ou mesmo eliminar) as operações de multiplicação, o algoritmo LAD pode ser vantajoso em aplicações tais como equalizadores adaptativos, canceladores de eco adaptativos e preditores adaptativos [73].

O algoritmo LAD foi generalizado em [74], artigo que propõe o algoritmo *least mean p-power* (LMP), o qual ajusta os parâmetros do filtro de modo a efetuar uma otimização estocástico do erro elevado à $p$-ésima potência, o que se correlaciona com a função custo MPE (do inglês, *mean p-power error*):

$$\mathcal{F}_{\text{LMP}}[e(k)] = \frac{1}{p}|e(k)|^p, \tag{37}$$

de sorte que o algoritmo LMP se degenera no algoritmo LAD quando o projetista adota a escolha $p = 1$. A solução ótima da função custo MPE pode diferir da solução de Wiener podendo, a depender da configuração, estar mais próxima dos verdadeiros parâmetros do sistema do que a solução de Wiener [74]. A vantagem é garantida quando o cenário de filtragem encontra-se contaminado com ruído gaussiano generalizado [75]. Alguns autores alegam que a hipótese de gaussianidade do ruído não é válida na maior parte das aplicações, o que forneceria um argumento adicional para a adoção de funções custo não quadráticas [76]. Ademais, a solução ótima acima aludida pode apresentar melhor taxa de convergência ou mais robustez a ruído impulsivo quando $p = 1$ (ou seja, quando o algoritmo LAD - ou sign-LMS - é escolhido) [74]. Em termos genéricos, o algoritmo LMP pode ser superior ao LMS quando o ruído é não gaussiano, em especial quando a cauda da sua distribuição não é muito pesada [77], [78]. O parâmetro $p$ deve ser escolhido a partir da distribuição do ruído $\nu(k)$; na prática, tal escolha pode ser resultante de otimização em tempo real, por meio de aprendizado por reforço [79].

O algoritmo LMP pode ser descrito por meio da seguinte equação de atualização [74]:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta p|e(k)|^{p-2}e(k)\boldsymbol{x}(k). \tag{38}$$

Comparando-se (38) com (35), observa-se que podemos considerar o algoritmo LMP como um algoritmo LMS com um passo de aprendizagem variante no tempo $\beta(k)$,

com:

$$\beta(k) = \beta p|e(k)|^{p-2}. \tag{39}$$

A equação (39) revela que o algoritmo LMP com $p > 2$ pode apresentar um passo de aprendizagem maior do que o do LMS (que adota a escolha $p = 2$) no início da adaptação (ou seja, quando $|e(k)|$ é alto). Esta característica pode ser vantajosa, já que implica aumento da taxa de convergência. Quando o algoritmo converge, o erro $|e(k)|$ pode ser pequeno, tornando o passo de aprendizagem menor do que o respectivo do LMS (especialmente para $p > 2$), implicando um menor desajuste assintótico.

O algoritmo LMP é considerado um algoritmo da classe LMS com uma não linearidade no erro. Um dos primeiros algoritmos dessa classe a serem propostos (e um caso particular do LMP) é o algoritmo *least mean fourth* (LMF) [80], proposto por Walach e Widrow [81], cuja equação de atualização é descrita por:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \boldsymbol{x}(k) e^3(k). \tag{40}$$

O algoritmo LMF apresenta, em algumas circunstâncias, uma taxa de convergência melhor que a do LMS. Alguns autores alegam que o algoritmo LMF apresenta uma melhor compromisso entre taxa de convergência e desempenho em regime permanente [82], embora apresente problemas de instabilidade quando os elementos dos vetores de entrada são não limitados [83]. O algoritmo LMF também apresenta problemas de estabilidade e apresenta grande sensibilidade ao passo de aprendizagem do algoritmo [84]. Este problema de instabilidade limita a capacidade de rastreamento do algoritmo [85]. A análise teórica do algoritmo LMF também é bem mais difícil do que a do algoritmo LMS [86]. Algumas análises teóricas concluem que o limite de estabilidade do algoritmo LMF se reduz à medida que a cauda da função densidade de probabilidade do ruído se reduz, para o mesmo desajuste em regime permanente [83].

## 2.2- Algoritmos LMLS e LLAD

Dadas as limitações supracitadas do algoritmo LMF, a referência de Sayin et al. Sayin, Vanli e Kozat [65] propôs uma função de custo logarítmica de onde o algo-

ritmo LMLS é derivado por gradiente estocástico. Esse algoritmo combinou a taxa de convergência maior do LMF com a maior estabilidade do LMS. Assim, através da mesma função de custo, também é derivado o algoritmo LLAD, o qual combina a maior robustez a ruídos impulsivos do algoritmo *least absolute difference* (LAD) com o melhor desempenho do LMS.

Como apresentado por Sayin, Vanli e Kozat [65], ambos os algoritmos LMLS e LLAD são derivados a partir da função de custo

$$\mathcal{J}[e(k)] \triangleq \mathcal{F}[e(k)] - \frac{1}{\alpha}\ln(1+\alpha\mathcal{F}[e(k)]), \tag{41}$$

onde $\alpha > 0$ é um parâmetro de projeto e $\mathcal{F}[e(k)]$ é a função de custo convencional dependente de $e(k)$ em geral, como $\mathbb{E}[e^2(k)]$, em que $\mathbb{E}[\cdot]$ representa o operador valor esperado. Aplicando a aproximação estocástica $\mathcal{F}[f(e(k))] \approx f(e(k))$, a equação de atualização de $\boldsymbol{w}(k)$ que minimiza $\mathcal{J}[e(k)]$ é dada por [65]:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta\boldsymbol{x}(k)\frac{\partial f(e(k))}{\partial e(k)}\frac{\alpha f(e(k))}{1+\alpha f(e(k))}, \tag{42}$$

onde $\beta$, como antes, é o fator de aprendizado, que fornece um compromisso entre taxa de convergência, desempenho em regime permanente e probabilidade de instabilidade. Escolhendo $f(e(k)) = e^2(k)$, obtêm-se o algoritmo LMLS

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta\frac{\alpha\boldsymbol{x}(k)e^3(k)}{1+\alpha e^2(k)} \tag{43}$$

e, assumindo $f(e(k)) = |e(k)|$, obtém-se o algoritmo LLAD

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta\frac{\alpha\boldsymbol{x}(k)e(k)}{1+\alpha|e(k)|}. \tag{44}$$

Cumpre reforçar que as equações de atualização (43) e (44) foram encontradas por meio da adoção da técnica de otimização via gradiente estocástico.

**Reposicionando os algoritmos LMLS e LLAD sob o princípio da distorção mínima**

Reinterpretar o problema de otimização por meio do princípio da distorção mínima pode enriquecer a descrição de um algoritmo derivado por outro paradigma [87], permitindo a inserção de estratégias "naturais" do paradigma do princípio da distorção mínima no algoritmo originalmente derivado por meio de otimização via gradiente estocástico.

Neste diapasão, apresentamos o seguinte teorema inédito:

**Teorema 1**. As equações de atualização (43) e (44) são soluções do seguinte problema de otimização com restrições

$$\begin{aligned} &\min_{\boldsymbol{w}(k+1)} \|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^2 \\ &\text{s.t. } e_p(k) = (1 - g\left[e(k)\right])\, e(k), \end{aligned} \tag{45}$$

onde o erro *a posteriori* $e_p(k)$ é definido por

$$e_p(k) \triangleq d(k) - \boldsymbol{w}^T(k+1)\boldsymbol{x}(k). \tag{46}$$

e a função $g[e(k)]$ é

$$g_{\text{LMLS}}\left[e(k)\right] \triangleq \beta \frac{\alpha\|\boldsymbol{x}(k)\|^2 e^2(k)}{1 + \alpha e^2(k)} \tag{47}$$

e

$$g_{\text{LLAD}}\left[e(k)\right] \triangleq \beta \frac{\alpha\|\boldsymbol{x}(k)\|^2}{1 + \alpha|e(k)|} \tag{48}$$

para o LMLS e o LLAD, respectivamente.

*Demonstração*: Por meio da técnica dos multiplicadores de Lagrange, o problema de otimização com restrições (45) pode ser equivalentemente reescrito como o seguinte problema de otimização *sem* restrições:

$$\min_{\boldsymbol{w}(k+1)} \|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^2 + \lambda\left[e_{\mathsf{p}}(k) - (1 - g[e(k)])e(k)\right], \tag{49}$$

onde $\lambda \in \mathbb{R}$ denota o multiplicador de Lagrange, a ser determinado.

Igualando ao vetor zero o gradiente de (49) com relação a $\boldsymbol{w}(k+1)$ dá origem a

$$2\boldsymbol{w}(k+1) - 2\boldsymbol{w}(k) - \lambda\boldsymbol{x}(k) = \boldsymbol{0}, \tag{50}$$

ou:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \frac{\lambda}{2}\boldsymbol{x}(k). \tag{51}$$

Utilizando-se (51) na restrição $e_p(k) = (1 - g\,[e(k)])\,e(k)$ de (45) nos leva a:

$$\underbrace{d(k) - \boldsymbol{w}^T(k)\boldsymbol{x}(k)}_{=e(k)} - \frac{\lambda}{2}\|\boldsymbol{x}(k)\|^2 = e(k) - g[e(k)]e(k), \tag{52}$$

o que implica que:

$$\frac{\lambda}{2} = \frac{g[e(k)]e(k)}{\|\boldsymbol{x}(k)\|^2}. \tag{53}$$

Utilizando-se as definições (47) e (48) no resultado (53) nos leva à demonstração pretendida.

*Comentários*: As funções $g_{\mathrm{LMLS}}[e(k)]$ e $g_{\mathrm{LLAD}}[e(k)]$ combinam medidas do erro de ordem superior e inferior, o que permite o uso de medidas estatísticas de ordem superior quando as perturbações são pequenas [88]. Pode-se observar que o algoritmo LMLS se comporta como o LMF quando o erro é pequeno, o que melhora sua taxa de convergência, e como o LMS quando o erro é grande, evitando os supracitados problemas de estabilidade do LMF [65]. De modo similar, o algoritmo LLAD preserva a robustez do LAD em situações de ruídos impulsivos, já que $g_{\mathrm{LLAD}}[e(k)]$ se aproxima de $g_{\mathrm{LAD}}[e(k)]$ para erros de grande magnitude, e apresenta melhor convergência do que o LAD, já que $g_{\mathrm{LLAD}}[e(k)]$ se aproxima de $g_{\mathrm{LMS}}[e(k)]$ quando o erro é pequeno. Por fim, nota-se que as escolhas

$$g_{\mathrm{LMS}}[e(k)] = \beta\|\boldsymbol{x}(k)\|^2, \tag{54}$$

$$g_{\mathrm{LMF}}[e(k)] = \beta\|\boldsymbol{x}(k)\|^2 e^2(k), \tag{55}$$

e

$$g_{\mathrm{LAD}}[e(k)] = \frac{\beta\|\boldsymbol{x}(k)\|^2 e^2(k)}{|e(k)|} \tag{56}$$

dão origem aos algoritmos tradicionais LMS, LMF e LAD, respectivamente [89].

O Teorema 1 apresenta uma obtenção alternativa para os algoritmos LMLS e LLAD. Esta derivação alternativa nos permite inserir de maneira inédita as estratégias de *i*) adaptação proporcional e *ii*) reúso de coeficientes nestes algoritmos. Antes de tal aplicação, viabilizada pelo Teorema 1 (que é uma das contribuições deste capítulo) iremos brevemente descrever estas estratégias que serão incorporadas aos algoritmos.

## 2.3- Revisão sobre Algoritmos Proporcionais

Os algoritmos LMS, LMF e LAD são agnósticos com relação a uma eventual esparsidade da solução ótima a identificar [90]. No mundo real, existem muitos exemplos de sistemas esparsos cujas respostas ao impulso contém uma grande porcentagem de coeficientes próximos de zero ou mesmo nulos [91]. A referência [92] cita como exemplos de aplicações nas quais as plantas a identificar são esparsas os casos de cancelamento de eco, canais de HDTV e canais *wireless* de múltiplos percursos. Quando tal esparsidade existe, uma estratégia útil para acelerar a taxa de convergência consiste na adaptação proporcional [93]. Existem diversos algoritmos de filtragem adaptativa proporcionais que são populares. A referência [93] cita como exemplos o LMS normalizado proporcional (PNLMS, do inglês *proportionate normalized LMS*) [94], o PNLMS melhorado (IPNLMS, do inglês *improved PNLMS*) [95], o IPNLMS-$\ell_0$ [96], o *memory proportionate affine projection algorithm* (MPAPA) [97], o PNLMS esparso em blocos [98], o LMS normalizado com promoção de esparsidade [99], o algoritmo normalizado proporcional com critério de máximo correntropia [100] e o algoritmo proporcional com critério de máxima versoria [101].

Os algoritmos proporcionais almejam redistribuir a "energia de atualização" aumentando individualmente o fator de aprendizagem para os coeficientes de maior magnitude. Tal redistribuição pode ser modelada por meio de uma matriz diagonal $\mathbf{\Lambda}(k)$, de modo que a atualização de um algoritmo proporcional pode ser descrita por [102]:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \frac{\mathbf{\Lambda}(k)\boldsymbol{x}(k)}{\boldsymbol{x}^T(k)\mathbf{\Lambda}(k)\boldsymbol{x}(k)} e(k), \tag{57}$$

onde a diferenciação entre os diversos algoritmos desta família depende da escolha dos elementos da diagonal principal da matriz $\mathbf{\Lambda}(k)$.

Em esquemas de adaptação proporcional, cada coeficiente tem seu fator de aprendizado ajustado de acordo com sua magnitude relativa. Dependendo da esparsidade da função de transferência a ser identificada, esses algoritmos podem atingir taxas de convergência maiores do que algoritmos tradicionais [103]. Entretanto, para canais dispersivos, o desempenho dos algoritmos proporcionais pode ser significativamente degradado [104]. Uma versão modificada do algoritmo proporcional apresentado por

Benesty e Gay [105] foi proposto por Liu, Fukumoto e Saiki [106] para estimar a esparsidade do canal a ser identificado, ajustando automaticamente a distribuição da energia de adaptação entre os vários coeficientes. Tal ajuste permite uma taxa de convergência comparável a algoritmos tradicionais para canais dispersivos (ou seja, não esparsos). Segundo a proposta de Liu, Fukumoto e Saiki [106], esses elementos são determinados a partir da estimativa da esparsidade do canal $\xi$, definida como

$$\xi(k) = (1-\eta)\xi(k-1) + \eta\xi_{\mathbf{w}}(k),\ 0 < \eta \ll 1, \tag{58}$$

onde $\eta$ é um fator de esquecimento e $\xi_{\mathbf{w}}(k)$ é definido como

$$\xi_{\mathbf{w}}(k) \triangleq \frac{N}{N-\sqrt{N}}\left(1 - \frac{\|\boldsymbol{w}(k)\|_1}{\sqrt{N}\|\boldsymbol{w}(k)\|_2}\right), \tag{59}$$

onde $N$ é o número de coeficientes do canal. Finalmente, a esparsidade estimada $\xi(k)$ é mapeada para o parâmetro $\gamma(k) \triangleq 2\xi(k) - 1$ e os elementos da diagonal principal de $\boldsymbol{\Lambda}(k)$ são dados por

$$\lambda_n(k) \triangleq \frac{1-\gamma(k)}{2N} + \frac{(1+\gamma(k))G(|w_n(k)|)}{2\|G(|\boldsymbol{w}(k)|)\|_1 + \epsilon}, \tag{60}$$

onde $\epsilon$ é uma constante de regularização e $G(|w_n(k)|)$ é definido por [106]

$$G(|w_n(k)|) \triangleq \begin{cases} 400|w_n(k)|, & |w_n(k)| < 0.005 \\ 8.51|w_n(k)| + 1.96, & \text{caso contrário.} \end{cases} \tag{61}$$

Embora diversas outras escolhas para a matriz $\boldsymbol{\Lambda}(k)$ possam ser adotadas, a proposta de [106] é bastante popular, sendo portanto mobilizada no novo algoritmo a ser proposto neste capítulo.

### 2.4- Reúso de Coeficientes

A existência de ruído impacta negativamente o desempenho dos algoritmos de filtragem adaptativa. O desvio quadrático médio do vetor adaptativo pode alcançar valores muito altos, mesmo em regime permanente, quando a razão sinal ruído é baixa [107]. Este

problema pode ser mitigado mediante o concurso de técnicas de *reúso de coeficientes* (RC), as quais propõem a minimização de uma soma ponderada das normas euclidianas das diferenças entre o vetor de pesos $\boldsymbol{w}(k+1)$ e os últimos $L$ vetores de coeficientes adaptativos $\{\boldsymbol{w}(k), \cdots, \boldsymbol{w}(k-L+1)\}$, sujeita a uma restrição linear na nova solução $\boldsymbol{w}(k+1)$ [108]. O parâmetro $L \in \mathbb{N}$ é determinado pelo projetista [107], [108]. Tais técnicas podem reduzir o desvio quadrático médio assintótico ao reduzir a variância do estimador adaptativo, apresentando impacto pequeno na taxa de convergência, especialmente quando a razão sinal ruído é pequena [108]–[115].

Assim, a equação de atualização do algoritmo RC-LMS pode ser obtida por meio da solução do seguinte problema de otimização [116]:

$$\min_{\boldsymbol{w}(k+1)} \mathcal{F}_{\mathsf{RC}}[\boldsymbol{w}(k+1)] \triangleq \sum_{l=0}^{L-1} \rho^l \|\boldsymbol{w}(k+1) - \boldsymbol{w}(k-l)\|^2$$

$$\text{sujeito a } e_p(k) = (1 - \beta||\boldsymbol{x}(k)||^2)\overline{e}(k), \tag{62}$$

onde $\rho \in (0, 1]$ é um parâmetro ajustável que estabelece o peso relativo entre vetores adaptativos mais antigos e os mais recentes e o erro *suavizado* $\overline{e}(k)$ é descrito por

$$\overline{e}(k) \triangleq d(k) - \theta(\rho) \sum_{l=0}^{L-1} \rho^l \boldsymbol{w}^T(k-l)\boldsymbol{x}(k), \tag{63}$$

com

$$\theta(\rho) \triangleq \begin{cases} \left(\frac{\rho-1}{\rho^L-1}\right), & \text{para } \rho \neq 1, \\ \frac{1}{L}, & \text{para } \rho = 1. \end{cases} \tag{64}$$

A solução de (62) (por meio da técnica dos multiplicadores de Lagrange) dá origem à equação de atualização do algoritmo RC-LMS [109]:

$$\boldsymbol{w}(k+1) = \theta(\rho) \sum_{l=0}^{L-1} \rho^l \boldsymbol{w}(k-l) + \beta\overline{e}(k)\boldsymbol{x}(k). \tag{65}$$

## 2.5- Algoritmos propostos: FX-IMPLMLS-RC e FX-IMPLLAD-RC

Sendo distúrbios com relação a uma configuração ideal, os sinais de ruído apresentam efeitos deletérios no desempenho de sistemas de processamento de sinais [117]. Um dos problemas dos sistemas ANC consiste na baixa taxa de convergência e no desempenho reduzido na presença de ruído impulsivo [118]. Nesta seção, nós conjugamos a reformulação dos algoritmos LMLS e LLAD (reposicionados pelo Teorema 1 como solucionadores de um problema de otimização modulado pelo princípio da distorção mínima) com as técnicas de adaptação proporcional e reúso de coeficientes. Tal combinação dá origem a um novo algoritmo, que obtém resultados superiores a algoritmos no estado da arte. Estes algoritmos consistem na principal contribuição deste capítulo.

Considere o erro *a posteriori* generalizado

$$\mathcal{G}[\overline{e}(k)] \triangleq (1 - g\,[\overline{e}(k)])\,\overline{e}(k), \tag{66}$$

com $\overline{e}(k)$ definido por

$$\overline{e}(k) \triangleq d(k) - \frac{1-\rho}{1-\rho^L} \sum_{l=0}^{L-1} \rho^l \boldsymbol{w}^T(k-l)\boldsymbol{x}(k), \tag{67}$$

onde $L$ é o comprimento da janela de reúso dos coeficientes pregressos[1] e $0 < \rho < 1$ é um parâmetro que equilibra o desempenho em regime permanente e a taxa de convergência[2]. A família de algoritmos adaptativos proporcionais com reúso do vetor de coeficientes proposta pode ser obtida através da solução do problema de otimização

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{l=0}^{L-1} \rho^l \left\| \boldsymbol{w}(k+1) - \boldsymbol{w}(k-l) \right\|^2_{\boldsymbol{\Lambda}^{-1}(k)} \\ \text{s.t.} \quad & e_p(k) = \mathcal{G}[\overline{e}(k)], \end{aligned} \tag{68}$$

onde $\|\boldsymbol{x}\|^2_{\boldsymbol{\Lambda}} \triangleq \boldsymbol{x}^T(k)\boldsymbol{\Lambda}\boldsymbol{x}(k)$, $\|\cdot\|^2$ é a norma euclidiana e $\boldsymbol{\Lambda}(k)$ é uma matriz diagonal cujos elementos de sua diagonal principal ajustam a taxa de aprendizado de cada coeficiente através dos princípios da adaptação proporcional [106], de modo que os coeficientes de

[1] Aumentar $L$ melhora o desempenho em regime permanente, o que é importante quando a razão sinal ruído é baixa [108], ao custo de redução da taxa de convergência.

[2] Escolher um $\rho$ pequeno equivale a valorizar os coeficientes mais recentes. No caso limite $\rho \to 1$, todos os $L$ últimos vetores têm igual importância.

maior magnitude sejam atualizados de forma proporcionalmente maior.

A técnica dos multiplicadores de Lagrange permite derivar uma função com um mínimo local coincide com o mínimo do problema original com restrições (68), dada por

$$\begin{aligned}\mathcal{F}[\boldsymbol{w}(k+1)] &= \frac{1}{2}\sum_{l=0}^{L-1}\rho^l \left\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k-l)\right\|^2_{\boldsymbol{\Lambda}^{-1}(k)} \\ &\quad + \lambda\left[e_p(k)-\left(1-g\left[\overline{e}(k)\right]\right)\overline{e}(k)\right].\end{aligned}$$

A minimização de $\mathcal{F}[\boldsymbol{w}(k+1)]$ em relação a $\boldsymbol{w}(k+1)$ resulta em

$$\begin{aligned}\nabla_{\boldsymbol{w}(k+1)}\mathcal{F}[\boldsymbol{w}(k+1)] &= \boldsymbol{\Lambda}^{-1}(k)\boldsymbol{w}(k+1)\left(\frac{1-\rho^L}{1-\rho}\right) \\ &\quad -\boldsymbol{\Lambda}^{-1}(k)\sum_{l=0}^{L-1}\rho^l\boldsymbol{w}(k-l)-\lambda\boldsymbol{x}(k)=\mathbf{0}.\end{aligned}$$

Conclui-se que a forma da equação de atualização é

$$\boldsymbol{w}(k+1)=\frac{1-\rho}{1-\rho^L}\sum_{l=0}^{L-1}\rho^l\boldsymbol{w}(k-l)+\frac{1-\rho}{1-\rho^L}\lambda\boldsymbol{\Lambda}(k)\boldsymbol{x}(k). \tag{69}$$

Aplicando (69) na restrição (66) e empregando (67) resulta em

$$\frac{1-\rho}{1-\rho^L}\lambda=\frac{g\left[\overline{e}(k)\right]\overline{e}(k)}{\left\|\boldsymbol{x}(k)\right\|^2_{\boldsymbol{\Lambda}(k)}}. \tag{70}$$

Combinando (69) e (70), obtêm-se a equação de atualização genérica

$$\boldsymbol{w}(k+1)=\frac{1-\rho}{1-\rho^L}\sum_{l=0}^{L-1}\rho^l\boldsymbol{w}(k-l)+\frac{\boldsymbol{\Lambda}(k)\boldsymbol{x}(k)g\left[\overline{e}(k)\right]\overline{e}(k)}{\left\|\boldsymbol{x}(k)\right\|^2_{\boldsymbol{\Lambda}(k)}}, \tag{71}$$

a qual dá origem a dois algoritmos *inéditos*: o FX-IMPLMLS-RC (caso adotemos a escolha $g_{\text{LMLS}}[e(k)]$ da Equação (47)) e o algoritmo FX-IMPLAD-RC (caso adotemos a escolha $g_{\text{LLAD}}[e(k)]$ da Equação (48)). Espera-se que estes algoritmos conjuguem o aumento da taxa de convergência (devido ao fato de utilizarem a matriz $\boldsymbol{\Lambda}(k)$, de acordo com o princípio da adaptação proporcional) a um melhor desempenho em regime permanente (devido à adoção do reúso de coeficientes). Adicionalmente, espera-se que os algoritmos resultantes apresentem robustez a ruído impulsivo, devido ao fato de incorporarem os algoritmos LMLS e LLAD.

Cabe ressaltar que o problema de otimização (68) é equivalente a uma programação

quadrática convexa com um subespaço afim unidimensional e, portanto, possui solução de forma fechada.

O custo computacional para se avaliar o termo

$$\boldsymbol{\phi}(k) \triangleq \overbrace{\frac{1-\rho}{1-\rho^L}}^{\triangleq\theta(\rho)} \sum_{l=0}^{L-1} \rho^l \boldsymbol{w}(k-l) \tag{72}$$

pode ser reduzido pela recursão

$$\boldsymbol{\phi}(k+1) = \theta(\rho)\boldsymbol{w}(k+1) + \rho\boldsymbol{\phi}(k) - \theta(\rho)\rho^L \boldsymbol{w}(k-L+1), \tag{73}$$

indicando que o custo computacional devido a este termo não aumenta com $L$.

É simples verificar que a escolha de $L = 1$, $\rho \to 1$ e $\boldsymbol{\Lambda}(k) = \boldsymbol{I}_N$ (isto é, a matriz identidade de ordem $N$), que leva a $\overline{e}(k) = e(k)$, em conjunto com (47) e (48), gera os algoritmos LMLS e LLAD, que podem ser considerados como soluções particulares do problema descrito por (68). Assim, expomos uma formulação bastante genérica, capaz de gerar os algoritmos inéditos (FX-IMPLMLS-RC e FX-IMPLLAD-RC).

## 2.6- Relação de conservação de energia

Argumentos de conservação de energia podem ser utilizados para se prever o desempenho de algoritmos adaptativos [119], [120]. Seja

$$\tilde{\boldsymbol{w}}(k+1) \triangleq \boldsymbol{w}^\star - \boldsymbol{w}(k), \tag{74}$$

o vetor desvio. A equação de atualização genérica (71) pode ser reescrita como

$$\tilde{\boldsymbol{w}}(k+1) = \theta(\rho) \sum_{l=0}^{L-1} \rho^l \tilde{\boldsymbol{w}}(k-l) - \frac{\boldsymbol{\Lambda}(k)\boldsymbol{x}(k)g(\cdot)e(k)}{\|\boldsymbol{x}(k)\|^2_{\boldsymbol{\Lambda}(k)}}. \tag{75}$$

Definindo as variáveis relacionadas ao erro

$$e_{\mathsf{p}}^{\boldsymbol{\Sigma}}(k) \triangleq \boldsymbol{x}^T(k)\boldsymbol{\Sigma}\tilde{\boldsymbol{w}}(k+1), \tag{76}$$

$$e_{\mathsf{a},l}^{\boldsymbol{\Sigma}}(k) \triangleq \boldsymbol{x}^T(k)\boldsymbol{\Sigma}\tilde{\boldsymbol{w}}(k-l), \text{para} l \in \{0,1,\ldots,L-1\}, \tag{77}$$

onde $\boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}$ é uma matriz simétrica arbitrária. Pré-multiplicando ambos os lados de (75) por $\boldsymbol{x}^T(k)\Sigma$ resulta em

$$e_{\mathsf{p}}^{\boldsymbol{\Sigma}}(k) = \theta(\rho)\sum_{l=0}^{L-1}\rho^l e_{\mathsf{a},l}^{\boldsymbol{\Sigma}}(k) - \frac{\|\boldsymbol{x}(k)\|^2_{\boldsymbol{\Sigma}\boldsymbol{\Lambda}(k)}}{\|\boldsymbol{x}(k)\|^2_{\boldsymbol{\Lambda}(k)}}g(\cdot)e(k). \tag{78}$$

Isolando $g(\cdot)e(k)$ em (78) e substituindo esse termo em (75), obtêm-se

$$\begin{aligned}\tilde{\boldsymbol{w}}(k+1) + \theta(\rho)\boldsymbol{\Lambda}(k)\boldsymbol{x}(k)\sum_{l=0}^{L-1}\rho^l\frac{e_{\mathsf{a},l}^{\boldsymbol{\Sigma}}(k)}{\|\boldsymbol{x}(k)\|^2_{\boldsymbol{\Sigma}\boldsymbol{\Lambda}(k)}} = \\ \theta(\rho)\sum_{l=0}^{L-1}\rho^l\tilde{\boldsymbol{w}}(k-l) + \frac{\boldsymbol{\Lambda}(k)\boldsymbol{x}(k)e_{\mathsf{p}}^{\boldsymbol{\Sigma}}(k)}{\|\boldsymbol{x}(k)\|^2_{\boldsymbol{\Sigma}\boldsymbol{\Lambda}}(k)}.\end{aligned} \tag{79}$$

Avaliando a energia de ambos os lados de (79), como proposto por Kim, Lee e Song [107] e Shin e Sayed [121] a seguinte identidade pode ser estabelecida:

$$\begin{aligned}\|\tilde{\boldsymbol{w}}(k+1)\|^2 + \theta(\rho)e_{\mathsf{p}}^{\boldsymbol{\Lambda}(k)}\sum_{l=0}^{L-1}\frac{\rho^l e_{\mathsf{a},l}^{\boldsymbol{\Sigma}}(k)}{\|\boldsymbol{x}(k)\|^2_{\boldsymbol{\Sigma}\boldsymbol{\Lambda}(k)}} \\ +\theta^2(\rho)\|\boldsymbol{x}(k)\|^2_{\boldsymbol{\Lambda}^2(k)}\sum_{l_1=0}^{L-1}\sum_{l_2=0}^{L-1}\frac{\rho^{l_1}\rho^{l_2}e_{\mathsf{a},l_1}^{\boldsymbol{\Sigma}(k)}e_{\mathsf{a},l_2}^{\boldsymbol{\Sigma}}(k)}{\|\boldsymbol{x}(k)\|^4_{\boldsymbol{\Sigma}\boldsymbol{\Lambda}(k)}} = \\ \theta^2(\rho)\sum_{l_1=0}^{L-1}\sum_{l_2=0}^{L-1}\rho^{l_1}\rho^{l_2}\tilde{\boldsymbol{w}}^T(k-l_1)\tilde{\boldsymbol{w}}(k-l_2) \\ +\theta(\rho)e_{\mathsf{p}}^{\boldsymbol{\Sigma}}(k)\sum_{l=0}^{L-1}\rho^l\frac{e_{\mathsf{a},l}^{\boldsymbol{\Sigma}(k)}(k)}{\|\boldsymbol{x}(k)\|^2_{\boldsymbol{\Sigma}\boldsymbol{\Lambda}(k)}} + \frac{\left[e_{\mathsf{p}}^{\boldsymbol{\Sigma}}(k)\right]^2\|\boldsymbol{x}(k)\|^2_{\boldsymbol{\Lambda}(k)}}{\|\boldsymbol{x}(k)\|^4_{\boldsymbol{\Sigma}\boldsymbol{\Lambda}(k)}},\end{aligned} \tag{80}$$

que pode ser usada para se realizar uma análise estocástica exata do transiente do algoritmo proposto utilizando procedimentos de técnicas baseadas em energia [121]. É importante notar que no caso da análise de regime permanente, a escolha $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}$ simplifica (80). O desenvolvimento desta relação é alheio ao escopo desta tese.

## 2.7- Resultados de simulações

A seguir, o desvio quadrático médio, definido por

$$\text{MSD } (\text{dB}) \triangleq \mathbb{E}\left\{10\log_{10}\|\boldsymbol{w}^{\star} - \boldsymbol{w}(k)\|^2\right\}, \tag{81}$$

é adotado como a métrica para avaliar o desempenho dos algoritmos propostos.

### 2.7.1 Identificação de sistema

Para medir o desempenho dos algoritmos propostos, foram realizadas simulações com dois grupos de algoritmos. Para ambos os grupos, foi utilizado o seguinte arranjo: *(i)* $\boldsymbol{x}(k)$ é um sinal branco gaussiano de média zero e variância 1; *(ii)* $\boldsymbol{w}^{\star}$ é um canal de largura 64 (modelo 1 em TSG [122]); *(iii)* ruído de medição branco gaussiano de média zero e variância $\sigma_{\nu}^2 = 0{,}01$.

O primeiro grupo de algoritmos foi composto do LMS, LMLS e do IMPLMLS-RC. O fator de regularização $\epsilon = 1 \times 10^{-6}$ e o fator de esquecimento $\eta = 0{,}1$ foi utilizado em todas as simulações. As Figuras 7 e 8 mostram o MSD obtido por 100 experimentos de Monte Carlo independentes para os três algoritmos. Os parâmetros de simulação estão listados nas Tabelas 2 e 3. Esses valores foram escolhidos para que todos os algoritmos tivessem um MSD de regime permanente similar, permitindo uma comparação justa. Ademais, resultados de simulação para $L = 5$ e $L = 10$ também estão presentes na Figura 7 para mostrar o efeito da largura da janela de reuso no desempenho do algoritmo proposto. Pode-se verificar que maiores valores de $L$ tendem a reduzir o MSD de regime permanente mantendo uma taxa de convergência competitiva em relação ao algoritmo LMLS. Resultados de simulação para $\rho = 0{,}7$ e $\rho = 0{,}1$, com $L = 2$, também são mostrados na Figura 8. Pode ser observado que à medida que $\rho$ diminui, a taxa de convergência melhora enquanto mantém um MSD de regime permanente comparável ao do LMS e do LMLS. A Tabela 4 indica a iteração que cada algoritmo alcançou a convergência em cada simulação.

| Algoritmo | $\beta$ | $\alpha$ | $\rho$ | $L$ |
|---|---|---|---|---|
| LMS | $1{,}2 \times 10^{-3}$ | – | – | – |
| LMLS | $2{,}5 \times 10^{-2}$ | 1 | – | – |
| IMPLMLS-RC | 0,45 | 4,5 | 0,45 | $2; 5; 10$ |

Tabela 2 – Parâmetros utilizados nas simulações da Figura 7.

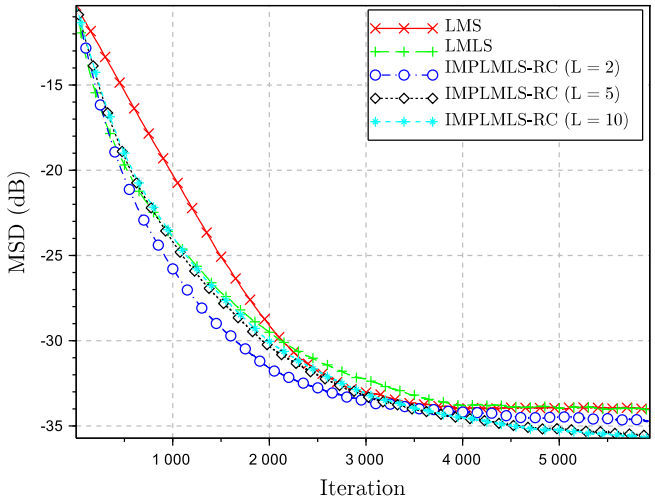



Figura 7 – Comparativo entre os algoritmos LMS, LMLS e IMPLMLS-RC para $\rho = 0{,}45$, $L = 2$, $L = 5$ e $L = 10$.

O segundo grupo de algoritmos consistiu do LAD, LLAD e IMPLLAD-RC. Um

| Algoritmo | $\beta$ | $\alpha$ | $\rho$ | $L$ |
|---|---|---|---|---|
| LMS | $1{,}2 \times 10^{-3}$ | – | – | – |
| LMLS | $2{,}5 \times 10^{-2}$ | 1 | – | – |
| IMPLMLS-RC | 0,45 | 4,5 | $0{,}1; 0{,}45; 0{,}7$ | 2 |

Tabela 3 – Parâmetros utilizados nas simulações da Figura 8.

| Algoritmo | Iteração |
|---|---|
| LMS | 4263 |
| LMLS | 4347 |
| IMPLMLS-RC $(L = 2; \rho = 0{,}45)$ | 4812 |
| IMPLMLS-RC $(L = 5; \rho = 0{,}45)$ | 5647 |
| IMPLMLS-RC $(L = 10; \rho = 0{,}45)$ | 5936 |
| IMPLMLS-RC $(L = 2; \rho = 0{,}1)$ | 3683 |
| IMPLMLS-RC $(L = 2; \rho = 0{,}7)$ | 5550 |

Tabela 4 – Iteração em que a convergência foi atingida nas simulações das Figuras 7 e 8.

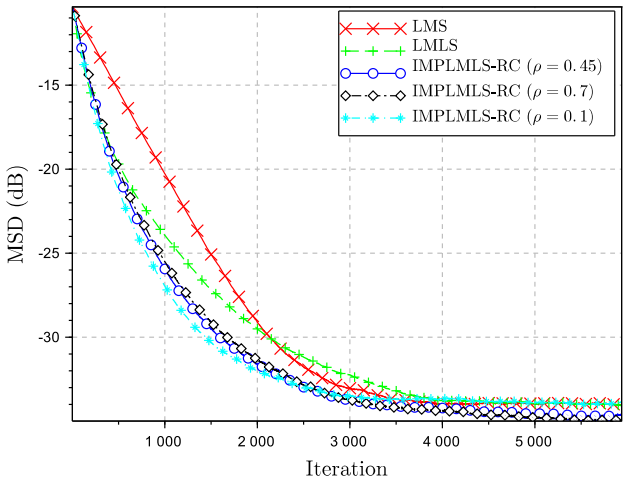



Figura 8 – Comparativoe entre os algoritmos LMS, LMLS e IMPLMLS-RC para $L = 2$, $\rho = 0{,}1$, $\rho = 0{,}45$ e $\rho = 0{,}7$.

ruído gaussiano impulsivo de média zero, variância $\sigma^2 = 1 \times 10^8$ e probabilidade de ocorrência de $1\%$ foi adicionado ao sinal desejado ($d(k)$) com as mesmas condições de simulação do primeiro grupo. As Figuras 9 e 10 mostram comparações entre os algoritmos do segundo grupo. Os parâmetros utilizados nessas simulações estão listados nas Tabelas 5 e 6. Esses valores foram escolhidos de modo a garantir um desempenho em regime permanente similar para todos os algoritmos. Resultados de simulações com $L = 5$ e $L = 10$ são apresentados na Figura 9. Pode ser observado que valores maiores de $L$ reduzem o MSD em regime permanente sem afetar de forma significativa a taxa de convergência. Resultados para $\rho = 0{,}1$ e $\rho = 0{,}7$ são mostrados na Figura 10. Nota-se que, para valores de $\rho$ maiores, há uma pequena melhora no MSD de regime permanente, enquanto que para valores menores de $\rho$ há um ganho insignificante na taxa de convergência, porém o regime permanente é degradado. A iteração em que cada algoritmo do segundo grupo convergiu nas simulações das Figuras 9 e 10 estão listadas na Tabela 7.

| **Algoritmo** | $\boldsymbol{\beta}$ | $\boldsymbol{\alpha}$ | $\boldsymbol{\rho}$ | $\boldsymbol{L}$ |
|---|---|---|---|---|
| LAD | $8 \times 10^{-4}$ | – | – | – |
| LLAD | $5 \times 10^{-3}$ | 2 | – | – |
| IMPLLAD-RC | 0,32 | 3 | 0,45 | $2; 5; 10$ |

Tabela 5 – Parâmetros utilizados nas simulações da Figura 9.

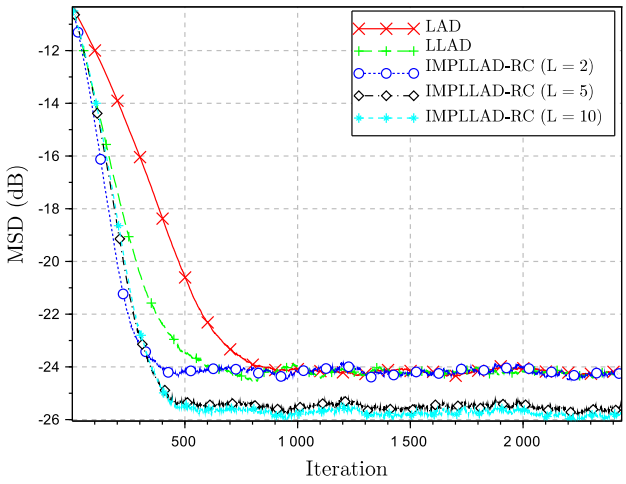



Figura 9 – Comparativo entre os algoritmos LAD, LLAD e IMPLLAD-RC para $\rho = 0{,}45$, $L = 2$, $L = 5$ e $L = 10$.

| **Algoritmo** | $\beta$ | $\alpha$ | $\rho$ | $L$ |
|---|---|---|---|---|
| LAD | $8 \times 10^{-4}$ | – | – | – |
| LLAD | $5 \times 10^{-3}$ | 2 | – | – |
| IMPLLAD-RC | 0,32 | 3 | $0{,}1; 0{,}45; 0.7$ | 2 |

Tabela 6 – Parâmetros utilizados nas simulações da Figura 10.

