

Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação
Departamento de Microondas e Óptica

**Análise de estabilidade de Lyapunov de algoritmos adaptativos
com contribuições ao estudo do critério de módulo constante**

Autor: Celso de Sousa Júnior

Orientador: Prof. Dr. João Marcos Travassos Romano
Co-orientador: Prof. Dr. Romis Ribeiro de Faissol Attux

Tese de Doutorado apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: Telecomunicações e Telemática.

Banca Examinadora:

Prof. Dr. João Marcos Travassos Romano (DMO/FEEC/UNICAMP)

Prof. Dr. Vítor Heloiz Nascimento (PSI/USP)

Prof. Dr. Charles Casimiro Cavalcante (GTEL/UFC)

Prof. Dr. João Bosco Ribeiro do Val (FEEC/UNICAMP)

Prof. Dr. Alim Pedro de Castro Gonçalves (FEEC/UNICAMP)

Campinas, 12 de agosto de 2011

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

So85a Sousa Junior, Celso de
 Análise de estabilidade de Lyapunov de algoritmos
 adaptativos com contribuições ao estudo do critério de
 módulo constante / Celso de Sousa Junior. --Campinas,
 SP: [s.n.], 2011.

 Orientadores: João Marcos Travassos Romano,
 Romis Ribeiro de Faissol Attux.

 Tese de Doutorado - Universidade Estadual de
 Campinas, Faculdade de Engenharia Elétrica e de
 Computação.

 1. Equalizadores (Eletrônica). 2. Estabilidade. 3.
 Lyapunov, Funções de. 4. Processamento digital de
 sinais. 5. Algoritmos on-line. I. Romano, João Marcos
 Travassos. II. Attux, Romis Ribeiro de Faissol. III.
 Universidade Estadual de Campinas. Faculdade de
 Engenharia Elétrica e de Computação. IV. Título.

Título em Inglês: Lyapunov stability analysis for adaptive algorithms with
 contributions to constant modulus criteria study

Palavras-chave em Inglês: Equalization (Electronics), Stability, Lyapunov Theory,
 Digital signal processing, Online algorithms

Área de concentração: Telecomunicações e Telemática

Titulação: Doutor em Engenharia Elétrica

Banca examinadora: Vítor Heloiz Nascimento, Charles Casimiro Cavalcante, João
 Bosco Ribeiro do Val, Alim Pedro de Castro Gonçalves

Data da defesa: 12/08/2011

Programa de Pós Graduação: Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE DOUTORADO

Candidato: Celso de Sousa Junior

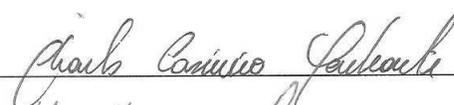
Data da Defesa: 12 de agosto de 2011.

Título da Tese: "Análise de estabilidade de Lyapunov de algoritmos adaptativos com contribuições ao estudo do critério de módulo constante".

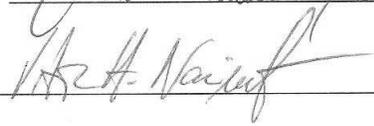
Prof. Dr. João Marcos Travassos Romano (Presidente):



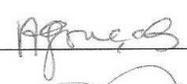
Prof. Dr. Charles Casimiro Cavalcante:



Prof. Dr. Vítor Heloiz Nascimento:



Prof. Dr. Alim Pedro de Castro Gonçalves:



Prof. Dr. João Bosco Ribeiro do Val:



Resumo

O problema de equalização adaptativa se vincula à busca por soluções iterativas que permitam reduzir ou eliminar os efeitos nocivos do canal de comunicação sobre um sinal transmitido de interesse. Uma vez que os sistemas adaptativos se baseiam em algoritmos capazes de ajustar os parâmetros de um filtro, pode-se considerar o conjunto equalizador / algoritmo adaptativo como um sistema dinâmico, o que termina por relacionar a possibilidade de obter uma solução satisfatória à noção de convergência.

A análise de convergência de algoritmos de equalização adaptativa se desenvolveu, tipicamente, considerando algumas hipóteses para viabilizar o tratamento matemático, mas nem sempre tais hipóteses são estritamente válidas. Um exemplo clássico nesse sentido é o uso da teoria da independência. Neste trabalho, buscamos uma abordagem distinta do estudo das condições de estabilidade de algoritmos de equalização clássica baseada na teoria de Lyapunov. Essa teoria é geralmente utilizada no estudo de sistemas não-lineares, e apresenta um amplo histórico de resultados sólidos na área de controle adaptativo. Isso motiva o uso no campo de processamento de sinais.

A primeira contribuição deste trabalho consiste em determinar, por meio da teoria de Lyapunov, a faixa de valores de passo de adaptação que garantem estabilidade do sistema de equalização para algoritmos baseados no critério de Wiener e para o algoritmo do módulo constante. A partir dos resultados para estabilidade, investigar-se-á também a região de convergência para os pesos do algoritmo LMS, o que trará uma produtiva relação com a idéia de *misadjustment*.

Como segunda linha de contribuição, será apresentada uma análise de um limitante inferior para o custo atingível e uma proposta de inicialização capaz de aumentar a probabilidade de convergência para o melhor ótimo gerado pelo critério para o algoritmo do módulo constante. Essa estratégia se baseia numa formulação do critério não-supervisionado de filtragem linear em termos da aplicação do critério de Wiener a uma estrutura polinomial. Os resultados obtidos revelam que a idéia é capaz de levar a um desempenho melhor que os do clássico método *center spike* e de uma estratégia de inicialização aleatória.

Palavras-chave: Equalização adaptativa, Estabilidade, Teoria de Lyapunov, SDA, LMS, CMA, Misadjustment, Convergência.

Abstract

The problem of adaptive equalization is related to the search for iterative solutions that allow the reduction or the elimination of the noxious effects of a communication channel on a transmitted signal of interest. Since adaptive systems are based on algorithms capable of adjusting the parameters of a filter, the combination between equalizer and learning algorithm can be considered to form a dynamical system, which relates the possibility of obtaining a satisfactory solution to the convergence issue.

The analysis of the convergence of adaptive equalization algorithms was developed, typically, considering certain simplifying hypotheses that, however, are not always strictly valid. A classical example that illustrates this assertion is the use of the so-called independence theory. In this work, it has been investigated a distinct approach to the study of stability conditions of classical methods based on Lyapunov theory. This theory is generally employed in the study of nonlinear systems, and presents a significant framework of sound results in the field of adaptive control, which motivates its use in the context of signal processing.

The first contribution of this work consists of determining, by means of Lyapunov theory, the range of step-size values that ensure stability of the equalization system for algorithms based on the Wiener criterion and for the constant modulus algorithm. Using the obtained stability results, the convergence region for the parameters estimated via LMS is also investigated, which establishes an interesting connection with the notion of misadjustment.

In a second line of study, we present an analysis of the lower bound for the attainable CM cost and an initialization heuristic capable of increasing the probability of convergence to the best optimum engendered by the constant modulus criterion. This strategy is based on a formulation of the unsupervised linear filtering criterion in terms of the application of the Wiener criterion to a polynomial structure. The obtained results reveal that the proposal is able to effectively lead to a performance level that is better than that achieved using the classical center spike method and a random approach.

Keywords: Adaptive equalization, Stability, Lyapunov Theory, SDA, LMS, CMA, Misadjustment, Convergence.

Agradecimentos

Gostaria de agradecer

a Deus, pela força de vontade e perseverança;

ao meu orientador, Prof. Dr. João Marcos Travassos Romano, pela orientação, atenção e oportunidade;

ao meu co-orientador, Prof. Dr. Romis Ribeiro de Faissol Attux, cujo apoio, amizade, paciência, leitura minuciosa do texto, discussões e ensinamentos ao longo desses anos foram essenciais para conclusão deste trabalho;

ao Prof. Dr. Ricardo Suyama, pelo apoio nos resultados e análise do algoritmo LMS e amizade;

ao Prof. Dr. Rafael Ferrari, pelas discussões sobre o uso de redes neurais do tipo RBF em filtragem não-supervisionada;

aos demais colegas do Laboratório de Processamento de Sinais para Comunicações - DPSCom, pelo convívio e amizade;

a Celi pelo apoio administrativo, e

a minha família pelo apoio incondicional.

À minha filha Isabela e minha esposa Laurizete

Sumário

Lista de Figuras	xvii
Lista de Tabelas	xxi
Glossário	xxiii
Lista de Símbolos	xxv
Trabalhos Publicados Pelo Autor	xxvii
1 Introdução	1
1.1 Alguns trabalhos relevantes para o estudo de estabilidade em equalização adaptativa .	3
1.2 Organização e contribuições do trabalho	5
2 Equalização Adaptativa Supervisionada e Não-Supervisionada	7
2.1 Problema de Transmissão Digital	7
2.2 Equalização Adaptativa	12
2.3 Equalização Adaptativa Supervisionada	14
2.3.1 Critério de Wiener	15
2.3.2 SDA (<i>Steepest Descent Algorithm</i>)	17
2.3.3 LMS (<i>Least Mean Square</i>)	18
2.4 Equalização Adaptativa Não-Supervisionada	22
2.4.1 CMA (<i>Constant Modulus Algorithm</i>)	23
2.5 Análise de Estabilidade de Algoritmos Adaptativos	25
2.5.1 Estabilidade do SDA	25
2.5.2 Estabilidade do LMS - Prova 1	27
2.5.3 Estabilidade do LMS - Prova 2	29
2.5.4 Estabilidade do CMA	32
2.6 Sumário	38
3 Estabilidade de Lyapunov	39
3.1 Definições	39
3.1.1 Caso Determinístico	40
3.1.2 Caso Estocástico	44
3.2 Primeiro Método de Lyapunov	45

3.3	Segundo Método de Lyapunov	49
3.3.1	Caso determinístico	49
3.3.2	Caso estocástico	52
3.4	Interpretação Geométrica	53
3.5	Análise de Convergência	56
3.6	Sumário	59
4	Análise da Estabilidade de Algoritmos Supervisionados Baseada no Segundo Método de Lyapunov	61
4.1	Análise do Algoritmo SDA	61
4.1.1	Análise de resultados	63
4.1.2	Extensão dos resultados	67
4.2	Análise do Algoritmo LMS Utilizando Abordagem de Lyapunov Determinística . . .	68
4.3	Análise do Algoritmo LMS Utilizando Abordagem de Lyapunov Estocástica	71
4.4	Análise de Convergência para LMS	74
4.4.1	Região de Convergência	74
4.4.2	Análise de convergência do LMS utilizando abordagem de Lyapunov determinística	78
4.5	Sumário	86
5	Análise da Estabilidade de Algoritmos Não-Supervisionados Baseada no Segundo Método de Lyapunov	89
5.1	Análise do algoritmo CMA de um coeficiente	89
5.1.1	Formulação para análise de estabilidade	90
5.2	Análise do algoritmo CMA de N coeficientes	97
5.3	Sumário	104
6	Inicialização do Algoritmo CMA	107
6.1	Nova formulação para o algoritmo CMA	107
6.2	Canal com 2 coeficientes	109
6.3	Equalizador com M coeficientes	111
6.4	Simulações	115
6.4.1	Caso A - Canal com dois coeficientes	115
6.4.2	Caso B - Canal com três coeficientes	117
6.4.3	Caso C - Canais com comprimento entre 2 e 5 coeficientes	118
6.5	Sumário	119
7	Conclusão	121
7.1	Perspectivas para trabalhos futuros	123
	Referências bibliográficas	124

A	Apêndice 1 - Resultados do uso de redes wavelets em equalização	131
A.1	Redes Wavelet	131
A.2	Adaptação Estrutural	134
A.3	Equalização com Rede Wavelet	134
A.4	Resultados de Simulação	136
A.5	Sumário	139

Lista de Figuras

1.1	Alguns pontos históricos e contribuições científicas significativas nas áreas de equalização adaptativa e controle adaptativo.	2
2.1	Canal de comunicação.	8
2.2	Esquema para filtragem.	8
2.3	Esquema para equalização.	9
2.4	Representação do equalizador linear e transversal.	11
2.5	Modelo de sistema de comunicação com equalizador adaptativo.	12
2.6	Equalizador adaptativo como sistema dinâmico.	14
2.7	Sistema de equalização supervisionada.	15
2.8	Diagrama de fluxo para o algoritmo SDA.	18
2.9	Diagrama de fluxo para o algoritmo LMS.	20
2.10	Sistema de equalização não-supervisionada.	23
2.11	Diagrama de fluxo para o algoritmo CMA.	24
3.1	Representação geométrica dos conjuntos envolvidos no Teorema 3.	40
3.2	Representação geométrica para estabilidade no sentido de Lyapunov.	41
3.3	Representação geométrica para instabilidade.	41
3.4	Representação geométrica para estabilidade uniforme.	42
3.5	Representação geométrica para uniformemente atrativo.	42
3.6	Função positiva definida e decrescente.	50
3.7	Análise de função decrescente.	51
3.8	Interpretação geométrica para teoria de Lyapunov.	55
3.9	Representação do princípio da invariância.	59
4.1	(a) Superfície de θ para canal $h(z) = 1 + 1,5z^{-1}$ e $d = 2$, sem ruído aditivo. (b) Mesma simulação com ruído aditivo branco e gaussiano (AWGN) com SNR = 20 dB - 100 tentativas de Monte Carlo.	64
4.2	(a) Superfície de θ para canal $h(z) = 1 + 1,5z^{-1}$ e $d = 2$, sem ruído aditivo. (b) Curva de nível da função custo (linha pontilhada) e θ , ou seja, o limitante superior do passo de adaptação (linha contínua).	65
4.3	(a) Superfície de θ para canal $h(z) = 1 - 1,5z^{-1}$ e $d = 2$, sem ruído aditivo. (b) Curva de nível da função custo (linha pontilhada) e θ , ou seja, o limitante superior do passo de adaptação (linha contínua).	66

4.4	(a) Superfície de θ para canal $h(z) = 1 + 0,6z^{-1}$ e $d = 0$, sem ruído aditivo. (b) Curva de nível da função custo (linha pontilhada) e θ , ou seja, o limitante superior do passo de adaptação (linha contínua).	66
4.5	(a) Superfície de θ para canal $h(z) = 1 - 0,6z^{-1}$ e $d = 0$, sem ruído aditivo. (b) Curva de nível da função custo (linha pontilhada) e θ , ou seja, o limitante superior do passo de adaptação (linha contínua).	67
4.6	Superfície que simboliza o limite para o passo de adaptação para equação (4.21). Canal $h(z) = 1 + 0,6z^{-1}$ e $d = 0$	71
4.7	Canal $h(z) = 1 + 0,6z^{-1}$ com atraso $d = 0$. Superfície que simboliza o passo de adaptação para equação (4.26).	72
4.8	Canal $h(z) = 1 + 0,6z^{-1}$ com atraso $d = 0$. Superfície que simboliza o passo de adaptação para equação (4.27).	73
4.9	Superfície representando o limitante superior do passo de adaptação e o ponto onde $\Delta V = 0$	74
4.10	(a) Evolução dos pesos do algoritmo LMS com análise de Lyapunov determinística. Canal $h(z) = 1 + 0,6z^{-1}$ com sinal s_n binário, $\mu = 0,3$. (b) Últimos passos da evolução dos pesos. (c) Curvas de nível de θ e J_{LMS}	76
4.11	(a) Evolução dos pesos do algoritmo LMS com análise de Lyapunov determinística. Canal $h(z) = 1 + 0,6z^{-1}$ com sinal $-1 + 2 * rand(1, N)$, $\mu = 0,3$. (b) Últimos passos da evolução dos pesos. (c) Curvas de nível de θ e J_{LMS}	77
4.12	Canal $h(z) = 1 + 0,6z^{-1}$ com atraso $d = 0$ e $\mu = 0,15$. (a) Superfície θ e respectivas curvas de nível. (b) Curvas de nível e lugar geométrico no sistema de coordenadas $\hat{w}_0 \times \hat{w}_1$. (c) Curvas de nível e lugar geométrico no sistema de coordenadas $\tilde{w}_0 \times \tilde{w}_1$	81
4.13	Canal $h(z) = 1 + 0,6z^{-1}$ com atraso $d = 0$, $\mu = 0,6$. (a) Superfície θ e respectivas curvas de nível. (b) Curvas de nível e lugar geométrico no sistema de coordenadas $\hat{w}_0 \times \hat{w}_1$. (c) Curvas de nível e lugar geométrico no sistema de coordenadas $\tilde{w}_0 \times \tilde{w}_1$	82
4.14	Canal $h(z) = 1 + 1,5z^{-1}$ com atraso $d = 2$, $\mu = 0,1$. (a) Superfície θ e respectivas curvas de nível. (b) Curvas de nível e lugar geométrico no sistema de coordenadas $\hat{w}_0 \times \hat{w}_1$. (c) Curvas de nível e lugar geométrico no sistema de coordenadas $\tilde{w}_0 \times \tilde{w}_1$	83
4.15	Canal $h(z) = 1 + 0,6z^{-1} + 1,5z^{-2}$ com atraso $d = 3$, $\mu = 0,1$. (a) Superfície θ e respectivas curvas de nível. (b) Curvas de nível e lugar geométrico no sistema de coordenadas $\hat{w}_0 \times \hat{w}_1$. (c) Curvas de nível e lugar geométrico no sistema de coordenadas $\tilde{w}_0 \times \tilde{w}_1$	84
4.16	Canal $h(z) = 1 + 0,6z^{-1}$ com atraso $d = 0$, $\mu = 1,2$. (a) Evolução dos pesos do algoritmo LMS. (b) Últimos 50 passos da evolução de pesos.	85
5.1	Esboço da equação (5.10) ($k = 2$, $A = 1$ e $\hat{w}_n = 0,3$).	92
5.2	Esboço da equação (5.10) ($k = 1,0$, $A = 1,1$ e $\hat{w}_n = 0,7$).	92
5.3	Esboço da equação (5.10) ($k = 1,6$, $A = 0,8$ e $\hat{w}_n = 0,2$).	93
5.4	Esboço da equação (5.10) ($k = 1,6$, $A = 0,8$ e $\hat{w}_n = -0,3$).	93
5.5	Análise das raízes ($k = 0,5$, $A = 3 \implies \overline{\hat{w}_{n,2}} = -2$ e $\underline{\hat{w}_{n,2}} = 2$).	96
5.6	Análise das raízes ($k = 1$, $A = 2 \implies \overline{\hat{w}_{n,2}} = -0,5$ e $\underline{\hat{w}_{n,2}} = 0,5$).	97
5.7	Norma das últimas 50 amostras do vetor de pesos $\hat{\mathbf{w}}_n$ para o canal $h(z) = 1 + 1,5z^{-1}$	102
5.8	Histograma para 100 realizações com o canal $h(z) = 1 + 1,5z^{-1}$	103