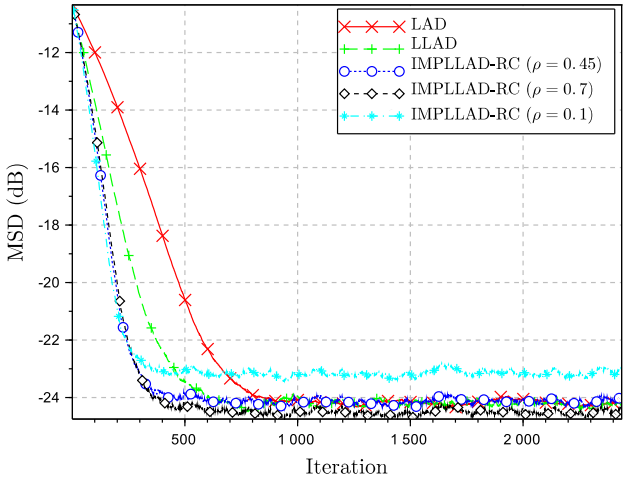



Figura 10 – Comparativo entre os algoritmos LAD, LLAD e IMPLLAD-RC para $L = 2$, $\rho = 0{,}1$, $\rho = 0{,}45$ and $\rho = 0{,}7$.

### 2.7.2 Controle Ativo de Ruído

Os algoritmos propostos neste capítulo até o momento se restringiram ao contexto de identificação adaptativa de sistemas. A transposição para técnicas de ANC é imediata.

| Algoritmo | Iteração |
|---|---|
| LAD | 973 |
| LLAD | 696 |
| IMPLLAD-RC $(L = 2; \rho = 0{,}45)$ | 593 |
| IMPLLAD-RC $(L = 5; \rho = 0{,}45)$ | 681 |
| IMPLLAD-RC $(L = 10; \rho = 0{,}45)$ | 714 |
| IMPLLAD-RC $(L = 2; \rho = 0{,}1$ | 500 |
| IMPLLAD-RC $(L = 2; \rho = 0{,}7)$ | 561 |

Tabela 7 – Iteração em que a convergência foi atingida nas simulações das Figuras 9 e 10

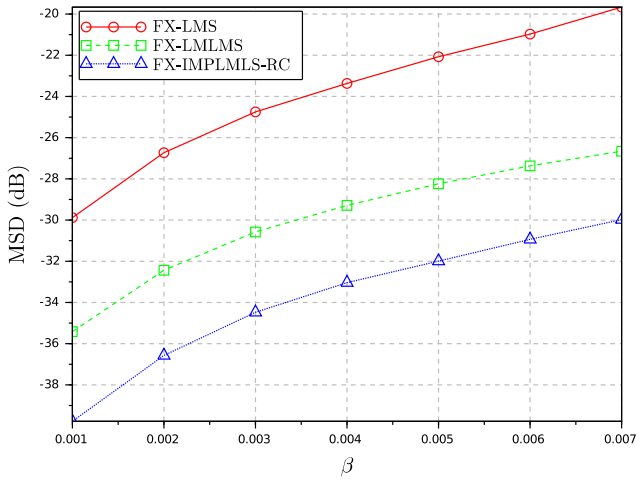



Figura 11 – MSD assintótico (em dB) para o experimento de ANC.

Assim, num cenário de ANC, a equação de atualização poderia ser assim descrita:

$$\boldsymbol{w}(k+1) = \frac{1-\rho}{1-\rho^L}\sum_{l=0}^{L-1}\rho^l \boldsymbol{w}(k-l) + \frac{\boldsymbol{\Lambda}(k)\boldsymbol{x}_{\mathrm{f}}(k) g\left[\overline{e}(k)\right]\overline{e}(k)}{\|\boldsymbol{x}_{\mathrm{f}}(k)\|^2_{\boldsymbol{\Lambda}(k)}}. \tag{82}$$

O cenário de ANC considerado utiliza a a função impulso como função de transferência ideal e $N = 20$.

A função de transferência do secundário é dada por

$$B(z) = 1 - 0.8z^{-1} + 0.6z^{-2} - 0.4z^{-3} + 0.2z^{-4},$$

e a variância do ruído aditivo branco e gaussiano é $\sigma_\nu^2 = 0.1$.

A Figura 11 apresenta o MSD assintótico para diferentes valores de $\beta$, com demais parâmetros apresentados na Tabela 8. Esta figura demonstra uma clara vantagem do

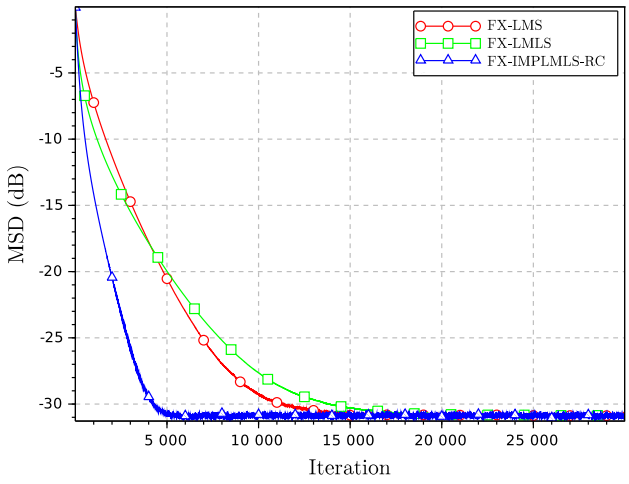



Figura 12 – Evolução do MSD (em dB) para o experimento de ANC para o FX-LMS (com $\beta = 0.000798$), FX-LMLS (com $\beta = 0.0028$) e FX-IMPLMLS-RC (com $\beta = 0.006$).

| **Algoritmo** | $\alpha$ | $\rho$ | $L$ |
|---|---|---|---|
| FX-LMS | – | – | – |
| FX-LMLS | 1 | – | – |
| FX-IMPLMLS-RC | 1 | 0.9 | 5 |

Tabela 8 – Parâmetros utilizados nas simulações das Figuras 11 e 12.

algoritmo proposto em regime permanente, particularmente devido à estratégia de reúso de coeficientes.

Já a Figura 12 apresenta as evoluções do MSD dos algoritmos quando os passos de aprendizagem são ajustados para obter valores de MSD similares em regime permanente. Os demais parâmetros empregados são mostrados na Tab. 8. Esta figura nos permite inferir que algoritmo proposto FX-IMPLMLS-RC apresenta a taxa de convergência mais rápida, devido à adoção do princípio de adaptação proporcional. Contudo, é importante destacar que para uma análise mais abrangente, seria necessário avaliar o comportamento dos algoritmos em cenários não esparsos, bem como em ambientes com variação temporal no sistema secundário ou sob diferentes tipos de ruído.

## 2.8- Comentários Conclusivos

Este capítulo recoloca dois algoritmos de filtragem adaptativa (LMLS e LLAD), tradicionalmente derivados por meio da técnica de gradiente estocástico, num paradigma determinístico, a partir do qual tais algoritmos são obtidos como solucionadores *exatos* de um problema de otimização local. Tal reformulação permitiu a incorporação da técnica de reúso de coeficientes a tais algoritmos, conferindo a eles maior robustez em cenários de razão sinal-ruído baixa. Por fim, o algoritmo resultante foi aplicado num contexto de cancelamento de ruído, o que ilustra como a transposição de ferramentas da área de identificação adaptativa de sistemas pode ser útil para melhorar técnicas de cancelamento ativo de ruído. Cumpre reforçar que as contribuições deste capítulo deram origem à publicação [67].

# 3- Cancelamento Ativo de Ruído com Gradiente Conjugado

Métodos de otimização de gradiente conjugado (CG, do inglês *conjugate gradient*) têm amplo uso nas áreas de cálculo matricial e otimização, especialmente em problemas de larga escala e sem restrições. Uma gama de técnicas baseadas em *Conjugate gradient* (CG) se utilizam do fato de que estas apresentam uma taxa de convergência superior à taxa de técnicas de gradiente tradicionais, além de apresentarem a propriedade de terminação finita para problemas quadráticos [123]. As técnicas de CG, em geral, minimizam uma função de custo executando pesquisas lineares apropriadas em direções vetoriais linearmente independentes, se aproveitando da ortogonalidade dos resíduos [124]. Mesmo para funções de custo não quadráticas, essas técnicas são efetivas, porém, neste caso, não há um limite para o número de iterações até a convergência no caso geral. Os algoritmos iterativos do tipo gradiente conjugado pertencem à família dos métodos de busca de linha exata, os quais são capazes de prover o máximo decremento da função custo a cada iteração [125]. É sabido que a otimização baseada em CG apresenta uma taxa de convergência superior à do *steepest descent* [126].

Recentemente, métodos CG foram aplicados a filtros adaptativos transversais projetados para tarefas de identificação de sistemas [127], [128], concentrando-se em obter estimativas de gradiente com maior acurácia. Tal extensão foi possível garantindo-se a conjugação par a par dos gradientes em cada passo de atualização do sistema. Essa característica relaxa a condição de conjugação entre as direções de busca, o que é impossível no contexto de filtragem adaptativa devido ao ruído de gradiente [123]. Nesse contexto, esquemas do tipo CG herdam tanto propriedades de algoritmos LMS quanto de algoritmos *Recursive Least Square* (RLS), no sentido de que apresentam maior taxa de convergência do que o LMS e complexidade computacional reduzida comparada ao RLS (que requer a inversão da matriz de auto-correlação do vetor de entrada) [127], [129].

Esse capítulo demonstra de maneira inédita que problemas de cancelamento ativo de ruídos também podem ser atacados por métodos CG, dado origem a toda uma nova família de soluções apta para resolver este problema, sob a ótica de processamento de sinais.

### 3.1- Identificação adaptativa de sistemas por meio de gradiente conjugado

Esta seção apresenta brevemente o algoritmo de CG aplicado à identificação de sistemas em filtragem adaptativa. Esta seção é fortemente baseada em [129]. Em termos quadráticos, assumindo uma solução ótima única para todos os instantes de tempo, a solução de Wiener $\boldsymbol{w}^\star$ para um algoritmo de filtragem adaptativa atende à seguinte igualdade:

$$\boldsymbol{R}\boldsymbol{w}^\star = \boldsymbol{b}, \tag{83}$$

onde $\boldsymbol{R} \in \mathbb{R}^{N\times N}$ é a matriz de autocorrelação do vetor de entrada[1]:

$$\boldsymbol{R} \triangleq \mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right], \tag{84}$$

e $\boldsymbol{b} \in \mathbb{R}^N$ é o vetor de correlação cruzada entre o vetor de entrada $\boldsymbol{x}(k)$ e o sinal $d(k)$:

$$\boldsymbol{b} \triangleq \mathbb{E}\left[\boldsymbol{x}(k)d(k)\right]. \tag{85}$$

Embora uma solução clássica para (83) seja o cômputo de $\boldsymbol{w}^\star = \boldsymbol{R}^{-1}\boldsymbol{b}$, o método CG permite obter uma solução sem a necessidade da inversão matricial, uma operação demasiado custosa. Se $\boldsymbol{R}$ e $\boldsymbol{b}$ são estimados por meio de uma janela deslizante ao longo dos dados, então o método CG pode ser interpretado como um método baseado em gradiente estocástico [129].

O algoritmo tradicional de CG (não adaptativo) pode ser descrito conforme descrito no Algoritmo 1 [130], onde $\overline{\alpha}(k)$ é o passo de aprendizagem, $\overline{\boldsymbol{\beta}}(k)$ garante a ortogonalidade em termos de $\boldsymbol{R}$ para o vetor de direção $\overline{\boldsymbol{p}}(k)$ e $\boldsymbol{g}(k)$ é o vetor de resíduo definido como:

$$\boldsymbol{g}(k) \triangleq \boldsymbol{b} - \boldsymbol{R}\boldsymbol{w}(k), \tag{86}$$

o qual é nulo quando a otimização termina (afinal, um $\boldsymbol{g}(k)$ nulo significa que a solução $\boldsymbol{w}(k)$ atende ao critério (83)).

A formulação descrita no Algoritmo 1 apresenta algumas propriedades desejáveis, entre as quais [129]:

[1] Observe que $\boldsymbol{R}$ equivale à matriz $\boldsymbol{R}_{i-j}$ definida em (29) quando $i = j$.

Algoritmo 1 – Algoritmo CG não adaptativo

**Entrada:** $\boldsymbol{w}(0)$, $g(0) = \boldsymbol{b}$, $\rho(0)$, $\overline{\boldsymbol{p}}(1) = \boldsymbol{g}(0)$
**Saída:** $\boldsymbol{w}(k)$
**Inicialização:** $\boldsymbol{w}(0) = \boldsymbol{0}$, $g(0) = \boldsymbol{b}$, $\rho(0) = \boldsymbol{g}^T(0)\boldsymbol{g}(0)$, $\overline{\boldsymbol{p}}(1) = \boldsymbol{g}(0)$

**para** $k = 1, 2, \ldots$ **faça**
- $\boldsymbol{v}(k) = \boldsymbol{R}\boldsymbol{p}(k)$
- $\overline{\alpha}(k) = \rho(k-1)/\overline{\boldsymbol{p}}^T(k)\boldsymbol{v}(k)$
- $\boldsymbol{w}(k) = \boldsymbol{w}(k-1) + \overline{\alpha}(k)\overline{\boldsymbol{p}}(k)$
- $\boldsymbol{g}(k) = \boldsymbol{g}(k-1) - \alpha(k)\boldsymbol{v}(k)$
- $\rho(k) = \boldsymbol{g}^T(k)\boldsymbol{g}(k)$
- $\overline{\beta}(k) = \rho(k)/\rho(k-1)$
- $\overline{\boldsymbol{p}}(k+1) = \boldsymbol{g}(k) + \overline{\beta}(k)\overline{\boldsymbol{p}}(k)$

1. Conjugação com relação a $\boldsymbol{R}$ dos vetores $\overline{\boldsymbol{p}}(k)$:

$$\boldsymbol{p}^T(k)\boldsymbol{R}\boldsymbol{p}(k) = 0, \text{ para todo } k \neq i \tag{87}$$

2. Ortogonalidade dos vetores de gradiente residuais:

$$\boldsymbol{g}^T(k)\boldsymbol{g}(i) = 0, \text{ para } i = 1, \ldots, k-1 \tag{88}$$

3. Atendimento ao teorema do subespaço expandido [131]:

$$\boldsymbol{g}^T(k)\overline{\boldsymbol{p}}(k) = 0, \text{ para } i \leq k. \tag{89}$$

4. A propriedade de término finito, por meio da qual $k_{\text{max}} \leq N$ é suficiente para que o algoritmo efetue a minimização pretendida.

5. A propriedade descendente:

$$\overline{\boldsymbol{p}}^T(k+1)\boldsymbol{g}(k) \geq 0. \tag{90}$$

Estas propriedades do algoritmo CG clássico (que atua a partir do conhecimento de $\boldsymbol{R}$ e de $\boldsymbol{b}$ ao longo de todos os dados) não são totalmente atendidas quando efetuamos aproximações no algoritmo para efeito de implementação em tempo real. Tal adaptação para operação em tempo real não é trivial. A seguir, apresentamos um algoritmo inédito para o problema de cancelamento automático de ruído por meio de uma adaptação

judiciosa do algoritmo CG para uma aplicação em tempo real.

## 3.2- Proposição de um método de cancelamento automático de ruído baseado em gradiente conjugado

Para aplicações em tempo real que demandam a estimativa da matriz $\tilde{\boldsymbol{R}}$ (vide (34)), é importante que tal matriz seja estimada iterativamente, à medida que os dados são adquiridos. Nessa tese, propomos a implementação de uma estimativa $\boldsymbol{R}_{\hat{s}\hat{s}}(k)$, que aproxima $\tilde{\boldsymbol{R}}$ por meio de uma solução de baixo custo computacional, via uma janela de dados com decaimento exponencial:

$$\boldsymbol{R}_{\hat{s}\hat{s}}(k) \triangleq \sum_{\tau=0}^{k} \gamma^{k-\tau} \hat{\boldsymbol{x}}_{\mathrm{f}}(k)\hat{\boldsymbol{x}}_{\mathrm{f}}^{T}(\tau) = \gamma \boldsymbol{R}_{\hat{s}\hat{s}}(k-1) + \hat{\boldsymbol{x}}_{\mathrm{f}}(k)\hat{\boldsymbol{x}}_{\mathrm{f}}^{T}(k), \tag{91}$$

onde $0 \ll \gamma < 1$ é o fator de esquecimento ajustável [28]. Nota-se que essa aproximação utiliza o vetor de entradas observável $\hat{\boldsymbol{x}}_{\mathrm{f}}(k)$, cuja notação enfatiza o fato de que o cômputo do sinal de entrada do algoritmo utiliza a planta estimada do secundário, e não a função de transferência real do secundário:

$$\hat{\boldsymbol{x}}_{\mathrm{f}}(k) = \sum_{i=0}^{\hat{M}-1} \hat{s}_i \boldsymbol{x}(k-i). \tag{92}$$

*Comentário*: Observe que em (92) não utilizamos os coeficientes reais $s_i$, mas os estimados, já que na prática temos acesso tão somente às estimativas destes coeficientes. A filtragem em (92) é inédita quando utilizada com o algoritmo CG.

De maneira similar, o vetor de correlação cruzada entre o vetor de entrada e a saída desejada é aproximado iterativamente (em tempo real) por:

$$\boldsymbol{b}_{\hat{s}}(k) \triangleq \sum_{\tau=0}^{k} \gamma^{k-\tau} \hat{\boldsymbol{x}}_{\mathrm{f}}(\tau)d(\tau) = \gamma \boldsymbol{b}_{\hat{s}}(k-1) + \hat{\boldsymbol{x}}_{\mathrm{f}}(k)d(k), \tag{93}$$

onde optamos por utilizar o mesmo parâmetro ajustável $\gamma$ que o adotado em (91), de sorte a evitar uma proliferação desnecessária do número de parâmetros ajustáveis.

Em esquemas baseados em CG, a equação de atualização de $\boldsymbol{w}(k)$ assume a

seguinte forma:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \alpha(k)\overline{\boldsymbol{p}}(k), \tag{94}$$

onde $\alpha(k) \in \mathbb{R}$ é o passo de aprendizado e $\overline{\boldsymbol{p}}(k) \in \mathbb{R}^N$ é a direção de atualização. No algoritmo CG tradicional, a direção de atualização deve satisfazer a condição de conjugação $\overline{\boldsymbol{p}}^T(k)\tilde{\boldsymbol{R}}\overline{\boldsymbol{p}}(k-1) = 0$. Na condição estocástica considerada, tal restrição é relaxada, resultando na conjugação de Markov:

$$\overline{\boldsymbol{p}}^T(k)\boldsymbol{R}_{\hat{s}\hat{s}}(k)\overline{\boldsymbol{p}}(k-1) = 0. \tag{95}$$

A partir da solução ótima, observa-se que

$$\boldsymbol{g}(k) \triangleq \boldsymbol{R}_{\hat{s}\hat{s}}(k)\boldsymbol{w}(k) - \boldsymbol{b}_{\hat{s}}(k) \tag{96}$$

é o resíduo da solução atual para o problema de otimização subjacente. Portanto, a atualização de $\overline{\boldsymbol{p}}(k)$ é dada por:

$$\overline{\boldsymbol{p}}(k) = -\boldsymbol{g}(k) + \beta(k)\overline{\boldsymbol{p}}(k-1), \tag{97}$$

onde $\beta(k)\overline{\boldsymbol{p}}(k-1)$ é introduzido para garantir a conjugação de Markov. Aplicando a Equação (97) na Equação (95) permite, após manipulações algébricas, encontrar a expressão para $\beta(k)$:

$$\beta(k) = \frac{\overline{\boldsymbol{p}}^T(k-1)\boldsymbol{R}_{\hat{s}\hat{s}}(k)\boldsymbol{g}(k)}{\overline{\boldsymbol{p}}^T(k-1)\boldsymbol{R}_{\hat{s}\hat{s}}(k)\overline{\boldsymbol{p}}(k-1)}. \tag{98}$$

Para que a descrição do algoritmo seja completa, é necessário definir uma expressão para o fator de aprendizagem $\alpha(k)$ da Equação (94). Pode ser observado que a solução ótima também é o mínimo da função quadrática

$$\begin{aligned} \mathcal{F}[\boldsymbol{w}(k)] &\triangleq \frac{1}{2}\boldsymbol{w}^T(k)\tilde{\boldsymbol{R}}\boldsymbol{w}(k) - \boldsymbol{w}^T(k)\tilde{\boldsymbol{p}} \\ &\approx \frac{1}{2}\boldsymbol{w}^T(k)\boldsymbol{R}_{\hat{s}\hat{s}}(k)\boldsymbol{w}(k) - \boldsymbol{w}^T(k)\boldsymbol{b}_{\hat{s}}(k), \end{aligned} \tag{99}$$

onde a aproximação foi necessária para efeitos de adaptação do algoritmo CG clássico para um contexto *real time*.

Aplicando a Equação (94) na Equação (99), derivando em relação a $\alpha(k)$, e em

seguida, igualando a zero, obtêm-se:

$$\alpha(k) = -\frac{\overline{\boldsymbol{p}}^T(k)\boldsymbol{g}(k)}{\overline{\boldsymbol{p}}^T(k)\boldsymbol{R}_{\hat{s}\hat{s}}(k)\overline{\boldsymbol{p}}(k)}. \tag{100}$$

O Algoritmo 2 resume os passos da técnica CG adaptativa proposta, a qual permite resolver problemas da cancelamento de ruído usando ferramentas de gradiente conjugado. Nota-se que um fator de regularização $\epsilon \in \mathbb{R}_+$ foi adicionado para garantir a estabilidade numérica do algoritmo.

Algoritmo 2 – Algoritmo adaptativo CG proposto para cancelamento de ruído

**Entrada:** $\gamma$, $\epsilon$, $\hat{\boldsymbol{x}}_{\mathrm{f}}(k)$, $d(k)$
**Saída:** $\boldsymbol{w}(k)$
**Inicialização:** $\boldsymbol{w}(0) = \boldsymbol{0}$, $\boldsymbol{R}_{\hat{s}\hat{s}}(-1) = \boldsymbol{0}$, $\boldsymbol{b}_{\hat{s}}(-1) = \boldsymbol{0}$, $\overline{\boldsymbol{p}}(-1) = \boldsymbol{0}$

**para** $k = 1, 2, \ldots$ **faça**
- $\boldsymbol{R}_{\hat{s}\hat{s}}(k) = \gamma\boldsymbol{R}_{\hat{s}\hat{s}}(k-1) + \hat{\boldsymbol{x}}_{\mathrm{f}}(k)\hat{\boldsymbol{x}}_{\mathrm{f}}^T(k)$ ▷ Atualiza a matriz de autocorrelação
- $\boldsymbol{b}_{\hat{s}}(k) = \gamma\boldsymbol{b}_{\hat{s}}(k-1) + \hat{\boldsymbol{x}}_{\mathrm{f}}(k)d(k)$ ▷ Atualização do vetor de correlação cruzada
- $\boldsymbol{g}(k) = \boldsymbol{R}_{\hat{s}\hat{s}}(k)\boldsymbol{w}(k) - \boldsymbol{b}_{\hat{s}}(k)$ ▷ Calcula o vetor gradiente
- $\beta(k) = \frac{\overline{\boldsymbol{p}}^T(k-1)\boldsymbol{R}_{\hat{s}\hat{s}}(k)\boldsymbol{g}(k)}{\overline{\boldsymbol{p}}^T(k-1)\boldsymbol{R}_{\hat{s}\hat{s}}(k)\overline{\boldsymbol{p}}(k-1)+\epsilon}$ ▷ Conjugação de Markov
- $\overline{\boldsymbol{p}}(k) = -\boldsymbol{g}(k) + \beta(k)\overline{\boldsymbol{p}}(k-1)$ ▷ Atualiza a direção de busca
- $\alpha(k) = -\frac{\overline{\boldsymbol{p}}^T(k)\boldsymbol{g}(k)}{\overline{\boldsymbol{p}}^T(k)\boldsymbol{R}_{\hat{s}\hat{s}}(k)\overline{\boldsymbol{p}}(k)+\epsilon}$ ▷ Calcula o passo de aprendizagem
- $\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \alpha(k)\boldsymbol{p}(k)$ ▷ Atualiza o filtro

### 3.3- Análise Estocástica

Diversas características de aprendizado de um algoritmo adaptativo, tais como taxa de convergência, polarização e estabilidade, podem ser elucidadas por meio de modelos estocásticos. Nesta seção, empreendemos uma análise estocástica do algoritmo proposto.

Em casos práticos, se tem acesso apenas à estimativa do caminho secundário $\hat{\boldsymbol{s}} \triangleq [\hat{s}_0 \ \hat{s}_1 \ \cdots \ \hat{s}_{M-1}]^T$. Denotam-se $\boldsymbol{R}_{ss}(k)$, $\boldsymbol{b}_s(k)$ e $\boldsymbol{x}_{\mathrm{f}}(k)$ como casos especiais de $\boldsymbol{R}_{\hat{s}\hat{s}}(k)$, $\boldsymbol{b}_{\hat{s}}(k)$ e $\hat{\boldsymbol{x}}_{\mathrm{f}}(k)$, respectivamente, supondo-se estimação perfeita, isto é, $\hat{\boldsymbol{s}} = \boldsymbol{s}$. Definindo a

matriz de entrada como:

$$\boldsymbol{X}(k) \triangleq \Big[\boldsymbol{x}(k) \quad \boldsymbol{x}(k-1) \quad \cdots \quad \boldsymbol{x}(k-M+1)\Big], \tag{101}$$

pode-se expressar, de forma compacta, $\boldsymbol{x}_{\mathrm{f}}(k) = \boldsymbol{X}(k)\boldsymbol{s}$ e $\hat{\boldsymbol{x}}_{\mathrm{f}}(k) = \boldsymbol{X}(k)\hat{\boldsymbol{s}}$.

Sem perda de generalidade, o vetor estimado $\hat{\boldsymbol{s}}$ pode ser decomposto como:

$$\hat{\boldsymbol{s}} = \boldsymbol{s} + \boldsymbol{\delta}, \tag{102}$$

onde $\boldsymbol{\delta} \triangleq [\delta_0 \; \delta_1 \; \cdots \; \delta_{M-1}]^T$ modela o erro de estimação do caminho secundário. A Equação (102) permite escrever:

$$\hat{\boldsymbol{x}}_{\mathrm{f}}(k) = \boldsymbol{X}(k)[\boldsymbol{s} + \boldsymbol{\delta}] = \boldsymbol{X}(k)\boldsymbol{s} + \boldsymbol{X}(k)\boldsymbol{\delta} = \boldsymbol{x}_{\mathrm{f}}(k) + \boldsymbol{x}_{\delta}(k). \tag{103}$$

Para se obter um modelo que descreva o comportamento em regime permanente, considera-se a seguinte representação matricial equivalente do algoritmo proposto:

$$\begin{bmatrix} \boldsymbol{w}(k) \\ \boldsymbol{g}(k) \\ \overline{\boldsymbol{p}}(k) \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \alpha(k)\boldsymbol{I} & \alpha(k)\beta(k-1)\boldsymbol{I} \\ -\hat{\boldsymbol{x}}_{\mathrm{f}}(k)\hat{\boldsymbol{x}}_{\mathrm{f}}^T(k) & \gamma\boldsymbol{I} - \alpha(k)\boldsymbol{R}_{\hat{s}\hat{s}}(k) & -\alpha(k)\beta(k-1)\boldsymbol{R}_{\hat{s}\hat{s}}(k) \\ \boldsymbol{0} & \boldsymbol{I} & \beta(k-1)\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}(k-1) \\ \boldsymbol{g}(k-1) \\ \overline{\boldsymbol{p}}(k-1) \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} \\ \hat{\boldsymbol{x}}_{\mathrm{f}}(k)d(k) \\ \boldsymbol{0} \end{bmatrix}. \tag{104}$$

Para tornar as manipulações matemáticas tratáveis, a seguinte hipótese estatística foi adotada:

*Hipótese da independência* (`HI`): As entidades aleatórias $\alpha(k)$, $\beta(k)$, $\overline{\boldsymbol{p}}(k)$, $\boldsymbol{w}(k)$, $\boldsymbol{g}(k)$, $\boldsymbol{b}_{\hat{s}}(k)$ e $\boldsymbol{R}_{\hat{s}\hat{s}}(k)$ são estatisticamente independentes.

*Nota*: A hipótese da independência é amplamente utilizada na área de aproximação estocástica (ver Lara, Tarrataca e Haddad [132] para mais detalhes).

Utilizando a `HI` e a transformada $\mathcal{Z}$ no sistema da Equação (104), obtêm-se as

seguintes recursões:

$$\boldsymbol{W}(z) = \boldsymbol{W}(z)z^{-1} + \overline{\alpha}\boldsymbol{P}(z), \tag{105}$$

$$z\boldsymbol{P}(z) = \overline{\beta}\boldsymbol{P}(z) + \boldsymbol{G}(z), \tag{106}$$

$$\boldsymbol{G}(z) = \frac{\overline{\boldsymbol{b}}_{\hat{s}} z}{z-1} - \boldsymbol{R}_{\hat{s}\hat{s}}\boldsymbol{W}(z), \tag{107}$$

onde $\overline{\alpha} \triangleq \mathbb{E}[\alpha(k)]$, $\overline{\beta} \triangleq \mathbb{E}[\beta(k)]$, $\overline{\boldsymbol{b}}_{\hat{s}} \triangleq \mathbb{E}[\boldsymbol{b}_{\hat{s}}(k)]$, $\overline{\boldsymbol{R}}_{\hat{s}\hat{s}} \triangleq \mathbb{E}[\boldsymbol{R}_{\hat{s}\hat{s}}(k)]$, $\boldsymbol{W}(z) \triangleq \mathcal{Z}\{\mathbb{E}[\boldsymbol{w}(k)]\}$, $\boldsymbol{P}(z) \triangleq \mathcal{Z}\{\mathbb{E}[\boldsymbol{p}(k)]\}$ e $\boldsymbol{G}(z) \triangleq \mathcal{Z}\{E[\boldsymbol{g}(k)]\}$. Após resolver as Equações (105), (106) e (107) para $\boldsymbol{W}(z)$, tem-se a solução:

$$\boldsymbol{W}(z) = [(z-1)(z-\overline{\beta})\boldsymbol{I} + \overline{\alpha}\overline{\boldsymbol{R}}_{\hat{s}\hat{s}} z]^{-1} \frac{\overline{\alpha}\overline{\boldsymbol{b}}_{\hat{s}} z^2}{z-1}. \tag{108}$$

Finalmente, aplicando-se o teorema do valor final, o valor esperado dos coeficientes do filtro em regime permanente $\boldsymbol{w}_\infty$ pode ser expresso por:

$$\begin{aligned} \boldsymbol{w}_\infty &\triangleq \lim_{k\to\infty} \mathbb{E}[\boldsymbol{w}(k)] = \lim_{z\to 1}(z-1)\boldsymbol{W}(z) = \overline{\boldsymbol{R}}_{\hat{s}\hat{s}}^{-1}\overline{\boldsymbol{b}}_{\hat{s}} \\ &= [\overline{\boldsymbol{R}}_{ss} + \overline{\boldsymbol{R}}_{\delta\delta} + 2\overline{\boldsymbol{R}}_{s\delta}]^{-1}[\overline{\boldsymbol{b}}_s + \overline{\boldsymbol{b}}_\delta], \end{aligned} \tag{109}$$

que é um resultado teórico importante, pois leva a concluir que o algoritmo proposto converge para a solução ótima se a estimativa do caminho secundário for perfeita (afinal, neste caso temos que $\overline{\boldsymbol{R}}_{\delta\delta}$, $\overline{\boldsymbol{R}}_{s\delta}$ e $\overline{\boldsymbol{b}}_\delta$ são todos nulos). Na próxima seção, a expressão de forma fechada da Equação (109) é validada por simulações.

### 3.4- Resultados de simulação

Nos resultados a seguir, o algoritmo proposto é comparado com os algoritmos FX-LMS e *Modified Filtered-X LMS* (MFX-LMS) [133]. Os coeficientes de $\boldsymbol{w}^\star$ são dados por:

$$\begin{aligned} W^\star(z) \;=\; & 1 + 0{,}8415z^{-1} + 0{,}4546z^{-2} + 0{,}047z^{-3} - 0{,}1892z^{-4} - 0{,}1918z^{-5} \\ & -0{,}0466z^{-6} + 0{,}0939z^{-7} + 0{,}1237z^{-8} + 0{,}0458z^{-9} - 0{,}0544z^{-10}. \end{aligned}$$

Os coeficientes do caminho secundário real são $\boldsymbol{s} = [1\ 0{,}8]^T$. O ruído aditivo $\nu(k)$ é amostrado de um processo gaussiano branco de média zero. Exceto quando indicado, o sinal de entrada é amostrado de um ruído branco gaussiano de variância um e média zero. Em todas as simulações, os parâmetros adotados para o algoritmo proposto, *Conjugate Gradient Filtered-X* (CG-FX), são $\epsilon = 10^{-6}$ e $\gamma = 0{,}99$. Vale notar que os algoritmos FX-LMS e MFX-LMS requerem uma escolha de parâmetros judiciosa para evitar problemas de estabilidade em diferentes cenários, enquanto o algoritmo proposto, sendo um esquema de passo variável [134], apresenta bom desempenho sem a necessidade de se modificar os parâmetros $\epsilon$ e $\gamma$.

A Figura 13 apresenta o comportamento médio dos coeficientes adaptativos para: (a) caminho secundário perfeitamente conhecido, isto é, $\hat{\boldsymbol{s}} = \boldsymbol{s}$; (b) caminho secundário estimado $\hat{\boldsymbol{s}} = [1{,}00242\ 0{,}80142]^T$. O número de ensaios de Monte Carlo independentes realizados é indicado por $K$. Essa figura sugere que o comportamento assimptótico previsto pela análise teórica da Seção 3.3 concorda com as curvas simuladas. Portanto, o algoritmo implementa uma estimativa sem viés se o caminho secundário for exatamente conhecido.

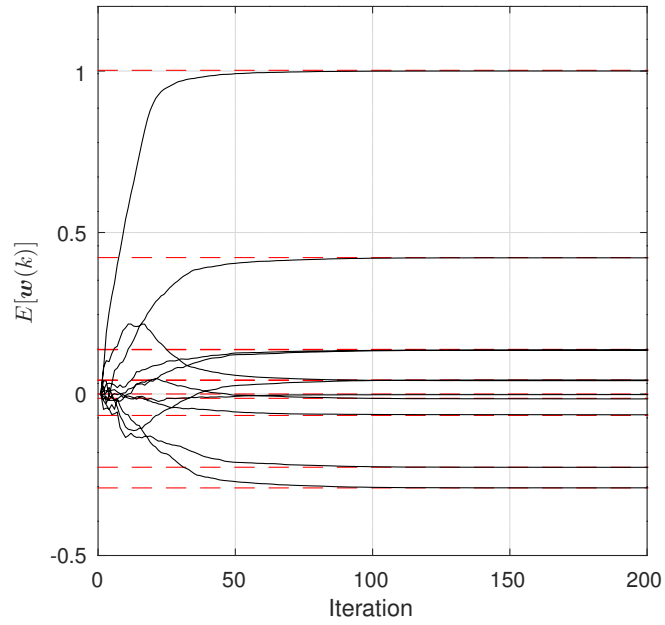
A Figura 14 mostra a evolução empírica do MSE e do *average noise reduction* (ANR) para o FX-LMS tradicional, MFX-LMS e o algoritmo proposto, onde

$$\text{ANR}(k) = 10\log_{10}\left(\frac{\mathbb{E}[e^2(k)]}{\mathbb{E}[d^2(k)]}\right). \tag{110}$$

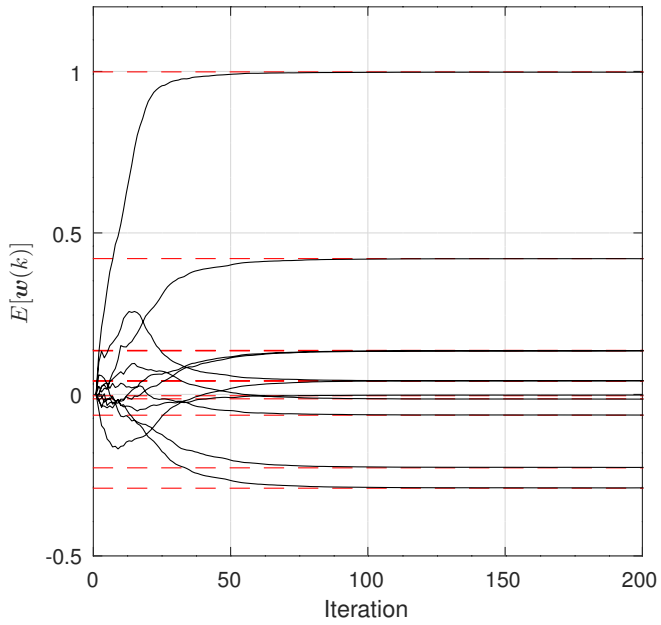
Da Figura 14 pode-se concluir que o algoritmo proposto apresenta uma capacidade de aprendizado adequada já que possui uma taxa de convergência maior e atinge melhor desempenho em regime permanente do que os métodos tradicionais. Essas vantagens não significam que o esquema elaborado contorna todos os compromissos envolvidos em filtragem adaptativa, pois o esquema proposto requer um maior esforço computacional.

Para avaliar o desempenho do algoritmo para entradas não-estacionárias, a Figura 15 mostra a evolução do MSE em um cenário onde a entrada é um sinal de fala[2] corrompido por um ruído branco gaussiano com variância $4 \times 10^{-4}$, com os demais parâmetros iguais à simulação anterior. É notável que a dinâmica de aprendizado do algoritmo proposto apresenta taxa de convergência superior e desempenho em regime permanente melhor.

[2]Obtido do arquivo `apdsqueaky-20151112-xve/b0110.wav` encontrado em `http://www.voxforge.org/home/downloads/speech/english/apdsqueaky-20151112-xve`.

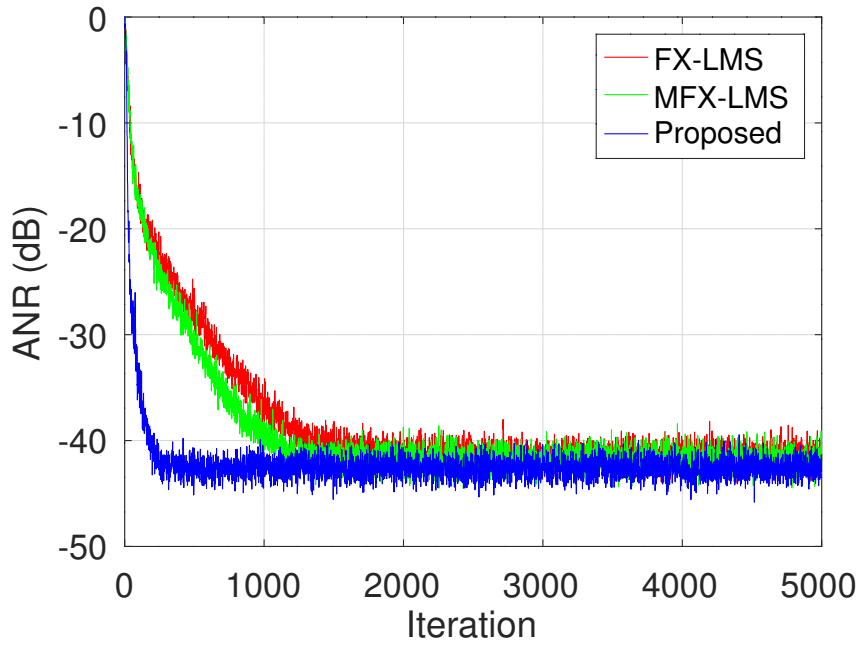



(a)

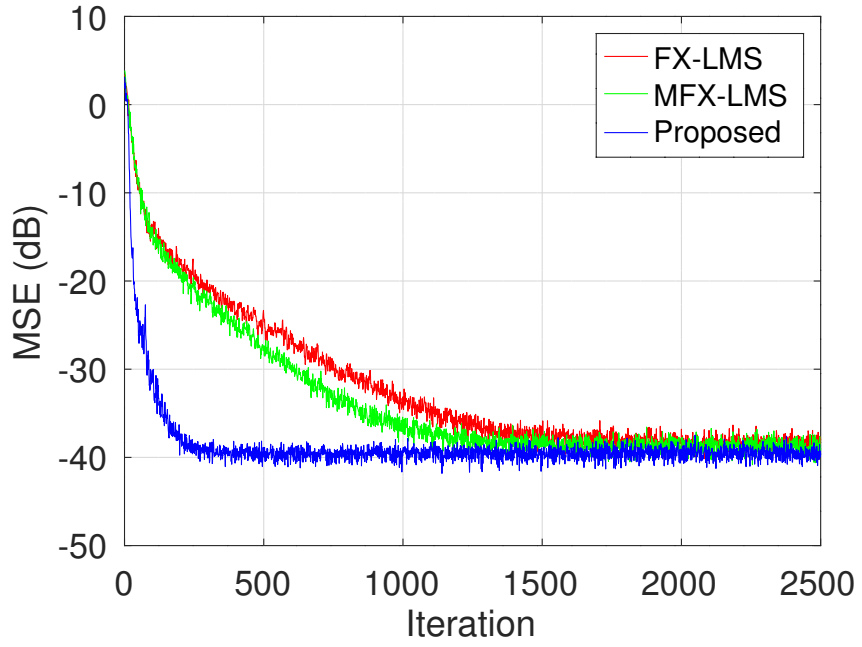



(b)

Figura 13 – Comportamento médio dos coeficientes do filtro adaptativo (em preto), com $K = 500$ e $\sigma_\nu^2 = 10^{-6}$. Linha vermelha tracejada: estimativa teórica utilizando a Equação (109). (a) $\hat{\boldsymbol{s}} = \boldsymbol{s}$. (b) $\hat{\boldsymbol{s}} \neq \boldsymbol{s}$




(a)




(b)

Figura 14 – Evolução de MSE e ANR ao longo das iterações, com entrada estacionária, ruído de medição branco, $K = 100$, $\beta_{\text{FX-LMS}} = \beta_{\text{MFX-LMS}} = 2.5 \times 10^{-2}$ e $\sigma_\nu^2 = 10^{-4}$

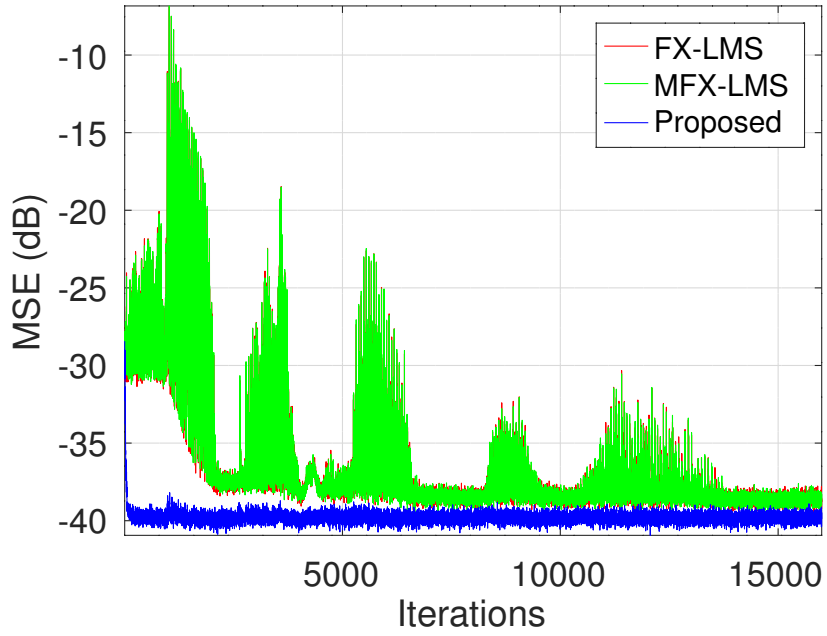



Figura 15 – Evolução do MSE (em dB) ao longo das iterações, com entrada de ruído branco e fala, $K = 100$, $\beta_{\text{FX-LMS}} = \beta_{\text{MFX-LMS}} = 3 \times 10^{-2}$ e $\sigma_\nu^2 = 10^{-4}$. As curvas do FX-LMS e MFX-LMS são indistinguíveis

### 3.5- Considerações Parciais

Neste capítulo, a técnica de otimização conhecida como gradiente conjugado foi pela primeira vez estendida para as técnicas de ANC. Ademais, um modelo estocástico referente ao desempenho do algoritmo foi desenvolvido, o qual demonstra a ausência de polarização assintótica quando o secundário é estimado corretamente. Por fim, o algoritmo proposto se mostrou vantajoso em alguns cenários, notadamente, quando sinais não estacionários (como sinais de voz) foram usados como entrada do cancelador.

# 4- Algoritmo SM-Sign-NLMS

## 4.1- Introdução

Algoritmos de filtragem adaptativa no estado da arte devem conjugar propriedades desejáveis, tais como alta taxa de convergência, capacidade de rastreamento, pouca necessidade de ajustar hiperparâmetros e baixo custo computacional. Combinar tais propriedades de maneira a dar origem a um algoritmo vantajoso em determinadas aplicações requer grande engenhosidade por parte do pesquisador. Neste capítulo, propomos um algoritmo inédito, intitulado SM-sign-NLMS, o qual apresenta um baixo custo computacional e robustez a ruído impulsivo, sem requerer o ajuste de um número excessivo de hiperparâmetros. Cumpre notar que interferências impulsivas podem ser encontradas em uma ampla faixa de aplicações, tais como predição de sinais, cancelamento de eco e rastreamento de posição, tendo um impacto bastante negativo nos algoritmos de adaptação que se valem de funções custo quadráticas [135].

A possibilidade de algoritmos adaptativos utilizarem a seleção de dados (oferecida pelas estratégias do tipo *Set-membership*) tem atraído bastante interesse, já que reduz a carga computacional dos processos de aprendizagem em geral [136]. Ademais, algoritmos do *Set-membership* evitam um aumento excessivo do custo computacional quando ocorre redundância de dados, o que é muito comum em cenários de *big data* [137]. Por essas razões, o algoritmo proposto (SM-sign-NLMS) mobiliza a estratégia *Set-membership*, combinando-a de maneira inédita com um uma não linearidade do tipo sign no sinal de erro, tornando o algoritmo resultante robusto a ruído impulsivo. O novo algoritmo é derivado a partir da solução de um problema inédito de otimização com restrições. Este capítulo deu origem ao artigo [138], o qual também contém contribuições dos próximos capítulos desta tese.

Este capítulo encontra-se estruturado do seguinte modo: a Seção 4.2 descreve a aplicação da filosofia *Set-membership* em filtragem adaptativa, enquanto que a Seção 4.3 apresenta o algoritmo SM-sign-NLMS proposto. Por fim, a Seção 4.4 compara, por meio de resultados de simulação, o algoritmo proposto com outros algoritmos constantes

na literatura, enquanto que a Seção 4.5 sintetiza o capítulo e estabelece sua conexão estreita com os capítulos seguintes. Cabe ressaltar que não existe uma seção específica neste capítulo para tratar dos algoritmos adaptativos do tipo *sign* (estratégia que também encontra-se incorporada no algoritmo proposto), porque tais algoritmos (também chamados de LAD, do inglês *least absolute difference*) já foram devidamente contemplados na Seção 2.1.

## 4.2- Algoritmos Adaptativos do tipo Set-Membership

A escolha de um algoritmo adaptativo (o qual requer maior ou menor complexidade computacional) a ser utilizado em determinado cenário depende das características de ruído que este cenário apresenta [139]. Recentemente, estratégias de censura adaptativa a dados têm se revelado promissoras para lidar com dados que devem ser processados. Tal censura, grosso modo, consiste em utilizar um "intervalo de censura" para descartar dados menos informativos de acordo com alguma regra estatística. Vale dizer: trata-se de selecionar adaptativamente, *e em tempo real*, apenas os dados que contenham informação relevante [140]. A família de técnicas do tipo *Set-membership* apresenta esta propriedade desejável de censura adaptativa, pois mobiliza a chamada *set estimation theory*, a qual viabiliza a incorporação de informação *a priori* acerca das incertezas e/ou do ruído [141]. Por vezes o processo de seleção de dados é obtido por meio de alguma conexão entre dois limiares, usada para classificar os dados como não inovadores/aberrações ou inovadores, em conjunção com uma probabilidade de atualização pré-estabelecida, a qual reduz a complexidade computacional do algoritmo adaptativo resultante [142].

Num contexto de identificação de sistemas, abordagens do tipo *Set-membership Filtering* (SMF) exploram a hipótese de um processo de ruído limitado $\nu(k)$ imerso num modelo linear nos parâmetros [23]:

$$d(k) = [\boldsymbol{w}^{\star}]^{T}\, \boldsymbol{x}(k) + \nu(k), \tag{111}$$

onde $d(k)$, como visto nos capítulos pregressos, é o sinal de referência, $\boldsymbol{w}^{\star} \in \mathbb{R}^{N}$ contém

o conjunto de coeficientes ideal (e desconhecido) que o filtro adaptativo pretende estimar e $x(k)$ abrange $N$ amostras consecutivas do sinal de entrada $x(k)$ na $k$-ésima iteração:

$$\boldsymbol{x}(k) \triangleq \begin{bmatrix} x(k) & x(k-1) & \dots & x(k-N+1) \end{bmatrix}^T, \tag{112}$$

onde assume-se uma estrutura transversal.

Em cada iteração, o filtro adaptativo gera uma amostra de saída $y(k) \in \mathbb{R}$, computada por meio de um produto interno:

$$y(k) = \boldsymbol{w}^T(k)\boldsymbol{x}(k), \tag{113}$$

onde $\boldsymbol{w}(k) \in \mathbb{R}^N$ contém os pesos adaptativos na $k$-ésima iteração:

$$\boldsymbol{w}(k) \triangleq \begin{bmatrix} w_0(k) & w_1(k) & \dots & w_{N-1}(k) \end{bmatrix}^T, \tag{114}$$

os quais devem ser adaptados de modo iterativo por uma regra específica de adaptação. A diferença entre o sinal de referência e a saída do filtro é o sinal de erro $e(k) \in \mathbb{R}$:

$$e(k) \triangleq d(k) - \boldsymbol{w}^T(k)\boldsymbol{x}(k). \tag{115}$$

Em termos gerais, a tarefa de identificação de sistemas é apresentada na Figura 16.

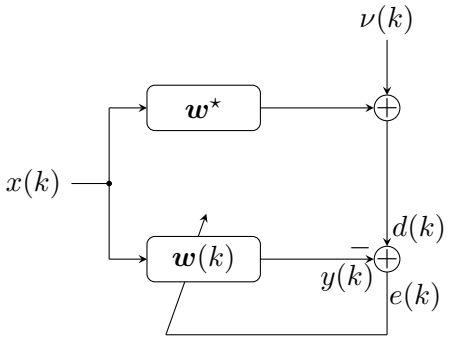



Figura 16 – Diagrama de blocos da estrutura de um algoritmo de filtragem adaptativa aplicado a uma tarefa de identificação de sistemas.

A técnica SMF pretende garantir um limiar pré-estabelecido na magnitude do erro dentro do horizonte temporal de relevo. Neste sentido, qualquer atualização do vetor adaptativo que provê um valor absoluto do erro menor do que o limiar ajustável

$\overline{\gamma}$ é considerada uma solução factível para o procedimento iterativo de aprendizagem. Adicionalmente, tal metodologia limita o erro no cenário *worst case* obtido pelo filtro, com um limitante de erro *a priori* [143], [144]. Tipicamente, o limiar $\overline{\gamma}$ depende da variância do ruído adaptativo $\sigma_\nu^2$ por meio da seguinte relação [145]

$$\overline{\gamma} = \sqrt{\tau \sigma_\nu^2}, \tag{116}$$

onde $\tau \in \mathbb{R}_+$ é um parâmetro ajustável. A efetividade das metodologias de SMF é modulada pela especificação de $\overline{\gamma}$ [146], uma tarefa que pode ser desafiadora em cenários reais devido à falta de conhecimento com relação ao ambiente e suas idiossincrassias dinâmicas [147]. A despeito disto, deve ser notado que a literatura em geral mostra que algoritmos do tipo SM são superiores às variantes clássicas, que não se valem da estratégia SM [148].

Considere $\mathcal{S}$ como o conjunto contendo os pares $\{\boldsymbol{x}(k), d(k)\}$ disponíveis para a estrutura adaptativa. O conjunto de restrição $\mathcal{H}_k$ é o conjunto de vetores de pesos adaptativos $\boldsymbol{w}$ que são consistentes com

$$\mathcal{H}_k \triangleq \{\boldsymbol{w} \in \mathbb{R}^N : |d(k) - \boldsymbol{w}^T(k)\boldsymbol{x}(k)| \leq \overline{\gamma}\}, \tag{117}$$

a qual define uma região delimitada por hiperplanos paralelos e sugere a utilização de mais conjuntos de restrição no mecanismo de atualização [23], [149]. O algoritmo SM-NLMS padrão resolve o seguinte problema de otimização [150]:

$$\begin{array}{c} \min_{\boldsymbol{w}(k+1)} \|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^2 \\ \text{sujeito a } \boldsymbol{w}(k+1) \in \mathcal{H}_k \end{array}, \tag{118}$$

onde o erro de saída é limitado por um limite pré-especificado (veja a Equação (117)). A solução não relaxada de (118) é

$$\boldsymbol{w}(k+1) = \begin{cases} \boldsymbol{w}(k) + \frac{\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2}\left[e(k) - \overline{\gamma}\right], & \text{se } |e(k)| > \overline{\gamma} \\ \boldsymbol{w}(k), & \text{caso contrário} \end{cases}, \tag{119}$$

a qual demanda menos atualizações para alcançar o regime permanente [150]. Utilizando a estratégia SMF, se torna possível reduzir a complexidade computacional em filtragem adaptativa, dado que atualizações dos coeficientes do filtro ocorrem apenas quando o

erro estimado supera o limiar pré-estabelecido [151].

O termo $\|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^2$ minimizado na Equação (118) deriva do princípio da perturbação mínima, um princípio conservador que assume que as atualizações prévias contém mais informação do que a atualmente oferecida pelo vetor de entrada atual, e por conseguinte a solução atual $\boldsymbol{w}(k)$ deve ser pouco perturbada num processo de aprendizagem bem projetado.

O algoritmo padrão sign-LMS (LAD) é tipicamente derivado de um procedimento de gradiente estocástico baseado para uma otimização de norma $\ell_1$ [152]:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \beta \nabla_{\mathbf{w}(k)} |e(k)| \tag{120}$$

$$\Rightarrow \boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \boldsymbol{x}(k) \text{sign}[e(k)], \tag{121}$$

onde o termo de atualização $\beta \boldsymbol{x}(k)\text{sign}[e(k)]$ é uma função "saturada" do ruído e os pesos adaptativos que substitui o sinal de erro por sua polaridade [153], [154]. O limitador do tipo *hard* na Equação (121) complicam a modelagem estocásticas do algoritmo resultante, frequentemente demandando uma substancial engenhosidade analítica para que uma análise teórica possa ser conduzida satisfatoriamente [155]. O fato de técnicas diferentes de otimização serem mobilizadas para a obtenção dos algoritmos SM e sign-LMS torna um desafio a proposição de um problema de otimização cuja solução dê origem a um algoritmo que conjugue a técnica SM com uma não linearidade do tipo sign no sinal de erro. O algoritmo proposto, discutido a seguir, cumpre este objetivo.

### 4.3- Algoritmo Proposto SM-Sign-NLMS

Como vimos, não é trivial projetar um filtro linear nos parâmetros cujo espaço de soluções factíveis para a atualização combine as habilidades de censura de dados da estratégia SM e a robustez contra ruído impulsivo oferecida pelas técnicas do tipo *sign*. Isto se deve ao fato de que enquanto a abordagem SM adota o critério de distorção mínima, algoritmos da modalidade *sign* usualmente adotam o método de otimização via gradiente estocástico em sua obtenção. Levando-se em consideração a estrutura e as restrições do problema e usando as conexões conceituais entre o princípio da distorção

mínima e o gradiente estocástico oferecidas por [5], [29], [31], [156], este problema é contornado no algoritmo proposto neste capítulo por meio do seguinte problema de otimização *inédito*:

$$\begin{aligned} &\min_{\boldsymbol{w}(k+1)} \tfrac{1}{2}\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|^2 \\ &\text{sujeito a } e_{\mathsf{p}}(k) = \begin{cases} e(k), & \text{se } |e(k)| \leq \overline{\gamma} \\ \left[1-\frac{\beta}{|e(k)|}\right] e(k), & \text{se } |e(k)| > \overline{\gamma} \end{cases}, \end{aligned} \tag{122}$$

onde $e_{\mathsf{p}}(k)$ denota o erro *a posteriori*, obtido após a atualização com o dado *atual* $\{d(k), \boldsymbol{x}(k)\}$:

$$e_{\mathsf{p}}(k) \triangleq d(k) - \boldsymbol{w}^T(k+1)\boldsymbol{x}(k). \tag{123}$$

Observe que a segunda restrição de (122) (isto é, $e_{\mathsf{p}}(k) = \left[1-\frac{\beta}{|e(k)|}\right] e(k)$) impõe *retornos decrescentes* no impacto que o erro $e(k)$ apresenta na magnitude da atualização. Tal característica é responsável pela robustez (desejável) contra ruído impulsivo, apresentada pelo algoritmo. Afinal, o aumento do módulo do erro diminui o termo $\frac{\beta}{|e(k)|}$, tornando o fator $\left[1-\frac{\beta}{|e(k)|}\right]$ próximo da unidade, o que implica que o erro *a posteriori* deve ser praticamente igual ao erro *a priori*, igualdade que é obtida modificando muito pouco a solução pregressa $\boldsymbol{w}(k)$. A solução de (122) pode ser apresentada como o resultado do Teorema 2:

**Teorema 2**. A solução de (122) define a seguinte equação de atualização *inédita*:

$$\boldsymbol{w}(k+1) = \begin{cases} \boldsymbol{w}(k), & \text{para } |e(k)| \leq \overline{\gamma} \\ \boldsymbol{w}(k) + \beta\frac{\mathsf{sign}[e(k)]}{\|\boldsymbol{x}(k)\|^2}\boldsymbol{x}(k), & \text{para } |e(k)| > \overline{\gamma} \end{cases}, \tag{124}$$

atualização que combinada as vantagens das estratégias *Set-membership* e *sign*.

*Demonstração*: Resolvendo (122) pela técnica dos multiplicadores de Lagrange, a solução de (122) pode ser convertida no seguinte problema equivalente e sem restrições:

$$\begin{aligned} \boldsymbol{F}\left[\boldsymbol{w}(k+1)\right] = &\frac{1}{2}\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|^2 \\ &+ \lambda\left\{e_{\mathsf{p}}(k) - \left[1-\frac{\beta}{|e(k)|}\right] e(k)\right\}. \end{aligned} \tag{125}$$

Zerando-se o gradiente de (125) nos leva a:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \lambda\boldsymbol{x}(k). \tag{126}$$

Se $|e(k)| \leq \overline{\gamma}$, a solução de (122), é trivial, pois arbitrar

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) \tag{127}$$

atende à restrição $e_{\mathsf{p}}(k) = e(k)$ e minimiza o termo $\frac{1}{2}\|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^2$. Esta possibilidade impõe a condição $\lambda = 0$ em (126) e explica a redução vantajosa da carga computacional, pois a atualização do peso adaptativo é evitada quando os dados atuais são redundantes. Adicionalmente, esta avaliação implícita do conteúdo de informação envolvido nos dados de entrada é também responsável pela habilidade de censura de dados do algoritmo resultante.

No caso $|e(k)| > \overline{\gamma}$, pode-se aplicar (126) na segunda restrição de (122), levando a:

$$\lambda = \beta \mathsf{sign}[e(k)]. \tag{128}$$

A combinação das Equações (126), (127) e (128) leva à Equação (124). □

*Comentário*: O paradigma *Set-Membership* assume que as medidas adquiridas pelos sensores são afetadas por um sinal de ruído limitado [157]. Esta característica prejudica o desempenho do algoritmo *Set-membership* quando o ruído no sinal desejado $d(k)$ apresenta características impulsivas. A adoção da função sign, inerente ao algoritmo proposto, foi judiciosamente mobilizada para contornar esta restrição, sem demandar o ajuste de parâmetros adicionais. Cabe ressaltar também que a normalização implicada pela divisão pelo fator $\|\boldsymbol{x}(k)\|^2$ (vide Equação (124)) reduz a amplificação do ruído do gradiente e facilita a escolha do passo de aprendizagem, em comparação com o algoritmo LMS, que não apresenta esta normalização [158].

Para os propósitos de descrição analítica do comportamento de aprendizado do algoritmo proposto (o que será nossa preocupação ao longo dos Capítulos 5, 6 e 8), uma descrição mais adequada da sua equação de atualização é

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \frac{\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2} f[e(k)], \tag{129}$$

onde

$$f[e(k)] = \begin{cases} 0, & \text{para } |e(k)| \leq \overline{\gamma}, \\ \mathsf{sign}[e(k)] & \text{para } |e(k)| > \overline{\gamma} \end{cases}. \tag{130}$$

O algoritmo proposto, assim como outros algoritmos que utilizam a técnica *Set-*

*membership*, utiliza uma projeção métrica em um conjunto convexo fechado (projeção convexa). Desafortunadamente, a não linearidade da projeção convexa coloca um desafio à análise teórica. Nos capítulos seguintes, diferentes modelos estocásticos serão empregados com o objetivo de extrair percepções relevantes sobre o desempenho do algoritmo proposto. Neste capítulo, além de descrever como obter a equação de atualização do algoritmo proposto, iremos focar em estudos comparativos do algoritmo proposto com alternativas pretéritas, publicadas na literatura.

### 4.3.1 Complexidade computacional

A partir da perspectiva da complexidade computacional, o número de somas e/ou subtrações escalares do algoritmo proposto é de $2N + 2$ (por iteração). Já o número respectivo de multiplicações escalares (também por iteração) é de $2N + 1$, quando uma atualização é efetuada. Apenas uma divisão escalar por atualização é requerida. Note que tal complexidade é muito similar à do algoritmo LMS, o qual requer $2N$ somas/subtrações escalares por iteração e $2N + 1$ multiplicações escalares por iteração. De fato, na prática o custo computacional do algoritmo proposto é ainda menor do que o apresentado, já que ele não atualiza o vetor de pesos em todas as iterações.

## 4.4- Simulações

A proposição do algoritmo SM-sign-NLMS foi motivada por aumento de robustez em cenários de ruído impulsivo. Para avaliar o impacto da introdução da não linearidade na equação de atualização do algoritmo SM-NLMS, iremos nesta seção comparar o desempenho do algoritmo proposto com o algoritmo tradicional SM-NLMS, sob condições de ruído impulsivo. Assim, assume-se que a amostra de ruído $\nu(k)$ é obtida por meio do seguinte modelo:

$$\nu(k) = [1 - \omega(k)]\tilde{\nu}(k) + \omega(k)\hat{\nu}(k), \tag{131}$$

onde $\tilde{\nu}(k)$ é uma variável aleatória gaussiana de variância $\sigma_\nu^2$ pequena (simulando ruído de quantização ou de medição) e $\hat{\nu}(k)$ é uma variável aleatória de alta variância $\sigma_{\text{IMP}}^2$ (relativa a um ruído impulsivo). A variável aleatória $\omega(k)$, estatisticamente independente de $\tilde{\nu}(k)$ e de $\hat{\nu}(k)$, indica com probabilidade $P_{\text{imp}}$ a presença de ruído impulsivo:

$$\omega(k) = \begin{cases} 1, & \text{com probabilidade } P_{\text{imp}} \\ 0, & \text{com probabilidade } 1 - P_{\text{imp}} \end{cases} . \tag{132}$$

Dado o compromisso entre taxa de convergência e desempenho assintótico, uma comparação justa entre dois algoritmos distintos deve equiparar, para uma mesma métrica de avaliação (como MSE ou MSD), uma destas características (taxa de convergência ou desempenho assintótico), e comparar se um algoritmo supera o outro. Como a característica mais comumente equalizada na literatura é o desempenho assintótico, iremos nesta seção ajustar o passo de aprendizagem de ambos os algoritmos para que ambos apresentem o mesmo desempenho assintótico, e iremos avaliar qual deles apresenta melhor taxa de convergência.

Para efeitos de comparação, iremos testar três diferentes cenários, com suas principais características sendo apresentadas na Tabela 9. O número de ensaios de Monte Carlo para levantamento das curvas empíricas de aprendizado em todos os casos foi de $10^5$ ensaios independentes. O sinal de entrada em todos os casos é extraído de um processo estocástico gaussiano, com amostras independentes e identicamente distribuídas, de média zero e variância unitária.

| #Cenário | #Modelo | $N$ | $\tau$ | $\sigma_\nu^2$ | $\sigma_{\text{IMP}}^2$ | $P_{\text{imp}}$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 64 | $\{2, 5\}$ | $10^{-6}$ | $0,1$ | $0,1$ |
| 2 | 1 | 64 | $\{2, 5\}$ | $10^{-6}$ | $1$ | $0,1$ |
| 3 | 6 | 120 | $5$ | $10^{-6}$ | $0,1$ | $0,1$ |

Tabela 9 – Características principais dos cenários analisados. O #Modelo reflete o número da função de transferência estimada por medições, constante no relatório da ITU [122].

### 4.4.1 Cenário 1

As Figuras 17 e 18 apresentam a evolução do MSD (em dB) para o Cenário 1, para as escolhas $\tau = 2$ e $\tau = 5$, respectivamente. Observa-se em ambos os casos a superioridade do algoritmo proposto (SM-sign-NLMS), já que este apresenta uma taxa de convergência muito maior, com um desempenho assintótico equivalente. Observe que o algoritmo SM-NLMS precisa de 2800 iterações (aproximadamente) para obter um MSD de -20 dB, enquanto que o algoritmo proposto demanda aproximadamente um terço disso (em torno de 1000 iterações). Contudo, é importante avaliar o comportamento dos algoritmos sob condições menos favoráveis. Cenários com menor SNR, ruído não-gaussiano ou menor correlação entre os sinais podem impactar significativamente a taxa de convergência e a eficácia das atualizações estocásticas. As probabilidades de atualização podem ser visualizadas nas Figuras 19 e 20, para $\tau = 2$ e $\tau = 5$, respectivamente. Nota-se que os algoritmos convergem para uma probabilidade de atualização relativamente alta (superior a 93%), o que implica uma pouca redução de complexidade computacional na configuração adotada. Ainda assim, o algoritmo proposto reduz muito mais rapidamente a sua probabilidade de atualização, demonstrando uma habilidade superior na redução de custo computacional.

### 4.4.2 Cenário 2

A evolução do MSD no Cenário 2 pode ser vista na Figura 21 (para $\tau = 2$) e na Figura 22 (para $\tau = 5$). Novamente, o algoritmo proposto apresenta uma taxa de convergência bastante superior, alcançando o regime permanente com um número de iterações sendo aproximadamente um terço do número de iterações requeridas pelo algoritmo SM-NLMS. Já nas probabilidades de atualização, verificamos uma queda mais rápida no algoritmo proposto. Porém, observa-se em regime permanente que, embora sejam muito próximas, a probabilidade de atualização do algoritmo SM-NLMS é ligeiramente menor do que a do algoritmo proposto. De todo modo, neste cenário, a probabilidade de atualização é muito alta, podendo a redução de complexidade computacional engendrada

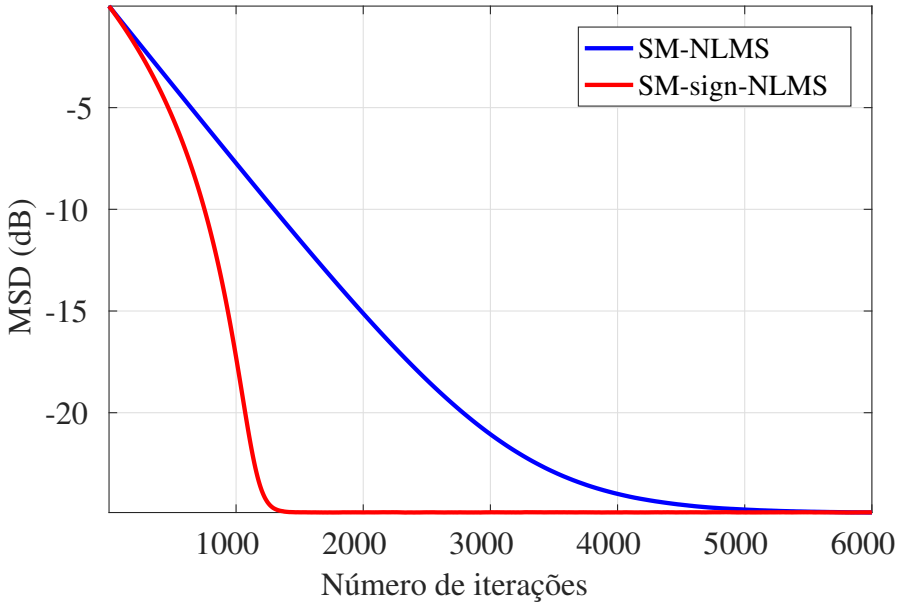



Figura 17 – Evolução do MSD (em dB) dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 1, para $\tau = 2$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,059$, $\beta_{\text{SM-sign}} = 0,0828$.

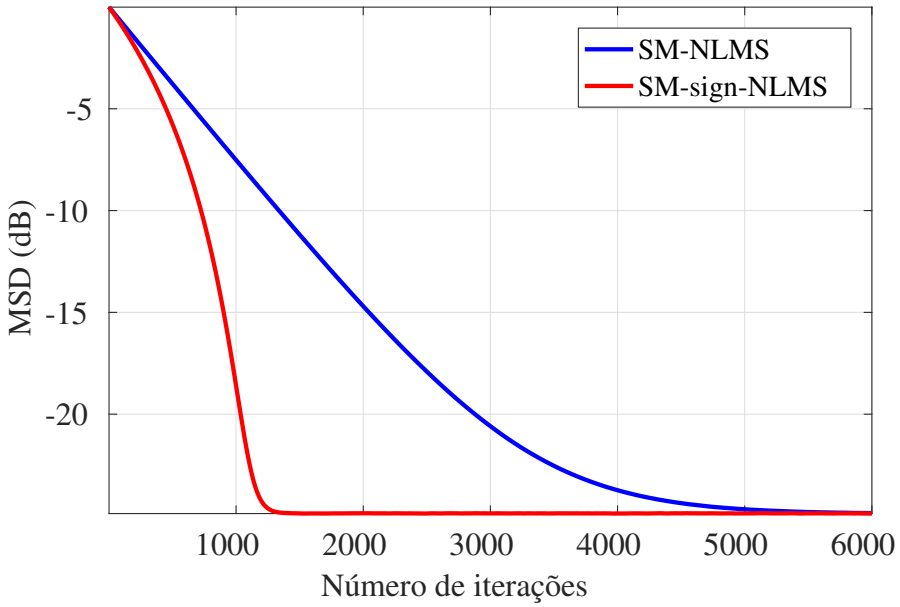



Figura 18 – Evolução do MSD (em dB) dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 1, para $\tau = 5$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,0575$, $\beta_{\text{SM-sign}} = 0,0862$.

pelos algoritmos Set-membership ser considerada desprezível.

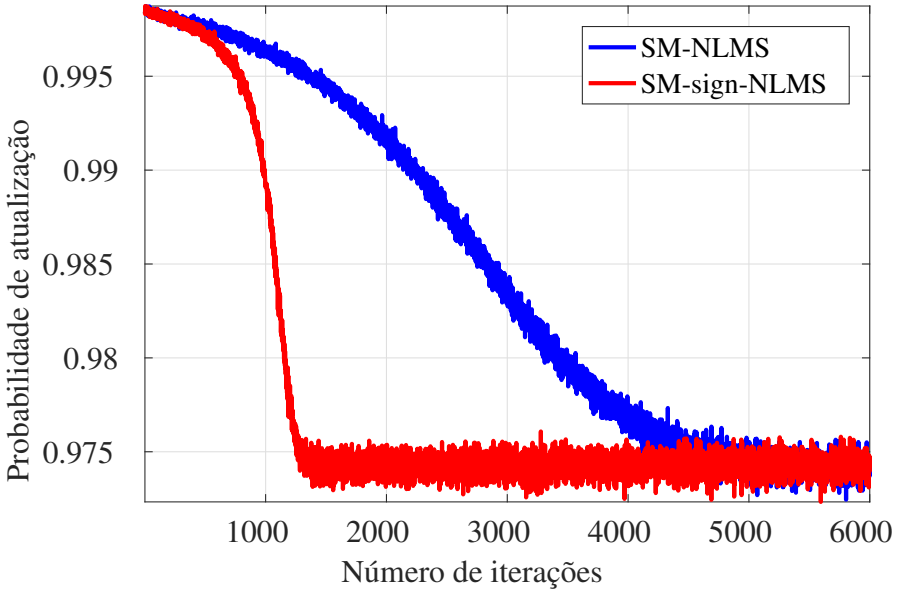



Figura 19 – Evolução da probabilidade de atualização dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 1, para $\tau = 2$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,059$, $\beta_{\text{SM-sign}} = 0,0828$.

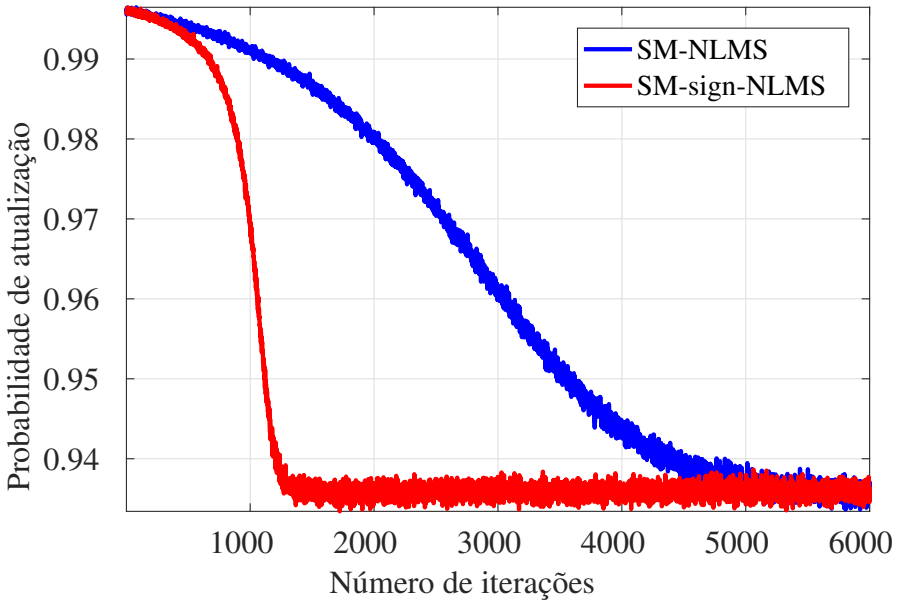



Figura 20 – Evolução da probabilidade de atualização dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 1, para $\tau = 5$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,0575$, $\beta_{\text{SM-sign}} = 0,0862$.

### 4.4.3 Cenário 3

No Cenário 3, o suporte da resposta ao impulso da planta a identificar tem mais coeficientes: 120. Cabe ressaltar que, neste cenário, almejamos equalizar o desempenho

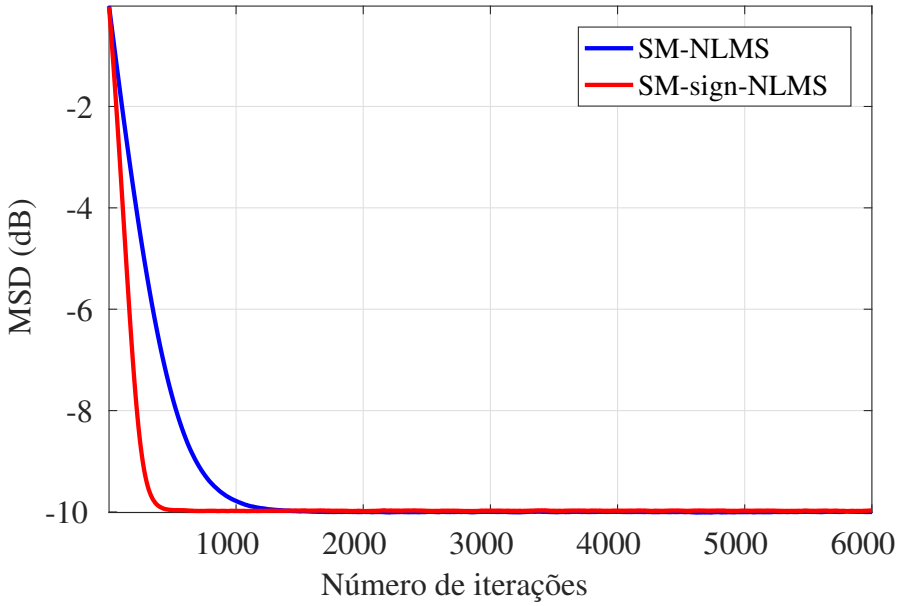



Figura 21 – Evolução do MSD (em dB) dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 2, para $\tau = 2$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,1761$, $\beta_{\text{SM-sign}} = 0,4542$.

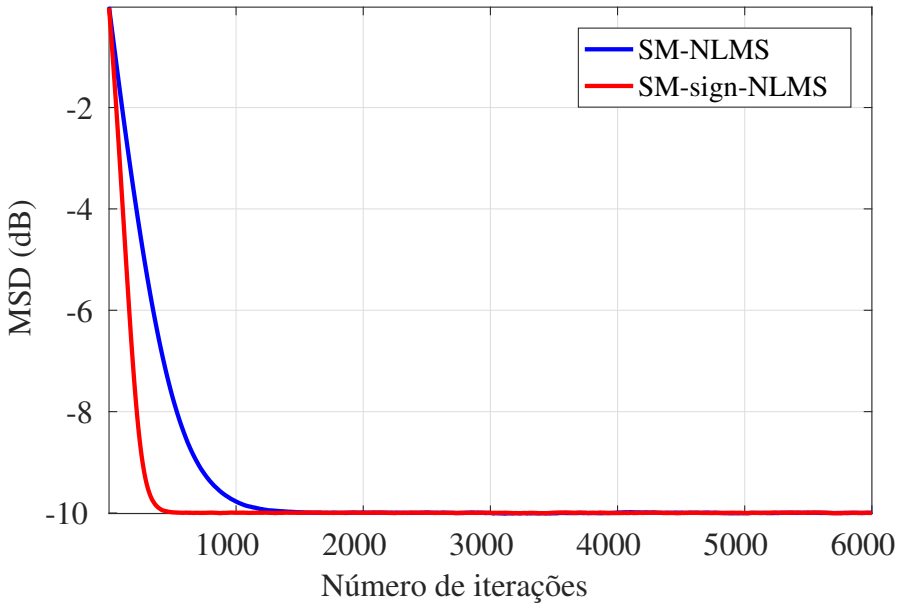



Figura 22 – Evolução do MSD (em dB) dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 2, para $\tau = 5$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,1753$, $\beta_{\text{SM-sign}} = 0,4563$.

assintótico do MSE, enquanto que nos cenários anteriores o objetivo era equalizar o MSD. A Figura 25 apresenta a evolução do MSE, equalizado para um desempenho em regime permanente (-20 dB). O algoritmo proposto apresenta um desempenho superior,

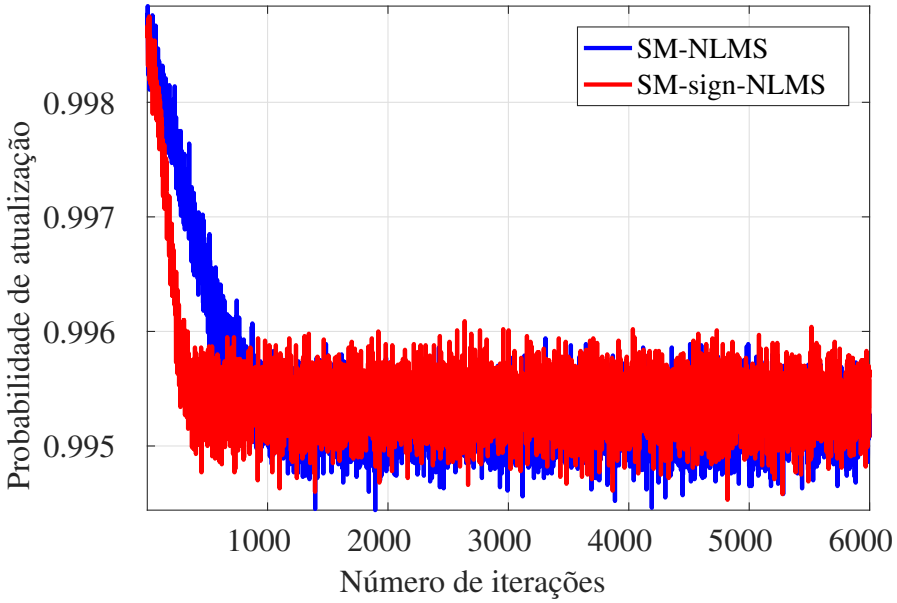



Figura 23 – Evolução da probabilidade de atualização dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 2, para $\tau = 2$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,1761$, $\beta_{\text{SM-sign}} = 0,4542$.