5.9	Norma das últimas 50 amostras do vetor de pesos \hat{w}_n para o canal $h(z) = 1 + 0,6z^{-1}$.	103
5.10	Histograma para 100 realizações com o canal $h(z) = 1 + 0,6z^{-1}$.	104
6.1	Esboço do equalizador adaptativo polinomial.	109
6.2	Formação do vetor de entrada para inicialização do algoritmo CMA, o que corresponde a transição entre as equações (6.19) e (6.20).	113
6.3	Valor da função custo para simulação com algoritmo de 2 coeficientes.	116
6.4	Valor da função custo para simulação com algoritmo de 3 coeficientes.	118
6.5	Valor da função custo para simulação com algoritmo de 2 a 5 coeficientes.	119
A.1	Representação de um nó da rede wavelet considerando apenas uma única entrada.	132
A.2	Representação da rede wavelet para múltiplas entradas.	133
A.3	Rede wavelet para equalização supervisionada.	135
A.4	Rede wavelet para equalização cega.	136
A.5	Rede wavelet com adaptação estrutural para o caso supervisionado.	137
A.6	Erro instantâneo da simulação da figura A.5.	138
A.7	Evolução do número de nós da simulação da figura A.5.	138
A.8	Rede wavelet com adaptação estrutural para o sistema não-supervisionado.	138
A.9	Diferença entre s_{n-d} e y_n para simulação da figura A.8 (caso não-supervisionado).	138
A.10	Evolução do número de nós da simulação da figura A.8.	139

Lista de Tabelas

6.1	Porcentagem de realizações que apresentam convergência global (%) utilizando a inicialização com a nova heurística, <i>center spike</i> e inicialização aleatória.	117
6.2	Porcentagem de realizações que apresentam convergência global (%) para canal com 3 coeficientes utilizando a inicialização com a nova heurística, <i>center spike</i> e inicialização randômica.	118
6.3	Porcentagem de realizações com convergência global (%) utilizando a inicialização com a nova heurística, <i>center spike</i> e inicialização randômica para o diversos comprimentos de canal.	119
A.1	Parâmetros das redes wavelet utilizadas nas simulações.	139

Glossário

AWGN - *Additive White Gaussian Noise*

BIBO - *Bounded-Input Bounded-Output*

BPSK - *Binary Phase Shift Keying*

CMA - *Constant Modulus Algorithm*

DD - *Decision-Directed Algorithm*

EQM - Erro Quadrático Médio

FFT - *Fast Fourier Transform* - Transformada Rápida de Fourier

FIR - *Finite Impulse Response*

i.i.d. - *independent and identically distributed* - independente e estatisticamente distribuída

IIR - *Infinite Impulse Response*

LMS - *Least Mean Square Algorithm*

MSE - *Mean Squared Error*

PAM - *Pulse Amplitude Modulation*

RLS - *Recursive Least Squares* - algoritmo mínimos quadrados recursivo

SDA - *Steepest Descent Algorithm*

SNR - *Signal-to-Noise Rate* - relação sinal ruído

Lista de Símbolos

$\{.\}$	sequência de amostras;
$*$	convolução de sequências discretas;
$\ .\ $	norma de um vetor ou de uma matriz. Neste trabalho é considerada apenas a norma-2 no espaço \mathcal{L}^2 ;
$abs(.)$	valor absoluto;
$det(.)$	determinante de uma matriz;
$diag(.)$	matriz diagonal;
$inf(.)$	operador infimum;
$sign(.)$	função sinal;
$sup(.)$	operador supremo;
$tr(.)$	traço de uma matriz;
\langle, \rangle	produto interno;
$.\ $	valor absoluto;
$(.)^H$	Hermitiano de um vetor ou matriz;
$(.)^*$	complexo conjugado;
$(.)^T$	transposto de matriz;
$E\{.\}$	operador esperança;
a	variável escalar;
\mathbf{a}	vetor ou matriz;
s_n	sinal transmitido ou sinal de referência;
d	número de amostras atrasadas, i.e. s_{n-d} ;
e_n	$s_{n-d} - y_n$: erro de estimação;
\mathbf{w}_n	pesos do filtro transversal no instante n ;
$\hat{\mathbf{w}}$	estimativa de \mathbf{w} ;
$\underline{\mathbf{w}}$	ponto de equilíbrio;
\mathbf{w}^*	ponto de equilíbrio;
\mathbf{w}_W	mínimo de Wiener;
\mathbf{p}	$E\{\mathbf{x}_n s_{n-d}^*\}$: vetor de correlação cruzada;
\mathbf{R}	$E\{\mathbf{x}_n \mathbf{x}_n^H\}$: matriz de correlação;
$\mathcal{L}^2(\mathbb{R})$	$\ f(t)\ _2 = \int f(t) ^2 dt < +\infty$: funções de energia finita;
$\mathcal{L}^p(\mathbb{R})$	espaço com norma $\ f(t)\ _p = (\int_0^\infty f(t) ^p dt)^{\frac{1}{p}} < +\infty$;
$\mathcal{L}^\infty(\mathbb{R})$	espaço com norma $\ f(t)\ _\infty = \sup_{t \geq 0} \ f(t)\ < +\infty$;

$\ell^p(\mathbb{Z})$	espaço de sequência, com $p > 0$, com norma $(\sum_n f_n ^p)^{\frac{1}{p}} < +\infty$;
$\ell^\infty(\mathbb{Z})$	espaço de sequência, com norma definida por $\ f_n\ = \sup_n f_n $;
\mathbb{C}	conjunto dos números complexos;
\mathbb{Z}	conjunto dos números inteiros;
\mathbb{N}	conjunto dos números naturais;
\mathbb{R}	conjunto dos números reais;
σ_s^2	variância do sinal s_n ;
Ω	espaço amostral;
\mathfrak{F}	família de eventos;
P	função probabilística;
\mathbf{Z}	espaço de dimensão n de vetores randômicos;
λ	autovalores da matriz de correlação \mathbf{R} ;
μ	passo de adaptação de algoritmos adaptativos;
θ	função que limita superiormente.

Trabalhos Publicados e Preparados Pelo Autor

1. Denis Gustavo Fantinato, Wesley Pavan, Celso de Sousa Júnior, Ricardo Suyama, Romis R. de F. Attux, João Marcos Travassos Romano. Lower Bound and Initialization Heuristic Derived from a Polynomial Formulation of the Constant Modulus Criterion. Submetido para *Signal Processing*.
2. Celso de Sousa Júnior, Romis R. de F. Attux, Ricardo Suyama, João Marcos T. Romano. Lyapunov-Based Stability Analysis of Supervised and Unsupervised Adaptive Algorithms. *IEEE International Symposium on Circuits and Systems (ISCAS2010)*, Paris, France, maio de 2010.
3. Celso de Sousa Júnior, Ricardo Suyama, Romis R. de F. Attux, João Marcos T. Romano, Análise de Estabilidade do LMS segundo a Teoria de Lyapunov, XXVI Simpósio Brasileiro de Telecomunicações (SBrT 2008), Rio de Janeiro, Brasil, setembro de 2008.
4. Celso de Sousa Júnior, Romis R. de F. Attux, João Marcos T. Romano, Esquemas para filtragem adaptativa baseados em redes wavelet, XXII Simpósio Brasileiro de Telecomunicações (SBrT 2005), Campinas, Brasil, setembro de 2005.

Capítulo 1

Introdução

A análise de estabilidade de algoritmos utilizados em filtragem adaptativa é um assunto que remonta à década de 60 [1]. Apesar de muitos algoritmos terem natureza estocástica, por vezes, recorre-se a análises de estabilidade e convergência que consideram versões determinísticas ou simplificações que implicam localidade ou restringem as características estatísticas dos sinais e modelos envolvidos. Consequentemente, surge a pergunta: como obter condições mais gerais de estabilidade de um algoritmo de filtragem adaptativa?

Uma técnica utilizada na engenharia de controle para o estudo de estabilidade de sistemas não-lineares é a teoria de Lyapunov. Essa teoria abrange um ramo que faz uso da linearização - o chamado primeiro método ou método indireto de Lyapunov - assim como uma formulação que utiliza o modelo do sistema sem simplificações, que é denominada segundo método ou método direto de Lyapunov. Tal arcabouço foi empregado pela primeira vez nos problemas de controle em 1966 no trabalho de Parks [2]. Mais tarde, o uso da teoria foi se estendendo a outros casos relevantes de controle adaptativo [3], chegando até mesmo a ser utilizada no controle baseado em redes neurais [4]. Vale ressaltar que a teoria de Lyapunov pode ser aplicada a sistemas estocásticos [5][6], merecendo destaque os sistemas markovianos [7].

Tendo em vista o sucesso das aplicações baseadas na teoria de Lyapunov no âmbito do problema de controle, temos, como primeira meta deste trabalho, o emprego da mesma abordagem para o caso de filtragem adaptativa, mais especificamente em equalização. Isso se justifica, num primeiro momento, pelo grande número de pontos de contato entre os problemas de controle adaptativo e de equalização adaptativa. De maneira resumida, podemos enumerar:

- ambos são problemas onde o ótimo pode mudar de localização devido a mudanças nas condições de modelamento ou variações devido ao ruído que podem ocorrer ao longo do processo de controle ou equalização;

- em geral, temos dinâmica não modelada e presença de ruído;
- existem algoritmos dedicados ao ajuste de parâmetros (do controlador ou equalizador) que necessitam de regras de projeto ou mesmo de inicialização.

Do ponto de vista histórico, podemos traçar um paralelo entre as áreas de controle adaptativo e equalização adaptativa. A figura 1.1 enumera alguns eventos relevantes a fim de comparar a evolução das duas áreas.

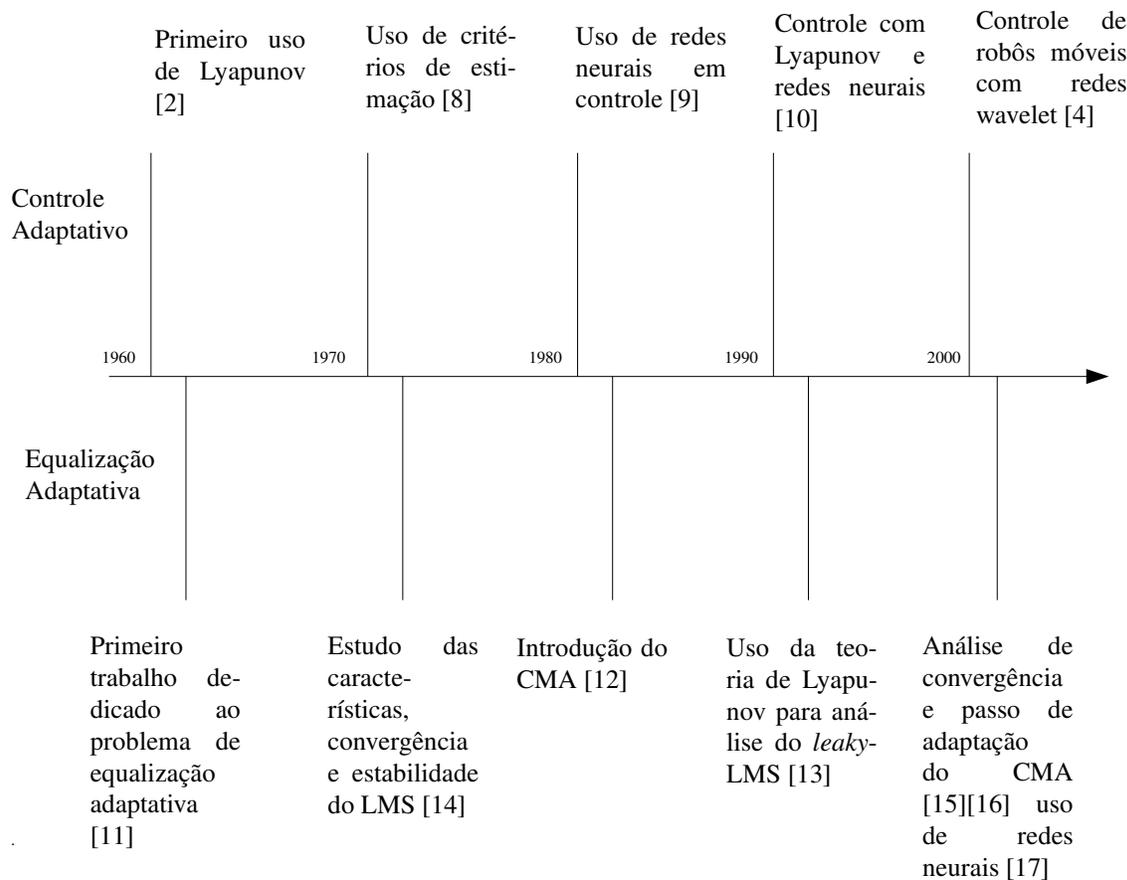


Fig. 1.1: Alguns pontos históricos e contribuições científicas significativas nas áreas de equalização adaptativa e controle adaptativo.

Outra justificativa muito forte para o uso da teoria de Lyapunov em filtragem adaptativa é o fato de que os algoritmos mais clássicos dessa área não tiveram sua estabilidade analisada a partir dessa teoria, apesar de técnicas mais específicas - (*Leaky-LMS* [13], *NAS-RIF* [18] e *Gradiente Conjugado* [19]) - já terem sido investigadas sob esse prisma. Em suma, como primeira e principal linha de

contribuição desta tese temos a análise das condições - em termos do passo de adaptação - para que algoritmos de equalização adaptativa sejam estáveis no sentido de Lyapunov. Buscar-se-á evitar o recurso a simplificações como a teoria da independência, visando assim obter resultados tão gerais quanto possível.

A seguir, discutiremos as referências que foram mais relevantes para desenvolvimento dessa primeira linha de contribuição.

1.1 Alguns trabalhos relevantes para o estudo de estabilidade em equalização adaptativa

Primeiramente, desejamos mencionar explicitamente o trabalho de Kundur e Hatzinakos [18], que faz uso da teoria de Lyapunov em uma versão determinística para o problema de filtragem aplicado ao tratamento de imagens. Apesar de ser um caso bastante específico, foi possível extrair um resultado muito relevante para a pesquisa desenvolvida: o uso da função custo como candidata a função de Lyapunov. Isso resolve um problema importante, pois, em geral, encontrar a função de Lyapunov não é uma tarefa elementar [20]. Por outro lado, vale frisar que a teoria de Lyapunov, mais especificamente o método direto desta teoria, evita a necessidade de buscar explicitamente pontos de equilíbrio do sistema dinâmico analisado [20][21], o que é uma vantagem, especialmente em contextos não-lineares [22].

No trabalho [23], são apresentadas as condições para que um algoritmo de equalização adaptativa (neste caso específico o LMS normalizado) tenha estabilidade assintótica exponencial. As restrições impostas por [23] são relativas à matriz de autocorrelação ou a matrizes formadas a partir do vetor composto pelo sinal de entrada do equalizador. Contudo, essa abordagem exige que se determine o menor autovalor da matriz de autocorrelação (ou de uma matriz equivalente à matriz de autocorrelação). Portanto, a técnica fica um tanto restrita a sistemas lineares ¹ e não faz alusão ao valor de passo de adaptação a ser considerado. Os resultados de [23] fazem uso do segundo método de Lyapunov na versão determinística e de elementos da teoria matemática de controle como o gramiano ².

Mais de uma década após o trabalho de Bitmead [23], surge o trabalho de Solo [24], no qual as condições para estabilidade do algoritmo LMS são relaxadas. Chega-se à conclusão que o LMS é exponencialmente estável quando o sinal apresenta momentos de quarta ordem limitados superiormente, sinais com excitação persistente e passo de adaptação pequeno. Mais uma vez, o valor do

¹Isso é razoável, pois, na década de 80, o problema de filtragem adaptativa não-supervisionada estava apenas começando a ser investigado, o que traria, como consequência, algoritmos e sistemas não-lineares.

²O conceito de gramiano é utilizado para o estudo de controlabilidade de sistemas lineares. Portanto, o vetor do sinal de entrada do filtro é utilizado para formar uma matriz que é considerada no uso do gramiano. Consequentemente, essa formulação apresenta exigência no modelo do canal para alcançar uma representação simples.

passo de adaptação fica indeterminado ou pelo menos não foi estudado a contento, pois o esforço ficou concentrado nas discussões sobre as características estatísticas do sinal transmitido.

No caso de uma versão modificada do algoritmo LMS chamada *leaky-LMS*, é possível encontrar as condições de estabilidade (como função do passo de adaptação) por intermédio do uso da teoria de Lyapunov e simplificações referentes ao sinal transmitido e modelo utilizado [25]. Veja que o algoritmo *leaky-LMS* apresenta performance melhor que a do LMS, principalmente se for considerado um ambiente com sinais não estacionários e sem excitação persistente. Mesmo assim, a teoria de Lyapunov é uma metodologia válida [26]. Podemos observar também a aplicação da teoria de Lyapunov com versão determinística para o caso de algoritmos de filtragem adaptativa baseados no gradiente conjugado [19]. Vale ressaltar que, como apenas o passo de adaptação pode ser ajustável (além, é claro, da própria inicialização), esse é o único parâmetro disponível para ajuste do algoritmo com o fim de obter uma condição de estabilidade.

Conseqüentemente, para apresentar uma visão um pouco mais ampla, este trabalho considerou como referência os valores clássicos determinados em [27] e [28] para o passo de adaptação do LMS. Vale ressaltar que, conforme mencionado em [28], uma das heurísticas utilizadas para determinar um limitante para o passo de adaptação faz uso do conceito de *direct-averaging* [29]. O método de *direct-averaging* considera que o resultado do sistema markoviano é praticamente o mesmo de um segundo sistema onde os parâmetros não são variáveis aleatórias. Portanto, é possível avançar consideravelmente na análise de estabilidade do algoritmo.

É bom lembrar que o conceito de Lyapunov é uma noção específica de estabilidade. Existem outros conceitos de estabilidade no estudo de sistemas dinâmicos ou mesmo na área de processamento digital de sinais (exemplo estabilidade BIBO - *Bounded-Input Bounded-Output*)³. Para os algoritmos de equalização adaptativa, existem ainda as discussões de outros conceitos como estabilidade no sentido médio quadrático e também de estabilidade *almost sure* [31]. Há também o conceito de convergência, o qual é diferente de estabilidade, estando relacionada à quantificação do grau de proximidade dos resultados dos algoritmos em relação a uma solução ótima⁴.

No caso específico do algoritmo LMS, tem-se uma faixa de passo de adaptação para a qual é possível observar convergência com probabilidade 1, mas divergência⁵ no sentido médio quadrático; uma segunda faixa para o passo onde temos divergência quase sempre (e também no sentido médio quadrático); e uma última faixa onde o algoritmo converge no sentido médio quadrático [25].

³Apenas como nota informativa, a referência [30, p.20] traz quatro definições formais para o conceito de estabilidade do ponto de vista de sistemas dinâmicos.

⁴Uma sequência de variáveis aleatórias $\{r_n\}$ converge "almost sure", ou com probabilidade 1, para r se para todo ponto w pertencente ao evento não nulo A , temos que $\lim_{n \rightarrow \infty} |r_n(w) - r(w)| = 0$ [32]. A convergência com relação a p -ésima média ($p > 0$) para r é dada por $\lim_{n \rightarrow \infty} E\{|r_n(w) - r(w)|^p\} \rightarrow 0$ [32].

⁵Divergência pode ser definida matematicamente como uma realização do algoritmo que leva $\lim_{n \rightarrow \infty} \|\hat{\mathbf{w}}_n - \mathbf{w}_{\text{ótimo}}\| = \infty$. Note que essa não é uma definição formal, mas suficiente para representar a idéia de divergência [33].

Contudo, podemos observar que a análise de estabilidade de algoritmos para equalização adaptativa vem tipicamente acompanhada de hipóteses para tornar o problema tratável do ponto de vista matemático, dentre as quais se destaca a teoria da independência [28].

Os resultados sobre as faixas de passo de adaptação do CMA que levam a situações de estabilidade são relativamente recentes. O primeiro trabalho que conhecemos que trata especificamente desse assunto é [16]. Nesse caso, é feita a análise estocástica do CMA com o auxílio de hipóteses interessantes em termos de tratabilidade, como a suposição de que a solução ótima é a própria solução *zero-forcing*. Outra observação é o uso de hipóteses com algum grau de similaridade com respeito à teoria da independência. Recentemente, um trabalho correlato foi publicado, mas fazendo uso de multiplicadores de Lagrange [34]. Contudo, julgamos que o resultado mais amplo até o momento continua sendo [16].

Na próxima seção, discutiremos a estrutura desta tese e mencionaremos as principais contribuições apresentadas.

1.2 Organização e contribuições do trabalho

No capítulo 2, descrevemos os resultados fundamentais da teoria de equalização adaptativa que serão de maior relevância para este trabalho. A teoria de Lyapunov é discutida no capítulo 3 como preparação para os capítulos 4 e 5, que descrevem as propostas e resultados obtidos no sentido de aplicá-la ao estudo de métodos clássicos de equalização.

No curso do desenvolvimento do trabalho, elaboramos uma análise para questões de "equalizabilidade" e inicialização para algoritmos não-supervisionados. Tomando como base uma estrutura de filtro polinomial [35], discutimos esse tópico no capítulo 6, concentrando-nos no estudo do algoritmo CMA.

O capítulo 7 apresenta as conclusões e perspectivas de trabalhos futuros e há um apêndice com alguns resultados ligados a equalização não-linear que foram obtidos no curso deste trabalho de doutoramento.

Em síntese, consideramos como principais contribuições do trabalho:

1. Estudo de estabilidade do algoritmo SDA por intermédio do segundo método de Lyapunov, vide capítulo 4;
2. Análise de estabilidade local do algoritmo LMS pelo segundo método de Lyapunov considerando versões estocástica e determinística, vide capítulo 4;
3. Estudo da região de convergência e *misadjustment* para o algoritmo LMS como uma consequência da análise de Lyapunov, vide capítulo 4;

4. Estudo da estabilidade local algoritmo CMA segundo o método de Lyapunov, vide capítulo 5;
5. Análise do critério CM e proposta de inicialização do algoritmo CMA para sinais bipolares, vide capítulo 6;
6. Análise de viabilidade para utilização de redes wavelets como equalizadores supervisionados e não-supervisionados, vide apêndice A.

Capítulo 2

Equalização Adaptativa Supervisionada e Não-Supervisionada

Neste capítulo, serão apresentados elementos teóricos sobre equalização adaptativa, bem como a notação a ser utilizada e alguns resultados fundamentais relativos ao estudo da estabilidade dos algoritmos SDA, LMS e CMA. Tudo isso servirá como referência para a análise, em capítulos futuros, dos resultados que formam o cerne desta tese.

2.1 Problema de Transmissão Digital

Até os anos 60, pode-se dizer que o processamento de sinais era predominantemente baseado em tecnologia analógica. Entretanto, a evolução dos computadores digitais e microprocessadores, juntamente com avanços teóricos importantes como a proposição da transformada rápida de Fourier (FFT), impulsionaram metodologias de processamento digital da informação. Desde então, essas metodologias têm desempenhado um papel fundamental na evolução de sistemas de tratamento da informação em domínios que vão da engenharia biomédica às telecomunicações [36].

Um problema muito importante em processamento digital de sinais corresponde ao de, a partir de um sinal de entrada, estimar outro sinal, considerado de interesse ou relevante para o cumprimento de outra tarefa (e.g. a identificação de um sistema desconhecido). Existem casos em que o processamento não é aplicado para obter um segundo sinal, mas sim para a caracterização do sinal de entrada.

Quando se lida com transmissão de sinais digitais, temos, essencialmente, um sinal a ser transmitido, o meio de comunicação ao qual o sinal será confiado e o sinal recebido. Portanto, dada a informação originalmente transmitida s_n , após o canal de comunicação (que pode corresponder a um par de fios de cobre, fibras óticas ou canais sem fio, entre outros meios), recebemos a informação x_n .

Tipicamente, o sinal recebido x_n não será igual ao sinal transmitido s_n ou a uma versão atrasada sua, o que dá origem a um problema crucial em comunicações. A figura 2.1 ilustra o processo.

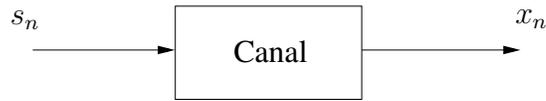


Fig. 2.1: Canal de comunicação.

O sinal transmitido s_n pode ser gerado de acordo com diferentes esquemas de codificação e modulação. Neste trabalho, o sinal transmitido obedecerá, como regra, às seguintes especificações:

- s_n é composto de amostras bipolares com valores $+1$ e -1 , como é o caso, por exemplo, em uma modulação do tipo BPSK em banda base;
- o sinal tem média nula, ou seja, $E\{s_n\} = 0$;
- s_n é composto de amostras independentes entre si e identicamente distribuídas, ou seja, é i.i.d.¹.

Uma das formas de mitigar os efeitos nocivos do canal é utilizar a técnica conhecida como filtragem. Em termos simples, o processo de filtragem consiste em construir algum tipo de estimativa combinando um conjunto de amostras de sinais disponíveis até um certo momento [28].

O elemento que executa a função de filtragem é conhecido como filtro. A figura 2.2 apresenta um esquema simples de filtragem, e já introduz a notação matemática a ser considerada. O filtro recebe a informação x_n e trabalha para obter - no contexto do problema que temos descrito - a estimativa do sinal transmitido original s_n . Como se trata de uma estimativa, o filtro produz como resultado \hat{s}_n .



Fig. 2.2: Esquema para filtragem.

¹i.i.d significa *independent and identically distributed*, o que quer dizer que as amostras do sinal são independentes e obedecem à mesma distribuição estatística.

O problema de filtragem no cenário que temos discutido corresponde à idéia de equalização. Dessa forma, o equalizador se destina a eliminar os efeitos nocivos e distorções causadas pelo meio de transmissão na mensagem enviada, ou seja, ele pode ser entendido como uma espécie de "modelo inverso" do canal. Portanto, a equalização é um caso particular de filtragem.

A figura 2.3 apresenta um esboço do sistema contendo o canal e o equalizador, sendo a_n o sinal de saída do canal, b_n o ruído aditivo, x_n o sinal de entrada do equalizador e y_n a saída do equalizador. Os parâmetros do canal e do equalizador são representados pelos vetores \mathbf{h}_n e \mathbf{w}_n , respectivamente. Note ainda que a figura 2.3 considera que a estrutura de equalização é linear, que corresponde a um compromisso interessante entre desempenho e tratabilidade matemática. Lidaremos, em geral, com esse caso, embora resultados apresentados aqui sejam extensíveis ao caso não-linear. Vale ressaltar que apesar das estruturas serem lineares, os algoritmos utilizados para equalização podem ser não-lineares, o que justifica o uso de ferramentas de análise de sistemas não-lineares, como a teoria de Lyapunov.

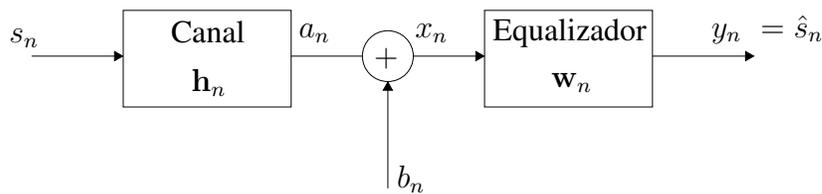


Fig. 2.3: Esquema para equalização.

Um efeito do canal particularmente nocivo em transmissão digital é aquele conhecido como interferência intersimbólica. Além dessa interferência, temos, no modelo da figura 2.3, o ruído aditivo b_n como representação de erros de medição ou dinâmica não modelada do canal. Ao utilizar um ruído aditivo, não contemplamos perturbações do tipo multiplicativas ou mesmo aspectos não-lineares, mas o modelo apresentado é o padrão em processamento de sinais para comunicações [28][8][37][21][22], sendo, portanto, suficientemente representativo para nossos fins.

Em um contexto linear, o sinal transmitido s_n é modificado via convolução com os elementos do vetor de coeficientes do canal h_n e adição do ruído b_n , produzindo o sinal recebido x_n :

$$\{x_n\} = \{s_n\} * \{h_n\} + \{b_n\}, \quad (2.1)$$

$$x_n = \sum_{k=-\infty}^{\infty} h_k s_{n-k} + b_n. \quad (2.2)$$

Se considerarmos um atraso de recuperação não-nulo d , isto é o tempo necessário para recuperar uma réplica da informação transmitida [28], podemos escrever:

$$x_n = h_d s_{n-d} + \sum_{k=-\infty, k \neq d}^{\infty} h_k s_{n-k} + b_n. \quad (2.3)$$

Como a resposta ao impulso do canal é geralmente não-trivial, o segundo termo da equação (2.3) acaba se tornando um indicativo de interferência entre os símbolos transmitidos (interferência intersimbólica) [38]. Portanto, o equalizador terá como objetivo tratar o sinal recebido x_n para diminuir ou eliminar a influência dos termos $\sum_{k=-\infty, k \neq d}^{\infty} h_k s_{n-k}$ e b_n sobre o sinal recebido. Caso o equalizador seja capaz de eliminar totalmente a influência desses dois termos, então diremos que ele terá atingido uma solução do tipo *zero-forcing* [28].

A fim de delimitar o escopo desta discussão, consideremos que o canal e o equalizador são filtros FIR representados, respectivamente, pelos vetores de parâmetros \mathbf{h}_n e \mathbf{w}_n , conforme mostra a figura 2.3. Os filtros FIR são representativos de canais de comunicação, e, do ponto de vista de equalização, em contraste com estruturas IIR, são intrinsecamente estáveis [36][39]. O vetor \mathbf{w}_n tem a seguinte estrutura:

$$\mathbf{w}_n = [w_{0,n} \quad w_{1,n} \quad w_{2,n} \quad \cdots \quad w_{M-1,n}]^T, \quad (2.4)$$

onde M é o número de elementos do vetor. Veja que o índice n no vetor mostra de maneira explícita que a solução é, como veremos adiante, adaptativa [27]. Em síntese, o elemento $w_{0,n}$ corresponde ao peso w_0 no instante n .

De maneira similar, o vetor correspondente aos coeficientes do canal, ou seja, \mathbf{h}_n , pode ser escrito da seguinte forma:

$$\mathbf{h}_n = [h_0 \quad h_1 \quad h_2 \quad \cdots \quad h_{N-1}]^T, \quad (2.5)$$

sendo N o comprimento do canal. Se tomarmos os elementos de \mathbf{h}_n como sendo uma sequência, teremos a resposta ao impulso do canal. Deste ponto em diante, faremos um abuso da notação no que se refere a $w_{0,n}$, utilizando apenas w_0 , para evitar um texto muito carregado.

Hipótese 1. *O vetor de entrada do equalizador, \mathbf{x}_n , é composto de versões atrasadas do sinal recebido x_n . A estrutura do vetor \mathbf{x}_n é a seguinte:*

$$\mathbf{x}_n = [x_n \quad x_{n-1} \quad x_{n-2} \quad \cdots \quad x_{n-L+1}]^T. \quad (2.6)$$

Outra maneira de expressar a equação (2.1) é utilizar a matriz de convolução do canal. Neste caso, a expressão dos sinais de saída do canal pode ser escrita em notação matricial:

$$\mathbf{x}_n = \begin{bmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_{n-L+1} \end{bmatrix}_{L \times 1} = \underbrace{\begin{bmatrix} h_0 & h_1 & \cdots & 0 \\ 0 & h_0 & h_1 & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{N-1} \end{bmatrix}}_H \begin{bmatrix} s_n \\ s_{n-1} \\ \vdots \\ s_{n-N+1} \end{bmatrix}_{N \times 1} + \mathbf{b}_n, \quad (2.7)$$

onde L é o comprimento do equalizador e N é o comprimento do canal.

O equalizador é um sistema linear com resposta ao impulso finita (FIR). Consequentemente,

$$y_n = \mathbf{w}_n^T \mathbf{x}_n. \quad (2.8)$$

A expressão do sinal y_n é equivalente a uma soma de convolução. Tal modelo de equalizador é ilustrado na figura 2.4.

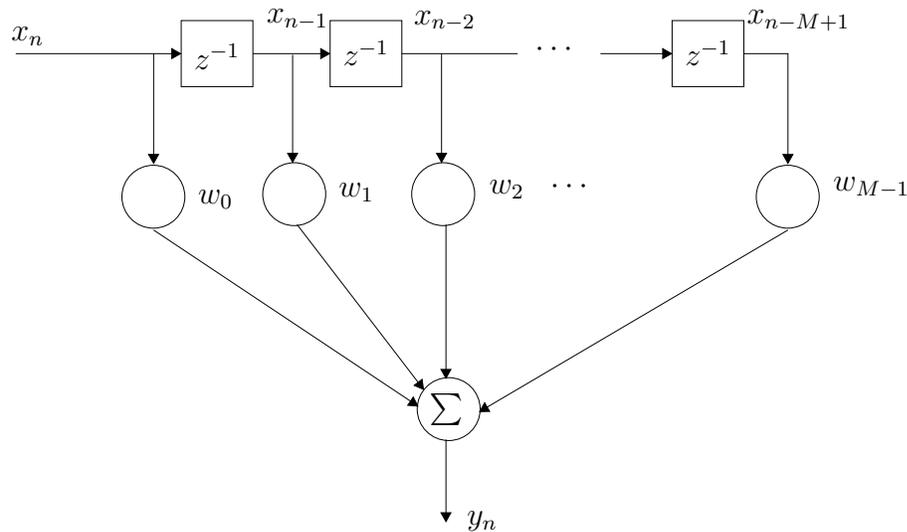


Fig. 2.4: Representação do equalizador linear e transversal.

Cada w_k ($k = 0, 1, \dots, M-1$) deverá ter um valor bem escolhido a fim de que se obtenha a correta equalização do sinal x_n . O projeto do equalizador corresponde, nesse caso, a calcular os valores de w_k , bem como a determinar o tamanho do vetor \mathbf{w}_n , ou seja, o parâmetro M .

O objetivo do equalizador é obter a melhor estimativa do sinal do transmitido \hat{s}_n . Portanto, o equalizador trabalha para eliminar os efeitos do canal de comunicação e ruído aditivo. Contudo, nem sempre o projeto de um equalizador pode ou deve ser feito em forma fechada. Surge, portanto, a possibilidade de trabalhar com uma classe de equalizadores que sejam capazes de se adaptar a fim de encontrar dinamicamente a melhor estimativa possível. Esse tipo de equalizador é conhecido como equalizador adaptativo.

Na próxima seção discutiremos a essência da idéia de equalização adaptativa. Apresentaremos o equalizador adaptativo, bem como uma descrição típica do modo pelo qual o equalizador pode ser ajustado.

2.2 Equalização Adaptativa

A grande vantagem dos sistemas adaptativos é a possibilidade de ajuste dinâmico de parâmetros e adaptação a novas situações de trabalho ao longo do processamento da informação recebida. Contudo, um cuidado extra que devemos ter com esse tipo de sistema é a boa especificação do objetivo a ser cumprido. Se este é bem definido e factível, há, certamente, maior chance de êxito no processo de equalização.

Partindo da equação (2.3), notamos que, além do sinal transmitido s_n , temos a influência do canal de comunicação h_n e do ruído aditivo b_n na composição do sinal recebido.

Caso o canal sofra alterações ao longo do tempo, ou mesmo caso tenhamos ruído de natureza não-estacionária, então as características do sinal de entrada do equalizador x_n serão alteradas continuamente. Portanto, como indicado anteriormente, surge a necessidade de adotar uma solução adaptativa para a compensação dos efeitos nocivos existentes.

Por conseguinte, o sistema da figura 2.3 evolui para o cenário da figura 2.5. Apesar da relativa simplicidade da representação da figura 2.5, podemos identificar os principais elementos relevantes para este trabalho: canal, equalizador e algoritmo adaptativo para ajuste do equalizador.

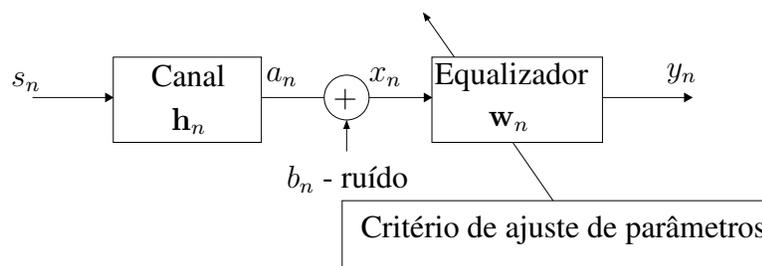


Fig. 2.5: Modelo de sistema de comunicação com equalizador adaptativo.

A próxima questão que surge é como ajustar os parâmetros do equalizador adaptativo. Para isso, em geral, parte-se de um critério de adaptação (vide figura 2.5), o qual consiste, em poucas palavras, na descrição matemática do objetivo que o processo de adaptação deve buscar. Em geral, os algoritmos de ajuste de parâmetros do equalizador são construídos com base em métodos de busca operando nas superfícies de custo originadas pelo critério de adaptação. O método mais usual é o *steepest descent*, embora o método de Newton também tenha um papel relevante [27]. No caso do método *steepest descent*, é utilizado o vetor gradiente para indicar a direção minimizante.

Podemos escrever a expressão de adaptação dos pesos do equalizador como

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{1}{2}\mu \nabla J(\mathbf{w}_n), \quad (2.9)$$

onde $J(\cdot)$ é a função custo que caracteriza cada critério de equalização adaptativa. Essa formulação pode ser descrita como algoritmo de busca do tipo gradiente para o caso discreto com adaptação *on-line* [40].

No caso em que não dispomos *a priori* de toda a informação estatística necessária, o vetor gradiente ∇J pode ter expressão exata desconhecida. No entanto, é possível utilizar estimativas, por exemplo uma estimativa instantânea de ∇J , para promover a adaptação. Isso fará com que o algoritmo determinístico se torne estocástico, pois estimativas como a instantânea contêm termos aleatórios que antes eram eliminados por intermédio do operador $E\{\cdot\}$.

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{1}{2}\mu \hat{\nabla} J(\mathbf{w}_n). \quad (2.10)$$

Observamos nas expressões (2.9) e (2.10) que existe um termo μ que controla a velocidade de adaptação dos parâmetros. Esse termo é conhecido com passo de adaptação.