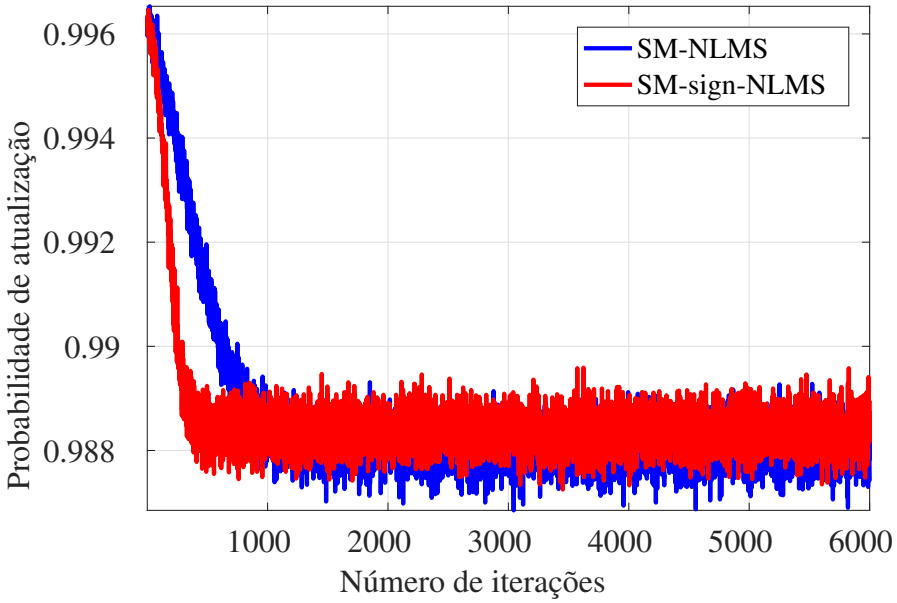



Figura 24 – Evolução da probabilidade de atualização dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 2, para $\tau = 5$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,1753$, $\beta_{\text{SM-sign}} = 0,4563$.

em termos de taxa de convergência, ao do algoritmo tradicional SM-sign-NLMS. Para efeitos de comparação com os cenários anteriores, a Figura 26 apresenta a evolução do MSD, a qual ilustra uma vantagem muito clara do algoritmo proposto, já que ele supera o

algoritmo SM-NLMS tanto em taxa de convergência quanto em desempenho assintótico. Por fim, a Figura 27 detalha a evolução da probabilidade de atualização dos algoritmos. Neste cenário, a vantagem do algoritmo proposto é muito grande: enquanto o algoritmo SM-NLMS apresenta uma probabilidade de atualização assintótica na faixa de 75%, o algoritmo proposto apresenta uma métrica muito melhor: em torno de 35%. Neste cenário, o algoritmo proposto claramente supera o clássico em *todas as três dimensões*: taxa de convergência, desempenho assintótico e probabilidade de atualização. Este tipo de vantagem multidimensional é muito importante na prática, sendo sua obtenção por algum algoritmo bastante desafiadora. Esta característica é particularmente interessante porque ocorre quando a planta desconhecida exige mais coeficientes adaptativos para ser identificada (o que se dá no Cenário 3), justamente a situação mais interessante e difícil, na prática.

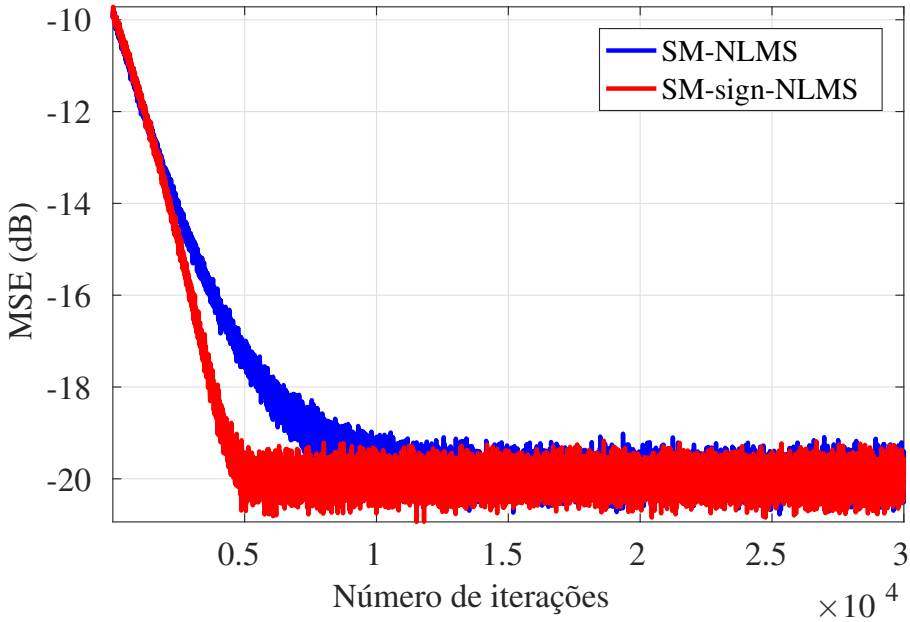



Figura 25 – Evolução do MSE (em dB) dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 3, para $\tau = 5$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,03$, $\beta_{\text{SM-sign}} = 0,01$.

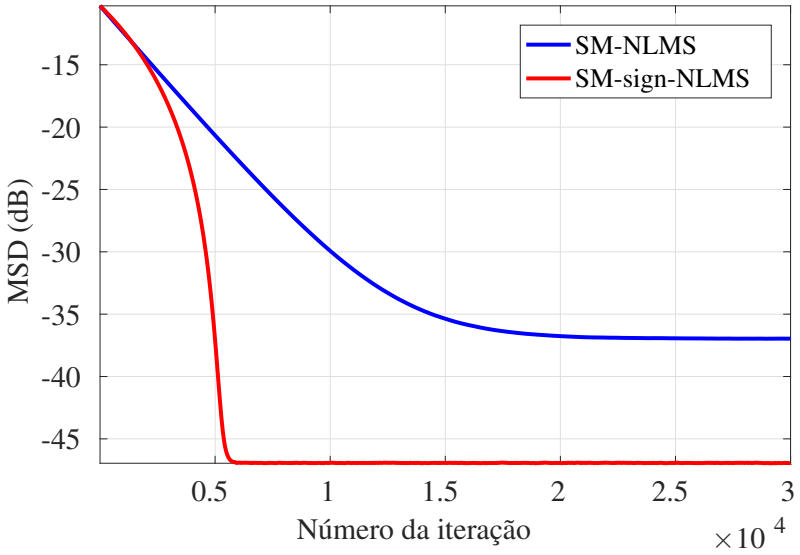



Figura 26 – Evolução do MSD (em dB) dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 3, para $\tau = 5$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,03$, $\beta_{\text{SM-sign}} = 0,01$.

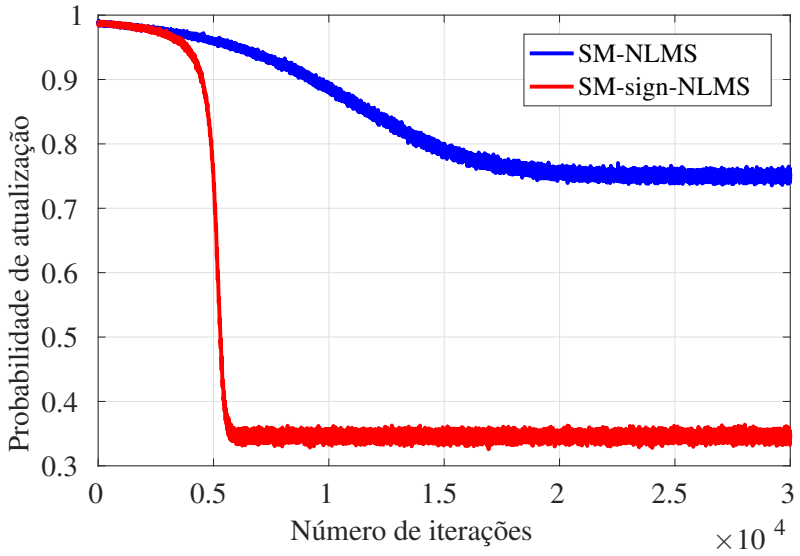



Figura 27 – Evolução da probabilidade de atualização dos algoritmos SM-NLMS e SM-sign-NLMS no Cenário 3, para $\tau = 5$. Curva azul: SM-NLMS, curva vermelha: SM-sign-NLMS. Os valores dos passos de aprendizagem, ajustados para equalização do desempenho em regime permanente, foram: $\beta_{\text{SM}} = 0,03$, $\beta_{\text{SM-sign}} = 0,01$.

### 4.5- Comentários Finais

Este capítulo apresenta uma contribuição desta tese: o algoritmo SM-sign-NLMS, o qual deu origem à publicação [138]. Por meio de simulações, ficou comprovada a

superioridade do algoritmo proposto frente ao algoritmo SM-NLMS, seja em termos de complexidade computacional, seja em termos de robustez a ruído impulsivo.

Entretanto, para que um novo algoritmo adaptativo possa ser adotado na indústria, é forçoso que haja garantias teóricas acerca de seu desempenho. Para isso, é comumente necessário o concurso de *modelos estocásticos*. Sendo inédito, o algoritmo SM-sign-NLMS necessita destes modelos, o que não é uma tarefa trivial, já que sua equação de atualização apresenta não linearidades sofisticadas, tais como a função sign$[\cdot]$ e a possibilidade de não atualização. Nos próximos capítulos, análises teóricas inéditas serão propostas para a modelagem deste algoritmo. Entre estas, podemos elencar a análise de regime permanente (objeto de estudo do Capítulo 5), a análise de transiente (objeto de atenção do Capítulo 6), bem como as análises de rastreamento e comprimento deficiente presentes no Capítulo 8. Importa aduzir que a versão com passos de aprendizagem variantes no tempo exposta no Capítulo 7 é consectária da análise de transiente delineada no Capítulo 6.

# 5- Análise de Regime Permanente do Algoritmo SM-Sign-NLMS

O algoritmo proposto SM-sign-NLMS apresenta duas não linearidades em sua equação de atualização: a normalização de dados (devido ao denominador $\|\boldsymbol{x}(k)\|^2$ em sua equação de atualização (124)) [159] e a não linearidade do erro [160]. A modelagem estocástica deste algoritmo é por conseguinte desafiadora, já que a ela cumpre considerar essas não linearidades simultaneamente [161], [162], tornando difícil a obtenção de critérios de projeto práticos para ele [163]. Logo, algumas hipóteses estocásticas serão forçosamente introduzidas para contornar tais não linearidades. Entre as hipóteses que poderiam ser adotadas, podemos elencar como exemplos da literatura pretérita: o princípio da separação, aproximações as mais diversas, um sinal de entrada branco, uma estrutura específica da distribuição dos vetores de entrada, um passo de aprendizagem pequeno, assumir que os filtros adaptativos têm comprimento elevado e aproximar algumas soluções por meio de integrais abelianas [162].

A escolha de tais hipóteses depende de experiência e de muita engenhosidade, já que hipóteses muito fortes podem distanciar as previsões teóricas do comportamento real do algoritmo, inutilizando o modelo teórico resultante. Por outro lado, hipóteses muito fracas podem ser insuficientes para lidar analiticamente com a sofisticação do processo de aprendizagem do algoritmo. Adicionalmente, determinadas combinações de hipóteses não levam a equações analíticas fechadas, as quais são desejáveis, já que revelam de maneira sucinta quais fatores concorrem para o desempenho do algoritmo.

Neste capítulo, argumentos de conservação de energia, os quais correspondem a uma das mais poderosas técnicas de análise teóricas do desempenho de algoritmos de filtragem adaptativa [164], são mobilizados na Seção 5.1 para obter uma predição para o desempenho assintótico do algoritmo. A abordagem via conservação de energia mostra que os erros *a priori* e *a posteriori* mantém um certo balanço de energia (conhecido como *relação de variância*), válido para todas iterações [165]. O modelo teórico obtido é objeto de uma exaustiva avaliação na Seção 5.2, a qual precede a seção de comentários finais do capítulo (Seção 5.3). As contribuições deste capítulo, assim como as do capítulo anterior e as dos capítulos seguintes, fazem parte da referência [138], diretamente derivada desta tese.

### 5.1- Análise de regime permanente

Considere o vetor de desvios descrito por

$$\tilde{\boldsymbol{w}}(k+1) \triangleq \boldsymbol{w}^{\star} - \boldsymbol{w}(k), \tag{133}$$

onde $\mathbf{w}^{\star} \in \mathbb{R}^N$ representa a solução ótima desconhecida. Usando a Equação (129) e a definição (133), temos:

$$\tilde{\boldsymbol{w}}(k+1) = \tilde{\boldsymbol{w}}(k) - \beta \frac{\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2} f[e(k)], \tag{134}$$

onde o objetivo de adotarmos a função $f[e(k)]$ é incorporar implicitamente a possibilidade de não atualização (vide Equação (130)), sem que precisemos modelar a probabilidade de atualização explicitamente em nosso modelo analítico.

A multiplicação pela esquerda de ambos os lados da Equação (134) por $\boldsymbol{x}^T(k)$ implica:

$$\overline{e}_{\mathsf{p}}(k) = e_{\mathsf{a}}(k) - \beta f[e(k)], \tag{135}$$

onde as seguintes métricas de erro são consideradas:

$$\overline{e}_{\mathsf{p}}(k) \triangleq \boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k+1), \tag{136}$$

$$e_{\mathsf{a}}(k) \triangleq \boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k). \tag{137}$$

*Comentário*: cabe observar que as métricas de erro definidas nas Equações (136) e (137) diferem das grandezas $e(k)$ (definida em (16)) e $e_{\mathsf{p}}(k)$ (definida em (15)), já que não dependem do sinal de ruído. Tal escolha foi motivada para não tornar a primeira parte da modelagem explicitamente dependente da distribuição do sinal de ruído.

A aplicação da Equação (135) na Equação (134) revela que

$$\tilde{\boldsymbol{w}}(k+1) + \frac{\boldsymbol{x}(k)e_{\mathsf{a}}(k)}{\|\boldsymbol{x}(k)\|^2} = \tilde{\boldsymbol{w}}(k) + \frac{\boldsymbol{x}(k)\overline{e}_{\mathsf{p}}(k)}{\|\boldsymbol{x}(k)\|^2}, \tag{138}$$

o que implica que as energias (isto é, a soma dos elementos ao quadrado dos vetores

envolvidos na igualdade) de ambos os lados são iguais:

$$\begin{aligned}&\left[\tilde{\boldsymbol{w}}^T(k+1)+\frac{\boldsymbol{x}^T(k)e_{\mathsf{a}}(k)}{\|\boldsymbol{x}(k)\|^2}\right]\left[\tilde{\boldsymbol{w}}(k+1)+\frac{\boldsymbol{x}(k)e_{\mathsf{a}}(k)}{\|\boldsymbol{x}(k)\|^2}\right]\\&=\left[\tilde{\boldsymbol{w}}^T(k)+\frac{\boldsymbol{x}^T(k)\overline{e}_{\mathsf{p}}(k)}{\|\boldsymbol{x}(k)\|^2}\right]\left[\tilde{\boldsymbol{w}}(k)+\frac{\boldsymbol{x}(k)\overline{e}_{\mathsf{p}}(k)}{\|\boldsymbol{x}(k)\|^2}\right].\end{aligned} \tag{139}$$

Após algumas manipulações, obtemos:

$$\|\tilde{\boldsymbol{w}}(k+1)\|^2+\frac{e_{\mathsf{a}}^2(k)}{\|\boldsymbol{x}(k)\|^2}=\|\tilde{\boldsymbol{w}}(k)\|^2+\frac{\overline{e}_{\mathsf{p}}(k)}{\|\boldsymbol{x}(k)\|^2}, \tag{140}$$

a qual provê uma descrição *sem aproximações* do fluxo de energia ao longo de cada iteração do algoritmo proposto. Note que a Equação (140) se mantém mesmo no caso de um ruído de medição colorido. O fundamento para (140) é baseado em uma relação básica de conservação de energia que foi originalmente estabelecida em [166], num contexto de análise de robustez de filtros adaptativos. Observe que tal relação é também satisfeita por algoritmos adaptativos clássicos que reutilizam dados [167], [168].

A aplicação do operador de valor esperado $\mathbb{E}\left[\cdot\right]$ em ambos os lados da Equação (140) resulta em

$$\mathbb{E}\left[\|\tilde{\boldsymbol{w}}(k+1)\|^2\right]+\mathbb{E}\left[\frac{e_{\mathsf{a}}^2(k)}{\|\boldsymbol{x}(k)\|^2}\right]=\mathbb{E}\left[\|\tilde{\boldsymbol{w}}(k)\|^2\right]+\mathbb{E}\left[\frac{\overline{e}_{\mathsf{p}}^2(k)}{\|\boldsymbol{x}(k)\|^2}\right]. \tag{141}$$

Assumindo que o algoritmo opera em regime permanente de forma estável permite obter a igualdade $\lim_{k\to\infty}\mathbb{E}\left[\|\tilde{\boldsymbol{w}}(k+1)\|^2\right]=\lim_{k\to\infty}\mathbb{E}\left[\|\tilde{\boldsymbol{w}}(k)\|^2\right]$. Logo, *em regime permanente*, vale a seguinte igualdade:

$$\mathbb{E}\left[\frac{e_{\mathsf{a}}^2(k)}{\|\boldsymbol{x}(k)\|^2}\right]=\mathbb{E}\left[\frac{\overline{e}_{\mathsf{p}}^2(k)}{\|\boldsymbol{x}(k)\|^2}\right]. \tag{142}$$

*Comentário*: a Equação (142) só é válida para o regime assintótico, que é o objetivo deste capítulo. No Capítulo 6, iremos desenvolver um outro modelo analítico especificamente projetado para elucidar o comportamento do algoritmo SM-sign-NLMS no regime transitório.

Quando o filtro apresenta um grande número de pesos e considerando que o denominador na Equação (142) é igual a $\|\boldsymbol{x}(k)\|^2$, o valor esperado da razão pode ser aproximado pela razão dos valores esperados. Esta é uma aproximação razoável, dado que em regime permanente os erros $e_{\mathsf{a}}(k)$ e $\overline{e}_{\mathsf{p}}(k)$ exibem baixa sensibilidade aos dados

de entrada, sendo também correlata ao chamado *princípio da separação* [169]. Logo, a identidade (142) se degenera em

$$\mathbb{E}\left[e_{\mathsf{a}}^2(k)\right] \approx \mathbb{E}\left[\overline{e}_{\mathsf{p}}^2(k)\right]. \tag{143}$$

A derivação de uma fórmula fechada que preveja o MSE em regime permanente requer alguns passos adicionais. Como os erros $\{\overline{e}_{\mathsf{p}}(k), e_{\mathsf{a}}(k), e(k)\}$ são relacionados por (veja as Equações (111) e (115)):

$$\overline{e}_{\mathsf{p}}(k) = e_{\mathsf{a}}(k) - \beta f[e(k)], \tag{144}$$

temos:

$$\overline{e}_{\mathsf{p}}^2(k) = e_{\mathsf{a}}^2(k) - 2\beta e_{\mathsf{a}}(k) f[e(k)] + \beta^2 f^2[e(k)]. \tag{145}$$

Após a aplicação da Equação (145) em (143), chegamos a:

$$2\mathbb{E}\left\{e_{\mathsf{a}}(k) f[e(k)]\right\} = \beta\mathbb{E}\left\{f^2[e(k)]\right\}. \tag{146}$$

A seguir, assumiremos que $e_{\mathsf{a}}(k)$ e $e(k)$ são conjuntamente gaussianos (uma hipótese razoável para filtros adaptativos longos [160]). Considerando

$$\xi \triangleq \mathbb{E}\left[e^2(k)\right] \tag{147}$$

e

$$\rho \triangleq e_{\mathsf{a}}(k) e(k). \tag{148}$$

e empregando o Teorema de Price [170], chegamos a:

$$\begin{aligned} \frac{\partial\mathbb{E}\left[e_{\mathsf{a}}(k) f[e(k)]\right]}{\partial\rho} &= \mathbb{E}\left[\frac{\partial e_{\mathsf{a}}(k)}{\partial e_{\mathsf{a}}(k)} \frac{\partial f[e(k)]}{\partial e(k)}\right] \\ = \int_{-\infty}^{\infty} \frac{[\delta[e(k)-\overline{\gamma}] + \delta[e(k)+\overline{\gamma}]}{\sqrt{2\pi\xi}} &\mathsf{exp}\left[-\frac{e^2(k)}{2\xi}\right] de(k) \\ = \frac{\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right] + \mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}{\sqrt{2\pi\xi}} &= \frac{2\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}{\sqrt{2\pi\xi}} \end{aligned}, \tag{149}$$

o que acarreta a identidade

$$\mathbb{E}\left\{e_{\mathsf{a}}(k) f[e(k)]\right\} = \sqrt{\frac{2}{\pi\xi}}\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right]\mathbb{E}\left[e_{\mathsf{a}}(k) e(k)\right]. \tag{150}$$

Assumindo que o erro é distribuído de acordo com uma distribuição gaussiana (o que é uma hipótese bastante utilizada na literatura; veja, por exemplo, [171]) resulta em

$$\begin{aligned}\mathbb{E}\left\{f^2[e(k)]\right\} &= 2\beta\int_{\overline{\gamma}}^{\infty}\frac{1}{\sqrt{2\pi\xi}}\mathsf{exp}\left[-\frac{e^2(k)}{2\xi}\right]de(k) \\ &= \frac{2}{\sqrt{\pi}}\int_{\frac{\overline{\gamma}}{\sqrt{\pi}}}^{\infty}\mathsf{exp}\left(-u^2\right)du = \mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right),\end{aligned} \tag{151}$$

onde

$$\mathsf{erfc}(x) \triangleq \frac{2}{\sqrt{\pi}}\int_{x}^{\infty}\mathsf{exp}(-t^2)dt. \tag{152}$$

A utilização das Equações (150), (151) e (146) permite-nos concluir:

$$2\sqrt{\frac{2}{\pi\xi}}\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right]\xi - \sqrt{\frac{2}{\pi\xi}}\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right]\sigma_\nu^2 = \beta\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right), \tag{153}$$

onde $\sigma_\nu^2$ denota a variância do ruído de medição $\nu(k)$, supostamente de média zero, i.i.d. e estatisticamente independente do sinal de entrada. A partir da Equação (153), pode-se obter a seguinte equação *de ponto fixo* para o erro quadrático médio assintótico:

$$\xi = \sigma_\nu^2 + \frac{\beta}{2}\frac{\sqrt{\frac{\pi\xi}{2}}\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right)}{\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}. \tag{154}$$

*Comentário*: A Equação (154) é a principal contribuição deste capítulo. Trata-se de uma equação não linear cuja solução positiva é uma aproximação teórica do MSE em regime permanente real do algoritmo proposto SM-sign-NLMS. Essa equação dá origem a uma estimativa teórica do desempenho assintótico do algoritmo $\xi$ por meio de um procedimento numérico simples: assumindo uma estimativa $\hat{\xi}_n$ para a grandeza $\xi$, esta pode ser refinada por meio de:

$$\hat{\xi}_{n+1} \leftarrow \sigma_\nu^2 + \frac{\beta}{2}\frac{\sqrt{\frac{\pi\hat{\xi}_n}{2}}\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\hat{\xi}_n}}\right)}{\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\hat{\xi}_n}\right]}. \tag{155}$$

A equação de refinamento (155) deve ser iterada um determinado número de vezes, até que a diferença $\left|\hat{\xi}_{n+1} - \hat{\xi}_n\right|$ seja desprezível. Após finalizado o procedimento iterativo, o valor de $\hat{\xi}_{n+1}$ reflete a estimativa teórica gerada pela equação analítica de ponto fixo (154).

Interessantemente, a Equação (154) implica que as estatísticas do sinal de en-

| #Cenário | #Modelo | $N$ | $\tau$ | $\sigma_\nu^2$ | $B(z)$ | $A(z)$ | $N_{\text{it}}$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 64 | 5 | $10^{-5}$ | 1 | 1 | $10^3$ |
| 2 | 1 | 64 | 5 | $10^{-5}$ | 1 | 1-0,75$z^{-1}$-0,19$z^{-2}$-0,09$z^{-3}$+0,5$z^{-4}$ | $10^3$ |
| 3 | 2 | 96 | $\{2,5\}$ | $10^{-6}$ | 1 | 1 | $4 \times 10^4$ |
| 4 | 3 | 96 | $\{2,5\}$ | $10^{-6}$ | 1-0,8$z^{-1}$ | 1 | $4 \times 10^4$ |

Tabela 10 – Características dos 4 cenários. Nos Cenários 3 e 4, o número de iterações $N_{\text{it}}$ foi maior para garantir o regime permanente, já que o número de coeficientes necessários para identificar a planta desconhecida é maior. #Modelo reflete o número da função de transferência estimada por medições, constante no relatório da ITU [122].

| #Cenário | $\tau$ | Máxima diferença (dB) |
|---|---|---|
| 1 | 5 | 0,100 |
| 2 | 5 | 0,293 |
| 3 | 2 | 0,069 |
| 3 | 5 | 0,069 |
| 4 | 2 | 0,014 |
| 4 | 5 | 0,031 |

Tabela 11 – Máxima diferença (em dB) entre as curvas teórica e simulada nos diferentes cenários.

trada (por exemplo, sua variância) não impactam o desempenho assintótico do algoritmo, embora possam afetar sua taxa de convergência (característica associada ao comportamento transitório do algoritmo). O mesmo raciocínio se aplica ao tamanho do filtro adaptativo, bem como aos valores específicos (e desconhecidos) dos coeficientes do filtro ótimo, os quais não influenciam o desempenho em regime permanente do algoritmo, segundo o modelo proposto.

## 5.2- Resultados

Para avaliar a qualidade da predição teórica engendrada pela Equação (154), nesta seção a predição por ela obtida será computada com estimativas oriundas de curvas simuladas com ensaios independentes de Monte Carlo. Como a Equação (154) não prevê nenhum impacto no desempenho assintótico de um eventual colorimento do sinal de entrada, iremos inserir, em alguns cenários, sinais de entrada coloridos. Assim, o sinal de entrada é obtido pelo colorimento de um sinal branco, gaussiano, de média zero e variância unitária pela função de transferência $\frac{B(z)}{A(z)}$, onde $A(z)$ e $B(z)$ denotam polinômios na variável $z$. Quando $A(z) = B(z) = 1$, não há colorimento no sinal de

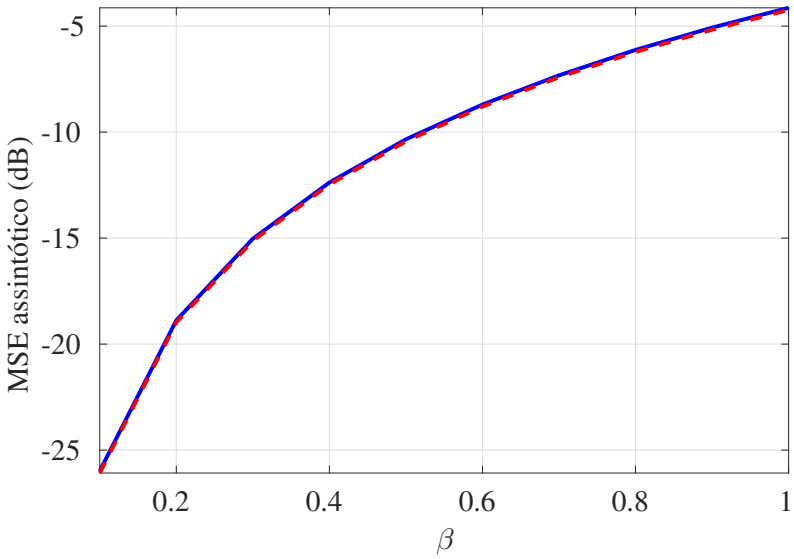



Figura 28 – Regime permanente teórico e simulado para o Cenário 1. Linha azul sólida: curva simulada; linha vermelha tracejada: predição teórica.

entrada. A Tabela 10 apresenta os detalhes dos quatro cenários simulados. As Figuras 28-33 comparam os valores teóricos (advindos da resolução numérica da Equação (154)), com os resultados de simulação.

A Tabela 11 mostra a maior diferença (em valor absoluto e em dB) entre as curvas teórica e simulada. Observa-se que a diferença é muito pequena, no mais das vezes inferior a 0,1 dB. O único caso em que a diferença é maior do que 0,1 dB é no Cenário 2, no qual o sinal de entrada é bastante colorido. Provavelmente tal discrepância se deve ao fato de que o princípio da separação é violado neste caso, já que um colorimento mais intenso introduz uma maior dependência estatística entre os coeficientes adaptativos e o sinal de entrada.

O número de iterações necessárias para resolver numericamente a Equação (154), por meio da Equação (155), é bastante estável, sendo necessárias entre 26 e 35 iterações para obtenção da estimativa teórica (para todos os valores de $\beta$, $\tau$ e $\sigma_\nu^2$ considerados), o que exige alguns décimos de segundo para seu cálculo em um computador de uso pessoal. Na nossa implementação, o processo iterativo termina quando a diferença $\left|\hat{\xi}_{n+1} - \hat{\xi}_n\right|$ é menor do que $10^{-10}$, em todos os casos. Ademais, impusemos o valor inicial $\hat{\xi}_0 = 10^{-2}$.

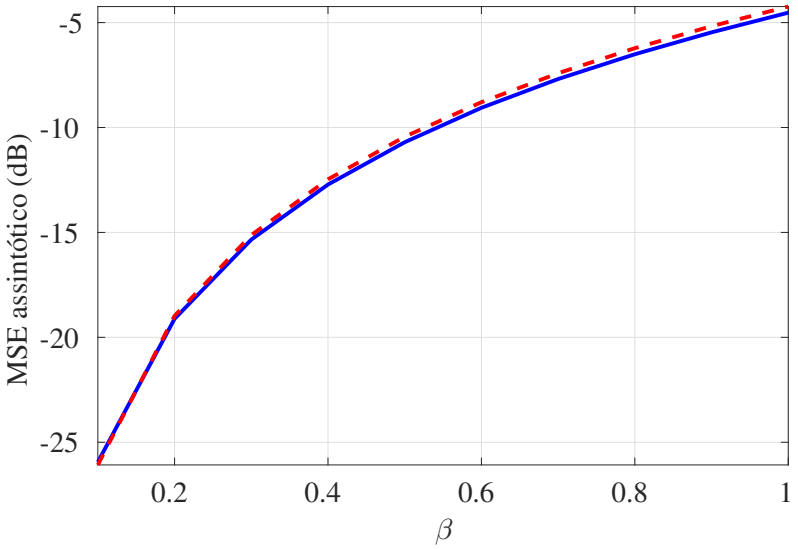



Figura 29 – Regime permanente teórico e simulado para o Cenário 2. Linha azul sólida: curva simulada; linha vermelha tracejada: predição teórica.

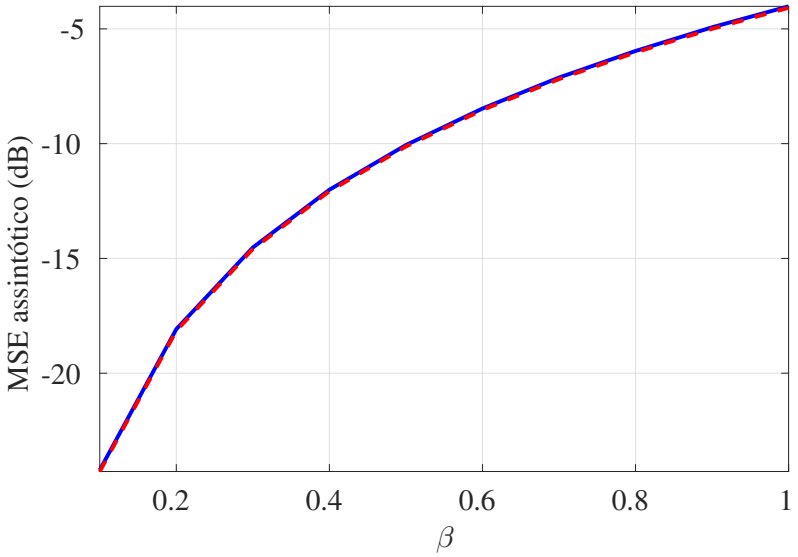



Figura 30 – Regime permanente teórico e simulado para o Cenário 3 ($\tau = 2$). Linha azul sólida: curva simulada; linha vermelha tracejada: predição teórica.

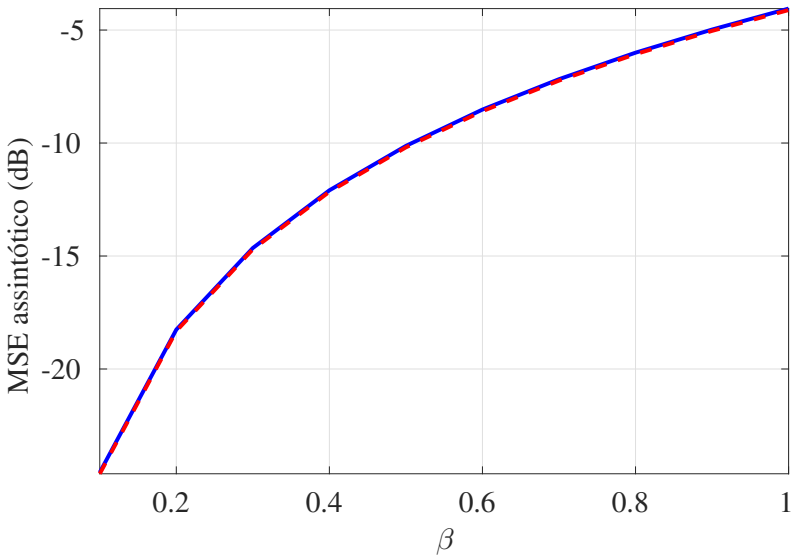



Figura 31 – Regime permanente teórico e simulado para o Cenário 3 ($\tau = 5$). Linha azul sólida: curva simulada; linha vermelha tracejada: predição teórica.

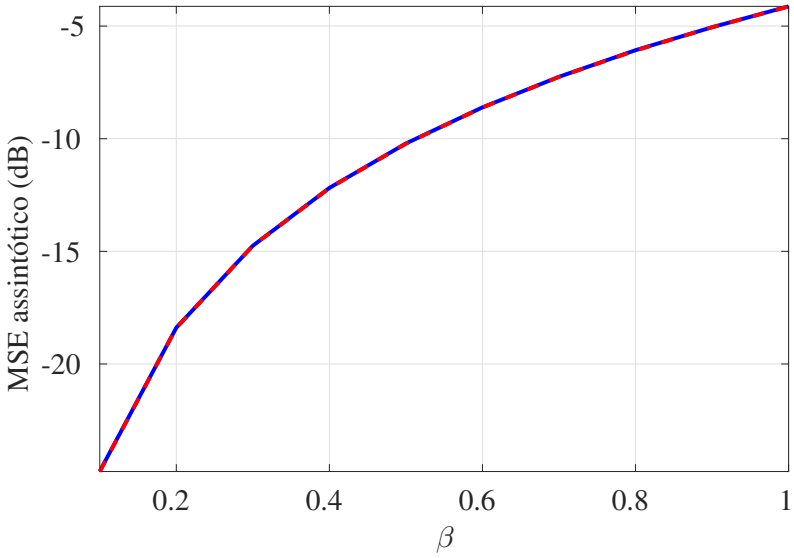



Figura 32 – Regime permanente teórico e simulado para o Cenário 4 ($\tau = 2$). Linha azul sólida: curva simulada; linha vermelha tracejada: predição teórica.

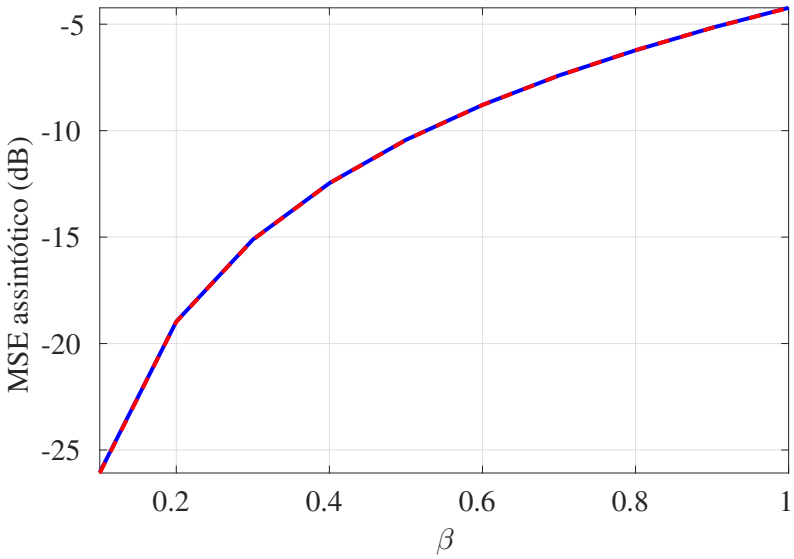



Figura 33 – Regime permanente teórico e simulado para o Cenário 4 ($\tau = 5$). Linha azul sólida: curva simulada; linha vermelha tracejada: predição teórica.

### 5.3- Comentários finais

Este capítulo contém uma análise de regime permanente do algoritmo proposto SM-sign-NLMS, a qual mobiliza argumentos de conservação de energia. O modelo estocástico proposto utiliza hipóteses estatisticamente razoáveis (tal como a de que os sinais $e_{\mathsf{a}}(k)$ e $e(k)$ são conjuntamente gaussianos, uma hipótese factível para filtros de comprimentos longos), dando origem a uma equação de ponto fixo concisa (Equação (154)), cuja solução numérica iterativa permite a extração de uma predição teórica para o desempenho assintótico do algoritmo. Tal predição foi extensamente aquilatada por meio de comparações com experimentos computacionais, apresentando elevada acurácia aos dados experimentais.

O regime permanente é um ponto de operação importante dos algoritmos de filtragem adaptativa; entretanto, é necessário também ter garantias teóricas acerca do número de iterações necessárias para que o algoritmo atinja o patamar assintótico. Afinal, filtros que apresentem um bom desempenho assintótico podem ser inúteis na maioria das aplicações, caso demandem um número excessivo de dados até que atinjam o regime permanente. Análises teóricas capazes de modelar o desempenho do algoritmo nas iterações que precedem o regime permanente são conhecidas na literatura como *análises de transiente*. O próximo capítulo efetua uma análise de transiente inédita do algoritmo

proposto SM-sign-NLMS.

# 6- Análise de Transiente do Algoritmo SM-Sign-NLMS

Assim como no caso da análise de regime permanente, para que possamos lidar com as características não lineares da equação de atualização do algoritmo SM-sign-NLMS num modelo teórico de seu comportamento transitório, é necessário empregar algumas hipóteses estatísticas. Neste capítulo, efetuaremos uma análise de transiente do algoritmo SM-sign-NLMS adotando um modelo simplificado para a estrutura do vetor de entrada do algoritmo. O modelo proposto dá origem a uma descrição do *comportamento modal* do aprendizado de segunda ordem do algoritmo. A descrição resultante permite a extração de novos *insights* acerca da dependência que a convergência do algoritmo SM-sign-NLMS apresenta com relação à distribuição dos autovalores da matriz de autocorrelação dos dados de entrada. O modelo teórico também será importante para o delineamento de um algoritmo SM-sign-NLMS com passo de aprendizagem variável, que será descrito no Capítulo 7. Neste capítulo, também são analisadas sob quais condições o modelo analítico de transiente é equivalente ao preconizado no Capítulo 5, para a região de operação em regime permanente.

O modelo a ser adotado neste capítulo utiliza a ubíqua *hipótese da independência* [172], a qual implica que os vetores $\boldsymbol{x}(k)$ são independentes e identicamente distribuídos. Tal hipótese, embora bastante popular, não é válida no caso de estruturas transversais, já que estas induzem um acoplamento determinístico entre vetores de entrada sucessivos. Adicionalmente, supõe-se que as distribuições radial e angular do vetor de entrada são estatisticamente independentes, e que esta última é uma variável aleatória discreta. Tais hipóteses foram judiciosamente projetadas para permitir a extração de equações analíticas de interesse, sem prejudicar em demasia a aderência do modelo final às curvas reais de aprendizado do algoritmo.

Este capítulo está estruturado do seguinte modo: a Seção 6.1 descreve a obtenção do modelo analítico para a região de transiente do algoritmo. O modelo é avaliado na Seção 6.2, a qual é sucedida pela Seção 6.3, que contém os comentários finais referentes a este capítulo.

## 6.1- Análise de Transiente do algoritmo SM-sign-NLMS

A decomposição em autovalores e autovetores da matriz de autocorrelação do vetor de entrada $\boldsymbol{R}$:

$$\boldsymbol{R} \triangleq \mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T = \sum_{i=0}^{N-1} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^T \tag{156}$$

dá protagonismo a duas grandezas essenciais para a análise a seguir: os autovalores $\lambda_i \in \mathbb{R}+$ (para $i \in \{0, 1, \ldots, N-1\}$) e os autovetores *ortonormais* $\boldsymbol{v}_i \in \mathbb{R}^N$ (para $i \in \{0, 1, \ldots, N-1\}$).