Hipótese 2. Consideramos que o passo de adaptação μ é sempre fixo e positivo. Ou seja,

$$0 < \mu < \mu_{max}. \quad (2.11)$$

O equalizador, cujos parâmetros se modificam ao longo do tempo devido ao algoritmo de adaptação, pode ser considerado um sistema dinâmico. Dessa forma, podemos escrever:

$$\mathbf{w}_{n+1} = f(\mathbf{w}_n, \dots), \quad (2.12)$$

Se $f(\cdot)$ é não-linear, o algoritmo pode originar um leque de comportamentos bastante amplo, abrangendo mesmo ciclos-limite e caos. Na figura 2.6, explicitamos a interpretação do equalizador adaptativo como um sistema dinâmico, o que será de suma importância nos próximos passos do trabalho.

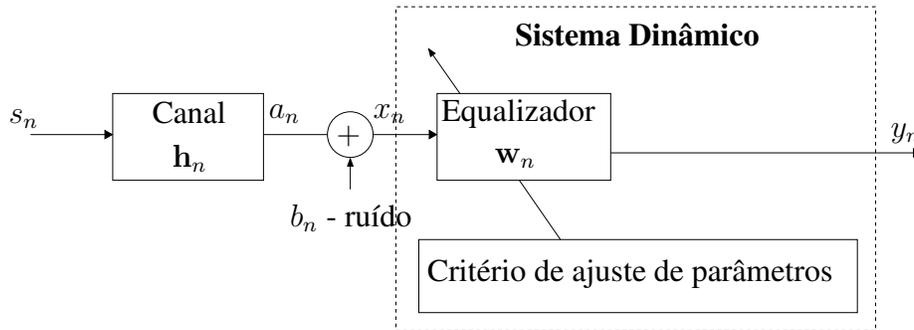


Fig. 2.6: Equalizador adaptativo como sistema dinâmico.

Na próxima seção discutiremos a técnica de equalização adaptativa supervisionada. Essa técnica considera que temos um sinal de referência para orientar o processo de ajuste do equalizador adaptativo. Consequentemente, temos um cenário propício para o desenvolvimento dos princípios teóricos e algoritmos chaves que serão considerados posteriormente. Vale ressaltar que a equalização supervisionada é um problema que encontra análogos nos estudos de engenharia, como o problema de controle adaptativo.

2.3 Equalização Adaptativa Supervisionada

Quando se lida com equalização supervisionada, supõe-se que o critério de equalização - e, conseqüentemente, o algoritmo dele derivado - é construído tendo por base um sinal de referência. Nesse caso, o cenário de adaptação é relativamente tratável de um ponto de vista analítico, o que será uma motivação interessante no prosseguimento da tese.

Fundamentalmente, consideraremos que a inversão do canal será buscada por meio de uma referência que corresponderá ao sinal transmitido ou a uma versão atrasada do mesmo. Assim, o diagrama da figura 2.5 pode tomar uma forma mais específica, conforme mostra a figura 2.7.

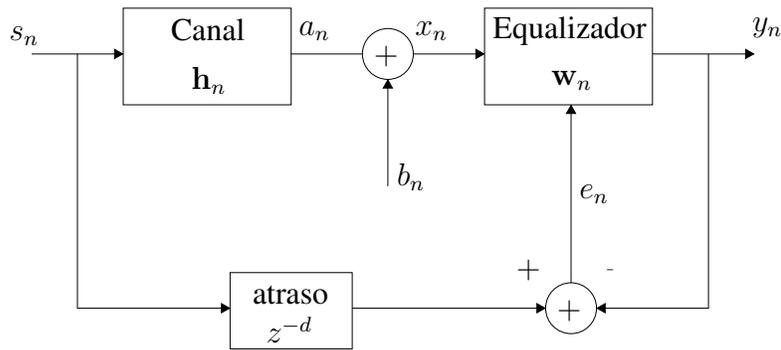


Fig. 2.7: Sistema de equalização supervisionada.

O sinal e_n representa o erro entre o sinal de referência s_{n-d} e o sinal equalizado y_n :

$$e_n = s_{n-d} - y_n. \quad (2.13)$$

Note que, na figura 2.7, temos o elemento z^{-d} , o que representa o atraso em d amostras para o sinal transmitido s_n .

Hipótese 3. *Os sinais abaixo são considerados estatisticamente estacionários no sentido amplo (ou seja, suas propriedades estatísticas são invariantes sob mudança no tempo):*

- s_n - sinal transmitido
- x_n - sinal de entrada do equalizador.

Um equalizador que opera em modo supervisionado tem acesso a um sinal de referência, e, conseqüentemente, orienta-se a construção de um critério supervisionado no sentido de fazer uso dessa informação. Na próxima seção discutiremos o critério de Wiener, que corresponde à base de estudo e análise de muitos resultados de equalização supervisionada e mesmo de equalização não-supervisionada.

2.3.1 Critério de Wiener

O critério de Wiener se vincula a uma formulação do problema de filtragem ótima tendo por base uma medida de erro quadrático médio. Neste trabalho, como indicado anteriormente, o critério será aplicado ao ajuste de um equalizador FIR [28].

Idealmente, a saída do equalizador deve ser uma cópia fiel da mensagem transmitida a menos de um possível atraso d e um fator de escala α , ou seja, $y_n = \alpha s_{n-d}$. Portanto, é natural que o ajuste dos

parâmetros do equalizador seja guiado pela idéia de se obter um sinal tão próximo quanto possível de s_{n-d} : isso justifica intuitivamente o critério de Wiener [28]. Matematicamente, os pesos ótimos do equalizador são tais que minimizam a seguinte função custo

$$J_{Wiener}(\mathbf{w}_n) = E \{ (s_{n-d} - y_n)^2 \} = E \{ (e_n)^2 \}, \quad (2.14)$$

A equação (2.14) pode ser desenvolvida tomando o quadrado do erro e considerando (2.8), levando à expressão a seguir

$$\begin{aligned} J_{Wiener}(\mathbf{w}_n) &= E \{ s_{n-d}^2 - s_{n-d} \mathbf{w}_n^T \mathbf{x}_n - s_{n-d} \mathbf{x}_n^T \mathbf{w}_n + \mathbf{w}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{w}_n \} \\ &= E \{ s_{n-d}^2 \} - E \{ s_{n-d} \mathbf{w}_n^T \mathbf{x}_n \} - E \{ s_{n-d} \mathbf{x}_n^T \mathbf{w}_n \} + E \{ \mathbf{w}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{w}_n \}. \end{aligned} \quad (2.15)$$

É bom lembrar que assumimos que s_{n-d} , \mathbf{x}_n são estritamente estacionários (hipótese 3). O vetor de pesos do equalizador é um valor fixo, portanto, não é uma grandeza estocástica. Definindo $\mathbf{R} = E \{ \mathbf{x}_n \mathbf{x}_n^T \}$ como a matriz de autocorrelação do sinal x_n e $\mathbf{p} = E \{ \mathbf{x}_n s_{n-d}^* \}$ como o vetor de correlação cruzada entre o sinal recebido e o sinal desejado [27], a equação (2.15) pode ser reescrita como

$$\begin{aligned} J_{Wiener}(\mathbf{w}_n) &= E \{ s_{n-d}^2 \} - \mathbf{w}_n^T E \{ s_{n-d} \mathbf{x}_n \} - E \{ s_{n-d} \mathbf{x}_n^T \} \mathbf{w}_n + \mathbf{w}_n^T E \{ \mathbf{x}_n \mathbf{x}_n^T \} \mathbf{w}_n \\ &= E \{ s_{n-d}^2 \} - \mathbf{w}_n^T \mathbf{p} - \mathbf{p}^T \mathbf{w}_n + \mathbf{w}_n^T \mathbf{R} \mathbf{w}_n. \end{aligned} \quad (2.16)$$

Para obter o mínimo erro quadrático médio, é necessário analisar o ponto em que o gradiente da função custo J_{Wiener} se anula

$$\nabla J_{Wiener} = 2\mathbf{R}\mathbf{w}_n - 2\mathbf{p} = 0. \quad (2.17)$$

Dessa forma, obtém-se a versão ótima do vetor parâmetros, também conhecida como solução de Wiener. A solução pode ser obtida analiticamente:

$$\mathbf{w}_W = \mathbf{R}^{-1} \mathbf{p}. \quad (2.18)$$

A matriz \mathbf{R} é tipicamente inversível pois é geralmente positiva definida [28].

É válido notar que, como o sinal transmitido tem média nula, a expressão da função custo (2.16)

pode tomar a seguinte forma:

$$J_{Wiener}(\mathbf{w}_n) = \sigma_s^2 - \mathbf{w}_n^T \mathbf{p} - \mathbf{p}^T \mathbf{w}_n + \mathbf{w}_n^T \mathbf{R} \mathbf{w}_n, \quad (2.19)$$

onde σ_s^2 é a variância do sinal transmitido. Note que o custo da formulação de Wiener tem o valor mínimo para o ponto \mathbf{w}_W , o que é dado pela expressão (2.18):

$$J_{Wiener, \min}(\mathbf{w}_n) = \sigma_s^2 - \mathbf{p}^T \mathbf{R}^{-1} \mathbf{p}. \quad (2.20)$$

Uma observação relevante sobre o critério de Wiener é que existe uma relação entre a solução ótima \mathbf{w}_W e o atraso d , o que se explica pelo fato de este último ser parte do cálculo do vetor de correlação cruzada \mathbf{p} . Consequentemente, para cada atraso temos, em tese, uma solução de Wiener diferente.

Na próxima seção, analisaremos o algoritmo SDA, o mais simples dentre os considerados neste trabalho. A análise do SDA nos auxiliará no desenvolvimento dos procedimentos de análise empregados, na formalização das técnicas de equalização adaptativa e no estabelecimento da notação que utilizaremos nos próximos capítulos.

2.3.2 SDA (*Steepest Descent Algorithm*)

O SDA é um algoritmo determinístico que pressupõe o emprego de um equalizador FIR supervisionado como mostrado na figura 2.7. O algoritmo é uma consequência direta do uso do método do gradiente clássico para minimizar a função custo de MSE (*Mean Squared Error*), estando, portanto, inserido em um contexto determinístico². Isso permite utilizar a representação da equação (2.19) [28]

$$J_{Wiener}(\mathbf{w}_n) = J_{SDA}(\mathbf{w}_n) = \sigma_s^2 - \mathbf{w}_n^T \mathbf{p} - \mathbf{p}^T \mathbf{w}_n + \mathbf{w}_n^T \mathbf{R} \mathbf{w}_n. \quad (2.21)$$

Note que a função custo associada ao SDA é exatamente a função custo do critério de Wiener, ou seja, o algoritmo SDA é fundamentalmente uma abordagem para minimizar o critério de Wiener. Para otimizar a função custo, é preciso chegar a uma expressão idêntica à equação (2.17)

$$\nabla J_{Wiener} = \nabla J_{SDA} = 2\mathbf{R}\mathbf{w}_n - 2\mathbf{p}. \quad (2.22)$$

²A rigor o SDA não necessita de treinamento.

Substituindo (2.22) na formulação geral para sistemas determinísticos com gradiente conhecido, ou seja, equação (2.9), temos a expressão geral de adaptação para o algoritmo SDA [28, p. 342]:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu(\mathbf{p} - \mathbf{R}\mathbf{w}_n). \quad (2.23)$$

A expressão (2.23) pode ser representada de um ponto de vista sistêmico conforme o diagrama de blocos da figura 2.8

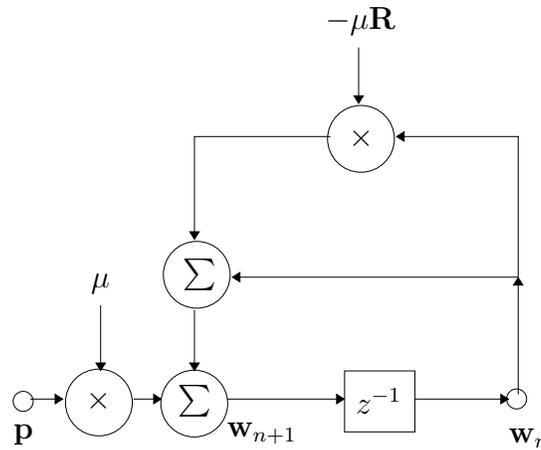


Fig. 2.8: Diagrama de fluxo para o algoritmo SDA.

Na figura 2.8, é possível observar que o SDA apresenta, como é a regra quando se lida com algoritmos adaptativos, laços de realimentação, o que cria uma dinâmica e a possibilidade de que o algoritmo se torne instável. O fator que determina a estabilidade do algoritmo, como veremos, é o passo de adaptação.

Não obstante a simplicidade e a solidez do SDA, é preciso ter em mente que, nem sempre, as estatísticas do sinal transmitido s_n e do sinal recebido x_n estão disponíveis. Portanto, a matriz de autocorrelação \mathbf{R} e mesmo o vetor de correlação cruzada \mathbf{p} não são, em geral, plenamente conhecidos. Para contornar essa situação, é preciso recorrer a estimativas desses dois parâmetros. Na próxima seção discutiremos o algoritmo LMS, a mais clássica solução nesse sentido.

2.3.3 LMS (*Least Mean Square*)

Assim como o algoritmo SDA, o LMS foi proposto com a finalidade de se obter iterativamente a solução de Wiener (vide equação (2.18)). Contudo, ao contrário do SDA, o LMS não requer o

conhecimento *a priori* das estatísticas de segunda ordem dos sinais envolvidos, como mostraremos a seguir.

Em um dado instante n , podemos considerar que a melhor estimativa instantânea para função custo $E\{e_n^2\}$ é o seu próprio argumento, ou seja, e_n^2 [27]. Como consequência direta, a matriz de autocorrelação e o vetor de correlação cruzada podem também tomar a forma de estimativas instantâneas [28, p. 370], ou seja,

$$\mathbf{R} = E\{\mathbf{x}_n \mathbf{x}_n^T\} \implies \hat{\mathbf{R}} = \mathbf{x}_n \mathbf{x}_n^T, \quad (2.24)$$

$$\mathbf{p} = E\{\mathbf{x}_n s_{n-d}\} \implies \hat{\mathbf{p}} = \mathbf{x}_n s_{n-d}. \quad (2.25)$$

Consequentemente, a função custo do LMS pode ser escrita com base no custo associado ao critério de Wiener (expressão (2.19)) [28]:

$$J_{LMS}(\hat{\mathbf{w}}_n) = \sigma_s^2 - \hat{\mathbf{w}}_n^T \hat{\mathbf{p}}_n - \hat{\mathbf{p}}_n^T \hat{\mathbf{w}}_n + \hat{\mathbf{w}}_n^T \hat{\mathbf{R}}_n \hat{\mathbf{w}}_n. \quad (2.26)$$

Note que o algoritmo LMS utiliza $\hat{\mathbf{w}}_n$, ou seja, uma estimativa de \mathbf{w}_n .

Da mesma maneira, a estimativa do vetor de gradiente da expressão (2.17) pode ser dada por

$$\nabla J_{LMS} = 2\hat{\mathbf{R}}\hat{\mathbf{w}}_n - 2\hat{\mathbf{p}} \quad (2.27)$$

$$= 2\mathbf{x}_n \mathbf{x}_n^T \hat{\mathbf{w}}_n - 2\mathbf{x}_n s_{n-d} \quad (2.28)$$

$$= 2\mathbf{x}_n (\mathbf{x}_n^T \hat{\mathbf{w}}_n - s_{n-d}) \quad (2.29)$$

$$= -2\mathbf{x}_n e_n. \quad (2.30)$$

A expressão de atualização dos pesos apresenta a seguinte forma [28]:

$$\hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n + \mu e_n^* \mathbf{x}_n. \quad (2.31)$$

Nessa expressão, o passo de adaptação μ tem mais uma vez um papel crucial, sendo a sua escolha capaz de influenciar na estabilidade e no desempenho do algoritmo. Isso é confirmado pela figura 2.9, que apresenta o diagrama de fluxo do LMS, o qual, da mesma maneira que no algoritmo SDA, possui um laço de realimentação.

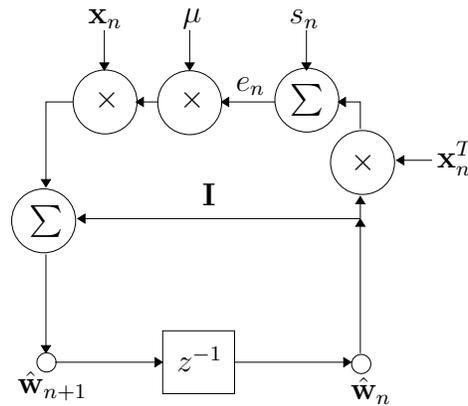


Fig. 2.9: Diagrama de fluxo para o algoritmo LMS.

Teoria da Independência

No caso do algoritmo LMS, é comum fazer referência a um conjunto de hipóteses simplificadoras conhecido como Teoria da Independência para as análises de convergência. Os quatro pontos fundamentais dessa teoria são [28]:

- O vetor de entrada do equalizador, ou seja, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ consiste de sequência de vetores estatisticamente independentes.
- No instante n , o vetor de entrada do equalizador, ou seja, \mathbf{x}_n , é estatisticamente independente de todas as amostras anteriores do sinal de referência, ou seja, s_1, s_2, \dots, s_{n-1} .
- No instante n , o sinal de referência ou sinal transmitido s_n é dependente dos termos do vetor de entrada do equalizador \mathbf{x}_n , mas estatisticamente independente de todas as amostras anteriores do sinal transmitido ou sinal de referência s_{n-1}, s_{n-2}, \dots .
- Os elementos do vetor de entrada do equalizador \mathbf{x}_n e o sinal de referência s_n consistem de variáveis aleatórias com distribuição gaussiana para todo n .

Considerando a equação (2.31), o vetor $\hat{\mathbf{w}}_{n+1}$ depende apenas de:

- Amostras anteriores do vetor de entrada do equalizador, isto é $\mathbf{x}_n, \mathbf{x}_{n-1}, \dots, \mathbf{x}_1$.
- Amostras anteriores do sinal de referência ou sinal transmitido, ou seja, s_n, s_{n-1}, \dots, s_1 .
- Valor inicial do vetor de pesos, $\hat{\mathbf{w}}_0$

Conseqüentemente, $\hat{\mathbf{w}}_{n+1}$ é independente de \mathbf{x}_{n+1} e s_{n+1} .

Veja que, no âmbito da filtragem no tempo, a teoria da independência é uma aproximação que ignora as dependências estatísticas inerentes ao modelo de canal / equalizador e ao processo de adaptação. Essa consideração será relevante quando tratarmos da análise de estabilidade do algoritmo LMS (vide seção 2.5). Vale ressaltar que a teoria é bastante usada na literatura clássica de filtragem adaptativa [28][27], o que se justifica por seu efeito em termos de tratabilidade matemática.

Na seção abaixo, discutiremos a definição de *misadjustment*, que corresponde a uma espécie de referência ou medida da eficiência do algoritmo LMS em situação de regime.

Misadjustment

O *misadjustment* é definido matematicamente como

$$M = \frac{J_{ex}(\infty)}{J_{min}}, \quad (2.32)$$

onde

$$J_{ex}(\infty) = J(n) |_{n \rightarrow \infty} - J_{min}, \quad (2.33)$$

sendo que J_{min} é obtido no ponto de mínimo da função custo. No caso da formulação de Wiener, ou mesmo do algoritmo LMS, $J_{min} = J_{Wiener}$ corresponde ao valor da função custo no mínimo de Wiener.