Um modelo simples que adere aos momentos de primeira e de segunda ordens de $\boldsymbol{x}(k)$ é [173]:

$$\boldsymbol{x}(k) = s_k r_k \tilde{\boldsymbol{v}}(k), \tag{157}$$

no qual as variáveis aleatórias[1] *estatisticamente independentes* $s_k$ (referente ao *sinal*), $r_k$ (referente à distribuição *radial*) e $\tilde{\boldsymbol{v}}(k)$ (referente à distribuição *angular*, assumida como uma variável aleatória discreta) são distribuídas de acordo com:

$$\text{Prob}\{s_k = \pm 1\} = \frac{1}{2}, \tag{158}$$

$$r(k) \sim \|\boldsymbol{x}(k)\|, \tag{159}$$

$$\text{Prob}\{\tilde{\boldsymbol{v}}(k) = \boldsymbol{v}_i\} = \frac{\lambda_i}{\text{Tr}[\boldsymbol{R}]}, \tag{160}$$

onde a Equação (159) significa que $r(k)$ emula a distribuição do vetor original (isto é, derivada de uma estrutura transversal) $\boldsymbol{x}(k)$. Note que a gaussianidade de $\boldsymbol{x}(k)$ pode ser parcialmente inserida no modelo (já que sua distribuição radial pode emular a distribuição de um vetor aleatório $\boldsymbol{x}(k)$ qualquer), o que motiva a aplicação da chamada *heurística de Price* [25]. Esta heurística será adotada no modelo estocástico apresentado neste capítulo. Observe que o modelo simplificado implicado pela Equação (157) é capaz de responder a questões para as quais não há resposta disponível utilizando outros modelos possivelmente mais realísticos [173]. Adicionalmente, o modelo não é demasiado simplista, já que não exige que o sinal de entrada seja branco, uma hipótese comum na

[1] No caso de $\boldsymbol{v}(k)$, a expressão adequada é "vetor aleatório".

literatura. É uma vantagem do modelo proposto a incorporação do efeito de um eventual colorimento do sinal de entrada, já que este colorimento sabidamente leva a uma redução na taxa de convergência do algoritmo [174].

Para que possamos obter um modelo de segunda ordem para o algoritmo SM-sign-NLMS, considere a multiplicação de ambos os lados de (134) por suas versões transpostas. Para caracterizar o comportamento médio do algoritmo, aplicamos o operador de valor esperado na recursão resultante, obtendo:

$$\begin{aligned} \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k+1) &= \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) - \beta\mathbb{E}\left\{\frac{\tilde{\boldsymbol{w}}(k)\boldsymbol{x}^T(k)}{\|\boldsymbol{x}(k)\|^2} f[e(k)]\right\} && (161)\\ &\quad - \beta\mathbb{E}\left\{\frac{\boldsymbol{x}(k)\tilde{\boldsymbol{w}}^T(k)}{\|\boldsymbol{x}(k)\|^2} f[e(k)]\right\} && (162)\\ &\quad + \beta^2\mathbb{E}\left\{\frac{\boldsymbol{x}(k)\boldsymbol{x}^T(k)}{\|\boldsymbol{x}(k)\|^4} f^2[e(k)]\right\}, && (163) \end{aligned}$$

onde

$$\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) \triangleq \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\right] \tag{164}$$

é a matriz de autocorrelação dos desvios. Esta matriz tem um papel fundamental na predição do MSE, porque permite o cômputo de

$$\tilde{\lambda}_i(k) \triangleq \boldsymbol{v}_i^T \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\boldsymbol{v}_i, \tag{165}$$

a partir do qual, com a hipótese da independência e com as hipóteses comuns para o ruído de medição, o MSE pode ser obtido:

$$\begin{aligned} \xi(k) \triangleq \mathbb{E}\left[e^2(k)\right] &= \sigma_\nu^2 + \mathbb{E}\left[\|\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\|^2\right] \\ &= \sigma_\nu^2 + \mathsf{Tr}\left[\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{R}_{\tilde{\boldsymbol{w}}}\boldsymbol{V}\right] \\ &= \sigma_\nu^2 + \sum_{i=0}^{N-1}\lambda_i\tilde{\lambda}_i(k). \end{aligned} \tag{166}$$

A Equação (166) apresenta uma implicação importante: a predição da dinâmica de $\tilde{\lambda}_i(k)$ permite estimar a evolução do MSE, que é o objetivo final da análise de transiente. Logo, ao multiplicar a Equação (163) pela esquerda por $\boldsymbol{v}_i^T$ e pela direita por $\boldsymbol{v}_i$ nos leva

a:

$$\begin{aligned}\tilde{\lambda}_i(k+1) &= \tilde{\lambda}_i(k) \underbrace{-\beta\mathbb{E}\left\{\frac{\boldsymbol{v}_i^T\tilde{\boldsymbol{w}}(k)\boldsymbol{x}^T(k)\boldsymbol{v}_i}{\|\boldsymbol{x}(k)\|^2}f[e(k)]\right\}}_{\text{①}} \\ &\quad \underbrace{-\beta\mathbb{E}\left\{\frac{\boldsymbol{v}_i^T\boldsymbol{x}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{v}_i}{\|\boldsymbol{x}(k)\|^2}f[e(k)]\right\}}_{\text{②}} \\ &\quad \underbrace{+\beta^2\mathbb{E}\left\{\frac{\boldsymbol{v}_i^T\boldsymbol{x}(k)\boldsymbol{x}^T(k)\boldsymbol{v}_i}{\|\boldsymbol{x}(k)\|^4}f^2[e(k)]\right\}}_{\text{③}}.\end{aligned} \tag{167}$$

O cômputo analítico dos termos ①-③ é uma tarefa desafiadora, especialmente devido à presença dos termos não lineares $f[e(k)]$ e $f^2[e(k)]$. Para lidar com esta complexidade, duas hipóteses simplificadoras são adotadas: o modelo de Slock (vide Equações (158)-(160)) e a heurística de Price (discutida em [25]). Concisamente, a heurística de Price utiliza o teorema de Price [175] (o qual é estritamente válido apenas quando variáveis aleatórias gaussianas encontram-se envolvidas nas esperanças estatísticas) quando o vetor de entrada é gerado de acordo com o modelo descrito pelas Euações (158)-(160).

Assim, utilizando tais hipóteses, o termo ① em (167) pode ser expresso como:

$$\begin{aligned}\text{①} &= -\beta\sqrt{\frac{2}{\pi\xi}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]\mathbb{E}\left[\frac{\boldsymbol{v}_i^T\tilde{\boldsymbol{w}}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{v}_i}{\|\boldsymbol{x}(k)\|^2}\right] \\ &= -\beta\sqrt{\frac{2}{\pi\xi}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]\mathbb{E}\left\{\sum_{j=1}^{N}\frac{\lambda_j}{\mathsf{Tr}[\boldsymbol{R}]}\frac{\mathbb{E}[s_k^2]\mathbb{E}[r_k^2]\boldsymbol{v}_i^T\tilde{\boldsymbol{w}}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{v}_j\boldsymbol{v}_j^T\boldsymbol{v}_i}{\mathbb{E}\left[s_k^2\right]\mathbb{E}\left[r_k^2\right]\boldsymbol{v}_j^T\boldsymbol{v}_j}\right\}\end{aligned} \tag{168}$$

A expressão em (168) pode ser simplificada, usando os fatos de que *i*) os autovetores $\boldsymbol{v}_j$ são ortonormais; *ii*) $\mathbb{E}[s_k^2]=1$ (vide Equação (158)) e *iii*) $\mathbb{E}\left[r_k^2\right]=\mathsf{Tr}\left[\boldsymbol{R}\right]$ (vide Equação (159)). Daí:

$$\text{①} = -\beta\sqrt{\frac{2}{\pi\xi}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]\frac{\lambda_i}{\mathsf{Tr}[\boldsymbol{R}]}\tilde{\lambda}_i(k). \tag{169}$$

enquanto que os termos Ⓘ e Ⓘ podem ser simplificados para

$$\begin{aligned}
\text{Ⓘ} &= -\beta\mathbb{E}\left[\frac{\boldsymbol{v}_i^T\boldsymbol{x}(k)\tilde{\boldsymbol{w}}^T(k)f[e(k)]\boldsymbol{v}_i}{\|\boldsymbol{x}(k)\|^2}\right] \\
&= -\beta\sqrt{\frac{2}{\pi\xi}}\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right]\mathbb{E}\left[\frac{\boldsymbol{v}_i^T\boldsymbol{x}(k)\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)\boldsymbol{v}_i}{\|\boldsymbol{x}(k)\|^2}\right], \\
&= -\beta\sqrt{\frac{2}{\pi\xi}}\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right]\frac{\lambda_i}{\mathsf{Tr}[\boldsymbol{R}]}\tilde{\lambda}_i(k),
\end{aligned} \tag{170}$$

$$\begin{aligned}
\text{Ⓘ} &\approx \beta^2\mathbb{E}\left[\frac{\boldsymbol{v}_i^T\boldsymbol{x}(k)\boldsymbol{x}^T(k)\boldsymbol{v}_i}{\|\boldsymbol{x}(k)\|^4}\right]\mathbb{E}\left\{f^2[e(k)]\right\} \\
&= \beta^2\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right)\sum_{j=0}^{N-1}\frac{\lambda_j}{\mathsf{Tr}[\boldsymbol{R}]}\frac{\mathbb{E}[s_k^2]\mathbb{E}\left[r_k^2\right]\boldsymbol{v}_i^T\boldsymbol{v}_j\boldsymbol{v}_j^T\boldsymbol{v}_i}{\mathbb{E}\left[s_k^4\right]\mathbb{E}\left[r_k^4\right]\boldsymbol{v}_j^T\boldsymbol{v}_j}. \\
&= \beta^2\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\pi}}\right)\frac{\lambda_i}{\mathbb{E}[r^4]}
\end{aligned} \tag{171}$$

A combinação das Equações (167)- (171) dá origem à recursão seguinte:

$$\begin{aligned}
\tilde{\lambda}_i(k+1) &= \tilde{\lambda}_i(k) - 2\sqrt{\frac{2}{\pi\xi}}\beta\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right]\frac{\lambda_i}{\mathsf{Tr}[\boldsymbol{R}]}\tilde{\lambda}_i(k) \\
&\quad + \beta^2\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right)\frac{\lambda_i}{\mathbb{E}[r^4]},
\end{aligned} \tag{172}$$

onde se observa que as dinâmicas dos modos $\tilde{\lambda}_i(k)$ são acopladas, já que $\xi$ depende de todos os $\tilde{\lambda}_i(k)$ (veja (166)). Usando (172) e (166), é possível prever a evolução do MSE. Logo, um modelo para o comportamento transitório do algoritmo é obtido, conforme desejado.

O modelo proposto não é restrito à região de transiente, já que em sua derivação não há nenhuma hipótese que seja inválida no regime assintótico. Neste caso, assumindo que o algoritmo opera de maneira estável, temos que *em regime permanente* $\tilde{\lambda}_i(k+1) = \tilde{\lambda}_i(k)$ e, assumindo operação assintótica:

$$\tilde{\lambda}_i(k) = \frac{\beta}{2}\sqrt{\frac{\pi\xi}{2}}\frac{\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right)}{\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}\frac{\mathsf{Tr}[\mathbf{R}]}{\mathbb{E}\left[r^4\right]}, \tag{173}$$

o que implica que em regime permanente:

$$\xi = \sigma_\nu^2 + \sum_{i=0}^{N-1}\lambda_i\tilde{\lambda}_i(k) = \sigma_\nu^2 + \frac{\beta}{2}\frac{\sqrt{\frac{\pi\xi}{2}}\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right)}{\mathsf{exp}\left(-\frac{\overline{\gamma}^2}{2\xi}\right)}\frac{\{\mathsf{Tr}[\boldsymbol{R}]\}^2}{\mathbb{E}\left[r^4\right]}, \tag{174}$$

uma identidade similar à que foi obtida no Capítulo 5, com argumentos de balanceamento de energia (veja a Equação (154)).

*Comentário*: É possível elucidar sob quais condições as Equações (174) e (154) são equivalentes. Assumindo que o vetor de entrada deriva de um processo gaussiano branco, temos:

$$\mathbb{E}\left[\|\boldsymbol{x}(k)\|_2^n\right] = 2^{\frac{n-2}{2}}\sigma_x^n N\frac{\Gamma\left(\frac{N+n}{2}\right)}{\Gamma\left(\frac{N+2}{2}\right)}, \tag{175}$$

onde $\sigma_x^2$ é a variância de $x$ e

$$\Gamma(x) \triangleq \int_0^\infty t^{x-1}\mathsf{exp}(-t)dt. \tag{176}$$

Logo,

$$\frac{\{\mathsf{Tr}[\boldsymbol{R}]\}^2}{\mathbb{E}\left[r^4\right]} = \frac{N}{2}\frac{\Gamma\left(\frac{N+2}{2}\right)}{\Gamma\left(\frac{N+4}{2}\right)}. \tag{177}$$

Como

$$\lim_{N\to\infty}\frac{N}{2}\frac{\Gamma\left(\frac{N+2}{2}\right)}{\Gamma\left(\frac{N+4}{2}\right)} = 1, \tag{178}$$

ambas as equações (isto é, (174) e (154)), a despeito de serem obtidas com hipóteses estatísticas bastante diferentes, são equivalentes quando $N$ assume valores elevados. De certo modo, isso é coerente com o fato de que o modelo de regime permanente apresentado no Capítulo 5 supõe que os sinais $e_{\mathsf{a}}(k)$ e $e(k)$ são conjuntamente gaussianos, o que é uma hipótese adequada apenas para filtros adaptativos de elevado comprimento [160]. Ou seja, o modelo de regime permanente no Capítulo 5 se aproxima da realidade quando $N$ apresenta um valor elevado, tornando-o equivalente ao modelo de transiente engendrado pela Equação (174), quando assume-se que o algoritmo opera em regime permanente.

## 6.2- Resultados

A combinação das Equações (172) e (166) dá origem a um modelo estocástico inédito para a predição do MSE do algoritmo SM-sign-NLMS. Nesta seção, iremos cotejar as predições teóricas com resultados advindos de simulação. Para esse propósito, foram construídos 4 cenários, especificados na Tabela 12. Os diferentes cenários contemplam

| #Cenário | #Modelo | $N$ | $\beta$ | $\tau$ | $\sigma_\nu^2$ | $B(z)$ | $N_{\text{it}}$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 64 | 0,5 | 2 | $10^{-6}$ | 1 | $8 \times 10^3$ |
| 2 | 2 | 96 | 0,4 | 2 | $10^{-6}$ | $1 - 0,8z^{-1} + 0,2z^{-2}$ | $4 \times 10^4$ |
| 3 | 3 | 96 | 0,2 | 5 | $10^{-5}$ | $1 - 0,8z^{-1}$ | $1 \times 10^5$ |
| 4 | 4 | 128 | 0,7 | 5 | $10^{-4}$ | $1 - 0,6z^{-1} + 0,2z^{-2}$ | $2 \times 10^4$ |

Tabela 12 – Especificação dos cenários adotados para avaliação da predição teórica de transiente. Todas as plantas tiveram seus coeficientes multiplicados por 100. A coluna #Modelo reflete o número da função de transferência estimada por medições, constante no relatório da ITU [122].

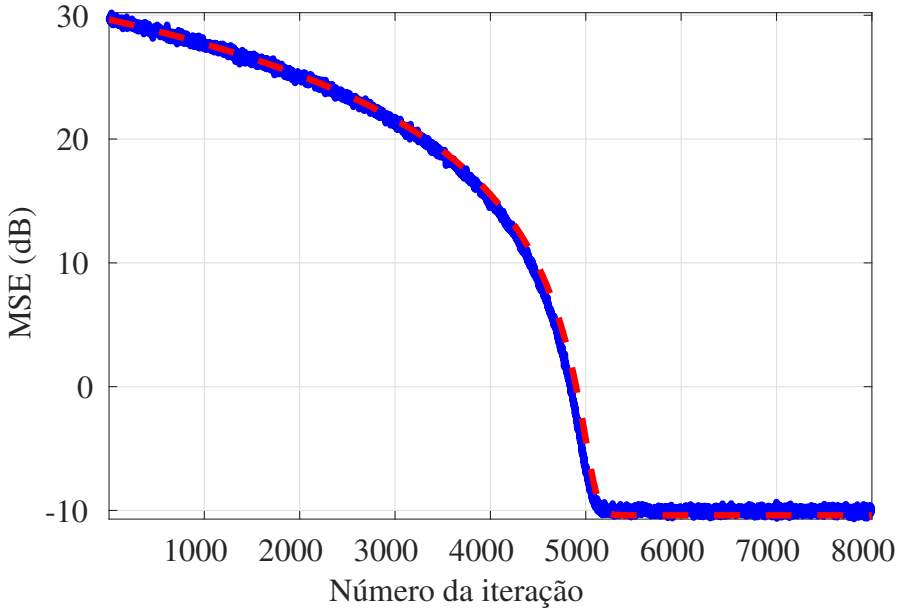



Figura 34 – Evolução do MSE (em dB) teórico (em vermelho tracejado) e empírico (em azul sólido), referente ao Cenário 1.

diferentes plantas a identificar, valores de $\beta$, $\tau$ e $\sigma_\nu^2$ distintos, assim como o filtro FIR $B(z)$ que, tendo como entrada um ruído gaussiano de média zero e branco, gera o sinal de entrada (potencialmente colorido, quando $B(z)$ não é unitário) do filtro adaptativo. As diferentes configurações dos cenários servem para ilustrar a capacidade do modelo estocástico de efetuar predições em diversas configurações. Em todos os cenários, foi computada a média de 1000 ensaios independentes de Monte Carlo, para efeito de cálculo da curva simulada. O ruído $\nu(k)$ é advindo de um processo gaussiano i.i.d. de média zero e variância $\sigma_\nu^2$.

As Figuras 34-37 apresentam as curvas de evolução do MSE (em dB) simuladas e as suas respectivas previsões teóricas. No mais das vezes, observa-se uma aderência da previsão teórica à curva simulada nas primeiras iterações, assim como em regime permanente. A maior diferença em valores absolutos (em dB) entre a curva teórica e

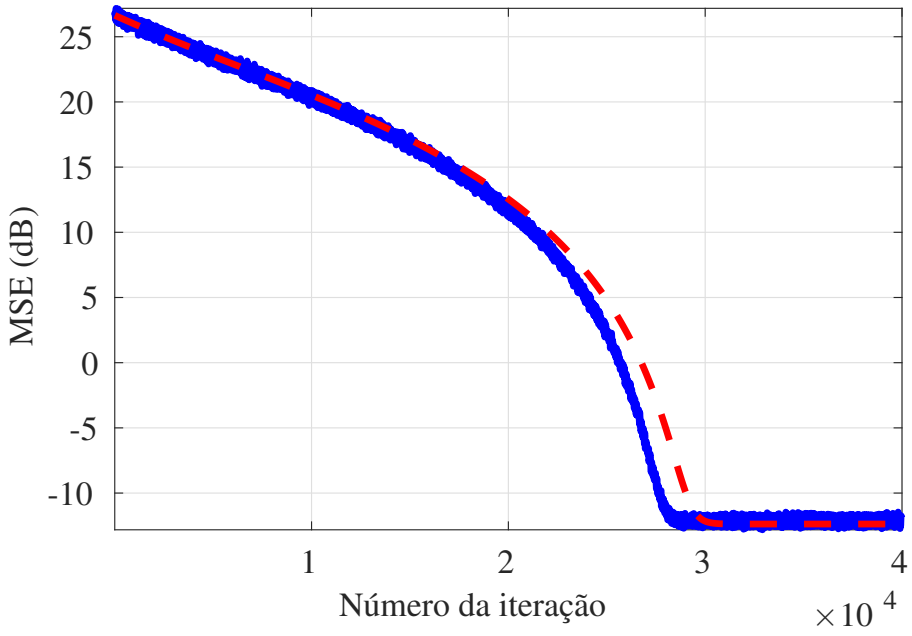



Figura 35 – Evolução do MSE (em dB) teórico (em vermelho tracejado) e empírico (em azul sólido), referente ao Cenário 2.

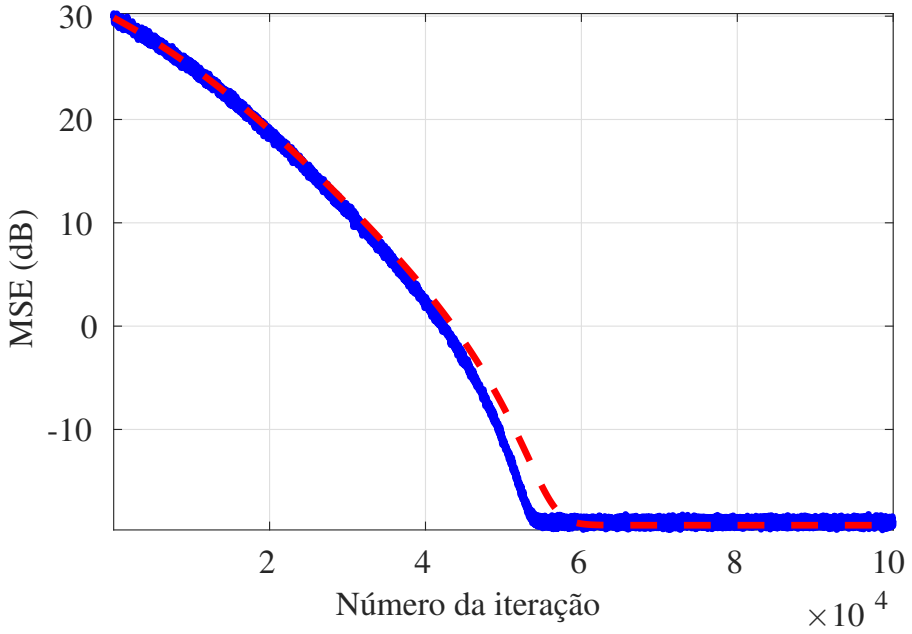



Figura 36 – Evolução do MSE (em dB) teórico (em vermelho tracejado) e empírico (em azul sólido), referente ao Cenário 3.

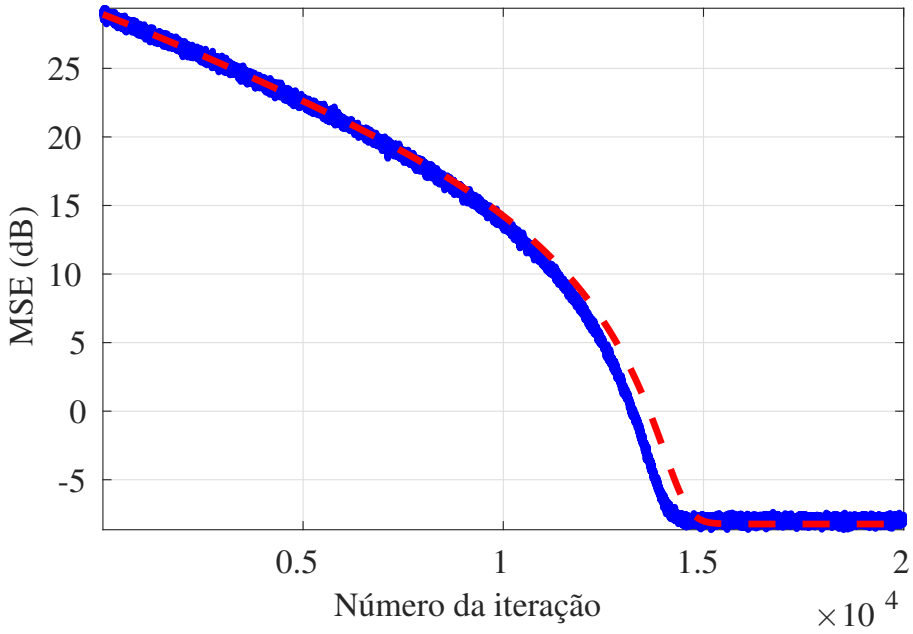



Figura 37 – Evolução do MSE (em dB) teórico (em vermelho tracejado) e empírico (em azul sólido), referente ao Cenário 4.

a simulada (ao longo de todas as iterações) é de 2,95 dB (no Cenário 1), 7,05 dB (no Cenário 2), 5,59 dB (no Cenário 3) e 4,29 dB (no Cenário 4). Observa-se uma maior discrepância nos cenários em que o sinal de entrada é colorido. Pode-se justificar este fato evocando-se o sofisticado acoplamento estatístico entre o sinal de entrada e os coeficientes adaptativos, o qual é apenas parcialmente capturado pelo modelo teórico proposto (afinal, ele pressupõe independência estatística entre o vetor $\boldsymbol{x}(k)$ e o vetor de coeficientes adaptativos $\boldsymbol{w}(k)$). Tal acoplamento estatístico é mais intenso quando o sinal de entrada é colorido, já que existe maior dependência estatística entre vetores de entrada consecutivos que alimentam o filtro adaptativo.

A Tabela 13 apresenta o número de iterações necessárias $N_{\text{iter}}^{\Delta}$ (tanto nas simulações quanto nas predições teóricas) para que o MSE inicial caia em 5, 10, 15 e 20 dB. Também se computa a taxa de convergência (em dB/iteração), tanto para a previsão teórica quanto para o resultado simulado. Para todos os cenários, observa-se uma razoável aderência do modelo teóricos às curvas simuladas. Em termos percentuais, observa-se que a taxa de convergência $R_{\text{iter}}^{\Delta}$ é estimada com um erro percentual máximo de 3,27% (no Cenário 1), de 5,83% (no Cenário 2), de 3,26% (no Cenário 3) e de 15,70% (no Cenário 4). Já o número de iterações necessárias para que o MSE caia em $\Delta_{\text{MSE}}$ dBs é estimada com um erro percentual máximo de 3,38% (no Cenário 1), de 6,2% (no Cenário 2), de 3,37% (no Cenário 3) e 18,62% (no Cenário 4). Cabe notar que os maiores erros (em termos

| **Cenário 1** | | | | |
|---|---|---|---|---|
| $\Delta_{\text{MSE}}$ (em dB) | $N_{\text{iter}}^{\Delta}$ (simul.) | $N_{\text{iter}}^{\Delta}$ (teór.) | $R_{\text{iter}}^{\Delta}$ (simul.) | $R_{\text{iter}}^{\Delta}$ (teór.) |
| 5 | 2215 | 2163 | 0,0022573 | 0,0023116 |
| 10 | 3281 | 3392 | 0,0030479 | 0,0029481 |
| 15 | 4016 | 4096 | 0,0037351 | 0,0036621 |
| 20 | 4432 | 4505 | 0,0045126 | 0,0044395 |
| **Cenário 2** | | | | |
| $\Delta_{\text{MSE}}$ (em dB) | $N_{\text{iter}}^{\Delta}$ (simul.) | $N_{\text{iter}}^{\Delta}$ (teór.) | $R_{\text{iter}}^{\Delta}$ (simul.) | $R_{\text{iter}}^{\Delta}$ (teór.) |
| 5 | 8093 | 8154 | 0,00061782 | 0,0006132 |
| 10 | 14689 | 15600 | 0,00068078 | 0,00064103 |
| 15 | 19922 | 20783 | 0,00075294 | 0,00072174 |
| 20 | 23257 | 24123 | 0,00085996 | 0,00082908 |
| **Cenário 3** | | | | |
| $\Delta_{\text{MSE}}$ (em dB) | $N_{\text{iter}}^{\Delta}$ (simul.) | $N_{\text{iter}}^{\Delta}$ (teór.) | $R_{\text{iter}}^{\Delta}$ (simul.) | $R_{\text{iter}}^{\Delta}$ (teór.) |
| 5 | 10354 | 10489 | 0,00048291 | 0,00047669 |
| 10 | 18329 | 18723 | 0,00054558 | 0,0005341 |
| 15 | 25158 | 25880 | 0,00059623 | 0,0005796 |
| 20 | 31393 | 32451 | 0,00063708 | 0,00061631 |
| **Cenário 4** | | | | |
| $\Delta_{\text{MSE}}$ (em dB) | $N_{\text{iter}}^{\Delta}$ (simul.) | $N_{\text{iter}}^{\Delta}$ (teór.) | $R_{\text{iter}}^{\Delta}$ (simul.) | $R_{\text{iter}}^{\Delta}$ (teór.) |
| 5 | 3388 | 4019 | 0,0014758 | 0,0012441 |
| 10 | 6990 | 7486 | 0,0014306 | 0,0013358 |
| 15 | 9840 | 10133 | 0,0015244 | 0,0014803 |
| 20 | 11436 | 11899 | 0,0017489 | 0,0016808 |

Tabela 13 – Comparação da estimativa da taxa de convergência com relação ao valor simulado. $N_{\text{iter}}^{\Delta}$: número de iterações necessárias (pelo modelo teórico e por simulações) para que haja uma queda de $\Delta_{\text{MSE}}$ dB em relação ao MSE inicial. $R_{\text{iter}}^{\Delta}$ (em dB/iteração): valor estimado da razão entre $\Delta$ (queda do MSE, em dB, a partir do MSE inicial) e $N_{\text{iter}}^{\Delta}$. Observe que $R_{\text{iter}}^{\Delta}$ relaciona-se com a taxa de convergência do algoritmo.

percentuais) concentram-se no Cenário 4, justamente o mais desafiador, por ter um sinal de entrada colorido e ter o maior comprimento da planta desconhecida a identificar. Cabe entretanto ressaltar que a maior discrepância entre o MSE teórico e a curva simulada do MSE teórico ocorreu no Cenário 2, o que ilustra que uma discrepância maior neste quesito não se traduz necessariamente numa estimativa menos acurada da taxa de convergência.

### 6.3- Comentários finais

Neste capítulo, uma análise de transiente inédita do algoritmo SM-sign-NLMS foi empreendida, mobilizando o chamado *modelo de Slock*. Este modelo admite uma estrutura estocástica particular para os sucessivos vetores de entrada $\boldsymbol{x}(k)$, os quais apresentam uma distribuição radial contínua estatisticamente independente da sua distribuição angular a qual, assume-se, é discreta. O modelo proposto estabele que as variáveis de estado $\tilde{\lambda}_i(k)$ evoluem consoante a Equação (173), as quais permitem estimar a evolução do MSE por meio da Equação (166). As condições que garantem a equivalência entre o modelo analítico proposto e o modelo de regime permanente oferecido no Capítulo 5 são matematicamente elucidadas. As predições teóricas foram comparadas com curvas simuladas, obtidas por ensaios independentes de Monte Carlo. O modelo teórico proposto foi capaz de estimar a taxa de convergência do algoritmo, com um erro percentual máximo (em dB/iteração) de 15,70%. As contribuições deste capítulo também fazem parte da publicação [138]. No próximo capítulo, o modelo estocástico proposto para o comportamento de transiente do algoritmo SM-sign-NLMS será explicitamente empregado para a derivação de uma versão inédita do algoritmo SM-sign-NLMS com passos de aprendizagem variantes no tempo, o que é importante para lidar de maneira efetiva com o compromisso entre taxa de convergência e desempenho assintótico.

# 7- Algoritmo SM-Sign-NLMS com Passo de Aprendizagem Variável

Este capítulo contempla a derivação de uma versão inédita do algoritmo SM-sign-NLMS que utiliza um passo de aprendizagem variante no tempo, de modo a lidar de maneira efetiva com o compromisso entre um MSE assintótico pequeno e uma alta taxa de convergência. Em termos gerais, opta-se por passos de aprendizagem mais elevados no início do processo de aprendizagem (para garantir uma taxa de convergência maior), os quais vão se reduzindo ao longo das iterações para que, reduzindo a variância dos pesos estimados, melhorem o desempenho assintótico do algoritmo.

Em termos estatísticos, a escolha de um fator de aprendizagem ótimo requer que se tenham disponíveis informações (ou seja, momentos estatísticos) que na prática são impossíveis de obter nas aplicações em geral. Assim, os algoritmos que utilizam fatores de aprendizado variantes no tempo utilizam aproximações instantâneas para os momentos estatísticos de interesse para a determinação destes passos ao longo das iterações, de modo a obter uma solução sub-ótima com desempenho superior ao das versões que utilizam um parâmetro $\beta$ fixo [165].

Este capítulo encontra-se estruturado do seguinte modo: a Seção 7.1 descreve a ideia básica subjacente a algoritmos adaptativos com passos de aprendizagem variantes no tempo, enquanto que a versão VSS (do inglês *variant step size*) proposta nesta tese para o algoritmo SM-sign-NLMS é descrita na Seção 7.2. As sequências de passo de aprendizagem só foram obtidas mediante o aporte do modelo teórico descrito no Capítulo 6. O desempenho do algoritmo proposto (intitulado VSS-sign-NLMS) é avaliado na Seção 7.3, por meio de comparações com a versão padrão do algoritmo, com passo de aprendizagem fixo ao longo das iterações. Por fim, a Seção 7.4 contém os comentários finais deste capítulo. Importa notar que as contribuições deste capítulo também constam no artigo [138], diretamente egresso desta tese.

## 7.1- Algoritmos com passos de aprendizagem variantes no tempo

Algoritmos do tipo *Variable Step Size* (VSS) classicamente oferecem uma melhora significativa do aprendizado de algoritmos de filtragem adaptativa, caso comparados com implementações nas quais $\beta$ (também chamado de "constante de *feedback*" [176]) assume valores fixos ao longo das iterações [177]. As variantes VSS também apresentam vantagens quando as estatísticas do sinal de entrada são não estacionárias, porque possibilitam um rastreamento do passo de aprendizagem ótimo, mediante um modesto aumento da complexidade computacional [178]. Ademais, essas variantes por vezes são insensíveis a eventuais disparidades dos autovalores da matriz de autocorrelação do vetor de entrada [176]. Tais variantes são uma extensão de ideias anteriores na área de aproximação estocástica para variar o passo de aprendizagem no método *steepest descent* [176].

O tempo de convergência de um algoritmo de filtragem adaptativa é inversamente proporcional a $\beta$. Em cenários não estacionários, por conseguinte, é importante adotar valores elevados de $\beta$; infelizmente, tal configuração leva a uma perda de desempenho [176].

Ainda que modelos puramente teóricos, que assumem conhecimento perfeito das estatísticas conjuntas dos sinais de entrada e de referência, possam fornecer uma sequência ótima de fatores de aprendizagem para todas as iterações [179], na prática um projetista de um algoritmo VSS deve adotar escolhas judiciosas (as quais demandam grande engenhosidade), pois um algoritmo VSS deve *i*) utilizar apenas informações disponíveis em tempo real (ou seja, não se assume informações estatísticas *a priori*, já que uma das principais vantagens dos algoritmos reside em não demandarem tal tipo de conhecimento); *ii*) apresentar baixa complexidade computacional e *iii*) demandar o ajuste de poucos hiperparâmetros adicionais (ou, caso tais hiperparâmetros adicionais sejam necessários, apresentar um desempenho pouco sensível aos valores de seus hiperparâmetros). Por meio de uma combinação de aproximações específicas, o mecanismo de VSS proposto neste capítulo apresenta todas estas vantagens.

Um das primeiras (e mais populares) variantes VSS do algoritmo LMS foi proposta em [180]. O algoritmo resultante ajusta o passo de aprendizagem na $k$-ésima iteração

$\beta(k)$ por meio da seguinte equação:

$$\beta'(k+1) = \vartheta_1 \beta(k) + \vartheta_2 e^2(k), \tag{179}$$

na qual

$$0 < \vartheta_1 < 1, \vartheta_2 > 0 \tag{180}$$

com o termo $\beta'(k+1)$ sendo objeto de limitações (mínima e máxima), as quais podem ser formalizadas como

$$\beta(k+1) = \begin{cases} \beta_{\text{max}}, & \text{se } \beta'(k+1) > \beta_{\text{max}} \\ \beta_{\text{min}}, & \text{se } \beta'(k+1) < \beta_{\text{min}} \\ \beta'(k+1), & \text{caso contrário.} \end{cases}, \tag{181}$$

onde $\beta_{\text{max}}$ costumeiramente é um valor próximo ao passo de aprendizagem utilizado na configuração na qual o algoritmo não é do tipo VSS, sendo escolhido de modo a garantir que o algoritmo convirja [181]. A Equação (179) mostra que a escolha do passo de aprendizagem depende de uma recursão não linear, um padrão comum em mecanismos de VSS [182].

Cabe ressaltar uma grande desvantagem do algoritmo descrito pelas Equações (179)-(181): a necessidade de ajustar um grande número de hiperparâmetros ($\vartheta_1$, $\vartheta_2$, $\beta_{\text{min}}$ e $\beta_{\text{max}}$).

Como outros exemplos de algoritmos dotados de mecanismos de VSS que exigem o ajuste de um número excessivo de parâmetros adicionais são o proposto em [183]:

$$\beta(k) = \vartheta_3 \left\{ \frac{3}{1 + \exp\left[-\vartheta_4 |\arctan(e(k))|\right]^{\vartheta_5}} - 1 \right\}, \tag{182}$$

onde $\vartheta_3$, $\vartheta_4$ e $\vartheta_5$ são parâmetros adicionais, e o algoritmo proposto em [184], no qual $\beta(k)$ é escolhido de acordo com:

$$\beta(k) = \frac{\vartheta_6 \arctan(e^{\vartheta_7}(k))}{\arctan[e^{\vartheta_7}(k)] + \vartheta_8}, \tag{183}$$

cujos parâmetros adicionais são $\vartheta_6$, $\vartheta_7$ e $\vartheta_8$. A determinação dos valores destes novos parâmetros é difícil, com seleções pouco adequadas eliminando o benefício potencial do mecanismo de VSS [185].

Observe em (179) a dependência de $\beta$ com relação ao erro *a priori* (uma dependência bastante comum na literatura; veja, por exemplo, [174]). O erro *a priori* é naturalmente computado ao longo do processo adaptativo, sendo por isso uma grandeza facilmente disponível para dar suporte ao procedimento de variação do passo de aprendizagem. Ademais, pode-se observar que os fatores $\vartheta_1$ e $\vartheta_2$ em (179) servem como fatores de memória relativos ao último passo e ao último erro, respectivamente [181]. Um valor típico para $\vartheta_2$ é $7,65 \times 10^{-4}$ [180], o que implica que apenas uma pequena fração do quadrado do erro é levada em consideração [181].

### 7.2- Algoritmo proposto: VSS-SM-sign-NLMS

Nesta seção, iremos utilizar a análise de transiente apresentada no Capítulo 6 para obter uma versão VSS do algoritmo proposto SM-sign-NLMS. A métrica de desempenho que utilizaremos para efeitos de otimização do passo de aprendizagem será o MSE. Tal escolha é motivada pelo fato de que temos facilmente acesso na prática a uma estimativa estocástica do erro quadrático, por meio da informação residente em $e^2(k)$.

Pela Equação (166), vemos que o MSE é minimizado quando as variáveis de estado $\tilde{\lambda}_i(k)$ são reduzidas ao máximo. Admitindo uma estratégia VSS, considere um passo de aprendizagem variante no tempo $\beta(k)$ em (172). A minimização de $\tilde{\lambda}_i(k+1)$ pode ser imposta via

$$\begin{aligned}\frac{\partial\tilde{\lambda}_i(k+1)}{\partial\beta(k)} &= -2\sqrt{\frac{2}{\pi\xi(k)}}\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]\frac{\lambda_i}{\mathsf{Tr}[\mathbf{R}]}\tilde{\lambda}_i(k) \\ &\quad + 2\beta(k)\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi(k)}}\right)\frac{\lambda_i}{\mathbb{E}\left[r^4\right]} = 0. \end{aligned} \tag{184}$$

Logo, de modo a maximizar a taxa de convergência do algoritmo, a seguinte escolha teórica poderia ser efetuada:

$$\beta_{\mathsf{ótimo}}(k) = \frac{\mathbb{E}[r^4]}{\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi(k)}}\right)}\sqrt{\frac{2}{\pi\xi(k)}}\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]\frac{\tilde{\lambda}_i(k)}{\mathsf{Tr}[\mathbf{R}]}. \tag{185}$$

Desafortunadamente, a Equação (185) não é uma escolha factível na prática, já

que $\tilde{\lambda}_i(k)$ e $\text{Tr}[\mathbf{R}]$ não são observáveis. Logo, algumas simplificações devem ser adotadas para que tenhamos um algoritmo prático. Idealmente, tais simplificações devem gerar uma estimativa suficientemente próxima de $\beta_{\text{ótimo}}(k)$ (descrito na Equação (185)), de modo a promover a eficácia do algoritmo resultante.

Como primeira hipótese simplificadora, assumiremos que o sinal de entrada é branco. A referência [180] argumenta que tal hipótese pode ser justificada em identificação de sistemas lineares porque sinais de entrada brancos são ótimos, devendo ser utilizados sempre que possível; ademais, um sinal de entrada branco cria um limite potencial para o desempenho do algoritmo em transiente. Assumindo tal hipótese, podemos escrever:

$$\tilde{\lambda}_i(k) \approx \sigma_x^2 \sigma_{\tilde{w}}^2 \approx \frac{\xi - \sigma_\nu^2}{N}, \tag{186}$$

onde $\sigma_{\tilde{w}}^2$ é a variância dos desvios $\tilde{w}_i^2(k)$, suposta constante ao longo dos coeficientes adaptativos. Esta hipótese é necessária, já que na prática estimar a variância de cada coeficiente adaptativo aumentaria em demasia o custo computacional do algoritmo.

Combinando (186) com a aproximação $\mathbb{E}[r^4] \approx N^2 \sigma_x^2$ nos leva a:

$$\beta(k) = \sqrt{\frac{2}{\pi \xi(k)}} \frac{\exp\left[-\frac{\bar{\gamma}^2}{2\xi(k)}\right]}{\text{erfc}\left(\frac{\bar{\gamma}}{\sqrt{2\xi(k)}}\right)} \left[\xi(k) - \hat{\sigma}_\nu^2\right], \tag{187}$$

onde $\hat{\sigma}_\nu^2$ é uma estimativa de $\sigma_\nu^2$. Note que $\sigma_\nu^2$ já é automaticamente estimado na abordagem *Set-membership* (veja (116)) e $\xi(k)$ pode ser estimado em tempo real via

$$\hat{\xi}(k+1) = \overline{\lambda}\hat{\xi}(k) + (1 - \overline{\lambda})e^2(k), \tag{188}$$

onde

$$\overline{\lambda} = 1 - \frac{1}{KN}, \tag{189}$$

com $K \geq 2$. De fato, o algoritmo com passo de aprendizagem variante no tempo é bastante insensível à escolha específica do parâmetro $K$, de modo que imporemos $K = 2$ para evitar o ajuste de um parâmetro adicional [186]. O algoritmo proposto (chamado de VSS-SM-sign-NLMS) é descrito no Algoritmo 3, o qual utiliza somente informações disponíveis na prática para o ajuste do passo de aprendizagem $\beta(k)$. Ademais, o algoritmo proposto propositalmente não exige o ajuste de parâmetros adicionais (já que impomos a condição $\overline{\lambda} = 1 - \frac{1}{2N}$). Ademais, os algoritmos do tipo SM já demandam uma estimativa

da variância do ruído $\sigma_\nu^2$. Tal variância pode até mesmo ser conhecida *a priori* em alguns casos, quando por exemplo o ruído deriva de um erro de quantização, o qual é dependente do número de bits utilizado (que é objeto de escolha do projetista). Por estas razões, não consideramos a estimativa $\hat{\sigma}_\nu^2$ um parâmetro adicional a ajustar no algoritmo proposto.

Algoritmo 3 – Algoritmo VSS-SM-sign-NLMS proposto.

**Entrada:** $\overline{\gamma}$, $\beta$, $\hat{\sigma}_\nu^2$,$x(k)$,$d(k)$,$\overline{\lambda}$
**Saída:** $\boldsymbol{w}(k)$
**Inicialização:** $\boldsymbol{w}(0) = \mathbf{0}$, $\hat{\xi}(0) = 0$,$\overline{\lambda} = 1 - \frac{1}{2N}$

**para** $k = 1, 2, \ldots$ **faça**
- $e(k) = d(k) - \boldsymbol{w}^T(k)\boldsymbol{x}(k)$ ▷ Calcula o erro
- $\hat{\xi}(k) = \overline{\lambda}\hat{\xi}(k-1) + (1-\overline{\lambda})e^2(k)$ ▷ Atualiza a estimativa $\hat{\xi}(k)$
- $\beta(k) = \sqrt{\frac{2}{\pi\hat{\xi}(k)}} \frac{\exp\left[-\frac{\overline{\gamma}^2}{2\hat{\xi}(k)}\right]}{\operatorname{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\hat{\xi}(k)}}\right)} \left[\hat{\xi}(k) - \hat{\sigma}_\nu^2\right]$ ▷ Calcula $\beta(k)$
- Calcule $f[e(k)]$ de acordo com a Equação (130)
- $\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta(k)\frac{\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2} f[e(k)]$ ▷ Atualize a estimativa do filtro

### 7.3- Resultados

Tipicamente, a avaliação de algoritmos do tipo VSS ocorre por meio de comparações com as versões dos algoritmos com um passo de aprendizagem fixo [180]. Assim, iremos avaliar a proposta VSS por meio de comparações com o algoritmo SM-sign-NLMS com $\beta$ constante. As comparações foram efetuadas ao longo de 5 cenários, especificados na Tabela 14. Os valores de passo aprendizagem (para os algoritmo padrão, no qual este é mantido constante ao longo das iterações) testados foram $\beta \in \{0,1; 0,5; 1\}$. Em todos os cenários, foram efetuados 1000 ensaios independentes de Monte Carlo. Os algoritmos foram executados por 1,5 $\times 10^4$ iterações. O ruído aditivo é gaussiano, branco, de média zero e variância $\sigma_\nu^2$. Para avaliar o impacto de uma estimativa imprecisa da variância do ruído, assumimos que o algoritmo utiliza o valor $\hat{\sigma}_\nu^2$ para a variância do ruído, sendo que $\hat{\sigma}_\nu^2$ pode ser igual (no caso de uma estimativa perfeita) ou diferente (no caso de uma estimativa imprecisa) à variância real do ruído $\sigma_\nu^2$. O sinal de entrada é obtido filtrando-se um sinal branco, gaussiano, de média zero e variância unitária pelo filtro $\frac{B(z)}{A(z)}$. A Tabela 15 apresenta os filtros de colorimentos para os cenários considerados.

| #Cenário | #Modelo | $N$ | $\tau$ | $\sigma_\nu^2$ | $\hat{\sigma}_\nu^2$ |
|---|---|---|---|---|---|
| 1 | 3 | 96 | 3 | $10^{-6}$ | $10^{-6}$ |
| 2 | 3 | 96 | 3 | $10^{-6}$ | $2 \times 10^{-6}$ |
| 3 | 4 | 128 | 3 | $10^{-6}$ | $3 \times 10^{-6}$ |
| 4 | 5 | 96 | 5 | $10^{-5}$ | $10^{-5}$ |
| 5 | 5 | 96 | 5 | $10^{-4}$ | $10^{-4}$ |

Tabela 14 – Especificações dos cenários testados. A coluna #Modelo reflete o número da função de transferência estimada por medições, constante no relatório da ITU [122].

| #Cenário | $B(z)$ | $A(z)$ |
|---|---|---|
| 1 | 1 | $1 - 0,75z^{-1} - 0,19z^{-2} - 0,09z^{-3} + 0,5z^{-4}$ |
| 2 | 1 | $1 - 0,75z^{-1} - 0,19z^{-2} - 0,09z^{-3} + 0,5z^{-4}$ |
| 3 | 1 | $1 - 0,75z^{-1} - 0,19z^{-2}$ |
| 4 | 1 | $1 - 0,75z^{-1}$ |
| 5 | 1 | $1$ |

Tabela 15 – Características do filtro de colorimento do sinal de entrada para os Cenários 1-5.

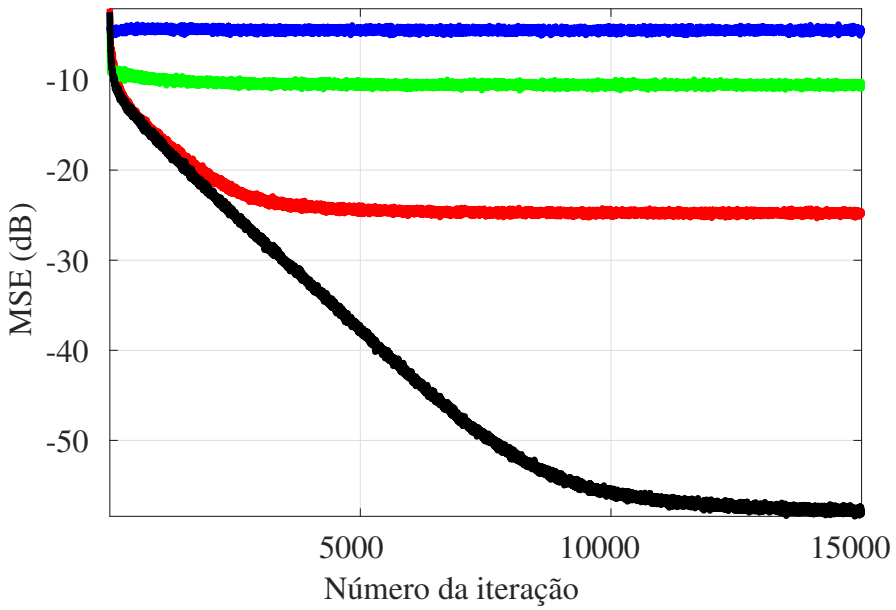



Figura 38 – Evolução do MSE (em dB) para o Cenário 1. Em vermelho: algoritmo SM-sign-NLMS com $\beta = 0,1$. Em verde: algoritmo SM-sign-NLMS com $\beta = 0,5$. Em azul: algoritmo SM-sign-NLMS com $\beta = 1$. Em preto: algoritmo proposto VSS-SM-sign-NLMS.

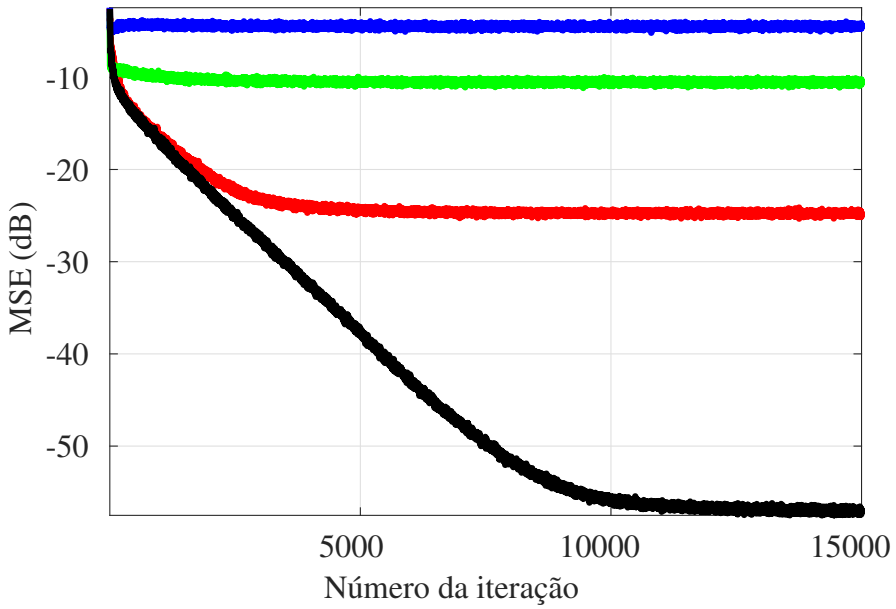



Figura 39 – Evolução do MSE (em dB) para o Cenário 2. Em vermelho: algoritmo SM-sign-NLMS com $\beta = 0,1$. Em verde: algoritmo SM-sign-NLMS com $\beta = 0,5$. Em azul: algoritmo SM-sign-NLMS com $\beta = 1$. Em preto: algoritmo proposto VSS-SM-sign-NLMS.

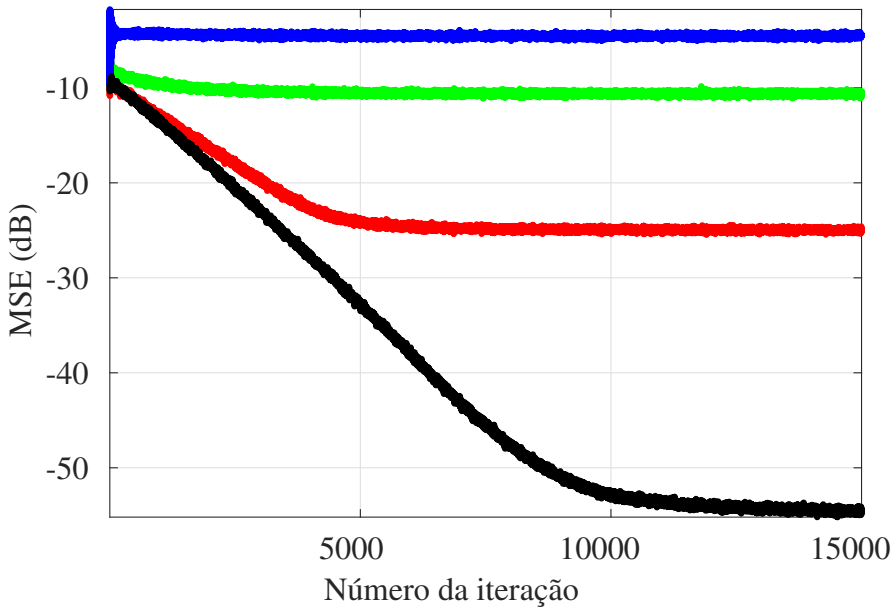



Figura 40 – Evolução do MSE (em dB) para o Cenário 3. Em vermelho: algoritmo SM-sign-NLMS com $\beta = 0,1$. Em verde: algoritmo SM-sign-NLMS com $\beta = 0,5$. Em azul: algoritmo SM-sign-NLMS com $\beta = 1$. Em preto: algoritmo proposto VSS-SM-sign-NLMS.

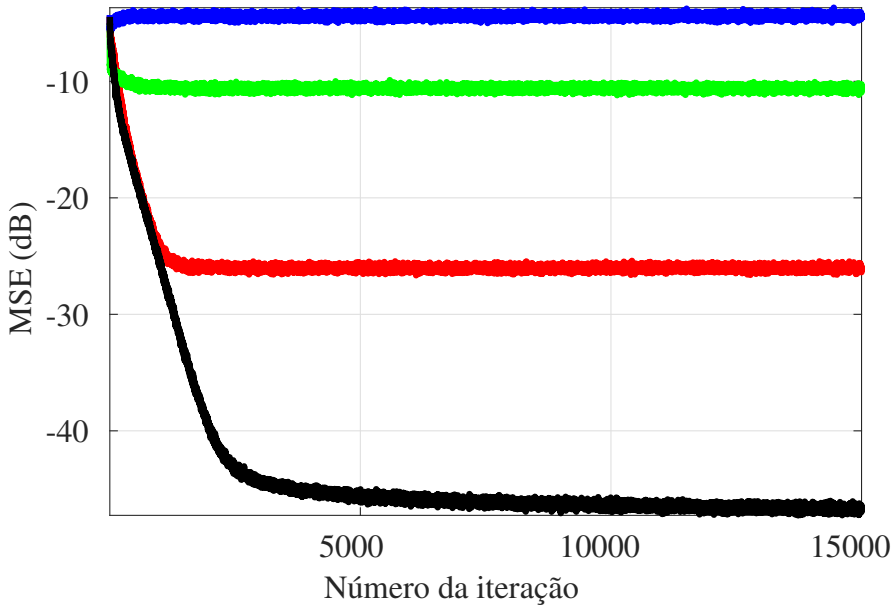



Figura 41 – Evolução do MSE (em dB) para o Cenário 4. Em vermelho: algoritmo SM-sign-NLMS com $\beta = 0,1$. Em verde: algoritmo SM-sign-NLMS com $\beta = 0,5$. Em azul: algoritmo SM-sign-NLMS com $\beta = 1$. Em preto: algoritmo proposto VSS-SM-sign-NLMS.

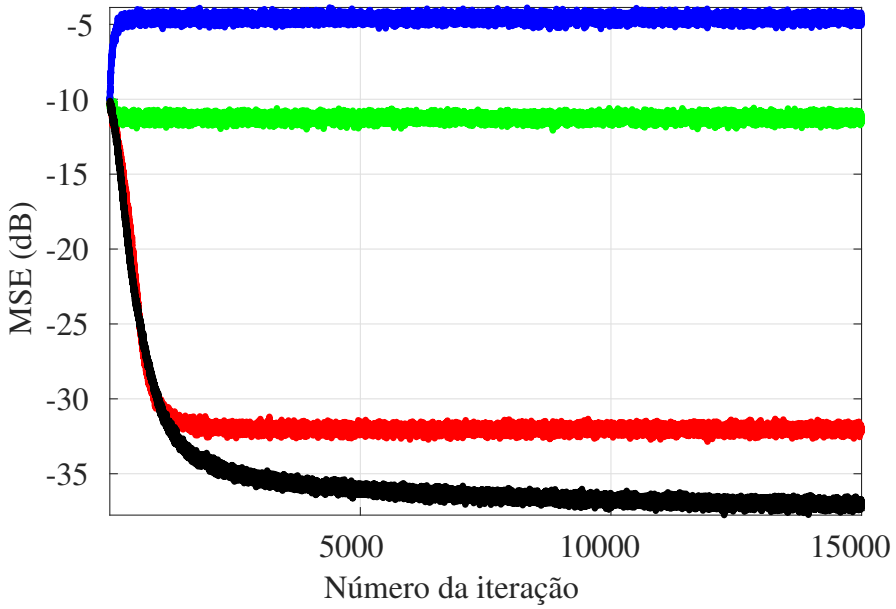



Figura 42 – Evolução do MSE (em dB) para o Cenário 5. Em vermelho: algoritmo SM-sign-NLMS com $\beta = 0,1$. Em verde: algoritmo SM-sign-NLMS com $\beta = 0,5$. Em azul: algoritmo SM-sign-NLMS com $\beta = 1$. Em preto: algoritmo proposto VSS-SM-sign-NLMS.

| #Cenário | $\beta = 1$ | $\beta = 0,5$ | $\beta = 0,1$ | Algoritmo VSS |
|---|---|---|---|---|
| 1 | -4,47 | -10,53 | -24,75 | -57,73 |
| 2 | -4,47 | -10,53 | -24,75 | -56,99 |
| 3 | -4,53 | -10,62 | -24,96 | -54,51 |
| 4 | -4,40 | -10,60 | -26,01 | -46,67 |
| 5 | -4,58 | -11,24 | -32,03 | -37,04 |

Tabela 16 – MSE em regime permanente (em dB) do algoritmo SM-sign-NLMS com $\beta$ constante e do algoritmo proposto VSS-SM-sign-NLMS. A estimativa do MSE assintótico foi efetuada por meio da média das últimas 1000 iterações.

As Figuras 38-42 apresentam a evolução do MSE para os 5 cenários sob consideração. Observa-se que o algoritmo com $\beta$ fixo apresenta um claro compromisso entre taxa de convergência e desempenho assintótico, com este último melhorando à medida que $\beta$ se reduz, às custas de uma menor taxa de convergência. Cabe ressaltar que a escolha $\beta = 1$ implica um valor alto de $\beta$, o qual implica uma alta variabilidade dos coeficientes, de modo que o algoritmo efetivamente não aprende, embora não se tenha observado divergência (uma situação catastrófica) em nenhum dos cenários. Em todos os cenários, o algoritmo proposto apresenta uma taxa de convergência superior ou, em alguns poucos casos, equivalente nas primeiras iterações (comparando-se com o algoritmo tradicional com o valor de $\beta = 0,1$). De todo modo, o algoritmo VSS proposto supera claramente a versão tradicional do algoritmo, em todos os cenários, em especial em regime permanente. A Tabela 16 quantifica esta conclusão. A partir dela, por exemplo, podemos concluir que no Cenário 1 o algoritmo VSS proposto supera a melhor versão tradicional (com $\beta = 0,1$) em 32,98 dB (em termos de MSE). Também cabe observar que uma estimativa pouco acurada da variância do ruído não tem impacto significativo no desempenho da versão VSS, o que é algo relevante, dado que na prática tal estimativa é desafiadora.

Por sua vez, as Figuras 43-47 apresentam a evolução da média (computada nos ensaios de Monte Carlo) de $\beta$ ao longo das iterações, para os 5 cenários considerados. Nas primeiras iterações, observa-se uma oscilação do passo de aprendizagem. Tal oscilação se deve ao fato de serem exigidas algumas iterações para que a estimativa $\hat{\xi}(k)$ se estabilize. Desconsiderando tal oscilação (que é bastante localizada, nas primeiras iterações), observamos que o passo de aprendizagem se comporta exatamente como esperado: valores altos nas primeiras iterações (para maximizar a taxa de convergência), com uma convergência assintótica para valores muito baixos (de modo a reduzir a variabilidade dos coeficientes e melhorar o desempenho assintótico). Observa-se que

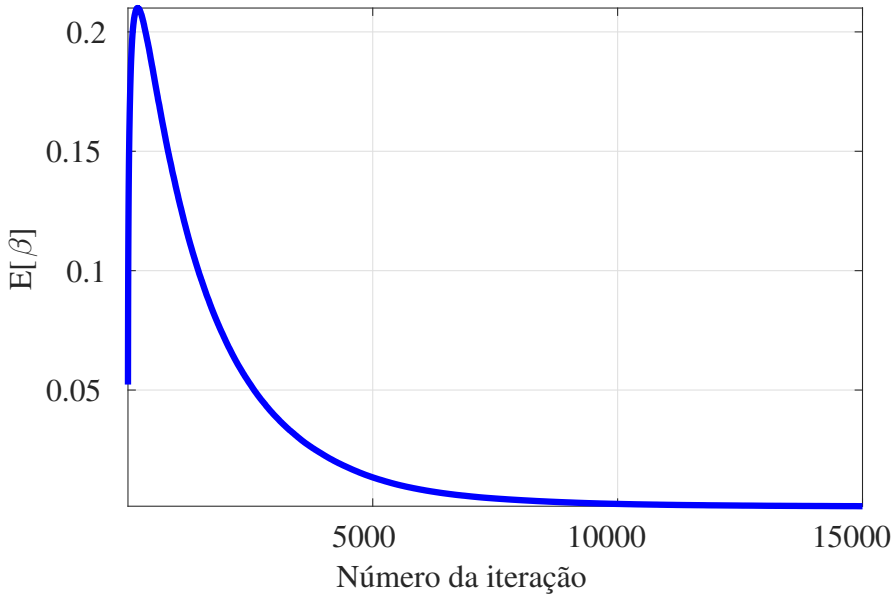



Figura 43 – Evolução do valor médio do passo de aprendizagem $\beta(k)$ para o algoritmo VSS-SM-sign-NLMS, no Cenário 1.

nos Cenários 4 e 5, pelo “fator de decaimento” observa-se haver uma convergência mais rápida para valores pequenos de $\beta$ nos Cenários 4 e 5. Estes cenários são justamente os que apresentam os sinais de entrada menos coloridos (no caso do Cenário 5, o sinal de entrada é branco, inclusive), o que implica uma maior taxa de convergência no aprendizado. Com uma maior taxa de convergência, o algoritmo pode reduzir mais rapidamente o passo de aprendizagem.

## 7.4- Comentários Finais

Neste capítulo, a análise de transiente empreendida no Capítulo 6 foi empregada para a obtenção de um algoritmo prático e dotado de um mecanismo de VSS capaz de conjuminar uma alta taxa de convergência com um bom desempenho assintótico. O algoritmo resultante utiliza expressamente as características de atualização do algoritmo SM-sign-NLMS (proposto no Capítulo 5), sem exigir o ajuste de um número adicional de parâmetros (o que é uma restrição bastante desafiadora no projeto de mecanismos de VSS). O algoritmo proposto, chamado de VSS-SM-sign-NLMS, claramente supera a versão do algoritmo com fator de aprendizagem fixo.

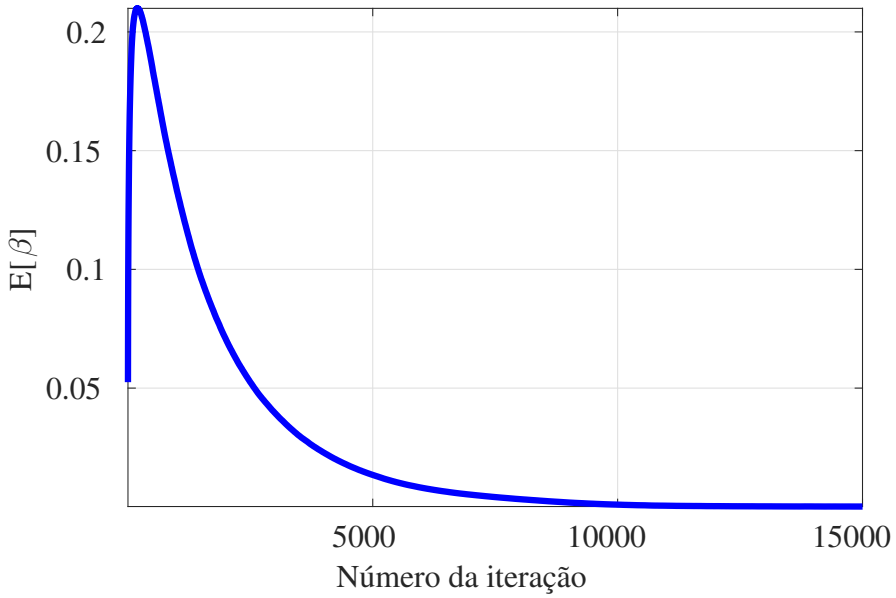



Figura 44 – Evolução do valor médio do passo de aprendizagem $\beta(k)$ para o algoritmo VSS-SM-sign-NLMS, no Cenário 2.

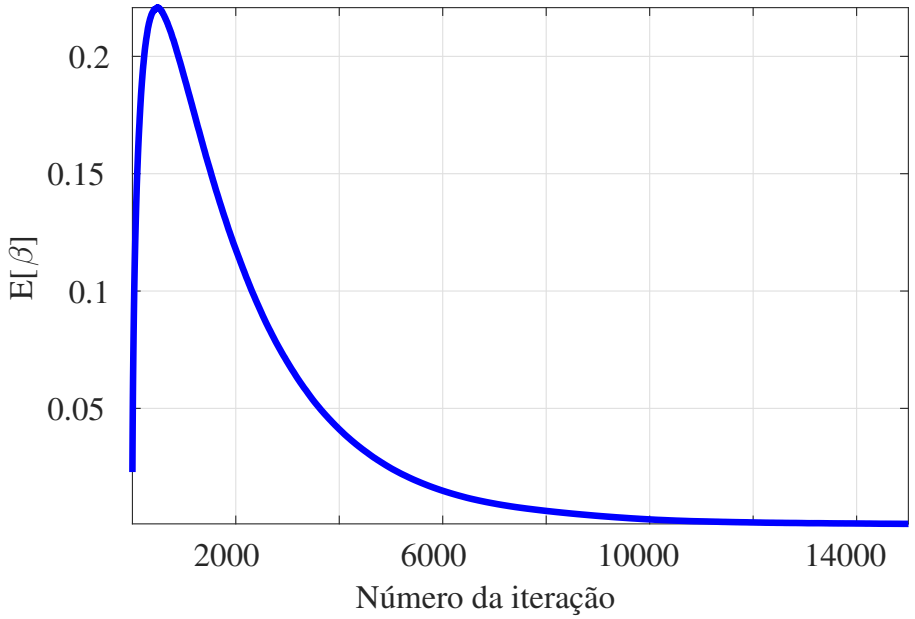



Figura 45 – Evolução do valor médio do passo de aprendizagem $\beta(k)$ para o algoritmo VSS-SM-sign-NLMS, no Cenário 3.

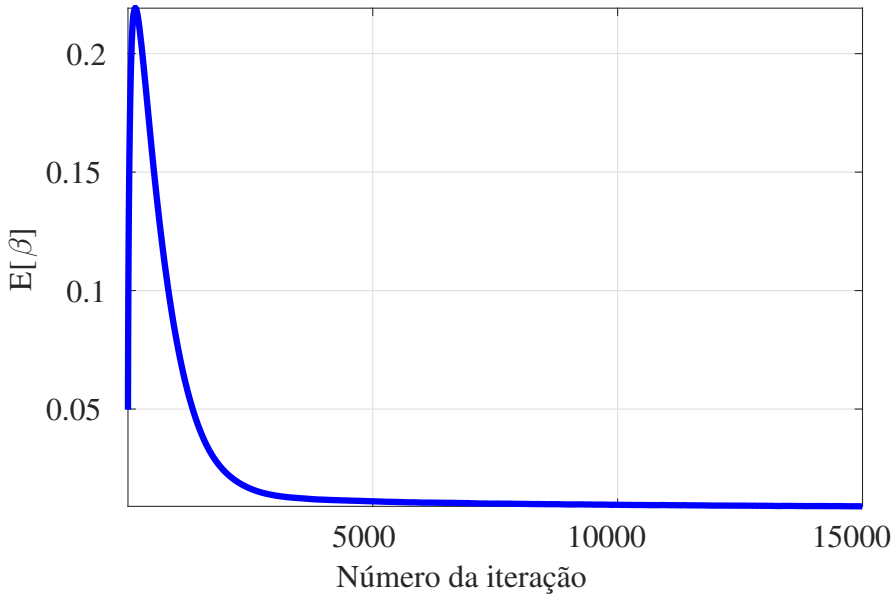



Figura 46 – Evolução do valor médio do passo de aprendizagem $\beta(k)$ para o algoritmo VSS-SM-sign-NLMS, no Cenário 4.

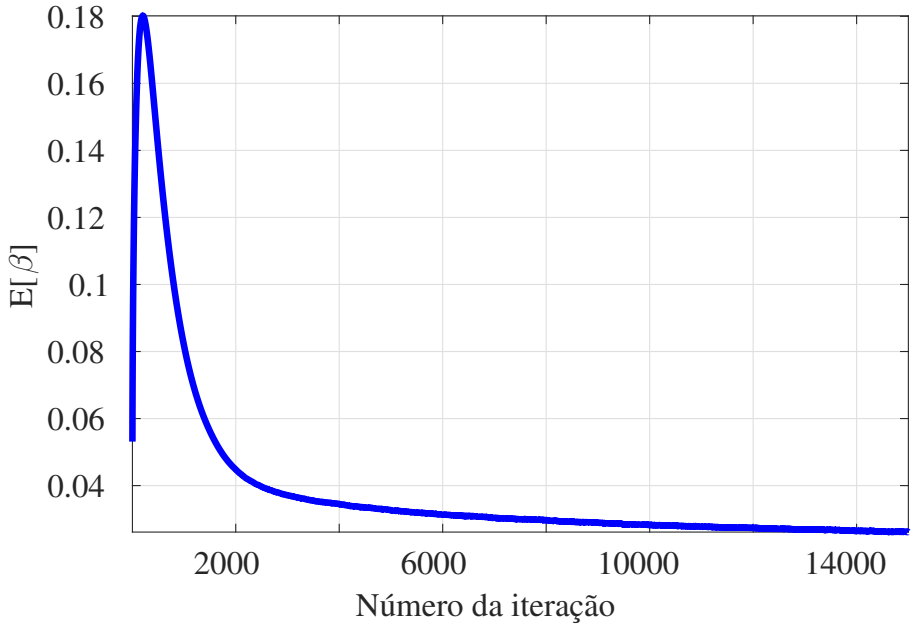



Figura 47 – Evolução do valor médio do passo de aprendizagem $\beta(k)$ para o algoritmo VSS-SM-sign-NLMS, no Cenário 5.

O algoritmo SM-sign-NLMS foi objeto de modelagens estocásticas nos Capítulos 5 e 6. Estas modelagens são importantes para descrever o comportamento de aprendizagem do algoritmo, bem como oferecer garantias teóricas e mesmo a sua extensão para versões VSS (o que foi efetuado neste capítulo). No próximo capítulo, novas análises teóricas são propostas, focadas em prever o desempenho do algoritmo em dois cenários comuns na prática: *i*) quando a planta a identificar é variante no tempo; e *ii*) quando o comprimento do filtro é insuficiente.

# 8- Análise de Rastreamento e Comprimento Deficiente do Algoritmo SM-Sign-NLMS

## 8.1- Introdução

Os Capítulos 5, 6 e 7 analisaram os casos nos quais os filtros adaptativos encontram-se num ambiente estacionário, com o comprimento do filtro adaptativo coincidindo com o tamanho da resposta ao impulso da planta a emular. Na prática, entretanto, os filtros podem operar em contextos mais sofisticados ou mais desafiadores. O primeiro destes contextos ocorre quando a planta a estimar é variante no tempo; justamente, uma das principais motivações para a adoção de filtros adaptativos reside no fato de que tais plantas variam ao longo do tempo, o que excluiria a adoção de técnicas *offline* que levariam em conta simultaneamente todos os dados disponíveis. Outro cenário que pode ocorrer na prática é o de um filtro adaptativo com *comprimento deficiente*, no qual o filtro opera com um número subótimo de coeficientes. Tal configuração pode ocorrer devido ao desconhecimento do comprimento ótimo, ou mesmo devido a restrições de complexidade computacional.

Este capítulo apresenta análises estocásticas inéditas (as quais também constituem parte da publicação apresentada em [138]) que modelam o comportamento de aprendizagem do algoritmo SM-sign-NLMS nos casos de uma planta variante no tempo (Seção 8.2) e de um filtro cujo comprimento é deficiente (Seção 8.3). As modelagens analíticas são objeto de escrutínio na Seção 8.4, cuja metodologia de avaliação consiste na comparação com simulações que se valem de diversos ensaios independentes de Monte Carlo. Os comentários finais do capítulo constam na Seção 8.5.

## 8.2- Rastreamento

A habilidade de operar em uma configuração não estacionária é uma das propriedades mais desejáveis de algoritmos de filtragem adaptativa [172]. O comportamento de aprendizagem de um filtro adaptativo é mais sofisticado neste caso, já que plantas variantes no tempo provocam um "*lag*" no processo de aprendizagem adaptativo. Logo, as capacidades de rastreamento de um algoritmo são realçadas quando o passo de aprendizagem apresenta valores elevados, enquanto que a variância dos coeficientes do estimador adaptativo diminui quando $\beta$ assume valores pequenos. Isto implica que o desempenho assintótico é otimizado quando este compromisso é levado em consideração.

Para efeitos da modelagem pretendida, a estacionariedade da planta ideal é explicitamente perturbada mediante a introdução do seguinte modelo de caminhada aleatória estocástica multivariada de Markov:

$$\boldsymbol{w}^{\star}(k+1) = \boldsymbol{w}^{\star}(k) + \boldsymbol{q}(k), \tag{190}$$

onde $\boldsymbol{q}(k)$ é um vetor aleatório de média zero, considerado estatisticamente independente das demais variáveis aleatórias. Este vetor também apresenta a seguinte matriz de covariância:

$$\boldsymbol{Q} \triangleq \mathbb{E}\left[\boldsymbol{q}(k)\boldsymbol{q}^{T}(k)\right] = \sigma_q^2 \boldsymbol{I}_N. \tag{191}$$

*Comentário*: O modelo de caminhada aleatória é popular na área de filtragem adaptativa. Sua adequabilidade é discutida em maiores detalhes em [187]. Cabe notar que perturbações de primeira ordem (como a modelada na Equação (190)) são encontradas em aplicações tais como cancelamento de eco acústico, comunicações, controle e sistemas de transmissão [165], [172], de sorte que sua adoção neste capítulo não pode ser atribuída tão somente à sua conveniência analítica.

Sob a Equação (190), uma nova formulação pode ser encontrada para a Equação (134):

$$\tilde{\boldsymbol{w}}(k+1) = \tilde{\boldsymbol{w}}(k) - \beta \frac{\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2} f[e(k)] + \boldsymbol{q}(k). \tag{192}$$

De modo similar, a Equação (163) pode ser reescrita da seguinte forma:

$$\begin{aligned} \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k+1) &= \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) - \beta\mathbb{E}\left\{\frac{\tilde{\boldsymbol{w}}(k)\boldsymbol{x}^T(k)}{\|\boldsymbol{x}(k)\|^2} f[e(k)]\right\} + \boldsymbol{Q} \\ &\quad - \beta\mathbb{E}\left\{\frac{\boldsymbol{x}(k)\tilde{\boldsymbol{w}}^T(k)}{\|\boldsymbol{x}(k)\|^2} f[e(k)]\right\} \\ &\quad + \beta^2\mathbb{E}\left\{\frac{\boldsymbol{x}(k)\boldsymbol{x}^T(k)}{\|\boldsymbol{x}(k)\|^4} f^2[e(k)]\right\}. \end{aligned} \tag{193}$$

Logo, uma atualização é necessária para a Equação (194), quando a não estacionariedade markoviana é considerada:

$$\begin{aligned} \tilde{\lambda}_i(k+1) &= \tilde{\lambda}_i(k) + \beta^2(k)\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi(k)}}\right)\frac{\lambda_i}{\mathbb{E}\left[r^4\right]} + \sigma_q^2 \\ &\quad - 2\sqrt{\frac{2}{\pi\xi(k)}}\beta(k)\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]\frac{\lambda_i}{\mathsf{Tr}[\boldsymbol{R}]}\tilde{\lambda}_i(k). \end{aligned} \tag{194}$$

A partir de diversas manipulações matemáticas (similares àquelas que deram origem à Equação (174)), chega-se à seguinte equação de ponto fixo, capaz de descrever o comportamento assintótico do algoritmo na configuração não estacionária considerada:

$$\xi = \sigma_\nu^2 + \frac{\beta\sqrt{\pi\xi}}{2\sqrt{2}}\frac{\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right)}{\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}\frac{\{\mathsf{Tr}[\mathbf{R}]\}^2}{\mathbb{E}\left[r^4\right]} + \frac{N\sqrt{\pi\xi}}{2\sqrt{2}\beta}\frac{\mathsf{Tr}[\boldsymbol{R}]}{\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}\sigma_q^2, \tag{195}$$

onde o termo *aditivo* inédito $\frac{N\sqrt{\pi\xi}}{2\sqrt{2}\beta}\frac{\mathsf{Tr}[\boldsymbol{R}]}{\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}\sigma_q^2$ contém a contribuição analítica da perturbação markoviana $\boldsymbol{q}(k)$ para o desempenho assintótico do algoritmo.

*Comentários*: É intuitivo que o termo aditivo $\frac{N\sqrt{\pi\xi}}{2\sqrt{2}\beta}\frac{\mathsf{Tr}[\boldsymbol{R}]}{\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}\sigma_q^2$, obtido pelo modelo estocástico, seja proporcional à variância dos elementos da perturbação markoviana, já que esta variância de certo modo incorpora o grau de não estacionariedade do vetor ótimo a identificar. Adicionalmente, também a divisão por $\beta$ neste termo também é intuitivo, já que aumentar $\beta$ aumenta a capacidade de rastreamento do algoritmo (já que incrementa a taxa de convergência). A qualidade da previsão teórica resultante da solução numérica da equação de ponto fixo (195) será avaliada na Seção 8.4.

## 8.3- Análise de Comprimento Deficiente

Na prática, o tamanho da planta ideal é maior do que o comprimento do filtro adaptativo. Por vezes, o projetista intencionalmente adere a tal configuração deficiente (ou sub-ótima) para lidar com limitações computacionais [188] ou para aumentar a taxa de convergência [189]. Logo, é importante caracterizar o desempenho do algoritmo sob operação sub-ótima [190]. A seguir uma análise teórica inédita para este caso, referente ao algoritmo proposto SM-sign-NLMS, será descrita. A modelagem estocástica será dividida em duas partes: *i)* análise de primeira ordem (voltada para o comportamento médio dos coeficientes); e *ii)* análise de segunda ordem (voltada para estatísticas de segunda ordem do estimador adaptativo).

### 8.3.1 Comportamento de primeira ordem

Na configuração de comprimento deficiente, o sinal desejado pode ser descrito como

$$d(k) = [\boldsymbol{w}^{\star}]^T \boldsymbol{x}(k) + [\overline{\boldsymbol{w}}^{\star}]^T \overline{\boldsymbol{x}}(k) + \nu(k), \tag{196}$$

onde $\overline{\boldsymbol{w}}^{\star} \in \mathbb{R}^L$ e

$$\overline{\boldsymbol{x}}(k) \triangleq \Big[x(k-N) \quad x(k-N-1) \quad \ldots \quad x(k-N-L+1)\Big]^T, \tag{197}$$

onde o vetor ideal tem comprimento $N + L$ (com $N$ denotando o comprimento do filtro adaptativo, como nos capítulos pretéritos).

Nesta configuração, a recursão (134) é ainda válida quando o erro é escrito como

$$e(k) = \tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k) + [\overline{\boldsymbol{w}}^{\star}]^T \overline{\boldsymbol{x}}(k) + \nu(k). \tag{198}$$

Logo, usando as Equações (134) e (198), e aplicando o operador de valor esperado e a heurística de Price em conjunção com a hipótese de independência estatística do ruído nos leva a:

$$\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] = \mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right] - \beta\sqrt{\frac{2}{\pi\xi(k)}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]\boldsymbol{\mathcal{A}} \tag{199}$$

onde

$$\boldsymbol{\mathcal{A}} \triangleq \mathbb{E}\left\{\frac{\boldsymbol{x}(k)\boldsymbol{x}^T(k)}{\|\boldsymbol{x}(k)\|^2}\tilde{\boldsymbol{w}}(k) + \frac{\overline{\boldsymbol{x}}^T(k)\overline{\boldsymbol{w}}^{\star}\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2} + \frac{\boldsymbol{x}(k)\nu(k)}{\|\boldsymbol{x}(k)\|^2}\right\}. \tag{200}$$

Usando a aproximação $\mathbb{E}\left[\frac{1}{\|\boldsymbol{x}(k)\|^2}\right] \approx \frac{1}{N\sigma_x^2}$ dá origem a:

$$\begin{aligned}\mathbb{E}\left[\tilde{\boldsymbol{w}}(k+1)\right] &= \boldsymbol{\mathcal{B}}\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right] \\ &\quad - \beta\sqrt{\frac{2}{\pi\xi(k)}}\exp\left(-\frac{\overline{\gamma}^2}{2\xi(k)}\right)\boldsymbol{\vartheta},\end{aligned} \tag{201}$$

onde

$$\boldsymbol{\mathcal{B}} \triangleq \left(\boldsymbol{I} - \beta\sqrt{\frac{2}{\pi\xi(k)}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]\frac{\boldsymbol{R}}{N\sigma_x^2}\right), \tag{202}$$

e

$$\boldsymbol{\vartheta} \triangleq \frac{\mathbb{E}\left[\overline{\boldsymbol{x}}^T(k)\overline{\boldsymbol{w}}^{\star}\boldsymbol{x}(k)\right]}{N\sigma_x^2}. \tag{203}$$

Assumindo como entrada um sinal branco (e, por conseguinte, $\boldsymbol{\vartheta} = \boldsymbol{0}$), a Equação (202) implica que o algoritmo SM-sign-NLMS é não enviesado, no sentido de que o vetor de pesos adaptativo converge *na média* para os primeiros $N$ elementos da planta ideal:

$$\mathbb{E}\left[\tilde{\boldsymbol{w}}(\infty)\right]_{\text{white}} = \boldsymbol{0} \Rightarrow \mathbb{E}\left[\boldsymbol{w}(\infty)\right] = \boldsymbol{w}^{\star}. \tag{204}$$

O fato de o algoritmo adaptativo proposto implementar uma estimação não tendenciosa é importante. Tal previsão teórica será avaliada nos resultados deste capítulo (na Seção 8.4). Cabe ressaltar que o modelo teórico somente prevê tal ausência de tendenciosidade no caso de sinais de entrada brancos.