Considerando que o algoritmo converge no sentido da média quadrática, ou seja, que os autovalores de matriz de correlação têm magnitude menor que 1 [28, p. 400], podemos quantificar, para o LMS, o *misadjustment* através de expressões fechadas como [28] [41] (isto é, considerando um passo de adaptação μ suficiente pequeno se comparado com λ_{max})

$$M = \sum_{i=1}^M \frac{\mu \lambda_i}{2 - \mu \lambda_i} \approx \frac{\mu}{2} \sum_{i=1}^M \lambda_i = \frac{\mu}{2} tr(\mathbf{R}), \quad (2.34)$$

onde λ_i são autovalores da matriz de autocorrelação.

O *misadjustment* possui algumas propriedades relevantes:

- M cresce linearmente em função do número de coeficientes do equalizador;
- M tem um relação direta com o passo de adaptação do algoritmo μ .
- o *misadjustment* representa um medida adimensional da eficiência do algoritmo em situação de regime.

Consideramos neste trabalho sinais transmitidos com média nula e espectralmente brancos. Se não fosse esse o caso, a definição clássica de *misadjustment* deveria ser atualizada [42] a fim de se considerarem os efeitos das flutuações do sinal transmitido no sinal de saída.

Com isto, concluímos nossa exposição preliminar das metodologias supervisionadas relevantes para este trabalho. Agora, trataremos do caso em que não é possível dispor de um sinal de referência, ou seja, do caso não-supervisionado.

2.4 Equalização Adaptativa Não-Supervisionada

Metodologias de equalização não-supervisionada são diferentes daquelas expostas no caso supervisionado, como a presente na figura 2.7. Em geral, consideram-se dois tipos de abordagens básicas [28]:

- Algoritmos baseados em estatísticas de ordem superior - aqui podemos enumerar dois sub-grupos:
 - Algoritmos baseados em estatísticas implícitas, ou seja, que exploram de forma indireta as estatísticas de ordem superior do sinal recebido. Neste grupo temos a importante classe dos algoritmos de Bussgang.
 - Algoritmos baseados explicitamente em estatísticas de ordem superior - fundamentalmente cumulantes - ou que utilizam suas transformadas discretas de Fourier (poliespectros). Os cumulantes e poliespectros podem ser interpretados como generalizações da função de autocorrelação e da densidade espectral de potência [28].
- Algoritmos baseados em cicloestacionariedade - esses algoritmos consideram as estatísticas cicloestacionárias de segunda ordem do sinal recebido. A propriedade de cicloestacionariedade surge em sinais modulados como resultado de variações na amplitude, fase, ou frequência de uma portadora senoidal.

Neste trabalho, lidaremos fundamentalmente com a classe de Bussgang, cuja principal diferença relativamente ao caso supervisionado é o uso de uma estimativa do sinal de erro e_n baseada no próprio sinal de saída y_n . Na figura 2.10, apresentamos a representação de um esquema de equalização não-supervisionada pertencente a essa classe.

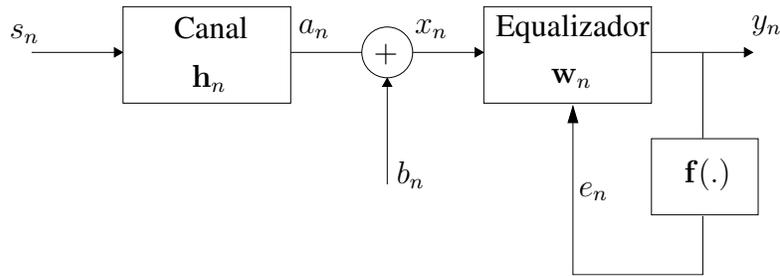


Fig. 2.10: Sistema de equalização não-supervisionada.

Embora haja outros algoritmos populares na classe de Bussgang (por exemplo, o algoritmo de decisão direta), trataremos, neste trabalho, apenas o CMA. Esse algoritmo é objeto de interesse de vários pesquisadores [15], principalmente devido à sua robustez (em comparação com outras abordagens cegas, como o algoritmo DD) no que se refere à existência de mínimos espúrios, e também possui importantes conexões teóricas com a abordagem de Shalvi-Weinstein [43].

2.4.1 CMA (*Constant Modulus Algorithm*)

O CMA se enquadra no molde dos algoritmos de Godard, os quais trabalham com o conceito de dispersão independente da recuperação de fase da portadora. Em um caso geral, a função custo de um algoritmo de Godard pode ser enunciada como [28]

$$J_{Godard} = E\{|y_n|^p - R_p\}^2, \quad (2.35)$$

onde $p \in \mathbb{N}^*$ e

$$R_p = \frac{E\{|s_n|^{2p}\}}{E\{|s_n|^p\}}. \quad (2.36)$$

A expressão de atualização do vetor de pesos do equalizador linear é dada por

$$\hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n + \mu e_n^* \mathbf{x}_n, \quad (2.37)$$

sendo o sinal de erro representado por

$$e_n = y_n |y_n|^{p-2} (R_p - |y_n|^p). \quad (2.38)$$

O algoritmo CMA é o caso particular do algoritmo de Godard para $p = 2$ - o que leva à ideia de preservação do módulo constante [44]. Em tal caso, a função custo do CMA pode ser escrita como

$$J_{CMA}(\hat{\mathbf{w}}_n) = E\{|y_n|^2 - R_2\}^2, \quad (2.39)$$

onde

$$R_2 = \frac{E\{|s_n|^4\}}{E\{|s_n|^2\}}. \quad (2.40)$$

O sinal de erro no caso do CMA é dado pela expressão

$$e_n = y_n(R_2 - |y_n|^2). \quad (2.41)$$

O diagrama de fluxo para o algoritmo CMA é apresentado na figura 2.11. Ressaltamos a presença, como é de praxe, do laço de realimentação, o qual pode, mais uma vez, levar a uma situação de instabilidade dependendo do passo escolhido.

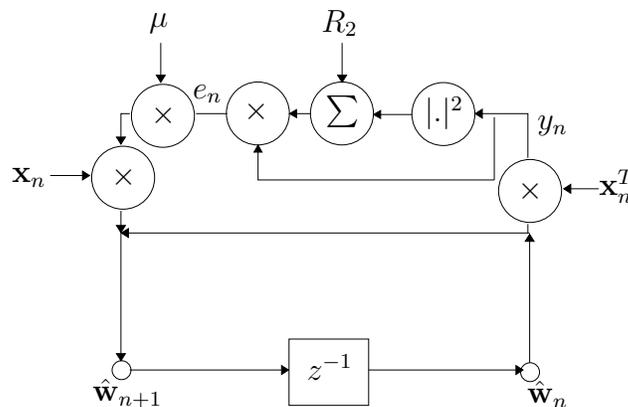


Fig. 2.11: Diagrama de fluxo para o algoritmo CMA.

Desde o trabalho de Ding [45], sabe-se que o CMA pode convergir para mínimos locais subótimos. Consequentemente, caso ocorra uma inicialização inadequada, pode-se não obter o melhor aproveitamento possível do potencial da estrutura de filtragem. Além disso, uma inicialização insatisfatória pode também levar a uma maior lentidão até que se atinja a vizinhança de um mínimo[46].

Portanto, a performance do algoritmo CMA depende fortemente de uma estratégia adequada de inicialização. Algumas idéias nesse sentido são:

- inicialização próxima a uma boa solução de Wiener - vale ressaltar que aqui devemos ter algum conhecimento prévio das estatísticas do sinal transmitido e recebido, o que nem sempre é viável.
- inicialização aleatória uniforme - essa abordagem se caracteriza por não fazer uso de informação *a priori*.

- inicialização por intermédio da técnica *center-spike*.

A idéia do *center spike* foi introduzida por Foschini [47], e consiste em inicializar o vetor de pesos do equalizador com apenas um elemento não-nulo - o elemento central. Por exemplo:

$$\mathbf{w}_{n=0} = [0 \ 0 \ 1 \ 0 \ 0]_{1 \times 5}^T. \quad (2.42)$$

Conseqüentemente, temos uma opção de inicialização que depende apenas do tamanho do vetor de pesos do equalizador. Vale ressaltar que esse tipo de técnica tem como vantagem a sua independência relativamente a maiores informações acerca do canal.

Tendo realizado uma exposição inicial das técnicas de equalização que serão discutidas, estamos aptos a discutir alguns resultados clássicos acerca de sua convergência, resultados estes que serão importantes nas análises subsequentes.

2.5 Análise de Estabilidade de Algoritmos Adaptativos

Estabilidade, de certa forma, é uma questão mais crítica que inicialização. Caso a inicialização seja ruim, o algoritmo apresentará um problema de desempenho, mas terá ainda a possibilidade de atingir um desempenho razoável. Contudo, caso o algoritmo se torne instável, será atingida uma situação irreversível.

É a preocupação com as condições de estabilidade de algoritmos adaptativos que serve de ponto de partida para este trabalho, lembrando que existem vários conceitos de estabilidade, como estabilidade ao redor do ponto de equilíbrio, estabilidade de órbitas periódicas e estabilidade de entrada-saída [22].

Aqui, lidaremos predominantemente com estabilidade em torno do ponto de equilíbrio. Nos estudos de sistemas não-lineares, esse tipo de estabilidade pode ser investigado através da teoria de Lyapunov, uma ferramenta de suma importância na área de controle adaptativo. Contudo, nota-se que não existem muitas aplicações da teoria em equalização adaptativa, embora haja claramente potencial para isso: preencher mais essa lacuna será uma motivação essencial dos próximos capítulos.

Antes, porém, de lidarmos com a abordagem de Lyapunov, trataremos de expor alguns resultados clássicos de estabilidade para os algoritmos SDA, LMS e CMA, os quais serão referências para comparações e observações descritas posteriormente.

2.5.1 Estabilidade do SDA

A fim de analisar a estabilidade do algoritmo SDA, vamos definir o erro entre o vetor ótimo e o vetor de parâmetros estimados [27],

$$\tilde{\mathbf{w}}_n = \hat{\mathbf{w}}_n - \mathbf{w}_W. \quad (2.43)$$

Note que \mathbf{w}_W corresponde à solução de Wiener ($\mathbf{w}_W = \mathbf{R}^{-1}\mathbf{p}$).

Escrevendo a equação (2.23) em função da solução de Wiener, temos

$$\hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n + \mu(\mathbf{R}\mathbf{w}_W - \mathbf{R}\hat{\mathbf{w}}_n). \quad (2.44)$$

A expressão (2.44) pode ainda ser reescrita em termos do erro $\tilde{\mathbf{w}}_n$

$$\tilde{\mathbf{w}}_{n+1} = (\mathbf{I} - \mu\mathbf{R})\tilde{\mathbf{w}}_n. \quad (2.45)$$

A equação (2.45) é obtida no sistema de coordenadas $\tilde{\mathbf{w}}_n$ ³. Seja a matriz unitária \mathbf{Q} ⁴ que obedece a propriedade de transformação de similaridade unitária [28], isto é, $\mathbf{Q}^T\mathbf{R}\mathbf{Q} = \mathbf{\Lambda}$, onde $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_M)$. Considerando $\tilde{\mathbf{w}}_n = \mathbf{Q}\check{\mathbf{w}}_n$, podemos escrever a equação (2.45) como:

$$\mathbf{Q}\check{\mathbf{w}}_{n+1} = (\mathbf{I} - \mu\mathbf{R})\mathbf{Q}\check{\mathbf{w}}_n. \quad (2.46)$$

Multiplicando ambos os lados por \mathbf{Q}^{-1}

$$\begin{aligned} \check{\mathbf{w}}_{n+1} &= \mathbf{Q}^{-1}(\mathbf{I} - \mu\mathbf{R})\mathbf{Q}\check{\mathbf{w}}_n \\ &= (\mathbf{I} - \mu\mathbf{\Lambda})\check{\mathbf{w}}_n, \end{aligned} \quad (2.47)$$

onde a matriz de autovalores obedece a $\mathbf{\Lambda} = \mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}$. Com essas operações, garantimos que não existe acoplamento na dinâmica que atualiza o vetor de pesos, o que facilita a análise.

A solução para o SDA pode ser escrita como

$$\check{\mathbf{w}}_n = (\mathbf{I} - \mu\mathbf{\Lambda})^n \check{\mathbf{w}}_0, \quad (2.48)$$

onde $\check{\mathbf{w}}_0$ é o ponto onde o algoritmo é inicializado. O algoritmo será estável quando

$$\lim_{n \rightarrow \infty} (\mathbf{I} - \mu\mathbf{\Lambda})^n = \mathbf{0}. \quad (2.49)$$

³Veja que a operação $\tilde{\mathbf{w}}_n = \hat{\mathbf{w}}_n - \mathbf{w}_W$ corresponde a uma translação do sistema de coordenadas $\hat{\mathbf{w}}_n$ para $\tilde{\mathbf{w}}_n$.

⁴Matriz unitária é aquela que apresenta como propriedade $\mathbf{Q}^{-1} = \mathbf{Q}^T$. A multiplicação pela matriz \mathbf{Q} é equivalente a uma rotação em torno dos eixos principais [27]. A matriz \mathbf{Q} tem como suas colunas um conjunto de autovetores ortogonais associados com os autovalores da matriz \mathbf{R} [28]. Neste sentido, essa operação vai diagonalizar a matriz \mathbf{R} , ou seja, $\mathbf{Q}^T\mathbf{R}\mathbf{Q} = \mathbf{\Lambda}$, sendo $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_M)$, onde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$ são os autovalores da matriz \mathbf{R} .

A equação (2.50) representa todos os termos da diagonal de Λ . Escrevendo de maneira matricial, temos:

$$\begin{bmatrix} \lim_{n \rightarrow \infty} (1 - \mu\lambda_0)^n & 0 & \dots & 0 \\ 0 & \lim_{n \rightarrow \infty} (1 - \mu\lambda_1)^n & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \lim_{n \rightarrow \infty} (1 - \mu\lambda_L)^n \end{bmatrix} = \mathbf{0}. \quad (2.50)$$

Para garantir a estabilidade e convergência do sistema, cada termo da matriz da equação (2.50) deve ser, em módulo, menor do que 1 [27]. Consequentemente,

$$|1 - \mu\lambda_i| < 1 \implies 0 < \mu < \frac{2}{\lambda_i} \quad (2.51)$$

Portanto, a condição geral de estabilidade do algoritmo SDA é

$$0 < \mu < \frac{2}{\lambda_{max}} \quad (2.52)$$

onde λ_{max} é o maior autovalor da matriz de autocorrelação \mathbf{R} .

A expressão (2.52) será a base para comparação com os resultados posteriormente obtidos via método de Lyapunov.

Na próxima seção, descreveremos condições de estabilidade para o algoritmo LMS.

2.5.2 Estabilidade do LMS - Prova 1

Descreveremos nesta seção uma importante abordagem para análise das condições de estabilidade do algoritmo LMS do ponto de vista do passo de adaptação. A metodologia considerada seguirá o mesmo desenvolvimento exposto em [27].

Seja a expressão (2.31) na forma expandida com os termos de erro da equação (2.13)

$$\hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n + \mu(s_{n-d} - \hat{\mathbf{w}}_n^T \mathbf{x}_n)^* \mathbf{x}_n \quad (2.53)$$

$$= \hat{\mathbf{w}}_n + \mu(s_{n-d}^* \mathbf{x}_n - \mathbf{x}_n \mathbf{x}_n^T \hat{\mathbf{w}}_n). \quad (2.54)$$

Aplicando o operador esperança em ambos os lados da expressão (2.54), temos:

$$E\{\hat{\mathbf{w}}_{n+1}\} = E\{\hat{\mathbf{w}}_n\} + \mu(E\{s_{n-d}^* \mathbf{x}_n\} - E\{\mathbf{x}_n \mathbf{x}_n^T \hat{\mathbf{w}}_n\}). \quad (2.55)$$

Para um sinal de entrada do equalizador que corresponda a um processo estacionário, é esperado que, após um número suficiente de iterações, a grandeza $E\{\hat{\mathbf{w}}_n\}$ convergirá para a solução ótima de Wiener [27].

Considerando a solução de Wiener (equação (2.18)), a notação da expressão (2.25) e utilizando a teoria da independência, ou seja, assumindo que $\hat{\mathbf{w}}_n$ e \mathbf{x}_n são independentes

$$E\{\hat{\mathbf{w}}_{n+1}\} = E\{\hat{\mathbf{w}}_n\} + \mu(\mathbf{p} - \mathbf{R}E\{\hat{\mathbf{w}}_n\}) \quad (2.56)$$

$$= (\mathbf{I} - \mu\mathbf{R})E\{\hat{\mathbf{w}}_n\} + \mu\mathbf{R}\mathbf{w}_W. \quad (2.57)$$

É possível obter uma expressão simplificada para o LMS subtraindo \mathbf{w}_W de ambos os lados da equação (2.57) (vide também a notação da equação (2.43))

$$E\{\tilde{\mathbf{w}}_{n+1}\} = (\mathbf{I} - \mu\mathbf{R})E\{\tilde{\mathbf{w}}_n\}. \quad (2.58)$$

Da mesma forma que a análise do algoritmo SDA (capítulo 2.5.1), vamos utilizar a transformação similaridade unitária [28], ou seja, a expressão (2.58) será modificada em ambos os lados pela matriz unitária \mathbf{Q}

$$\mathbf{Q}E\{\check{\mathbf{w}}_{n+1}\} = (\mathbf{I} - \mu\mathbf{R})\mathbf{Q}E\{\check{\mathbf{w}}_n\}. \quad (2.59)$$

Multiplicando ambos os lados por \mathbf{Q}^{-1}

$$E\{\check{\mathbf{w}}_{n+1}\} = \mathbf{Q}^{-1}(\mathbf{I} - \mu\mathbf{R})\mathbf{Q}E\{\check{\mathbf{w}}_n\} \quad (2.60)$$

$$= (\mathbf{I} - \mu\mathbf{\Lambda})E\{\check{\mathbf{w}}_n\} \quad (2.61)$$

$$= (\mathbf{I} - \mu\mathbf{\Lambda})^n E\{\check{\mathbf{w}}_0\}, \quad (2.62)$$

onde $\check{\mathbf{w}}_0$ é o ponto onde o algoritmo é inicializado. O algoritmo LMS convergirá, no sentido da média quadrática, para o ponto de equilíbrio quando

$$\lim_{n \rightarrow \infty} (\mathbf{I} - \mu \mathbf{\Lambda})^n = \mathbf{0}. \quad (2.63)$$

ou seja, a condição geral de estabilidade na média do algoritmo LMS é

$$0 < \mu < \frac{2}{\lambda_{max}} \quad (2.64)$$

onde λ_{max} é o maior autovalor da matriz \mathbf{R} ou o maior termo da matriz diagonal $\mathbf{\Lambda}$.