### 8.3.2 Análise de Segunda Ordem

O comportamento de segunda ordem do algoritmo é mais desafiador. Utilizando a Equação (198), temos:

$$\begin{aligned} e^2(k) &= \tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k) + 2\tilde{\boldsymbol{w}}^T(k)\boldsymbol{x}(k)\overline{\boldsymbol{x}}^T(k)\overline{\boldsymbol{w}}^\star \\ &\quad + \left[\overline{\boldsymbol{w}}^\star\right]^T \overline{\boldsymbol{x}}(k)\overline{\boldsymbol{x}}^T(k)\overline{\boldsymbol{w}}^\star + \nu^2(k) + \mathcal{O}[\nu(k)]. \end{aligned} \tag{205}$$

Após a aplicação do operador de valor esperado e das hipóteses estatísticas usuais, temos:

$$\mathbb{E}\left[e^2(k)\right] = \sigma_\nu^2 + \left[\overline{\boldsymbol{w}}^\star\right]^T \boldsymbol{R}_{N-M}\overline{\boldsymbol{w}}^\star + 2\boldsymbol{b}^T\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right] + \mathsf{Tr}\left\{\boldsymbol{R}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right\}, \tag{206}$$

onde

$$\boldsymbol{b} \triangleq \mathbb{E}\left\{\left[\overline{\boldsymbol{x}}^T(k)\overline{\boldsymbol{w}}^\star\right]\boldsymbol{x}(k)\right\}, \tag{207}$$

$$\boldsymbol{R}_{N-M} \triangleq \mathbb{E}\left[\overline{\boldsymbol{x}}(k)\overline{\boldsymbol{x}}^T(k)\right]. \tag{208}$$

Ao multiplicar a Equação (134) por sua transposta e aplicando o operador de valor esperado com as hipóteses estatísticas usuais nos leva a, após algumas manipulações matemáticas:

$$\begin{aligned} \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k+1) &= \boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) - g\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\boldsymbol{R} - g\boldsymbol{R}\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k) \\ &\quad - g\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right]\mathbb{E}\left\{\boldsymbol{x}^T(k)\left\{\left[\overline{\boldsymbol{w}}^\star\right]^T\overline{\boldsymbol{x}}(k)\right\}\right\} \\ &\quad + \frac{\beta^2}{\mathbb{E}\left[r^4\right]}\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi(k)}}\right), \end{aligned} \tag{209}$$

onde

$$g \triangleq \frac{\beta}{\mathbb{E}\left[\|\boldsymbol{x}(k)\|^2\right]}\sqrt{\frac{2}{\pi\xi(k)}}\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]. \tag{210}$$

Assumindo um sinal de entrada branco e utilizando a aproximação $\mathbb{E}\left[\|\boldsymbol{x}(k)\|^2\right] = N\sigma_x^2$ e

então aplicando o operador de traço na Equação (209) gera:

$$\theta(k+1) = \theta(k) - \frac{2\beta}{N}\sqrt{\frac{2}{\pi\xi(k)}}\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]\theta(k) + \frac{N\beta^2}{\mathbb{E}\left[r^4\right]}\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi(k)}}\right), \tag{211}$$

onde

$$\theta(k) \triangleq \mathsf{Tr}\left[\boldsymbol{R}_{\tilde{\boldsymbol{w}}}(k)\right], \tag{212}$$

é o desvio quadrático médio (MSD) do algoritmo.

Assumindo um sinal de entrada branco e gaussiano simplifica a recursão para:

$$\theta(k+1) = \theta(k) - \frac{2\beta}{N}\sqrt{\frac{2}{\pi\xi(k)}}\mathsf{exp}\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]\theta(k) + \frac{\beta^2\Gamma\left(\frac{N+2}{2}\right)}{2\sigma_x^4\Gamma\left(\frac{N+4}{2}\right)}\mathsf{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi(k)}}\right), \tag{213}$$

onde $\theta(k)$ e $\xi(k)$ são relacionados por:

$$\xi(k) = \sigma_\nu^2 + \sigma_x^2\|\overline{\boldsymbol{w}}^\star\|^2 + \sigma_x^2\theta(k). \tag{214}$$

As Equações (213) e (214) especificam um par *acoplado* de equações de ponto fixo. Isto implica que o MSD e o MSE assintóticos podem ser teoricamente estimados por meio de um refinamento iterativo de uma estimativa inicial de $\theta(k)$. A cada iteração do procedimento numérico de resolução das Equações (213) e (214), cada estimativa do MSD $\hat{\theta}(k)$ leva a uma nova estimativa do MSE $\hat{\xi}(k)$ por meio da Equação (214), com as estimativas atuais de $\hat{\xi}(k)$ e $\hat{\theta}(k)$ podendo ser mobilizadas para a atualização da $\hat{\theta}(k)$ por meio da Equação (213). Logo, as Equações (213) e (214) oferecem estimativas teóricas tanto para o MSD em regime permanente quanto para o MSE em regime permanente do algoritmo proposto SM-sign-NLMS. Note que $g$ apresenta um forte acoplamento estocástico com os vetores aleatórios $\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right]$ e $\mathbb{E}\left\{\boldsymbol{x}^T(k)\left\{\left[\overline{\boldsymbol{w}}^\star\right]^T\overline{\boldsymbol{x}}(k)\right\}\right\}$, o qual, por simplicidade, foi negligenciado na modelagem estocástica, para facilitar a determinação das equações analíticas. Esta hipótese de independência desafortunadamente restringe a análise para configurações nas quais o passo de aprendizagem é pequeno. O distanciamento com relação ao funcionamento real do algoritmo quando tal condição não é satisfeita será avaliado na seção a seguir (Seção 8.4).

| # Cenário | $\tau$ | $\sigma_\nu^2$ | $\sigma_q^2$ | $B(z)$ |
|---|---|---|---|---|
| 1 | 4 | $10^{-6}$ | $10^{-5}$ | 1 |
| 2 | 5 | $5 \times 10^{-6}$ | $2 \times 10^{-5}$ | 1 |
| 3 | 3 | $5 \times 10^{-7}$ | $8 \times 10^{-6}$ | 1 |
| 4 | 4 | $10^{-6}$ | $10^{-5}$ | $1 - 0,8z^{-1} + 0,2z^{-2}$ |

Tabela 17 – Especificação dos Cenários analisados para avaliação da capacidade de rastreamento do algoritmo SM-sign-NLMS.

**8.4- Resultados**

### 8.4.1 Rastreamento

Para avaliar a qualidade da previsão teórica ofertada pela Equação (195), iremos comparar sua solução com resultados obtidos por meio de simulações. A planta desconhecida a identificar é a quarta constante no relatório da ITU [122]. O sinal de entrada do filtro é obtido pela filtragem de um sinal gaussiano, branco, de média zero e variância unitária pelo filtro $B(z)$. O sinal de ruído $\nu(k)$ é branco, gaussiano e de variância $\sigma_\nu^2$. A perturbação markoviana $\boldsymbol{q}(k)$ é um vetor aleatório gaussiano, branco e de variância $\sigma_q^2$. A Tabela 17 especifica os quatro cenários testados. Em todos os cenários, o algoritmo SM-sign-NLMS foi executado por $2 \times 10^4$ iterações e o MSE em regime permanente foi computado por meio da média das últimas 1000 iterações da curva de MSE computada por meio de 1000 ensaios independentes de Monte Carlo.

As Figuras 48-51 detalham as curvas simuladas, bem como as obtidas pelo modelo teórico proposto. Observa-se uma grande aderência entre as curvas teórica e simulada. O comportamento observado também é bastante coerente com o discutido na literatura: para valores pequenos de $\beta$, há uma perda de desempenho, dada uma maior "lentidão" do algoritmo em acompanhar a variação da planta desconhecida. Valores altos de $\beta$, por sua vez, também são prejudiciais ao desempenho assintótico, já que aumentam a variância dos coeficientes adaptativos. Existe assim um valor de $\beta$ ótimo, capaz de otimizar o desempenho assintótico. Em todos os cenários testados, o modelo teórico proposto foi capaz de prever o desempenho assintótico para todos os valores de $\beta$, fornecendo com acurácia o valor de $\beta$ que otimiza o desempenho assintótico.

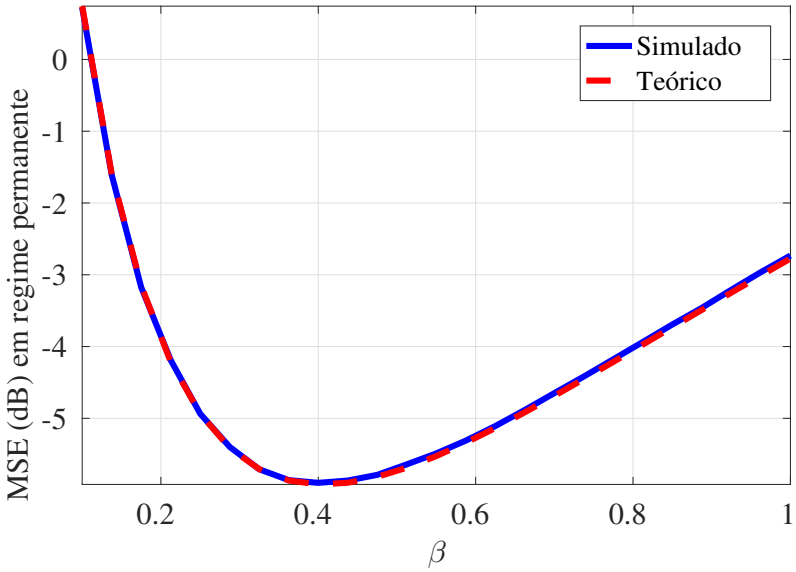



Figura 48 – MSE (em dB) em regime permanente em termos de $\beta$, para o Cenário 1. Em azul sólido, curva obtida por simulações. Em vermelho tracejado, previsão teórica engendrada pela resolução numérica da Equação (195).

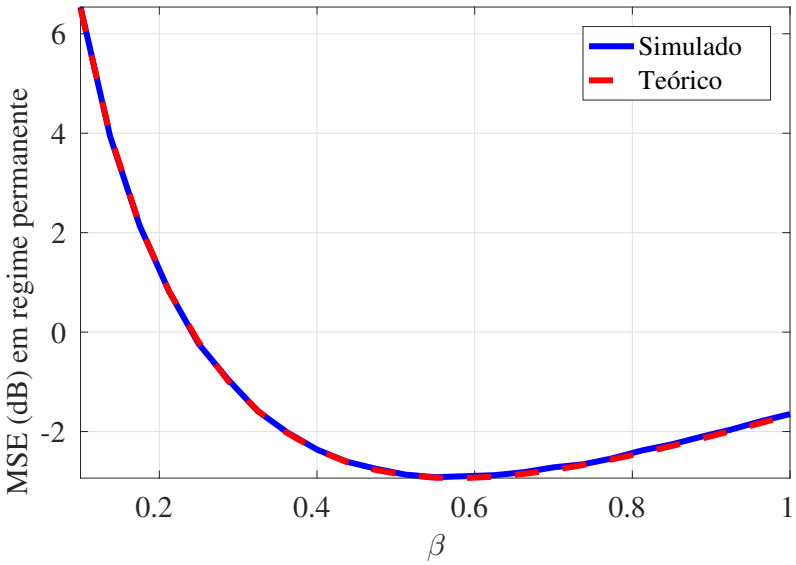



Figura 49 – MSE (em dB) em regime permanente em termos de $\beta$, para o Cenário 2. Em azul sólido, curva obtida por simulações. Em vermelho tracejado, previsão teórica engendrada pela resolução numérica da Equação (195).

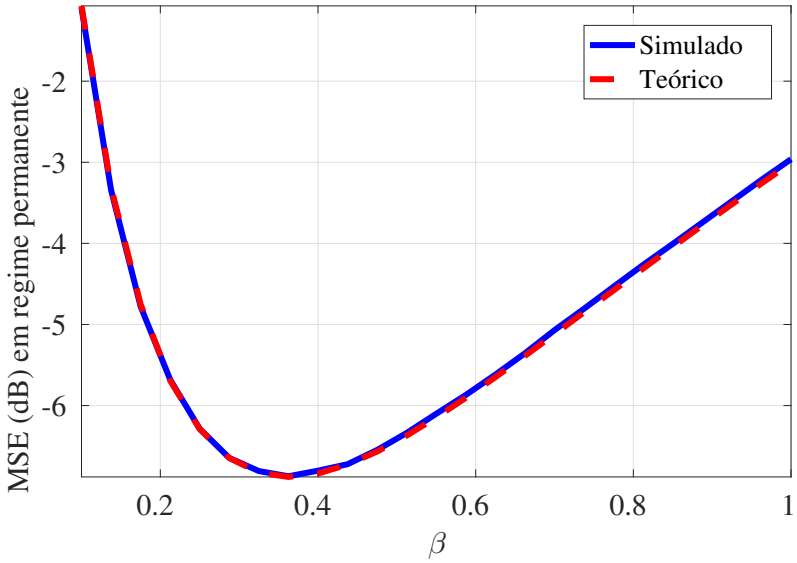



Figura 50 – MSE (em dB) em regime permanente em termos de $\beta$, para o Cenário 3. Em azul sólido, curva obtida por simulações. Em vermelho tracejado, previsão teórica engendrada pela resolução numérica da Equação (195).

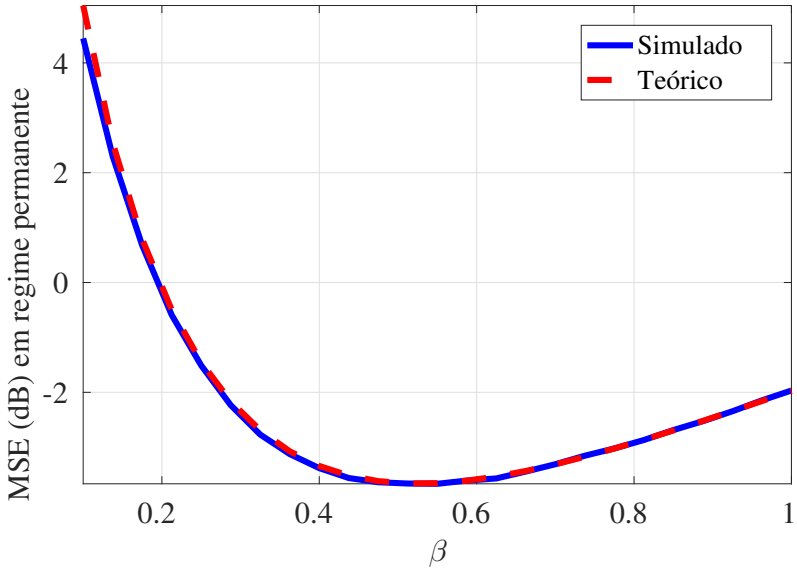



Figura 51 – MSE (em dB) em regime permanente em termos de $\beta$, para o Cenário 4. Em azul sólido, curva obtida por simulações. Em vermelho tracejado, previsão teórica engendrada pela resolução numérica da Equação (195).

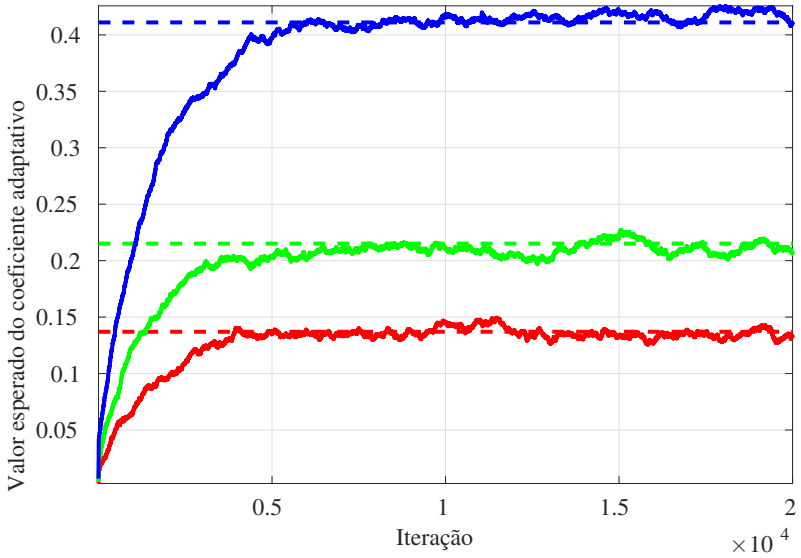



Figura 52 – Evolução da média dos coeficientes adaptativos. As retas tracejadas horizontais sinalizam o coeficiente correspondente na resposta ao impulso da planta desconhecida. As curvas representam a evolução da média de $w_1(k)$ (em vermelho), $w_4(k)$ (em verde) e $w_8(k)$ (em azul). As curvas foram computadas por meio de $10^4$ ensaios independentes de Monte Carlo.

### 8.4.2 Comprimento Deficiente

Nesta seção, pretendemos avaliar os achados teóricos referente à configuração de comprimento deficiente. O comportamento de primeira ordem (falta de tendenciosidade) utiliza hipóteses menos fortes (afinal, análises de estatísticas de primeira ordem exigem menos recursos analíticos). Assim, utilizaremos uma simulação para corroborar a ausência de viés assintótico do estimador adaptativo implementado pelo algoritmo SM-sign-NLMS. Seja um cenário no qual $\tau = 2$, $\sigma_\nu^2 = 10^{-6}$, $\beta = 0,5$, e o filtro adaptativo apresenta um comprimento de 20 coeficientes (inferior à resposta ao impulso desconhecida, a qual tem 96 coeficientes, sendo a quinta planta apresentada no estudo técnico da ITU [122]). Tanto o sinal de entrada (de variância unitária) quanto o ruído aditivo são advindos de processos estocásticos brancos e gaussianos. A Figura 52 mostra que os coeficientes adaptativos tendem para os valores dos coeficientes respectivos da resposta ao impulso da planta a identificar, conforme esperado pelo modelo teórico (vide Equação (204)).

Uma análise das estatísticas de segunda ordem (necessárias para estimar o MSE) exige mais cuidado, já que requer a mobilização de mais hipóteses estatísticas. Assim,

| #Cenário | #Modelo | $N_{\text{total}}$ | $\beta$ | $\tau$ | $\sigma_\nu^2$ | $N_{\text{it}}$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 64 | 0,001 | 2 | $10^{-6}$ | $4 \times 10^4$ |
| 2 | 2 | 96 | 0,002 | 3 | $10^{-5}$ | $3 \times 10^4$ |
| 3 | 3 | 96 | 0,003 | 4 | $10^{-5}$ | $2 \times 10^4$ |
| 4 | 4 | 128 | 0,005 | 5 | $10^{-4}$ | $2 \times 10^5$ |
| 5 | 4 | 128 | 0,01 | 4 | $10^{-4}$ | $15 \times 10^4$ |
| 6 | 4 | 128 | 0,1 | 5 | $10^{-4}$ | $5 \times 10^3$ |

Tabela 18 – Especificações dos 6 cenários de simulação adotados para avaliação da predição teórica de desempenho assintótico no caso de comprimento deficiente. $N_{\text{total}} \triangleq N+L$ denota o comprimento da planta ótima a identificar. $N_{\text{it}}$ denota o número de iterações executadas. A coluna #Modelo reflete o número da planta estimada por medições, constante no relatório da ITU [122].

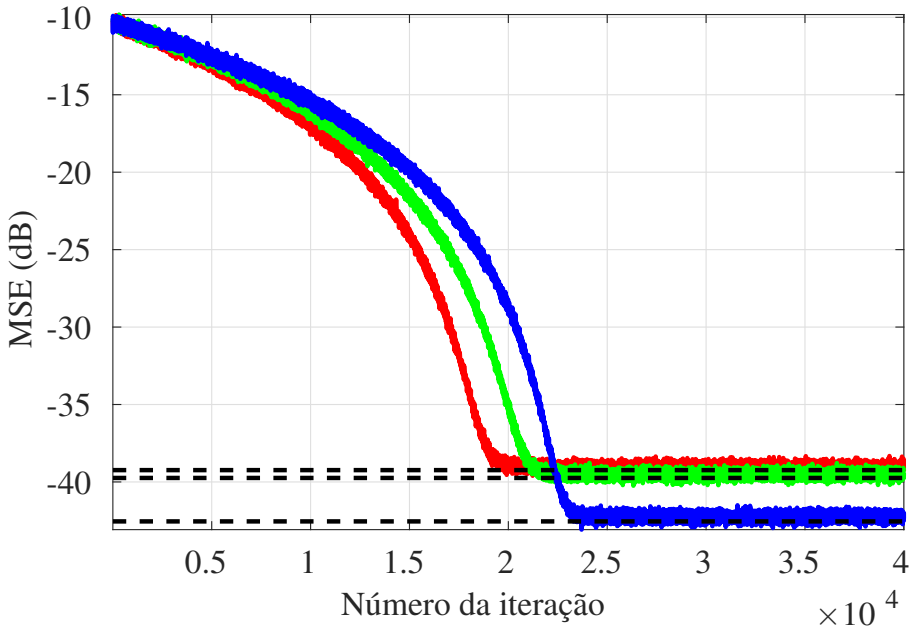



Figura 53 – Evolução do MSE ao longo das iterações, para o Cenário 1.

foram construídos 6 cenários de simulação para efeitos de comparação com a predição do MSE assintótico engendrada pela resolução numérica das Equações (213) e (214). A Tabela 18 especifica os cenários simulados. Ao longo de todos os cenários, o sinal de ruído aditivo é branco, gaussiano e de variância $\sigma_\nu^2$. O sinal de entrada é branco, gaussiano, de média zero e de variância unitária. Todas as curvas simuladas foram computadas por meio de 1000 ensaios independentes de Monte Carlo.

As Figuras 53-58 mostram a evolução do MSE (em dB) até o regime permanente, para diferentes valores de $N$ (comprimento do filtro adaptativo). Observe que os valores testados de $N$ são sempre inferiores a $N_{\text{total}}$ (veja a Tabela 18), de modo a caracterizar operação de comprimento deficiente. As retas tracejadas em preto sinalizam

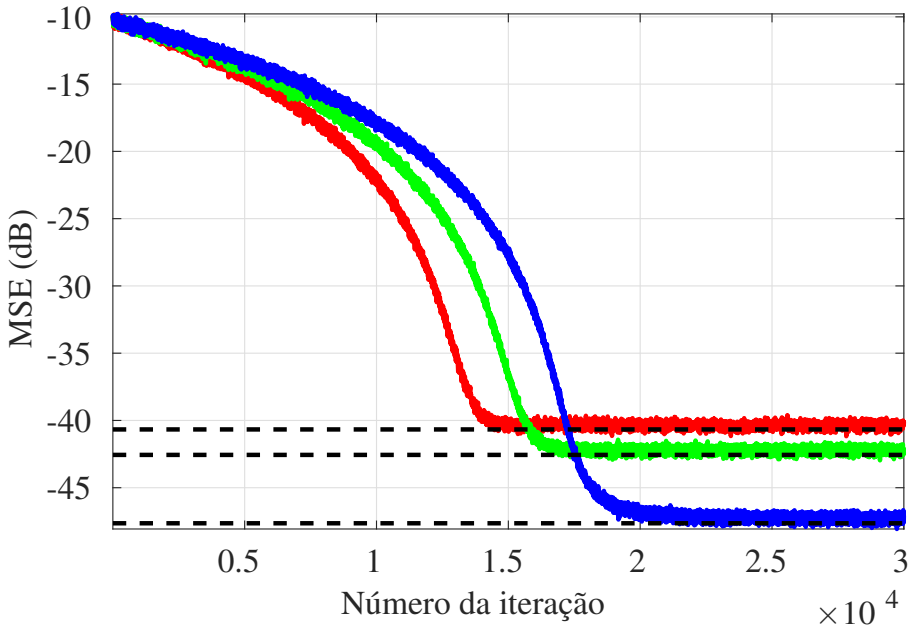



Figura 54 – Evolução do MSE ao longo das iterações, para o Cenário 2.

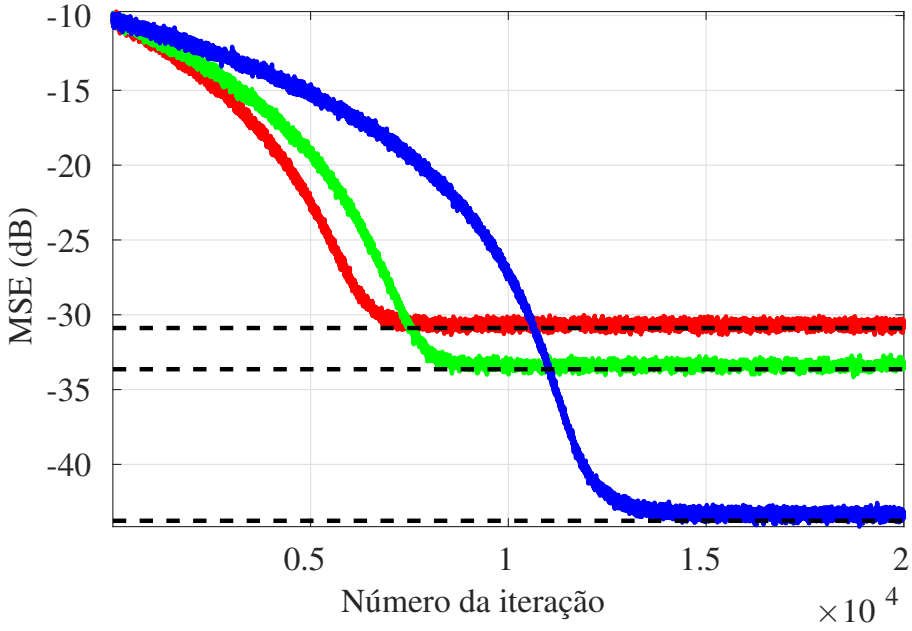



Figura 55 – Evolução do MSE ao longo das iterações, para o Cenário 3.

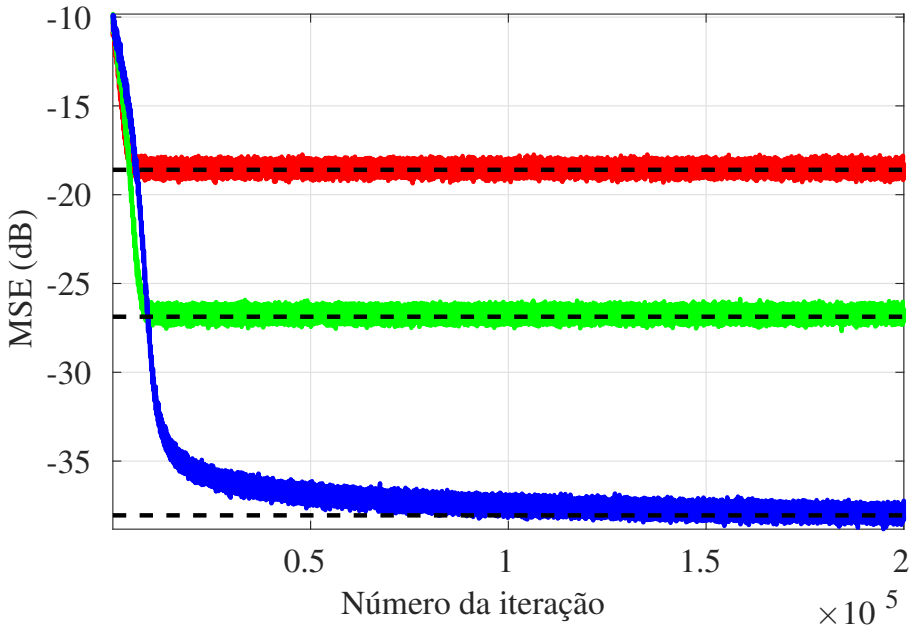



Figura 56 – Evolução do MSE ao longo das iterações, para o Cenário 4.

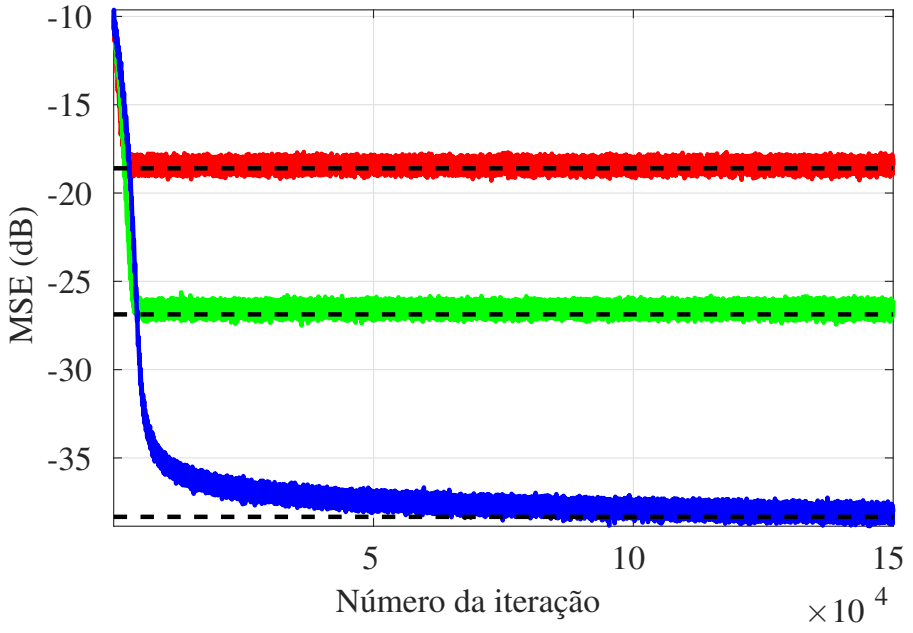



Figura 57 – Evolução do MSE ao longo das iterações, para o Cenário 5.

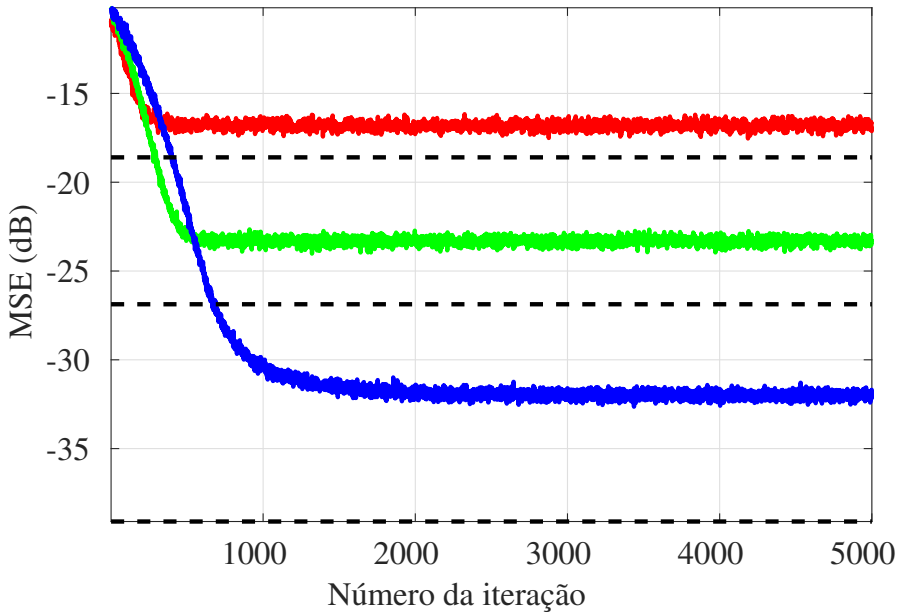



Figura 58 – Evolução do MSE ao longo das iterações, para o Cenário 6.

| #Cenário | $\beta$ | $N$ | $\text{MSE}_{\text{EMP}}$ (dB) | $\text{MSE}_{\text{TEOR}}$ (dB) | Erro (dB) |
|---|---|---|---|---|---|
| 1 | 0,001 | 50 | -39,00 | -39,24 | 0,23 |
| 1 | 0,001 | 55 | -39,52 | -39,74 | 0,22 |
| 1 | 0,001 | 60 | -42,26 | -42,55 | 0,29 |
| 2 | 0,002 | 70 | -40,38 | -40,67 | 0,29 |
| 2 | 0,002 | 80 | -42,22 | -42,56 | 0,34 |
| 2 | 0,002 | 90 | -47,27 | -47,64 | 0,37 |
| 3 | 0,003 | 50 | -30,67 | -30,88 | 0,21 |
| 3 | 0,003 | 60 | -33,35 | -33,64 | 0,29 |
| 3 | 0,003 | 90 | -43,33 | -43,77 | 0,44 |
| 4 | 0,005 | 50 | -18,51 | -18,60 | 0,09 |
| 4 | 0,005 | 80 | -26,71 | -26,87 | 0,16 |
| 4 | 0,005 | 120 | -37,95 | -38,06 | 0,11 |
| 5 | 0,01 | 50 | -18,40 | -18,60 | 0,19 |
| 5 | 0,01 | 80 | -26,57 | -26,87 | 0,30 |
| 5 | 0,01 | 120 | -38,11 | -38,34 | 0,23 |
| 6 | 0,1 | 50 | -16,78 | -18,60 | 1,81 |
| 6 | 0,1 | 80 | -23,30 | -26,87 | 3,57 |
| 6 | 0,1 | 120 | -31,99 | -39,11 | 7,12 |

Tabela 19 – Comparação entre o MSE em regime permanente obtido por meio de simulações ($\text{MSE}_{\text{EMP}}$) e o obtido pela resolução das Equações (213) e (214) ($\text{MSE}_{\text{TEOR}}$). O $\text{MSE}_{\text{EMP}}$ foi obtido pela média das últimas 1000 iterações da curva de MSE computada com 1000 ensaios independentes de Monte Carlo.

o valor teórico predito pela resolução numérica (predição teórica) das Equações (213) e (214). Observa-se uma maior aderência dos resultados teóricos aos valores simulados ns Cenários 1-4, nos quais os valores de $\beta$ são menores. À medida que o passo de aprendizagem ($\beta$) assume valores mais elevados, tal aderência vai se reduzindo. Tal conclusão é quantitativamente embasada por meio da Tabela 19. Observe, nesta tabela, por exemplo, que o erro no Cenário 1 (onde $\beta = 0,001$ é de no máximo 0,29 dB, enquanto que no Cenário 6 (onde $\beta = 0,1$) temos um erro de até 7,12 dB. Esta discrepância pode ser explicada pelo fato de o modelo teórico ter ignorado a dependência estatística de $g$ (definido na Equação (210)) com relação aos vetores aleatórios $\mathbb{E}\left[\tilde{\boldsymbol{w}}(k)\right]$ e $\mathbb{E}\left\{\boldsymbol{x}^T(k)\left\{\left[\overline{\boldsymbol{w}}^{\star}\right]^T\overline{\boldsymbol{x}}(k)\right\}\right\}$. Deste modo, a análise de comprimento deficiente de segunda ordem é a menos bem-sucedida de todos os modelos propostos, já que desafortunadamente é acurada apenas quando valores muito pequenos de $\beta$ são adotados.

Dessa forma, de fato, a utilização de um passo variável, como no algoritmo VSS-SM-sign-NLMS, tende a melhorar o desempenho transiente do estimador adaptativo e a promover uma melhor adaptação em cenários não estacionários. Entretanto, no contexto específico da análise de comprimento deficiente com modelagem de segunda ordem, as limitações observadas se devem majoritariamente à perda de representatividade do modelo estocástico adotado, bem como à natureza truncada do filtro, que restringe o espaço de parametrização

## 8.5- Comentários finais

Na prática, algoritmos adaptativos podem operar com um comprimento inferior ao necessário. Adicionalmente, podem atuar em contextos não estacionários, para os quais cumpre prever a perda de desempenho resultante. Este capítulo propôs modelos estocásticos inéditos para essas configurações. Primeiramente, demonstramos que a equação de ponto fixo para o regime permanente (174) é perturbada, num contexto de não estacionariedade markoviana de primeira ordem, por um termo aditivo diretamente proporcional à variância da perturbação markoviana e inversamente proporcional a $\beta$. No caso de comprimento deficiente, para sinais de entrada brancos, o modelo teórico prevê que o algoritmo SM-sign-NLMS implementa uma estimação não enviesada. Ademais,

seu desempenho (em termos de métricas quadráticas) em regime permanente pode ser obtido por meio da resolução numérica de um *par acoplado* de equações de ponto fixo. As estimativas teóricas foram extensamente comparadas com resultados de simulações. Observa-se que, devido às aproximações estocásticas adotadas, a predição do MSE no caso de operação em comprimento deficiente somente é acurada para valores pequenos do passo de aprendizagem.

# 9- Conclusões

Os algoritmos adaptativos LMLS e LLAD são versões com funções custo logarítmicas dos algoritmos LMF, LMS e LAD, os quais apresentam maior robustez a ruído impulsivo. Tais algoritmos são comumente derivados na literatura por meio do paradigma do gradiente estocástico. Nesta tese, por meio do Teorema 1 (demonstrado no Capítulo 2), tais algoritmos foram reposicionados como solucionadores de um problema de otimização convexa com restrições lineares, o qual pretende minimizar a distância entre a nova solução $\boldsymbol{w}(k+1)$ e a solução atual $\boldsymbol{w}(k)$. Tal minimização da distância implementa o chamado *princípio da distorção mínima*. Ao reposicionar tais algoritmos sob tal princípio, foi possível, pela primeira vez na literatura, incorporar metodologias tais como a adaptação proporcional (responsável por agilizar a taxa de convergência, no caso de um filtro ótimo esparso) quanto o reúso de coeficientes (responsável por aprimorar o desempenho assintótico do algoritmo). Assim, no Capítulo 2, num contexto de cancelamento automático de ruído, são propostos os algoritmos inéditos FX-IMPLMLS-RC e FX-IMPLLAD-RC, os quais conjugaram a uma taxa de convergência mais alta um melhor desempenho assintótico.

No Capítulo 3, a técnica de gradiente conjugado, costumeiramente associado a um processo de otimização *off-line*, foi judiciosamente relaxado para um contexto de otimização estocástica, também para algoritmos de cancelamento automático de ruído. O algoritmo proposto, intitulado CG-FX, foi objeto de uma análise estocástica de primeira ordem, a qual revelou que o algoritmo converge *na média* para a solução ótima quando a estimativa do secundário é perfeita. O algoritmo resultante apresenta uma taxa de convergência superior a de algoritmos no estado da arte, com tal vantagem se mantendo mesmo quando os sinais de entrada são sinais de voz reais.

O Capítulo 4 introduz um novo algoritmo, intitulado SM-sign-NLMS, o qual conjuga propriedades desejáveis tais como redução de complexidade computacional, robustez a ruído impulsivo e pouca necessidade de ajuste de hiperparâmetros adicionais. O algoritmo é cotejado com as versões clássicas de algoritmos, se revelando superior quando há ruído impulsivo no sinal de referência. Observa-se em alguns cenários que o algoritmo supera o algoritmo SM-NLMS nas três dimensões mais relevantes: taxa de convergência,

desempenho assintótico e custo computacional. Tal vantagem multidimensional é bastante rara, por ser bastante desafiadora de ser obtida.

O Capítulo 5 efetua uma análise estocástica do desempenho assintótico do algoritmo SM-sign-NLMS. A metodologia de análise fundamentou-se em argumentos de conservação de energia. Assim, foi encontrada uma equação de ponto fixo analítica cuja solução dá origem a uma estimativa teórica do desempenho do algoritmo. Tal estimativa teórica foi comparada com resultados de simulações, mostrando com relação a elas uma boa aderência.

O Capítulo 6 efetua uma análise de transiente do algoritmo proposto SM-sign-NLMS. O modelo estocástico mobilizado divorcia as distribuições radial e angular do vetor de entrada, entendendo-as como geradas por mecanismos estatisticamente independentes. Conjugando esta modelagem com o chamado modelo de Slock, foi possível determinar o comportamento modal do algoritmo. A dinâmica encontrada permite a extrapolação para previsões teóricas relativas à operação assintótica. Então, demonstra-se que tais previsões equivalem às do Capítulo 5, quando assumimos que o filtro apresenta comprimento elevado e o sinal de entrada é branco e gaussiano. As previsões teóricas relativas à dinâmica de comportamento do algoritmo foram confirmadas por meio de simulações computacionais.

O Capítulo 7 utiliza o modelo teórico apresentado no Capítulo 6 para propor uma variante VSS do algoritmo SM-sign-NLMS. Assim, o modelo teórico modula uma política de ajuste em tempo real do passo de aprendizagem do algoritmo. Tal política utiliza apenas informações disponíveis em tempo real para o algoritmo, na prática, requer o ajuste de poucos hiperparâmetros. A vantagem do algoritmo proposto VSS-SM-sign-NLMS é então confirmada por meio de simulações computacionais.