Contudo, no caso do algoritmo LMS, nem sempre é possível ter acesso a todos os autovalores da matriz \mathbf{R} . Na verdade, em princípio, a matriz de correlação \mathbf{R} não é necessariamente conhecida. Note ainda que λ_{max} é tipicamente menor que o traço da matriz \mathbf{R} , que é a soma dos elementos da diagonal de \mathbf{R} . Então, para convergência média quadrática do LMS consideramos que (esta é uma condição um pouco mais restritiva) [27]

$$\lambda_{max} < tr(\mathbf{R}) \implies 0 < \mu < \frac{2}{tr(\mathbf{R})}. \quad (2.65)$$

Veja que as condições das equações (2.64) e (2.65) não são condições necessárias para convergência do LMS, mas servem de guia para o ajuste do algoritmo.

Na próxima seção, apresentaremos uma segunda prova de estabilidade para o LMS.

2.5.3 Estabilidade do LMS - Prova 2

De maneira similar à prova de estabilidade da seção anterior, podemos tomar como base a expressão (2.31) e, expandindo a mesma com o auxílio do sinal da erro da equação (2.13), temos:

$$\hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n + \mu(s_{n-d} - \hat{\mathbf{w}}_n^T \mathbf{x}_n)^* \mathbf{x}_n. \quad (2.66)$$

Subtraindo de ambos os lados da equação a solução ótima de Wiener \mathbf{w}_W , e utilizando a notação da equação (2.43), temos:

$$\tilde{\mathbf{w}}_{n+1} = \tilde{\mathbf{w}}_n + \mu(s_{n-d} - \mathbf{x}_n^T \tilde{\mathbf{w}}_n - \mathbf{x}_n^T \mathbf{w}_W) \mathbf{x}_n. \quad (2.67)$$

Definindo $e_{0,n}$ como o erro estimado produzido na solução de Wiener ($e_{0,n} = s_{n-d} - \mathbf{x}_n^T \mathbf{w}_W$),

$$\tilde{\mathbf{w}}_{n+1} = \tilde{\mathbf{w}}_n - \mu \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n + \mu e_{0,n} \mathbf{x}_n \quad (2.68)$$

$$= (\mathbf{I} - \mu \mathbf{x}_n \mathbf{x}_n^T) \tilde{\mathbf{w}}_n + \mu e_{0,n} \mathbf{x}_n. \quad (2.69)$$

Neste momento, é importante apresentar um teorema conhecido como *direct-averaging*, que será o ponto chave para essa versão da prova de estabilidade do algoritmo LMS.

Teorema 1. *Seja um sistema dinâmico com modelo [29, pag. 106]*

$$\tilde{\mathbf{w}}_{n+1} = \tilde{\mathbf{w}}_n + \mu G(\tilde{\mathbf{w}}_n, \xi_n). \quad (2.70)$$

Seja

- $\tilde{\mathbf{w}}_n^e$ a função constante por partes que é igual $\tilde{\mathbf{w}}_n$ em $[ne, ne + e]$.

São dadas as seguintes hipóteses:

1. $\{\tilde{\mathbf{w}}_n^e, \xi_{n-1}^e, n \geq 0\}$ é um processo de Markov com a função de transição dada por

$$P^e(\tilde{\mathbf{w}}, \xi, n, l, A) = P\{(\tilde{\mathbf{w}}_n^e, \xi_{n-1}^e) \in A | \tilde{\mathbf{w}}_n^e = \tilde{\mathbf{w}}, \xi_{n-1}^e = \xi\}$$

onde ξ_{n-1}^e são valores em S , que é um espaço métrico;

2. para cada $n < T < \infty$ e um conjunto compacto Q , $\sup_{e, n, en \leq \infty} E\{ \sup_{\tilde{\mathbf{w}}_n \in Q} (|G(\tilde{\mathbf{w}}_n, \xi)|^2) \} < \infty$;
3. seja $P^e(\xi, l, B | \tilde{\mathbf{w}}_n) = P(\xi_n^e \in B | \xi_{n-1}^e = \xi, \tilde{\mathbf{w}}_n^e = \tilde{\mathbf{w}})$ a função de transição para um passo nos conjuntos Borel ⁵ B de S , e suponha que não exista dependência de n . Para cada $\tilde{\mathbf{w}}$ existe uma função de transição $P(\xi, l, \cdot | \tilde{\mathbf{w}})$ nos conjuntos Borel de S tais que para cada $f(\cdot)$ contínuo e limitado, $\int f(\xi) P(\xi, 1, d\xi | \tilde{\mathbf{w}})$ é contínuo em $(\tilde{\mathbf{w}}, \xi)$ e $\int f(\xi) P^e(\xi, 1, d\xi | \tilde{\mathbf{w}})$ converge uniformemente em cada conjunto compacto $(\tilde{\mathbf{w}}, \xi)$ quando $e \rightarrow 0$;
4. existe uma única medida de probabilidade invariante $P^{\tilde{\mathbf{w}}}(\cdot)$ correspondente a matriz de transição $P(\xi, l, \cdot | \tilde{\mathbf{w}}_n)$ e para cada conjunto compacto Q o conjunto de medidas invariantes $\{P^{\tilde{\mathbf{w}}}(\cdot), \tilde{\mathbf{w}}_n \in Q\}$ é tight ⁶;

⁵Um conjunto de Borel é qualquer conjunto em um espaço topológico que pode ser formado a partir de conjuntos abertos (ou equivalentemente, a partir de conjuntos fechados) através de um número de operações contáveis de união, intersecção e complemento.

⁶Uma sequência de probabilidades $\{\beta_k, k \in \mathbb{Z}\}$ é chamada *tight* se para cada $\epsilon > 0$, existe um conjunto compacto $C \subset X$ tal que $\lim_{k \rightarrow \infty} \inf(\beta_k(C)) \geq 1 - \epsilon$, sendo X a σ -álgebra de conjuntos C [30][29].

5. $\int P(\xi, l, d\xi|\tilde{\mathbf{w}})G(\tilde{\mathbf{w}}_n, \xi)$ é contínua em $(\tilde{\mathbf{w}}, \xi)$ e igual a (onde o limite é uniforme para cada conjunto compacto $(\tilde{\mathbf{w}}, \xi)$) $\lim_{e \rightarrow 0} \int G(\tilde{\mathbf{w}}, \xi)P^e(\xi, l, d\xi|\tilde{\mathbf{w}})$;

6. Ou S é compacto ou $\{\xi_n^e, n \geq 0\}$ é mutuamente independente para qualquer $e > 0$.

Seja a expressão diferencial

$$\frac{d\tilde{\mathbf{w}}}{dt} = \int P^{\tilde{\mathbf{w}}}(d\xi|\tilde{\mathbf{w}})G(\tilde{\mathbf{w}}_n, \xi) = E^{\tilde{\mathbf{w}}}\{G(\tilde{\mathbf{w}}_n, \xi)|\tilde{\mathbf{w}}_n\}, \quad (2.71)$$

que tem uma única solução em $[0, \infty)$ para cada condição inicial. $P^{\tilde{\mathbf{w}}}$ e $E^{\tilde{\mathbf{w}}}$ são as funções de probabilidade e esperança com o uso de processos estacionários com $\tilde{\mathbf{w}}$ fixo. Então, se $\tilde{\mathbf{w}}_{n=0}^e \rightarrow \tilde{\mathbf{w}}_0^e$, $\tilde{\mathbf{w}}^e \rightarrow \tilde{\mathbf{w}}(\cdot)$, onde $\tilde{\mathbf{w}}(\cdot)$ satisfaz (2.71) com $\tilde{\mathbf{w}}(0) = 0$. O lado direito da equação (2.71) é contínuo em $\tilde{\mathbf{w}}$.

Portanto, considerando

1- um passo de adaptação μ suficientemente pequeno

2- ruído aditivo $(\mu e_{0,n} \mathbf{x}_n)$ independente de $\tilde{\mathbf{w}}_n$

e por intermédio do teorema 1, é possível reescrever a equação a diferenças (2.69) como outra formulação que utiliza

$$E\{\mathbf{I} - \mu \mathbf{x}_n \mathbf{x}_n^T\} = \mathbf{I} - \mu \mathbf{R}, \quad (2.72)$$

ou seja, a equação (2.69) é "transformada" em:

$$\tilde{\mathbf{w}}_{n+1}^e = (\mathbf{I} - \mu \mathbf{R})\tilde{\mathbf{w}}_n^e + \mu e_{0,n}^e \mathbf{x}_n. \quad (2.73)$$

que é praticamente o mesmo que o modelo markoviano original para todo n . A matriz $(\mathbf{I} - \mu \mathbf{R})$ é a matriz de transição do sistema de Markov modificado. A evolução temporal desse sistema modificado oferece o mesmo resultado que o sistema original apenas para o caso de um passo de adaptação μ pequeno e decrescente [48].

Quando μ é muito pequeno, então as estatísticas de ordem superior (a partir da segunda ordem) podem ser ignoradas na expansão da expressão de $\tilde{\mathbf{w}}_{n+1}^e$, equação (2.73) [48].

Seja a matriz de correlação cruzada dada por

$$\mathbf{K}_n = E\{\tilde{\mathbf{w}}_n^e \tilde{\mathbf{w}}_n^{e,H}\}, \quad (2.74)$$

então, com base na equação (2.73) e considerando a definição dada pela expressão (2.74), podemos escrever que

$$\mathbf{K}_{n+1} = (\mathbf{I} - \mu\mathbf{R})\mathbf{K}_n(\mathbf{I} - \mu\mathbf{R}) + \mu^2 J_{min}\mathbf{R}, \quad (2.75)$$

O termo $\mu^2 J_{min}\mathbf{R}$ da expressão (2.75) é obtido através da multiplicação $(\mu e_{0,n}^e \mathbf{x}_n)(\mu e_{0,n}^e \mathbf{x}_n)^T$ e considerando a independência entre $e_{0,n}^e$ e \mathbf{x}_n . Veja que esse termo impede que $\mathbf{K}_n = 0$, ou seja, o vetor de pesos $\tilde{\mathbf{w}}_n^e$ fica flutuando em torno do zero.

Para garantir a estabilidade do sistema da equação (2.75) é necessário que $(\mathbf{I} - \mu\mathbf{R})$ seja analisado como na equação (2.60), ou seja, em linhas gerais, é necessário que

$$0 < \mu < \frac{2}{tr(\mathbf{R})} \quad (2.76)$$

para garantir a convergência no sentido da média quadrática da expressão (2.75), e, conseqüentemente, temos uma referência para questões de estabilidade.

Assim, temos o mesmo resultado que o da expressão (2.64), porém utilizando uma formulação diferente. Vale ressaltar que temos dois casos de análise distintos: estabilidade na média e estabilidade no sentido da média quadrática.

Veja que a abordagem utilizada para obter a equação (2.76), mesmo considerando a técnica de *direct-averaging*, recorreu ao uso da teoria da independência. A formulação utilizada nesta seção é uma demonstração clássica para a estabilidade do algoritmo LMS e a mesma foi extraída de [28] e [48].

A próxima seção apresentará um resultado importante sobre as condições de estabilidade do algoritmo CMA.

2.5.4 Estabilidade do CMA

A análise estocástica do CMA descrita nesta seção tomará como base o trabalho [16]. Contudo, é importante salientar que é possível obter as condições de estabilidade para versões específicas do CMA utilizando multiplicadores de Lagrange, conforme mencionado em [34]. O procedimento abaixo é semelhante ao que foi seguido por [49].

Seja o sinal de erro relacionado com o vetor de pesos do equalizador:

$$\tilde{\mathbf{w}}_n = \mathbf{w}_* - \hat{\mathbf{w}}_n. \quad (2.77)$$

onde \mathbf{w}_* é o vetor de pesos ótimo. Consequentemente, vamos enunciar a hipótese a seguir com relação à solução para o CMA.

Hipótese 4. *Consideraremos que a potência do ruído é suficientemente pequena para a solução zero-forcing ser um mínimo global. Portanto, podemos definir*

$$\mathbf{w}_*^T \mathbf{x}_n \approx \alpha s_{n-d}. \quad (2.78)$$

Vamos escrever o sinal de saída do equalizador como

$$\begin{aligned} y_n &= \hat{\mathbf{w}}_n^T \mathbf{x}_n = (\mathbf{w}_* - \tilde{\mathbf{w}}_n)^T \mathbf{x}_n \\ &\approx \alpha s_{n-d} - \tilde{\mathbf{w}}_n^T \mathbf{x}_n = \alpha s_{n-d} - e_{a,n}, \end{aligned} \quad (2.79)$$

onde $e_{a,n} = \tilde{\mathbf{w}}_n^T \mathbf{x}_n$.

Considerando sinais reais, podemos escrever o sinal de erro do algoritmo CMA a partir da equação (2.41) e (2.79)

$$\begin{aligned} e_n &= y_n(R_2 - y_n^2) \approx (\alpha s_{n-d} - e_{a,n})[R_2 - (\alpha s_{n-d} - e_{a,n})^2] \\ &= -\alpha^3 s_{n-d}^3 + e_{a,n}^3 + 3\alpha^2 s_{n-d}^2 e_{a,n} - 3\alpha s_{n-d} e_{a,n}^2 + (\alpha s_{n-d} - e_{a,n})R_2. \end{aligned} \quad (2.80)$$

Desconsiderando os termos de mais alta potência de $e_{a,n}$, e considerando que a condição inicial ($\hat{\mathbf{w}}_0$) é suficientemente próxima de \mathbf{w}_* no ponto em que $e_{a,n}$ é pequeno, então a equação (2.80) pode ser reescrita como

$$\begin{aligned} e_n &\approx -\alpha^3 s_{n-d}^3 + 3\alpha^2 s_{n-d}^2 e_{a,n} + (\alpha s_{n-d} - e_{a,n})R_2 \\ &= -\alpha^3 s_{n-d}^3 + (3\alpha^2 s_{n-d}^2 - R_2)e_{a,n} + \alpha s_{n-d} R_2. \end{aligned} \quad (2.81)$$

Portanto, podemos definir

$$\begin{aligned} \gamma_n &= 3\alpha^2 s_{n-d}^2 - R_2 \\ \beta_n &= -\alpha^3 s_{n-d}^3 + \alpha s_{n-d} R_2, \end{aligned} \quad (2.82)$$

e a equação (2.81) pode ser escrita como

$$e_n \approx \gamma_n e_{a,n} + \beta_n. \quad (2.83)$$

Tomando as equações (2.77) e (2.83), a expressão recursiva do CMA pode ser escrita como

$$\begin{aligned}\tilde{\mathbf{w}}_{n+1} &= \tilde{\mathbf{w}}_n + \mu e_n^* \mathbf{x}_n \\ &= (\mathbf{I} - \mu \gamma_n \mathbf{x}_n \mathbf{x}_n^T) \tilde{\mathbf{w}}_n - \mu \beta_n \mathbf{x}_n.\end{aligned}\quad (2.84)$$

Utilizaremos a expressão (2.84) multiplicada pelo seu transposto. Também aplicaremos o operador esperança, o que produz a equação:

$$\begin{aligned}\mathbf{S}_{n+1} &= E\{\tilde{\mathbf{w}}_{n+1} \tilde{\mathbf{w}}_{n+1}^T\} = +E\{\tilde{\mathbf{w}}_n \tilde{\mathbf{w}}_n^T\} + E\{\mu^2 \gamma_n^2 \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n \tilde{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T\} \\ &\quad + E\{(\mu^2 \beta_n \gamma_n \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n \mathbf{x}_n^T + \mu^2 \beta_n \gamma_n \mathbf{x}_n \tilde{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \\ &\quad - \mu \tilde{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n - \mu \tilde{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n) \gamma_n\} \\ &\quad - E\{\mu \beta_n \tilde{\mathbf{w}}_n \mathbf{x}_n^T\} + E\{\mu^2 \beta_n^2 \mathbf{x}_n \mathbf{x}_n^T\}.\end{aligned}\quad (2.85)$$

Faremos algumas considerações adicionais a fim de simplificar a expressão (2.85):

- os vetores $\tilde{\mathbf{w}}_n$ e \mathbf{x}_n são independentes, ou seja, $E\{\mathbf{x}_n \tilde{\mathbf{w}}_n^T \tilde{\mathbf{w}}_n \mathbf{x}_n^T\} = E\{\tilde{\mathbf{w}}_n^T \tilde{\mathbf{w}}_n\} E\{\mathbf{x}_n \mathbf{x}_n^T\} = \mathbf{S}_n \mathbf{R}$;
- o comprimento do canal é longo o suficiente para que β_n seja independente de \mathbf{x}_n ;
- β_n tem dependência direta de do sinal bipolar s_n , portanto vale o cálculo $E\{\beta_n\} = 0$;
- o comprimento do canal é longo o bastante para que as estatísticas de quarta ordem sejam aproximadas por aquelas de um vetor Gaussiano, ou seja, $E\{\mathbf{x}_n^T \mathbf{x}_n \mathbf{S}_n \mathbf{x}_n^T \mathbf{x}_n\} \approx 2\mathbf{R} \mathbf{S}_n \mathbf{R} + \mathbf{R} \text{tr}(\mathbf{R} \mathbf{S}_n)$.

$$\begin{aligned}\mathbf{S}_{n+1} &= E\{\tilde{\mathbf{w}}_{n+1} \tilde{\mathbf{w}}_{n+1}^T\} = \mathbf{S}_n + \mu^2 E\{\gamma_n^2\} E\{\mathbf{x}_n^T \mathbf{x}_n \mathbf{S}_n \mathbf{x}_n^T \mathbf{x}_n\} \\ &\quad - \mu E\{\gamma_n\} \mathbf{R} \mathbf{S}_n - \mu E\{\gamma_n\} \mathbf{S}_n \mathbf{R} + \mu^2 E\{\beta_n^2\} \mathbf{R} \\ &= \mathbf{S}_n + 2\mu^2 E\{\gamma_n^2\} \mathbf{R} \mathbf{S}_n \mathbf{R} + \mu^2 E\{\gamma_n^2\} \mathbf{R} \text{tr}(\mathbf{R} \mathbf{S}_n) \\ &\quad - \mu E\{\gamma_n\} \mathbf{R} \mathbf{S}_n - \mu E\{\gamma_n\} \mathbf{S}_n \mathbf{R} + \mu^2 E\{\beta_n^2\} \mathbf{R}.\end{aligned}\quad (2.86)$$

Multiplicando ambos os lados por \mathbf{Q}^{-1} e fazendo uso da transformação de similaridade unitária [28]:

$$\begin{aligned}
\mathbf{T}_{n+1} &= \mathbf{Q}^T \mathbf{S}_{n+1} \mathbf{Q} = \mathbf{T}_n + \mathbf{Q}^T 2\mu^2 E\{\gamma_n^2\} \mathbf{R} \mathbf{S}_n \mathbf{R} \mathbf{Q} + \mathbf{Q}^T \mu^2 E\{\gamma_n^2\} \mathbf{R} \text{tr}(\mathbf{R} \mathbf{S}_n) \mathbf{Q} \\
&\quad - \mathbf{Q}^T \mu E\{\gamma_n\} \mathbf{R} \mathbf{S}_n \mathbf{Q} - \mathbf{Q}^T \mu E\{\gamma_n\} \mathbf{S}_n \mathbf{R} \mathbf{Q} + \mathbf{Q}^T \mu^2 E\{\beta_n^2\} \mathbf{R} \mathbf{Q} \\
&= \mathbf{T}_n + 2\mu^2 E\{\gamma_n^2\} \mathbf{Q}^T \mathbf{R} \mathbf{S}_n \mathbf{R} \mathbf{Q} + \mu^2 E\{\gamma_n^2\} \mathbf{Q}^T \mathbf{R} \text{tr}(\mathbf{R} \mathbf{S}_n) \mathbf{Q} \\
&\quad - \mu E\{\gamma_n\} \mathbf{Q}^T \mathbf{R} \mathbf{S}_n \mathbf{Q} - \mu E\{\gamma_n\} \mathbf{Q}^T \mathbf{S}_n \mathbf{R} \mathbf{Q} + \mu^2 E\{\beta_n^2\} \mathbf{Q}^T \mathbf{R} \mathbf{Q} \\
&= \mathbf{T}_n + 2\mu^2 E\{\gamma_n^2\} \mathbf{\Lambda} \mathbf{T}_n \mathbf{\Lambda} + \mu^2 E\{\gamma_n^2\} \mathbf{\Lambda} \text{tr}(\mathbf{R} \mathbf{S}_n) \\
&\quad - \mu E\{\gamma_n\} \mathbf{T}_n \mathbf{\Lambda} - \mu E\{\gamma_n\} \mathbf{T}_n \mathbf{\Lambda} + \mu^2 E\{\beta_n^2\} \mathbf{Q}^T \mathbf{R} \mathbf{Q}.
\end{aligned} \tag{2.87}$$

Definindo $\mathbf{s}_n = \text{diag}(\mathbf{Q}^T \mathbf{S}_n \mathbf{Q})$, ou seja, os elementos da diagonal de $\mathbf{Q}^T \mathbf{S}_n \mathbf{Q}$ e $\Upsilon = [\lambda_1, \lambda_2, \dots, \lambda_M]^T$. A intenção é utilizar o teorema de diagonalização de Fisher [49]⁷.

$$\begin{aligned}
\mathbf{T}_{n+1} &= \mathbf{T}_n + 2\mu^2 E\{\gamma_n^2\} \mathbf{\Lambda} \mathbf{T}_n \mathbf{\Lambda} + \mu^2 E\{\gamma_n^2\} \mathbf{\Lambda} \text{tr}(\mathbf{\Lambda} \mathbf{T}_n) \\
&\quad - 2\mu E\{\gamma_n\} \mathbf{T}_n \mathbf{\Lambda} + \mu^2 E\{\beta_n^2\} \mathbf{\Lambda} \\
&= [\mathbf{I} + 2\mu^2 E\{\gamma_n^2\} \mathbf{\Lambda} \mathbf{\Lambda} + \mu^2 E\{\gamma_n^2\} \mathbf{\Lambda} \text{tr}(\mathbf{\Lambda}) \\
&\quad - 2\mu E\{\gamma_n\} \mathbf{\Lambda}] \mathbf{T}_n + \mu^2 E\{\beta_n^2\} \mathbf{\Lambda}.
\end{aligned} \tag{2.88}$$

Note que os termos da direita da equação são representados por escalares e matrizes diagonais. Portanto, temos uma matriz diagonal cujos elementos são representados na equação abaixo.

$$\mathbf{s}_{n+1} = \overbrace{[\mathbf{I} + 2\mu^2 E\{\gamma_n^2\} \mathbf{\Lambda}^2 + \mu^2 E\{\gamma_n^2\} \Upsilon \Upsilon^T - 2\mu E\{\gamma_n\} \mathbf{\Lambda}]}^{\mathbf{A}} \mathbf{s}_n + \mu^2 E\{\beta_n^2\} \Upsilon. \tag{2.89}$$

A matriz \mathbf{A} pode ser escrita como

$$\begin{aligned}
\mathbf{A} &= \mathbf{I} + 2\mu^2 E\{\gamma_n^2\} \mathbf{\Lambda}^2 + \mu^2 E\{\gamma_n^2\} \Upsilon \Upsilon^T - 2\mu E\{\gamma_n\} \mathbf{\Lambda} \\
&= (\mathbf{I} - 2\mu E\{\gamma_n\} \mathbf{\Lambda})^2 + \mu^2 (2E\{\gamma_n^2\} - E\{\gamma_n\}^2) \mathbf{\Lambda}^2 + \mu^2 E\{\gamma_n^2\} \Upsilon \Upsilon^T,
\end{aligned} \tag{2.90}$$

Como $E\{\gamma_n^2\} \leq E\{\gamma_n\}^2$, a matriz \mathbf{A} é positiva definida para qualquer $\mu > 0$. Portanto, é necessário apenas garantir que o maior autovalor de \mathbf{A} seja menor que 1 [16] (essa condição também está presente no primeiro método da teoria de Lyapunov e é conhecida como estabilidade de Schur para o caso discreto [50]).

Note que exigir que os autovalores da matriz \mathbf{A} sejam menores do que 1 equivale a dizer que

⁷O teorema de diagonalização de Fisher é enunciado como: Seja uma matriz \mathbf{K}_n que é diagonalizada por \mathbf{P} de maneira que $\mathbf{P} \mathbf{K}_n \mathbf{P} = \text{diag}(\Upsilon)$.

$\mathbf{B} = \mathbf{I} - \mathbf{A}$ é uma matriz positiva definida.

$$\mathbf{B} = -2\mu^2 E\{\gamma_n^2\} \mathbf{\Lambda}^2 - \mu^2 E\{\gamma_n^2\} \Upsilon \Upsilon^T + 2\mu E\{\gamma_n\} \mathbf{\Lambda}. \quad (2.91)$$

A matriz \mathbf{B} deve oferecer as seguintes propriedades para ser positiva definida (devemos encarar essas condições como necessárias e suficientes para a positividade):

- i Os autovalores de uma matriz hermitiana positiva definida devem ser todos positivos [51].
- ii O traço, o determinante e os mínimos principais de uma matriz positiva definida devem ser positivos [52]. Mais especificamente, $\det(\mathbf{A}_i) > 0, i = 1, \dots, n$.
- iii Os elementos da diagonal devem ser positivos [51].

Do ponto (i) e equação (2.90), temos:

$$2\mu E\{\gamma_n\} \lambda_i - 3\mu^2 E\{\gamma_n^2\} \lambda_i^2 > 0 \iff 0 < \mu < \frac{2E\{\gamma_n\}}{3E\{\gamma_n^2\} \lambda_i}, i = 1, \dots, M, \quad (2.92)$$

Na verdade, fazendo analogia com a análise clássica de estabilidade para o algoritmo LMS (vide equação (2.64)), a expressão (2.92) poderia ser reescrita como

$$0 < \mu < \frac{2E\{\gamma_n\}}{3E\{\gamma_n^2\} \lambda_{max}}, i = 1, \dots, M, \quad (2.93)$$

onde λ_{max} é o maior autovalor da matriz \mathbf{R} .

Contudo, conforme mencionado acima (item ii), o traço da matriz \mathbf{B} deve ser positivo. O traço da matriz de autocorrelação está sujeito à propriedade [28]

$$tr(\mathbf{R}) = \sum_{i=0}^M \lambda_i. \quad (2.94)$$

Portanto, podemos elaborar a seguinte condição a partir da equação (2.91):

$$tr(\mathbf{B}) = \sum_{i=0}^M (2\mu E\{\gamma_n\} \lambda_i - 3\mu^2 E\{\gamma_n^2\} \lambda_i^2) = \sum_{i=0}^M \lambda_i (2\mu E\{\gamma_n\} - 3\mu^2 E\{\gamma_n^2\} \lambda_i) > 0. \quad (2.95)$$

Note que os autovalores λ_i são todos positivos. Ou seja, para garantir a desigualdade da equação (2.95), é necessário obedecer à condição abaixo

$$(2\mu E\{\gamma_n\} - \sum_{i=0}^M 3\mu^2 E\{\gamma_n^2\} \lambda_i) > 0 \implies 2\mu(M+1)E\{\gamma_n\} - 3\mu^2 E\{\gamma_n^2\} \text{tr}(\mathbf{R}) > 0. \quad (2.96)$$

Portanto, uma condição para garantir que \mathbf{B} é positiva definida é

$$\mu < \frac{2E\{\gamma_n\}(M+1)}{3E\{\gamma_n^2\} \text{tr}(\mathbf{R})}. \quad (2.97)$$

Ainda no ponto (ii), podemos fazer uso da propriedade relativa ao determinante. Portanto,

$$\begin{aligned} \det(\mathbf{B}_i) > 0 &\implies \prod_{j=1}^i (2\mu E\{\gamma_n\} \lambda_j - 3\mu^2 E\{\gamma_n^2\} \lambda_j^2) > 0 \\ &\implies \prod_{j=1}^i \lambda_j \mu (2(E\{\gamma_n\} - \mu E\{\gamma_n^2\} \lambda_j) - \mu E\{\gamma_n^2\} \lambda_j) > 0. \end{aligned} \quad (2.98)$$

Os termos da equação (2.98) podem ser reescritos conforme abaixo:

$$\prod_{j=1}^i \overset{>0}{2\lambda_j \mu} \left(\overset{>0}{(E\{\gamma_n\} - \mu E\{\gamma_n^2\} \lambda_j)} \left(1 - \frac{\mu E\{\gamma_n^2\} \lambda_j}{2(E\{\gamma_n\} - \mu E\{\gamma_n^2\} \lambda_j)} \right) \right), \quad (2.99)$$

então é necessário estabelecer como condição

$$1 - \frac{\mu E\{\gamma_n^2\} \lambda_j}{2(E\{\gamma_n\} - \mu E\{\gamma_n^2\} \lambda_j)} > 0 \implies \frac{\mu E\{\gamma_n^2\} \lambda_j}{2(E\{\gamma_n\} - \mu E\{\gamma_n^2\} \lambda_j)} < 1. \quad (2.100)$$

Se considerarmos que $\text{tr}(\mathbf{R}) \geq \lambda_i$ ⁸ teremos uma condição suficiente mas não necessária [16] para simplificar a expressão

$$\frac{\mu E\{\gamma_n^2\} \lambda_j}{2(E\{\gamma_n\} - \mu E\{\gamma_n^2\} \lambda_j)} \leq \frac{\mu E\{\gamma_n^2\} \text{tr}(\mathbf{R})}{2(E\{\gamma_n\} - \mu E\{\gamma_n^2\} \text{tr}(\mathbf{R}))} < 1. \quad (2.101)$$

Consequentemente, tomando

⁸A propriedade com relação à matriz de autocorrelação é $\text{tr}(\mathbf{R}) = \sum_{i=1}^M \lambda_i$ [28].

$$0 < \mu < \frac{2E\{\gamma_n\}}{3E\{\gamma_n^2\}tr(\mathbf{R})}. \quad (2.102)$$

temos um condição que satisfaz os pontos (i), (ii) e (iii) bem como os resultados obtidos nas equações (2.92) e (2.97)

A expressão (2.102) oferece uma faixa de valores para o passo de adaptação que garante estabilidade no sentido médio e quadrático. Um resultado similar foi obtido em [34], considerando variáveis reais e para um caso mais genérico da expressão de algoritmos de Bussgang.

Chegamos ao fim desse capítulo após a análise e discussão dos critérios clássicos de estabilidade para os algoritmos de equalização adaptativa que serão considerados neste trabalho. A próxima seção apresenta um resumo dos principais tópicos discutidos até o momento. Esse sumário servirá também como preparação para os próximos capítulos.

2.6 Sumário

Neste capítulo, foram apresentadas as bases conceituais e teóricas do problema de equalização, que tem um papel central para o desenvolvimento da tese. A apresentação teve início por uma discussão acerca de aspectos relevantes do problema de transmissão digital, a qual levou a uma análise do modelo de canal proposto e das dificuldades a ele associadas. Em seguida, colocou-se a metodologia de equalização adaptativa como uma contramedida a essas dificuldades, sendo expostas, em seguida, abordagens supervisionadas (SDA e LMS) e não-supervisionadas (CMA) para aprendizado dos parâmetros da estrutura de filtragem. Uma análise desses métodos nos permitiu verificar que a obtenção de um adequado desempenho depende de vários fatores:

- modelo do canal de comunicação, o que engloba a ordem do canal e posição de pólos e zeros;
- intensidade e distribuição estatística do ruído;
- ponto de inicialização;
- distância entre o ponto de inicialização e ponto de equilíbrio;
- tamanho do passo de adaptação.

Nos próximos capítulos, será buscada uma nova perspectiva analítica baseada na teoria de Lyapunov, perspectiva esta que será embasada pelos conceitos fundamentais expostos no próximo capítulo.

Capítulo 3

Estabilidade de Lyapunov

A teoria de Lyapunov - presente principalmente na análise de estabilidade de sistemas não-lineares - pode ser dividida em dois métodos:

- Método indireto (ou primeiro método) - faz uso da linearização em torno do ponto de equilíbrio [3];
- Método direto (ou segundo método) - considera o uso de uma função que representa a energia do sistema dinâmico.

Neste trabalho, as principais contribuições apresentadas se relacionam ao segundo método de Lyapunov nas versões determinística e estocástica para sistemas dinâmicos discretos no tempo. A linearização de um sistema em torno de um ponto de equilíbrio seguida por uma análise de autovalores, metodologia fundamentalmente equivalente à exposta nas análises vistas no capítulo anterior, por sua vez, relaciona-se com o primeiro método de Lyapunov. Essa abordagem, no entanto, não é aplicável a todos os casos de análise de estabilidade, pois deixa de fora os sistemas que tem funções periódicas em sua descrição e mesmo sistemas com parâmetros que têm mudanças bruscas com parte da dinâmica [53]. Ademais, uma das grandes vantagens do método direto de Lyapunov é que não é necessário saber de antemão qual é o ponto de equilíbrio em questão [54]. Caso contrário, toda análise deveria vir acompanhada com as soluções do sistema de equações diferenciais ou equações à diferença. Consequentemente, pode ser possível evitar um esforço matemático considerável fazendo o uso desse método para análises de estabilidade.

3.1 Definições

Algumas definições, notas e hipóteses serão enunciadas a seguir para facilitar a interpretação deste trabalho.

3.1.1 Caso Determinístico

Definição 1. O ponto \underline{w} é um ponto de equilíbrio se uma vez que w_n é igual a \underline{w} , então w_n permanece igual a \underline{w} para todos os instantes de tempo futuros [20].

Definição 2. Considere o sistema discreto [55][21]

$$w_{n+1} = f_n(w_n) , w_n \in \mathbb{R}^N , f_n : \mathbb{R}^N \rightarrow \mathbb{R}^N . \quad (3.1)$$

O ponto de equilíbrio \underline{w} é estável (no sentido de Lyapunov) se, para cada $\epsilon > 0$ (vide B_ϵ na figura 3.1) e $n_0 > 0$, existe $\delta(\epsilon, n_0)$ (vide B_δ na figura 3.1) positivo tal que

$$\|w_{n=0} - \underline{w}\| < \delta(\epsilon, n_0) \Rightarrow \|f_n(w_n) - \underline{w}\| < \epsilon, \forall n > n_0 \in \mathbb{Z} . \quad (3.2)$$

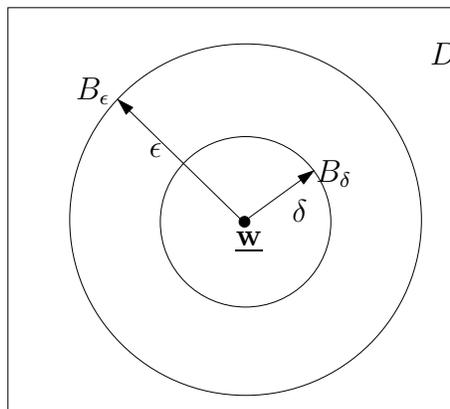


Fig. 3.1: Representação geométrica dos conjuntos envolvidos no Teorema 3.

Vide a figura 3.2 para uma representação geométrica simplificada do conceito de estabilidade. No caso do sistema estável no sentido de Lyapunov é possível inicializar o sistema em uma região definida por δ . Consequentemente, o sistema estável poderá até percorrer a área delimitada por ϵ , mas nunca poderá exceder os limites estabelecidos por ϵ quando o tempo n cresce infinitamente. O caso para instabilidade está retratado na figura 3.3.

Nota 1. A convergência de variáveis de estado não implica em estabilidade [20].

Nota 2. A definição de estabilidade utilizada neste trabalho apenas garante que o sistema ficará confinado em uma região em torno do ponto de equilíbrio. Ou seja, o sistema não "explode". Essa definição é um tanto relaxada, mas existem versões mais restritivas do caso de estabilidade as quais serão introduzidas abaixo.

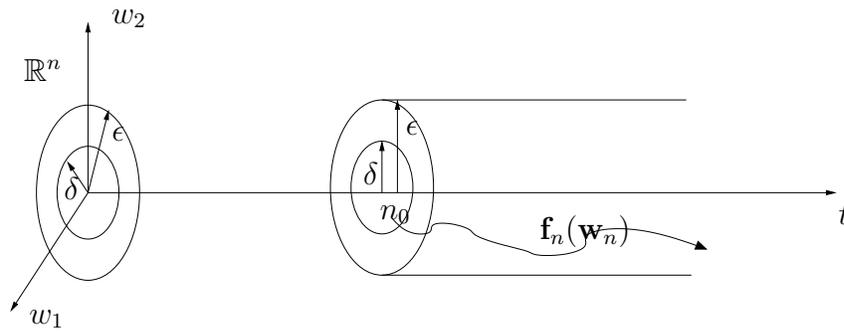


Fig. 3.2: Representação geométrica para estabilidade no sentido de Lyapunov.

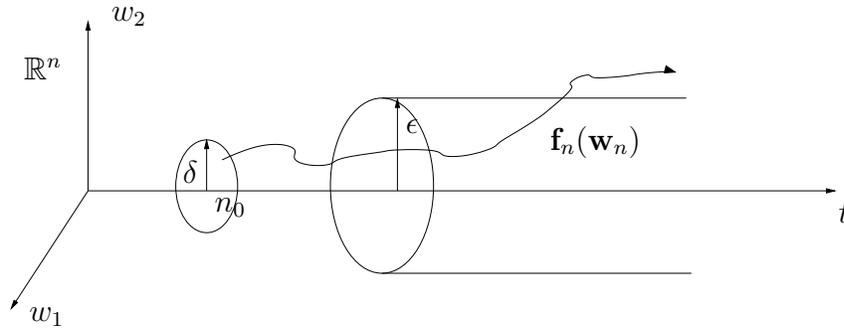


Fig. 3.3: Representação geométrica para instabilidade.

Definição 3. O ponto de equilíbrio $\underline{\mathbf{w}}$ é atrativo se, para cada $n_0 > 0$, existe $\zeta(n_0)$ tal que [21]

$$\|\mathbf{w}_{n_0} - \underline{\mathbf{w}}\| < \zeta(n_0) \Rightarrow \|\mathbf{f}_n(\mathbf{w}_n) - \underline{\mathbf{w}}\| \rightarrow \mathbf{0} \text{ quando } n \rightarrow \infty.$$

Definição 4. O ponto de equilíbrio $\underline{\mathbf{w}}$ na região B_δ é uniformemente estável se, para cada $\epsilon > 0$ e $n_0 > 0$, existe $\delta(\epsilon)$ positivo tal que (δ não depende de n_0) [21]

$$\|\mathbf{w}_{n_0} - \underline{\mathbf{w}}\| < \delta(\epsilon) \Rightarrow \|\mathbf{f}_n(\mathbf{w}_n) - \mathbf{f}_n(\underline{\mathbf{w}})\| < \epsilon, \forall n \in \mathbb{Z}.$$

A representação geométrica para esta definição está na figura 3.4.