Por fim, novos modelos teóricos são adotados no Capítulo 8 para efetuar análises estocásticas relativas ao aprendizado do algoritmo SM-sign-NLMS para o caso de plantas não estacionárias (sujeitas a uma perturbação markoviana de primeira ordem), bem como para o caso de um filtro adaptativo com comprimento subótimo. Para este último caso, um par de equações de ponto fixo acopladas foi encontrado, o qual revelou aderência às simulações sob a condição de um passo de aprendizagem pequeno.

Essa tese propõe alguns algoritmos, com um deles (SM-sign-NLMS) sendo objeto de diversos modelos estocásticos. A modelagem estocástica dos demais algoritmos (FX-IMPLMLS-RC, FX-IMPLLAD-RC e CG-FX) ainda não foi efetuada (exceto a análise

de primeira ordem do algoritmo CG-FX, que consta no Capítulo 3 desta tese). Portanto, consideramos tal modelagem promissora para trabalhos futuros, assim como o aprimoramento do modelo de segunda ordem de operação sob regime deficiente, de modo a estendê-lo para configurações onde o passo de aprendizagem assume valores elevados.

# Bibliografia


[1] V. K. Chauhan, *Stochastic optimization for large-scale machine learning*, en. London, England: CRC Press, nov. de 2021.

[2] J. Gu, Q. Zou, C. Deng e X. Wang, "A Novel Robust Online Extreme Learning Machine for the Non-Gaussian Noise," *Chinese Journal of Electronics*, v. 32, n. 1, pp. 130–139, 2023.

[3] Y. Huang, T. Z. Shabestary, A. Gruenstein e L. Wan, "Multi-microphone adaptive noise cancellation for robust hotword detection," 2019.

[4] S. Dixit e D. Nagaria, "LMS Adaptive Filters for Noise Cancellation: A Review.," *International Journal of Electrical & Computer Engineering (2088-8708)*, v. 7, n. 5, 2017.

[5] M. dos S. Vieitos, M. P. Tcheou, D. B. Haddad e M. H. C. Dias, "Improved Proportionate Constrained Normalized Least Mean Square for Adaptive Beamforming," *Circuits, Systems, and Signal Processing*, ago. de 2023.

[6] V. S. N. Junior, M. P. Tcheou, M. H. C. Dias e D. B. Haddad, "Constrained Least Mean Square Algorithm with Coefficient Reusing," *Circuits, Systems, and Signal Processing*, v. 40, n. 11, pp. 5705–5717, abr. de 2021, ISSN: 1531-5878.

[7] V. S. Nogueira, M. H. C. Dias, M. P. Tcheou e D. Barreto Haddad, "Constrained NLMS Algorithm with Variable Coefficient Reusing," em *2021 SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference (IMOC)*, 2021, pp. 1–3.

[8] M. M. Halimeh, T. Haubner, A. Briegleb, A. Schmidt e W. Kellermann, "Combining Adaptive Filtering And Complex-Valued Deep Postfiltering For Acoustic Echo Cancellation," em *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2021, pp. 121–125.

[9] M. V. Matsuo, E. V. Kuhn e R. Seara, "Stochastic analysis of the NLMS algorithm for nonstationary environment and deficient length adaptive filter," *Signal Processing*, v. 160, pp. 190–201, 2019.

[10] L. Lu, K.-L. Yin, R. C. de Lamare et al., “A survey on active noise control in the past decade—Part I: Linear systems,” *Signal Processing*, v. 183, p. 108 039, jun. de 2021.

[11] L. Aydin, H. S. Artem e S. Oterkus, ed., *Designing engineering structures using stochastic optimization methods*, en. London, England: CRC Press, abr. de 2022.

[12] S. M. Kuo, I. Panahi, K. M. Chung, T. Horner, M. Nadeski e J. Chyan, “Design of Active Noise Control Systems With the TMS320 Family,” Texas Instruments, rel. técn. SPRA042, jun. de 1996.

[13] C. Hansen, *Active control of noise and vibration*. Boca Raton: CRC Press, 2020, ISBN: 9780367655747.

[14] N. V. George e G. Panda, “Advances in active noise control: A survey, with emphasis on recent nonlinear techniques,” *Signal Processing*, v. 93, n. 2, pp. 363–377, 2013, ISSN: 0165-1684.

[15] S. Elliott e P. Nelson, “Active noise control,” *IEEE Signal Processing Magazine*, v. 10, n. 4, pp. 12–35, out. de 1993, ISSN: 1053-5888.

[16] S. J. Elliott, “Down with noise [active noise control],” *IEEE Spectrum*, v. 36, n. 6, pp. 54–61, 1999.

[17] D. Miljković, “Brief Introduction to Active Noise Control,” em *2023 46th MIPRO ICT and Electronics Convention (MIPRO)*, 2023, pp. 977–982.

[18] P. Lueg, “Process of Silencing Sound Oscillations,” US Patent 2 043 416, 9 de jun. de 1936.

[19] L. Ljung, *System identification* (Prentice Hall information and system sciences series), en, 2ª ed. Philadelphia, PA: Prentice Hall, dez. de 1998.

[20] B. Zhu, B. Wang, B. Cai, Y. Zhu, P. Chao e Z. Fang, “A variable step size least mean p-power adaptive filtering algorithm based on multi-moment error fusion,” *EURASIP Journal on Advances in Signal Processing*, v. 2023, n. 1, jul. de 2023.

[21] M. R. Petraglia, E. L. Marques e D. B. Haddad, “Low-complexity affine projection subband algorithm for robust adaptive filtering in impulsive noise,” em *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2016, pp. 1–5.

[22] R. L. Das e M. Narwaria, “Lorentzian Based Adaptive Filters for Impulsive Noise Environments,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, v. 64, n. 6, pp. 1529–1539, 2017.

[23] S. Gollamudi, S. Nagaraj, S. Kapoor e Y.-F. Huang, “Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size,” *IEEE Signal Processing Letters*, v. 5, n. 5, pp. 111–114, 1998.

[24] J. C. Burgess, “Active adaptive sound control in a duct: A computer simulation,” *The Journal of the Acoustical Society of America*, v. 70, n. 3, pp. 715–726, set. de 1981.

[25] N. N. Siqueira, M. R. Petraglia e D. B. Haddad, “A novel heuristic for the analysis of block sign LMS algorithm,” *Electronics Letters*, v. 59, n. 7, abr. de 2023.

[26] S. Haykin, *Least-mean-square adaptive filters*. Hoboken, N.J: Wiley-Interscience, 2003, ISBN: 9780471215707.

[27] B. Lin, R. He, L. Song e B. Wang, “Steady-state performance analysis for adaptive filters with error nonlinearities,” em *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 3093–3096.

[28] S. Haykin, *Adaptive filter theory*. Upper Saddle River, New Jersey: Pearson, 2014, ISBN: 013267145X.

[29] T. R. P. Gonzaga, V. d. S. N. Junior, A. L. F. de Barros, F. d. R. Henriques e D. B. Haddad, “Transient Analysis of the Block Least Mean Squares Algorithm,” *IEEE Communications Letters*, v. 25, n. 2, pp. 608–612, 2021.

[30] L. C. Resende, D. B. Haddad e M. R. Petraglia, “A Variable Step-Size NLMS Algorithm with Adaptive Coefficient Vector Reusing,” em *2018 IEEE International Conference on Electro/Information Technology (EIT)*, 2018, pp. 0181–0186.

[31] D. B. Haddad, M. R. Petraglia e A. Petraglia, “A unified approach for sparsity-aware and maximum correntropy adaptive filters,” em *2016 24th European Signal Processing Conference (EUSIPCO)*, 2016, pp. 170–174.

[32] M. R. Petraglia, D. B. Haddad e E. L. Marques, “Affine Projection Subband Adaptive Filter With Low Computational Complexity,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 63, n. 10, pp. 989–993, 2016.

[33] M. R. Petraglia, D. B. Haddad e E. L. Marques, “Normalized Subband Adaptive Filtering Algorithm With Reduced Computational Complexity,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 62, n. 12, pp. 1164–1168, 2015.

[34] S. M. Kuo e D. R. Morgan, “Active noise control: a tutorial review,” *Proceedings of the IEEE*, v. 87, n. 6, pp. 943–973, 1999.

[35] B. Huang, Y. Xiao, J. Sun e G. Wei, “A Variable Step-Size FXLMS Algorithm for Narrowband Active Noise Control,” *IEEE Transactions on Audio, Speech, and Language Processing*, v. 21, n. 2, pp. 301–312, fev. de 2013.

[36] J.-m. Song e P. Park, “An optimal variable step-size affine projection algorithm for the modified filtered-x active noise control,” *Signal Processing*, v. 114, pp. 100–111, set. de 2015.

[37] P. Ramos Lorente, R. Martín Ferrer, F. Arranz Martínez e G. Palacios-Navarro, “Modified Filtered-X Hierarchical LMS Algorithm with Sequential Partial Updates for Active Noise Control,” *Applied Sciences*, v. 11, n. 1, p. 344, 2021.

[38] S. Elliott, I. Stothers e P. Nelson, “A multiple error LMS algorithm and its application to the active control of sound and vibration,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 35, n. 10, pp. 1423–1434, 1987.

[39] Q. Shen e A. Spanias, “Time and frequency domain X block LMS algorithms for single channel active noise control,” *Proc. 2nd Int. Congr. Recent Developments in Air-and Structure-Borne Sound Vibration*, pp. 353–360, 1992.

[40] S. M. Kuo e J. Luan, “Cross-coupled filtered-X LMS algorithm and lattice structure for active noise control systems,” em *1993 IEEE International Symposium on Circuits and Systems*, 1993, 459–462 vol.1.

[41] J. Thi e D. R. Morgan, “Delayless subband active noise control,” em *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, 1993, 181–184 vol.1.

[42] M. Bouchard e S. Quednau, “Multichannel recursive-least-square algorithms and fast-transversal-filter algorithms for active noise control and sound reproduction systems,” *IEEE Transactions on Speech and Audio Processing*, v. 8, n. 5, pp. 606–618, 2000.

[43] Seon Joon Park, Jeong Hyeon Yun, Young Cheol Park e Dae Hee Youn, “A delayless subband active noise control system for wideband noise control,” *IEEE Transactions on Speech and Audio Processing*, v. 9, n. 8, pp. 892–899, 2001.

[44] Hui Lan, Ming Zhang e Wee Ser, “A weight-constrained FxLMS algorithm for feedforward active noise control systems,” *IEEE Signal Processing Letters*, v. 9, n. 1, pp. 1–4, 2002.

[45] A. Carini e G. L. Sicuranza, “Steady-state and transient analysis of multichannel filtered-x affine projection algorithms,” em *Proceedings. (ICASSP ’05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 4, 2005, iv/345–iv/348 Vol. 4.

[46] X. Sun, S. M. Kuo e G. Meng, “Adaptive algorithm for active control of impulsive noise,” *Journal of Sound and Vibration*, v. 291, n. 1-2, pp. 516–522, mar. de 2006.

[47] A. Carini e G. L. Sicuranza, “Optimal Regularization Parameter of the Multichannel Filtered- $x$ Affine Projection Algorithm,” *IEEE Transactions on Signal Processing*, v. 55, n. 10, pp. 4882–4895, 2007.

[48] D. Das, S. Kuo e G. Panda, “New block filtered-X LMS algorithms for active noise control systems,” *IET Signal Processing*, v. 1, n. 2, pp. 73–81, jun. de 2007.

[49] X. Sun e S. M. Kuo, “Active Narrowband Noise Control Systems Using Cascading Adaptive Filters,” *IEEE Transactions on Audio, Speech, and Language Processing*, v. 15, n. 2, pp. 586–592, 2007.

[50] D. Zhou e V. DeBrunner, “A New Active Noise Control Algorithm That Requires No Secondary Path Identification Based on the SPR Property,” *IEEE Transactions on Signal Processing*, v. 55, n. 5, pp. 1719–1729, 2007.

[51] L. Lu e H. Zhao, “Improved Filtered-x Least Mean Kurtosis Algorithm for Active Noise Control,” *Circuits, Systems, and Signal Processing*, v. 36, n. 4, pp. 1586–1603, ago. de 2016.

[52] A. Puri, S. V. Modak e K. Gupta, “Modal filtered-x LMS algorithm for global active noise control in a vibro-acoustic cavity,” *Mechanical Systems and Signal Processing*, v. 110, pp. 540–555, set. de 2018.

[53] D. Shi, B. Lam, W.-S. Gan e S. Wen, “Optimal Leak Factor Selection for the Output-Constrained Leaky Filtered-Input Least Mean Square Algorithm,” *IEEE Signal Processing Letters*, v. 26, n. 5, pp. 670–674, mai. de 2019.

[54] L. Silva, J. Souza, J. Colares, A. Lima e D. Haddad, “Bias-compensated FX-LMS algorithm,” *Electronics Letters*, v. 56, n. 24, pp. 1340–1343, nov. de 2020.

[55] J. Chen, M. Wu, C. Gong, X. Wang e J. Yang, “Steady-State Performance Analysis of the Distributed FxLMS Algorithm for Narrowband ANC System With Frequency Mismatch,” *IEEE Signal Processing Letters*, v. 29, pp. 1167–1171, 2022.

[56] D. Shi, W.-S. Gan, B. Lam e X. Shen, “A Frequency-Domain Output-Constrained Active Noise Control Algorithm Based on an Intuitive Circulant Convolutional Penalty Factor,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 31, pp. 1318–1332, 2023.

[57] H. Zhang, J. Zhang, F. Ma, P. N. Samarasinghe e H. Sun, “A Time-Domain Multi-Channel Directional Active Noise Control System,” em *2023 31st European Signal Processing Conference (EUSIPCO)*, 2023, pp. 376–380.

[58] Z. Luo, D. Shi, W.-S. Gan e Q. Huang, “Delayless Generative Fixed-Filter Active Noise Control Based on Deep Learning and Bayesian Filter,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 32, pp. 1048–1060, 2024.

[59] Q. Zhang, D. Lin, Y. Xiao, Y. Zheng e S. Wang, “Error Reused Filtered-X Least Mean Square Algorithm for Active Noise Control,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 32, pp. 397–412, 2024.

[60] L. S. Resende e J. C. M. Bermudez, “An efficient model for the convergence behavior of the FXLMS algorithm with gaussian inputs,” em *IEEE/SP 13th Workshop on Statistical Signal Processing, 2005*, 2005, pp. 97–102.

[61] P. Lara, K. da S. Olinto, F. R. Petraglia e D. B. Haddad, “Exact analysis of the least-mean-square algorithm with coloured measurement noise,” *Electronics Letters*, v. 54, n. 24, pp. 1401–1403, nov. de 2018.

[62] 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.

[63] O. J. Tobias, J. C. M. Bermudez, N. J. Bershad e R. Seara, "New results and insights for the filtered-X LMS algorithm," em *Conference Record of Thirty-Second Asilomar Conference on Signals, Systems and Computers (Cat. No.98CH36284)*, vol. 2, 1998, 1161–1165 vol.2.

[64] G. Lan, *First-order and stochastic optimization methods for machine learning*, en. Cham, Switzerland: Springer Nature, mai. de 2020.

[65] M. O. Sayin, N. D. Vanli e S. S. Kozat, "A Novel Family of Adaptive Filtering Algorithms Based on the Logarithmic Cost," *IEEE Trans. Signal Process.*, v. 62, n. 17, pp. 4411–4424, set. de 2014.

[66] A. H. Sayed, *Fundamentals of adaptive filtering* (Wiley - IEEE), en. Nashville, TN: John Wiley & Sons, mai. de 2003.

[67] J. V. G. de Souza, D. B. Haddad, F. da R. Henriques e M. R. Petraglia, "Novel Proportionate Adaptive Filters with Coefficient Vector Reusing," *Circuits, Systems, and Signal Processing*, v. 39, n. 5, pp. 2473–2488, set. de 2019.

[68] D. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 30, n. 4, pp. 578–586, 1982.

[69] M. M. U. Faiz e A. Zerguine, "A steady-state analysis of the -normalized sign-error least mean square (NSLMS) adaptive algorithm," em *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2011, pp. 538–541.

[70] S. Ben Jebara e H. Besbes, "A comparative study between pre-whitened based sign algorithms in time-varying environments," em *2005 12th IEEE International Conference on Electronics, Circuits and Systems*, 2005, pp. 1–4.

[71] E. Eweda, "Convergence analysis of the sign algorithm with badly behaved noise," *Signal Processing*, v. 83, n. 1, pp. 57–65, jan. de 2003, ISSN: 0165-1684.

[72] N. Neurohr, M. Schoebinger, E. Prete e A. Sanders, "Adaptive decision-feedback equalization for band-limited high-speed serial links," em *2005 IEEE International Symposium on Circuits and Systems*, 2005, 924–927 Vol. 2.

[73] K. J. Sik e Y. D. Hwan, “Improved multiplication free adaptive digital filter with FSE in fast data transmission,” em *ISIE 2001. 2001 IEEE International Symposium on Industrial Electronics Proceedings (Cat. No.01TH8570)*, vol. 1, 2001, 210–215 vol.1.

[74] S.-C. Pei e C.-C. Tseng, “Least mean p-power error criterion for adaptive FIR filter,” *IEEE Journal on Selected Areas in Communications*, v. 12, n. 9, pp. 1540–1547, 1994.

[75] L. Lu, G. Zhu, X. Yang, K. Zhou, Z. Liu e W. Wu, “Affine Projection Algorithm-Based High-Order Error Power for Partial Discharge Denoising in Power Cables,” *IEEE Transactions on Instrumentation and Measurement*, v. 69, n. 4, pp. 1821–1832, 2020.

[76] S. Peng, Z. Wu e B. Chen, “Constrained least mean p-power error algorithm,” em *2016 35th Chinese Control Conference (CCC)*, 2016, pp. 5160–5163.

[77] Q. Wu, Y. Li, Z. Jiang e Y. Zhang, “A Novel Hybrid Kernel Adaptive Filtering Algorithm for Nonlinear Channel Equalization,” *IEEE Jose*, v. 7, pp. 62 107–62 114, 2019.

[78] X. Hou, H. Zhao e X. Long, “An Innovative Transient Analysis of Adaptive Filter With Maximum Correntropy Criterion,” *IEEE Signal Processing Letters*, v. 29, pp. 1689–1693, 2022.

[79] M. Vu, Y. Akiyama e K. Slavakis, “Dynamic Selection of p-norm in Linear Adaptive Filtering via online Kernel-based Reinforcement Learning,” em *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[80] S. Mula, V. C. Gogineni e A. S. Dhar, “Algorithm and Architecture Design of Adaptive Filters With Error Nonlinearities,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, v. 25, n. 9, pp. 2588–2601, 2017.

[81] E. Walach e B. Widrow, “The least mean fourth (LMF) adaptive algorithm and its family,” *IEEE transactions on Information Theory*, v. 30, n. 2, pp. 275–283, 1984.

[82] O. M. Abdelrhman, Y. Dou e S. Li, “A Generalized Robust Logarithmic Family-Based Adaptive Filtering Algorithms,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 70, n. 8, pp. 3199–3203, 2023.

[83] X. Wang e J. Han, “Affine Projection Algorithm Based on Least Mean Fourth Algorithm for System Identification,” *IEEE Access*, v. 8, pp. 11 930–11 938, 2020.

[84] V. Anant e R. Priemer, “Adjacent errors of stochastic gradient algorithms under general error criteria,” em *Proceedings of 40th Midwest Symposium on Circuits and Systems. Dedicated to the Memory of Professor Mac Van Valkenburg*, vol. 2, 1997, 778–781 vol.2.

[85] E. Eweda, “A Stable Normalized Least Mean Fourth Algorithm With Improved Transient and Tracking Behaviors,” *IEEE Transactions on Signal Processing*, v. 64, n. 18, pp. 4805–4816, 2016.

[86] S. H. Cho e S. D. Kim, “Adaptive filters based on the high order error statistics,” em *Proceedings of APCCAS’96 - Asia Pacific Conference on Circuits and Systems*, 1996, pp. 109–112.

[87] F. T. Castoldi e M. L. R. de Campos, “Application of a Minimum-Disturbance Description to Constrained Adaptive Filters,” *IEEE Signal Processing Letters*, v. 20, n. 12, pp. 1215–1218, 2013.

[88] M. O. Sayin, N. D. Vanli e S. S. Kozat, “Improved Convergence Performance of Adaptive Algorithms Through Logarithmic Cost,” em *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, IEEE, Florence,Italy, mai. de 2014, pp. 4513–4517.

[89] A. H. Sayed, *Adaptive Filters*, 1ª ed. John Wiley & Sons, 2011.

[90] M. Bashir e A. Zerguine, “Transform domain LMF algorithm for sparse system identification under low SNR,” em *2015 49th Asilomar Conference on Signals, Systems and Computers*, 2015, pp. 221–224.

[91] Y. Wang, Z. Qin, J. Tao e M. Jiang, “A Variable Step-Size l0-PRLS Algorithm and its Application in Sparse Channel Estimations,” em *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, 2023, pp. 1–5.

[92] Rosalin, N. K. Rout e D. P. Das, “Proportionate Exponential Hyperbolic Sine Adaptive Filter for Sparse System Identification,” em *2023 International Conference on Communication, Circuits, and Systems (IC3S)*, 2023, pp. 1–4.

[93] J. Ni, Y. Xing, Z. Zhu e J. Chen, “Proportionate Total Adaptive Filtering Algorithms for Sparse System Identification,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, v. 54, n. 3, pp. 1840–1852, 2024.

[94] 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.

[95] J. Benesty e S. L. Gay, “An improved PNLMS algorithm,” em *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 2002, pp. II-1881-II–1884.

[96] 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.

[97] H. Zhao, Y. Yu, S. Gao, X. Zeng e Z. He, “Memory Proportionate APA with Individual Activation Factors for Acoustic Echo Cancellation,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 22, n. 6, pp. 1047–1055, 2014.

[98] 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.

[99] C.-H. Lee, B. D. Rao e H. Garudadri, “Proportionate Adaptive Filtering Algorithms Derived Using an Iterative Reweighting Framework,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 29, pp. 171–186, 2021.

[100] W. Ma, D. Zheng, Z. Zhang, J. Duan e B. Chen, “Robust proportionate adaptive filter based on maximum correntropy criterion for sparse system identification in impulsive noise environments,” *Signal, Image and Video Processing*, v. 12, n. 1, pp. 117–124, jun. de 2017, ISSN: 1863-1711.

[101] 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.

[102] S. Zhang, H. Chen e A. H. Sayed, “Design of a Tap-Amplitude-Based Block Proportional Adaptive Filtering Algorithm,” *IEEE Transactions on Signal Processing*, v. 71, pp. 2448–2459, 2023.

[103] M. R. Petraglia e D. B. Haddad, “New adaptive algorithms for identification of sparse impulse responses — Analysis and comparisons,” em *2010 7th International Symposium on Wireless Communication Systems*, 2010, pp. 384–388.

[104] D. B. Haddad e M. R. Petraglia, “Transient and steady-state MSE analysis of the IMPNLMS algorithm,” *Digital Signal Processing*, v. 33, pp. 50–59, out. de 2014.

[105] J. Benesty e S. L. Gay, “An improved PNLMS algorithm,” em *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 2002, pp. 1881–1884.

[106] L. Liu, M. Fukumoto e S. Saiki, “An improved mu-law proportionate NLMS algorithm,” em *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 3797–3800.

[107] S.-E. Kim, J.-W. Lee e W.-J. Song, “Steady-state analysis of the NLMS algorithm with reusing coefficient vector and a method for improving its performance,” em *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, mai. de 2011.

[108] H. Cho, C. W. Lee e S. W. Kim, “Derivation of a new normalized least mean squares algorithm with modified minimization criterion,” *Signal Processing*, v. 89, n. 4, pp. 692–695, abr. de 2009.

[109] L. C. Resende, R. M. Pimenta, N. N. Siqueira, F. Igreja, D. B. Haddad e M. R. Petraglia, “Analysis of the least mean squares algorithm with reusing coefficient vector,” *Signal Processing*, v. 202, p. 108 742, jan. de 2023, ISSN: 0165-1684.

[110] R. Pimenta, L. Resende, D. Haddad, M. Petraglia e N. Siqueira, “Novo algoritmo proporcional com reúso de coeficientes e seleção de dados,” em *Simpósio Brasileiro de Telecomunicações*, 2017, pp. 1–5.

[111] R. Pimenta, L. Resende, N. Siqueira, D. Haddad e M. Petraglia, “Algoritmo NLMS com reúso de coeficientes e compensação de viés para entradas ruidosas,” 2018, pp. 1–5.

[112] R. M. S. Pimenta, L. C. Resende, N. N. Siqueira, D. B. Haddad e M. R. Petraglia, “A New Proportionate Adaptive Filtering Algorithm with Coefficient Reuse and Robustness Against Impulsive Noise,” em *2018 26th European Signal Processing Conference (EUSIPCO)*, IEEE, set. de 2018.

[113] L. C. Resende, D. B. Haddad e M. R. Petraglia, "A Variable Step-Size NLMS Algorithm with Adaptive Coefficient Vector Reusing," em *2018 IEEE International Conference on Electro/Information Technology (EIT)*, IEEE, mai. de 2018.

[114] L. C. Resende, N. N. Siqueira, R. M. S. Pimenta, D. B. Haddad e M. R. Petraglia, "LMS algorithm with reuse of coefficients and robustness against impulsive noise," 1, vol. 1, 2019.

[115] L. Resende, D. B. Haddad, G. da R. Ferreira, P. H. Campelo e M. R. Petraglia, "Sparsity-aware reuse of coefficients normalised least mean squares," *Electronics Letters*, v. 55, n. 9, pp. 561–563, mai. de 2019.

[116] L. Resende, R. Pimenta, N. Norat, D.B.Haddad e M.R.Petraglia, "LMS Algorithm with Reuse of Coefficients and Robustness Against Impulsive Noise," XXII Congresso Brasileiro de Automática, set. de 2018, pp. 1–5.

[117] H. A. Khan, A. Mirza, A. Wakeel, M. Y. Umair, F. Jabeen e F. Afzal, "Robust and Least Complex Online Secondary Path Estimation in Broadband Feed-Forward Active Control of Impulsive Noise Using FxLMAT Family," *IEEE Access*, v. 10, pp. 58 667–58 676, 2022.

[118] Y. Pang, X. Wang, Y. Gao e S. Ou, "Filtered-X $q$-Gradient Least Mean Mixed-Norm Algorithm for Active Noise Control," em *2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT)*, 2023, pp. 689–693.

[119] T. Y. Al-Naffouri e A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Transactions on Signal Processing*, v. 51, n. 3, pp. 639–652, mar. de 2003.

[120] N. R. Yousef e A. H. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters," *IEEE Transactions on Signal Processing*, v. 49, n. 2, pp. 314–324, fev. de 2001.

[121] H.-C. Shin e A. H. Sayed, "Mean-square performance of a family of affine projection algorithms," *IEEE Transactions on Signal Processing*, v. 52, n. 1, pp. 90–102, jan. de 2004.

[122] I. TSG, "15, Digital network echo cancellers (recommendation)," tech. rep., Tech. Rep. G. 168, ITU-T, rel. técn., 2004.

[123] N. A. Ahmad, “A Globally Convergent Stochastic Pairwise Conjugate Gradient-Based Algorithm for Adaptive Filtering,” *IEEE Signal Processing Letters*, v. 15, pp. 914–917, 2008.

[124] O. Diene e A. Bhaya, “Adaptive filtering algorithms designed using control Liapunov functions,” *IEEE Signal Processing Letters*, v. 13, n. 4, pp. 224–227, 2006.

[125] C.-L. Stanciu, J. Benesty, C. Paleologu, R.-L. Costea, L.-M. Dogariu e S. Ciochină, “Decomposition-Based Wiener Filter Using the Kronecker Product and Conjugate Gradient Method,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 32, pp. 124–138, 2024.

[126] O. Axelsson, *Iterative Solution Methods*. Cambridge, England: Cambridge University Press, mar. de 1996.

[127] C. .-. Lee, B. D. Rao e H. Garudadri, “A Sparse Conjugate Gradient Adaptive Filter,” *IEEE Signal Processing Letters*, v. 27, pp. 1000–1004, 2020.

[128] P. S. R. Diniz, M. O. K. Mendonca, J. O. Ferreira e T. N. Ferreira, “Data-Selective Conjugate Gradient Algorithm,” em *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 707–711.

[129] Pi Sheng Chang e A. N. Willson, “Analysis of conjugate gradient algorithms for adaptive filtering,” *IEEE Transactions on Signal Processing*, v. 48, n. 2, pp. 409–418, 2000.

[130] G. H. Golub e C. F. Van Loan, *Matrix Computations* (Johns Hopkins Studies in the Mathematical Sciences), en, 4ª ed. Baltimore, MD: Johns Hopkins University Press, fev. de 2013.

[131] D. G. Luenberger e Y. Ye, *Linear and nonlinear programming* (International Series in Operations Research & Management Science), en. New York, NY: Springer, nov. de 2010.

[132] P. Lara, L. D. Tarrataca e D. B. Haddad, “Exact expectation analysis of the deficient-length LMS algorithm,” *Signal Processing*, v. 162, pp. 54–64, set. de 2019.

[133] J. Guo, F. Yang e J. Yang, “Mean-Square Performance of the Modified Filtered-x Affine Projection Algorithm,” *Circuits, Systems, and Signal Processing*, v. 39, n. 8, pp. 4243–4257, fev. de 2020.

[134] P. Lara, F. Igreja, L. D. T. J. Tarrataca, D. B. Haddad e M. R. Petraglia, “Exact Expectation Evaluation and Design of Variable Step-Size Adaptive Algorithms,” *IEEE Signal Processing Letters*, v. 26, n. 1, pp. 74–78, 2019.

[135] M. Z. A. Bhotto, M. O. Ahmad e M. N. S. Swamy, “Robust Shrinkage Affine-Projection Sign Adaptive-Filtering Algorithms for Impulsive Noise Environments,” *IEEE Transactions on Signal Processing*, v. 62, n. 13, pp. 3349–3359, 2014.

[136] T. N. Ferreira, W. A. Martins, M. V. S. Lima e P. S. R. Diniz, “Convex Combination of Constraint Vectors for Set-membership Affine Projection Algorithms,” em *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 4858–4862.

[137] P. S. R. Diniz e H. Yazdanpanah, “Data censoring with set-membership algorithms,” em *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2017, pp. 121–125.

[138] J. V. G. De Souza, F. D. R. Henriques, N. N. Siqueira, L. Tarrataca, F. A. A. Andrade e D. B. Haddad, “Stochastic Modeling of the Set-Membership- Sign-NLMS Algorithm,” *IEEE Access*, v. 12, pp. 32 739–32 752, 2024.

[139] P. T. L. Pereira, G. Paim, E. A. C. da Costa, S. J. M. de Almeida e S. Bampi, “ReAdapt: A Reconfigurable Datapath for Runtime Energy-Quality Scalable Adaptive Filters,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, v. 70, n. 1, pp. 327–339, 2023.

[140] A. O. Sarp, E. C. Mengüç, M. Peker e B. Ç. Güvenç, “Data-Adaptive Censoring for Short-Term Wind Speed Predictors Based on MLP, RNN, and SVM,” *IEEE Systems Journal*, v. 16, n. 3, pp. 3625–3634, 2022.

[141] M. V. S. Lima, P. S. R. Diniz e H. Yazdanpanah, “Set-Membership Constrained Frequency-Domain Algorithm,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 68, n. 2, pp. 797–801, 2021.

[142] C. G. Tsinos e P. S. R. Diniz, “Data-selective LMS-Newton and LMS-Quasi-Newton Algorithms,” em *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 4848–4852.

[143] S. Nagaraj, S. Gollamudi, S. Kapoor e Y.-F. Huang, “BEACON: an adaptive set-membership filtering technique with sparse updates,” *IEEE Transactions on Signal Processing*, v. 47, n. 11, pp. 2928–2941, 1999.

[144] L. Guo e Y.-F. Huang, “Frequency-Domain Set-Membership Filtering and Its Applications,” *IEEE Transactions on Signal Processing*, v. 55, n. 4, pp. 1326–1338, 2007.

[145] M. V. Lima e P. S. Diniz, “Fast learning set theoretic estimation,” em *21st European Signal Processing Conference (EUSIPCO 2013)*, 2013, pp. 1–5.

[146] A. Flores e R. C. de Lamare, “Sparsity-aware set-membership adaptive algorithms with adjustable penalties,” em *2017 22nd International Conference on Digital Signal Processing (DSP)*, 2017, pp. 1–5.

[147] R. C. de Lamare e P. S. R. Diniz, “Set-Membership Adaptive Algorithms Based on Time-Varying Error Bounds for CDMA Interference Suppression,” *IEEE Transactions on Vehicular Technology*, v. 58, n. 2, pp. 644–654, 2009.

[148] M. V. S. Lima e P. S. R. Diniz, “On the steady-state MSE performance of the set-membership NLMS algorithm,” em *2010 7th International Symposium on Wireless Communication Systems*, 2010, pp. 389–393.

[149] S. Werner e P. Diniz, “Set-membership affine projection algorithm,” *IEEE Signal Processing Letters*, v. 8, n. 8, pp. 231–235, 2001.

[150] M. Z. A. Bhotto e A. Antoniou, “A set-membership nlms algorithm with adaptive error bound,” em *2008 Canadian Conference on Electrical and Computer Engineering*, 2008, pp. 002 007–002 010.

[151] H. Yazdanpanah e P. S. R. Diniz, “New Trinion and Quaternion Set-Membership Affine Projection Algorithms,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 64, n. 2, pp. 216–220, 2017.

[152] F. Huang, J. Zhang e S. Zhang, “Combined-Step-Size Affine Projection Sign Algorithm for Robust Adaptive Filtering in Impulsive Interference Environments,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 63, n. 5, pp. 493–497, 2016.

[153] E. Eweda, “Transient performance degradation of the LMS, RLS, sign, signed regressor, and sign-sign algorithms with data correlation,” *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, v. 46, n. 8, pp. 1055–1062, 1999.

[154] E. Eweda, “Tracking analysis of the sign algorithm without the Gaussian constraint,” *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, v. 45, n. 1, pp. 115–122, 1998.

[155] E. Eweda, “Almost sure convergence of a decreasing gain sign algorithm for adaptive filtering,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 36, n. 10, pp. 1669–1671, 1988.

[156] N. N. Siqueira, T. P. Ferreira, L. C. Resende, R. M. S. Pimenta, D. B. Haddad e M. R. Petraglia, “Transient Analysis of the Set-Membership LMS Algorithm,” *IEEE Communications Letters*, v. 25, n. 4, pp. 1298–1302, 2021.

[157] F. Farina, A. Garulli e A. Giannitrapani, “An Interpolatory Algorithm for Distributed Set Membership Estimation in Asynchronous Networks,” *IEEE Transactions on Automatic Control*, v. 67, n. 10, pp. 5464–5470, 2022.

[158] H. Yazdanpanah, M. V. S. Lima e P. S. R. Diniz, “On the robustness of the set-membership NLMS algorithm,” em *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2016, pp. 1–5.

[159] T. Al-Naffouri e A. Sayed, “Transient analysis of data-normalized adaptive filters,” *IEEE Transactions on Signal Processing*, v. 51, n. 3, pp. 639–652, 2003.

[160] T. Al-Naffouri e A. Sayed, “Transient analysis of adaptive filters with error nonlinearities,” *IEEE Transactions on Signal Processing*, v. 51, n. 3, pp. 653–663, 2003.

[161] N. Takahashi e I. Yamada, “Steady-State Mean-Square Performance Analysis of a Relaxed Set-Membership NLMS Algorithm by the Energy Conservation Argument,” *IEEE Transactions on Signal Processing*, v. 57, n. 9, pp. 3361–3372, 2009.

[162] T. Y. Al-Naffouri e M. Moinuddin, “Exact tracking analysis of the NLMS algorithm for correlated Gaussian inputs,” em *21st European Signal Processing Conference (EUSIPCO 2013)*, 2013, pp. 1–5.

[163] N. Verhoeckx e T. Claasen, “Some Considerations on the Design of Adaptive Digital Filters Equipped with the Sign Algorithm,” *IEEE Transactions on Communications*, v. 32, n. 3, pp. 258–266, 1984.

[164] W. M. Wee, M. Yamagishi e I. Yamada, “Mean-square performance of the hyperslab-based adaptive projected subgradient method,” em *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 6384–6388.

[165] A. Khalili, A. Rastegarnia, A. Farzamnia, S. Sanei e T. A. H. Alghamdi, “Tracking Analysis of Maximum Versoria Criterion Based Adaptive Filter,” *IEEE Access*, v. 12, pp. 30 747–30 753, 2024.

[166] M. Rupp e A. Sayed, “A time-domain feedback analysis of filtered-error adaptive gradient algorithms,” *IEEE Transactions on Signal Processing*, v. 44, n. 6, pp. 1428–1439, 1996.

[167] H.-C. Shin e A. Sayed, “Mean-square performance of a family of affine projection algorithms,” *IEEE Transactions on Signal Processing*, v. 52, n. 1, pp. 90–102, 2004.

[168] H.-C. Shin, W.-J. Song e A. Sayed, “Mean-square performance of data-reusing adaptive algorithms,” *IEEE Signal Processing Letters*, v. 12, n. 12, pp. 851–854, 2005.

[169] K. Pärlin, T. Riihonen, V. Le Nir e M. Adrat, “Estimating and Tracking Wireless Channels Under Carrier and Sampling Frequency Offsets,” *IEEE Transactions on Signal Processing*, v. 71, pp. 1053–1066, 2023.

[170] A. Papoulis e S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, Fourth. Boston: McGraw Hill, 2002, ISBN: 0071122567 9780071122566 0073660116 9780073660110 0071226613 9780071226615.

[171] R. Arablouei e K. Doğançay, “Tracking performance analysis of the set-membership NLMS Adaptive Filtering Algorithm,” em *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, 2012, pp. 1–6.

[172] 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.

[173] D. Slock, “On the convergence behavior of the LMS and the normalized LMS algorithms,” *IEEE Transactions on Signal Processing*, v. 41, n. 9, pp. 2811–2825, 1993.

[174] M. Ferrer, M. de Diego e A. Gonzalez, “Low Cost Variable Step-Size LMS With Maximum Similarity to the Affine Projection Algorithm,” *IEEE Open Journal of Signal Processing*, v. 5, pp. 82–91, 2024.

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

[176] R. Harris, D. Chabries e F. Bishop, “A variable step (VS) adaptive filter algorithm,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 34, n. 2, pp. 309–316, 1986.

[177] J. Evans e B. Liu, “A CMOS implementation of a variable step size digital adaptive filter,” em *International Conference on Acoustics, Speech, and Signal Processing,*, 1989, 2489–2492 vol.4.

[178] P. Diniz e J. Cousseau, “Optimal convergence factor for Gauss-Newton algorithms and its application to an adaptive parallel realization,” em *SBT/IEEE International Symposium on Telecommunications*, 1990, pp. 420–424.

[179] P. Lara, F. Igreja, L. D. T. J. Tarrataca, D. B. Haddad e M. R. Petraglia, “Exact Expectation Evaluation and Design of Variable Step-Size Adaptive Algorithms,” *IEEE Signal Processing Letters*, v. 26, n. 1, pp. 74–78, 2019.

[180] R. Kwong e E. Johnston, “A variable step size LMS algorithm,” *IEEE Transactions on Signal Processing*, v. 40, n. 7, pp. 1633–1642, 1992.

[181] M. Garcia-Hernandez, “A Variable-Step LMS-Based Adaptive Filtering Algorithm Using Standard Deviation,” em *2023 IEEE 9th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2023, pp. 146–150.

[182] N. J. Bershad e J. C. Bermudez, “A switched variable step size NLMS adaptive filter,” *Digital Signal Processing*, v. 101, p. 102 730, jun. de 2020, ISSN: 1051-2004.

[183] S. Yang, N. Li e X. Zhang, “Improved Variable Step Size LMS Algorithm Based on Arctangent Function,” em *2021 China Automation Congress (CAC)*, 2021, pp. 4666–4669.

[184] D. Liu, H. Cha e B. Wang, "A Novel Variable step length LMS Algorithm based on Arctangent Compound Function," em *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, vol. 5, 2021, pp. 1691–1696.

[185] D. G. Tiglea, R. Candido e M. T. M. Silva, "A Variable Step Size Adaptive Algorithm With Simple Parameter Selection," *IEEE Signal Processing Letters*, v. 29, pp. 1774–1778, 2022.

[186] J. Benesty, H. Rey, L. Vega e S. Tressens, "A Nonparametric VSS NLMS Algorithm," *IEEE Signal Processing Letters*, v. 13, n. 10, pp. 581–584, 2006.

[187] A. H. Sayed, *Adaptive Filters* (Wiley - IEEE), en. New York, NY: Wiley-IEEE Press, out. de 2011.

[188] S. J. M. de Almeida, M. H. Costa e J. C. M. Bermudez, "A stochastic model for the deficient length Pseudo Affine Projection adaptive algorithm," em *2009 17th European Signal Processing Conference*, 2009, pp. 1715–1719.

[189] D. Xu, B. Yin, W. Wang e W. Zhu, "Variable Tap-Length LMS Algorithm Based on Adaptive Parameters for TDL Structure Adaption," *IEEE Signal Processing Letters*, v. 21, n. 7, pp. 809–813, 2014.

[190] P. Lara, L. D. Tarrataca e D. B. Haddad, "Exact expectation analysis of the deficient-length LMS algorithm," *Signal Processing*, v. 162, pp. 54–64, 2019, ISSN: 0165-1684.