Nota 3. No caso do ponto de equilíbrio uniformemente estável, a região definida por $\delta(\epsilon)$ não depende de n_0 , que é a variável temporal do sistema [37]. Se o sistema é inicializado no instante n_0 ou n'_0 em uma região definida por δ , então o sistema ainda continua confinado na região definida por ϵ . Portanto, é possível reproduzir o comportamento qualitativo do sistema em qualquer momento uma vez que a variável temporal não tem influência. O caso mais simples deste tipo de estabilidade é o sistema invariante no tempo.

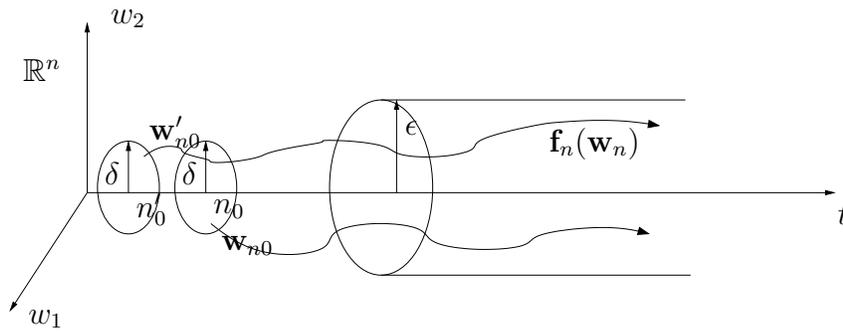


Fig. 3.4: Representação geométrica para estabilidade uniforme.

Definição 5. O ponto de equilíbrio \underline{w} é uniformemente atrativo se existe ζ positivo e $n_0 > 0$ tal que [21]

$$\|\mathbf{w}_{n_0} - \underline{w}\| < \zeta \Rightarrow \|\mathbf{f}_n(\mathbf{w}_n) - \underline{w}\| \rightarrow \mathbf{0} \text{ quando } n \rightarrow \infty, \text{ uniformemente em } n_0 \text{ e } \mathbf{w}_n.$$

De maneira equivalente, \underline{w} é uniformemente atrativo se existe um $\zeta > 0$, tal que para cada $\epsilon > 0$ existe um $m = m(\epsilon)$ tal que

$$\|\mathbf{w}_n - \underline{w}\| < \zeta, n_0 > 0 \Rightarrow \|\mathbf{f}_n(\mathbf{w}_n) - \underline{w}\| < \epsilon, \forall n > m(\epsilon).$$

A figura 3.5 apresenta a representação geométrica para a definição de um ponto uniformemente atrativo. Veja que o caráter uniforme implica em ter um sistema com comportamento que independe do instante de inicialização. A característica de atração significa que a longo do tempo o sistema tende a ficar cada vez mais próximo do ponto de equilíbrio. Após $m(\epsilon)$ (que é uma unidade temporal, vide figura 3.5) o sistema está a uma distância ϵ do ponto de equilíbrio. Estendendo $m(\epsilon) \rightarrow \infty$ então a distância ϵ seria reduzida a zero.

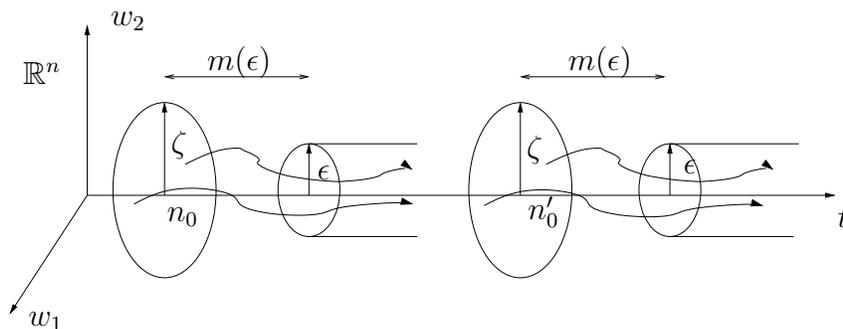


Fig. 3.5: Representação geométrica para uniformemente atrativo.

Definição 6. O ponto de equilíbrio $\underline{\mathbf{w}}$ é assintoticamente estável se é estável e atrativo [21].

Nota 4. No caso do ponto de equilíbrio assintoticamente estável, não existe dependência de δ com relação a ϵ [37].

Note que o sistema convergirá exatamente para região em torno do ponto de equilíbrio no caso de estabilidade assintótica [56], contudo região de atração e velocidade de convergência podem depender do tempo inicial n_0 [20].

Definição 7. O ponto de equilíbrio $\underline{\mathbf{w}}$ é uniformemente assintoticamente estável se é uniformemente estável e uniformemente atrativo [21].

Definição 8. O ponto de equilíbrio $\underline{\mathbf{w}}$ é globalmente uniformemente assintoticamente estável se é uniformemente estável, e para cada $\delta, \epsilon > 0$, existe um $m = m(\delta, \epsilon)$ tal que [21]

$$\|\mathbf{w}_n - \underline{\mathbf{w}}\| < \delta, n_0 \geq 0 \Rightarrow \|\mathbf{f}_n(\mathbf{w}_n) - \underline{\mathbf{w}}\| < \epsilon, \forall n > m.$$

podendo δ ser arbitrariamente grande [37].

Definição 9. Uma função $V(\mathbf{w}_n, n)$ é localmente positiva definida quando $V(\underline{\mathbf{w}}) = 0$ e $V(\mathbf{w}_n, n) > 0, \forall n \geq 0, \forall \mathbf{w} \neq \underline{\mathbf{w}}$ [20]. De maneira análoga, $V(\mathbf{w}_n, n)$ é positiva semi-definida se $V(\underline{\mathbf{w}}) = 0$ e $V(\mathbf{w}_n, n) \geq 0, \forall n \geq 0, \forall \mathbf{w} \neq \underline{\mathbf{w}}$ [37][20].

Definição 10. Se as condições da definição anterior valem para todo o espaço de estados, então $V(\mathbf{w}_n)$ é dita globalmente positiva definida.

Nota 5. A função $V(\mathbf{w}_n, n)$ ¹ é localmente positiva definida se existe uma função $V_0(\|\mathbf{w}_n\|)$ de classe K tal que $V(\underline{\mathbf{w}}, n) = 0$ and $V(\mathbf{w}_n, n) \geq V_0(\|\mathbf{w}_n\|), \forall n \geq 0$ [20]².

Portanto, a função variante no tempo é localmente positiva definida se ela domina uma função invariante no tempo que é localmente positiva definida.

Definição 11. $V(\mathbf{w}_n, n)$ é decrescente se $V(\underline{\mathbf{w}}) = 0$ e existe uma função $V_1(\|\mathbf{w}_n\|)$ de classe K e uma constante $r > 0$ tal que

$$V_1(\|\mathbf{w}_n\|) \geq V(\mathbf{w}_n, n), \forall n \geq 0.$$

¹A variável de tempo n é deixada de maneira explícita para deixar claro que temos um sistema variante no tempo.

²Uma função $\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ pertence à classe K se

- $\alpha(0) = 0$
- $\alpha(p) > 0, \forall p > 0$
- $\alpha(b) > \alpha(a), \forall b > a.$

[22] [20, p. 107].

Ou em outras palavras $V(\mathbf{w}_n, n)$ é limitada superiormente pela função invariante no tempo $V_1(\|\mathbf{w}_n\|)$ [20].

Definição 12. A norma $\|\mathbf{x}\|$ de um vetor \mathbf{x} é uma função de valor real com as seguintes propriedades [22]:

- $\|\mathbf{x}\| \geq 0$ para todo $\mathbf{x} \in \mathbb{R}^n$, com $\|\mathbf{x}\| = 0$ se e somente se $\mathbf{x} = 0$.
- $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$, para todo $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
- $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$, para todo $\alpha \in \mathbb{R}$ e $\mathbf{x} \in \mathbb{R}^n$.

A norma de um vetor pode ainda ser interpretada como o tamanho ou comprimento do vetor [56].

Definição 13. Para funções temporais, é definida a seguinte norma \mathcal{L}^p

$$\|f\|_p = \left(\int_0^\infty |f(t)|^p dt \right)^{\frac{1}{p}},$$

para $p \in [1, \infty)$, geralmente dizemos que $f \in \mathcal{L}^p$ quando $\|f\|_p$ existe, ou seja, quando $\|f\|_p < \infty$ [56]. Contudo, para sistemas discretos (espaço $\ell^p(\mathbb{Z})$), a norma é definida como

$$\|f\|_p = \left(\sum_n |f_n|^p \right)^{\frac{1}{p}}.$$

Definição 14. A norma \mathcal{L}^∞ é definida como

$$\|f\|_\infty = \sup |f(t)|, t \geq 0,$$

e geralmente dizemos que $f \in \mathcal{L}^\infty$ quando $\|f\|_\infty$ existe [56]. E para o caso discreto (espaço $\ell^\infty(\mathbb{Z})$) a norma é dada por

$$\|f\|_\infty = \sup_n |f_n|.$$

3.1.2 Caso Estocástico

Para o caso estocástico, seja a tripla $\{\Omega, \mathfrak{R}, P\}$ que caracteriza um espaço de probabilidade, onde Ω é o espaço amostral, \mathfrak{R} é a família de eventos e P é a função de probabilidade. Considere ainda \mathfrak{Z} um espaço de dimensão n de vetores aleatórios. $\mathcal{L}^p(\Omega)$ é o conjunto de elementos \mathbf{z} do espaço \mathfrak{Z} que satisfazem $E\{\|\mathbf{z}(\omega)\|^p\} < \infty$ [6].

Considere o sistema estocástico

$$\mathbf{z}_{n+1}(\omega) = \mathbf{f}_n(\mathbf{z}_n(\omega), \omega), \omega \in \Omega, n \in \mathbb{N}, \mathbf{f}_n : \mathbb{R}^N \times \Omega \rightarrow \mathbb{R}^N. \quad (3.3)$$

Seja o conjunto \mathcal{A} um subconjunto de \mathfrak{Z} .

Definição 15. A solução do sistema descrito pela equação (3.3) é estável em probabilidade com respeito a \mathcal{A} se para todo $\epsilon > 0$, $\eta > 0$, $n_0 \in \mathbb{N}$ existe $\delta(\eta, \epsilon) > 0$, tal que $\mathbf{z}_0(\omega) \in \mathcal{A}$ e $P(\omega \| \|\mathbf{z}_0(\omega)\|_p > \delta(\eta, \epsilon)) < \delta(\eta, \epsilon)$ então $P(\omega \| \|\mathbf{z}_n(\omega)\|_p > \eta) < \epsilon$ para $n \geq n_0$ [57].

Nota 6. De maneira intuitiva, o sistema estocástico estável deveria ser tal qual a maioria das amostras da sequência $\{z_n\}$ (ou seja o conjunto de amostras com probabilidade arbitrariamente próxima a 1) tenha $\|\mathbf{z}_n\|$ limitada [57].

Definição 16. A solução do sistema descrito pela equação (3.3) é estável com probabilidade 1 se para todo $\epsilon > 0$, $\eta > 0$, $n_0 \in \mathbb{N}$ se existe $\delta(\eta, \epsilon) > 0$ tal que $\|\mathbf{z}_0(\omega)\|_p < \delta(\eta, \epsilon)$ então $P(\omega | \sup \|\mathbf{z}_n(\omega)\|_p > \eta) < \epsilon$ para $n \geq n_0$ [5].

Definição 17. A solução do sistema descrito pela equação (3.3) é estável em $\mathcal{L}^p(\Omega)$, $1 < p < \infty$, com respeito a $\mathcal{A} = \{\mathbf{z}, \mathbf{z} \in \mathcal{L}^2(\Omega), \|\mathbf{z}\|_2 \leq H\}$ se para todo $\epsilon > 0$, $n_0 \in \mathbb{N}$, existe $\delta > 0$ tal que $\mathbf{z}_0(\omega) \in \mathcal{A} \cap \mathcal{L}^p(\Omega)$ e $\|\mathbf{z}_0(\omega)\|_p < \delta$ então $\|\mathbf{z}_n(n_0, x_0, \omega)\|_p < \epsilon$ para $n \geq n_0$. [6].

Definição 18. A solução do sistema dado pela equação (3.3) é uniformemente estável em $\mathcal{L}^p(\Omega)$, $1 < p < \infty$, com relação ao subconjunto \mathcal{A} se a definição 17 é válida e δ não depende de n_0 [6].

3.2 Primeiro Método de Lyapunov

A teoria de estabilidade de Lyapunov conta com o método indireto, ou primeiro método de Lyapunov, em que a linearização é utilizada para obter resultados acerca das propriedades do sistema dinâmico em torno do ponto de equilíbrio. Vale ressaltar que a essência desse método corresponde ao tipo de metodologia que foi aplicada à análise de algoritmos adaptativos no capítulo anterior.

Dado o sistema dinâmico autônomo da equação (3.1), pode-se enunciar que a linearização em torno do ponto de equilíbrio é dada por [21]

$$\mathbf{z}_{n+1} = \mathbf{A}\mathbf{z}_n, \quad (3.4)$$

onde $\mathbf{z}_n = \mathbf{w}_n - \underline{\mathbf{w}}$. A matriz \mathbf{A} é definida como

$$\mathbf{A} = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{w}_n} \right|_{\mathbf{w}_n = \underline{\mathbf{w}}}. \quad (3.5)$$

Supondo que seja verdade

$$\lim_{\|\mathbf{w}_n\| \rightarrow \underline{\mathbf{w}}} \sup \frac{\|\mathbf{f}_n(\mathbf{w}_n) - \mathbf{A}\mathbf{w}_n\|}{\|\mathbf{w}_n\|} = 0, \quad (3.6)$$

então a equação (3.4) é conhecida como linearização de (3.1) em torno do ponto de equilíbrio \underline{w} . O teorema abaixo apresenta o resultado de estabilidade para esse caso.

Teorema 2. *Seja o sistema dinâmico autônomo da equação (3.1) onde \mathbf{f} é de classe C^1 e definindo-se³ a matriz \mathbf{A} conforme a equação (3.5), então o ponto de equilíbrio \underline{w} é assintoticamente estável se os autovalores de \mathbf{A} tem magnitude menor que 1.*

Vale ressaltar que o método de linearização permite apenas análises locais e de sistemas com matriz de termos constantes para avaliação dos autovalores. Caso contrário, o segundo método de Lyapunov se torna mais adequado para os estudos de estabilidade.

Exemplo 1: Seja o sistema de equações referentes ao modelo do pêndulo invertido

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\frac{g}{l}\sin(x_1) - \frac{k}{m}x_2\end{aligned}$$

onde é possível observar dois pontos de equilíbrio fundamentais ($x_1 = 0, x_2 = 0$) e ($x_1 = \pi, x_2 = 0$). O sistema de equações contínuo no tempo pode ser linearizando considerando que devemos, no caso deste trabalho, passar os sistema para coordenadas discretas,

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) \implies \mathbf{z}_{n+1} = \mathbf{G}\mathbf{z}_n,$$

sendo $\mathbf{G} = e^{\mathbf{A}T}$, T é o intervalo de amostragem⁴

$$\begin{aligned}\mathbf{A} &= \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}_n} \right|_{\mathbf{x}_n=(x_1=0, x_2=0)} = \begin{bmatrix} 0 & 1 \\ -\frac{g}{l} & -\frac{k}{m} \end{bmatrix}. \\ \mathbf{A}' &= \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}_n} \right|_{\mathbf{x}_n=(x_1=\pi, x_2=0)} = \begin{bmatrix} 0 & 1 \\ \frac{g}{l} & -\frac{k}{m} \end{bmatrix}.\end{aligned}$$

É necessário levar o sistema para o domínio discreto, uma vez que a teoria enunciada neste capítulo trata os sistemas discretos no tempo. Contudo, existe a versão contínua do primeiro método de Lyapunov na literatura: vide, por exemplo, [22] e [20]. A versão discreta da matriz \mathbf{A} é apresentada logo a seguir:

³Uma função é de classe C^1 se é contínua, diferenciável e a derivada é contínua.

⁴O cálculo de $e^{\mathbf{A}T}$ pode ser realizado por intermédio da transformada de Laplace, ou seja, $e^{\mathbf{A}T} = L^{-1}((s\mathbf{I} - \mathbf{A})^{-1})$, onde L^{-1} é a transformada inversa de Laplace [58].

$$A_{dis} = \begin{bmatrix} \frac{e^{-\frac{kT}{2m}} \left(\frac{kl^2 m \sin\left(\frac{\sqrt{l(4gm^2-k^2l)}T}{2lm}\right)}{\sqrt{l(4gm^2-k^2l)}} + lm \cos\left(\frac{\sqrt{l(4gm^2-k^2l)}T}{2lm}\right) \right)}{lm} & \frac{2gm e^{-\frac{kT}{2m}} \sin\left(\frac{\sqrt{l(4gm^2-k^2l)}t}{2lm}\right)}{\sqrt{l(4gm^2-k^2l)}} \\ \frac{2lm e^{-\frac{kT}{2m}} \sin\left(\frac{\sqrt{l(4gm^2-k^2l)}T}{2lm}\right)}{\sqrt{l(4gm^2-k^2l)}} & \frac{e^{-\frac{kT}{2m}} \left(lm \cos\left(\frac{\sqrt{l(4gm^2-k^2l)}T}{2lm}\right) - \frac{kl^2 m \sin\left(\frac{\sqrt{l(4gm^2-k^2l)}T}{2lm}\right)}{\sqrt{l(4gm^2-k^2l)}} \right)}{lm} \end{bmatrix}$$

$$A'_{dis} = \begin{bmatrix} \frac{e^{-\frac{kT}{2m}} \left(\frac{kl^2 m \sinh\left(\frac{\sqrt{l(4gm^2+k^2l)}T}{2lm}\right)}{\sqrt{l(4gm^2+k^2l)}} + lm \cosh\left(\frac{\sqrt{l(4gm^2+k^2l)}T}{2lm}\right) \right)}{lm} & \frac{2gm e^{-\frac{kT}{2m}} \sinh\left(\frac{\sqrt{l(4gm^2+k^2l)}T}{2lm}\right)}{\sqrt{l(4gm^2+k^2l)}} \\ \frac{2lm e^{-\frac{kT}{2m}} \sinh\left(\frac{\sqrt{l(4gm^2+k^2l)}T}{2lm}\right)}{\sqrt{l(4gm^2+k^2l)}} & \frac{e^{-\frac{kT}{2m}} \left(lm \cosh\left(\frac{\sqrt{l(4gm^2+k^2l)}T}{2lm}\right) - \frac{kl^2 m \sinh\left(\frac{\sqrt{l(4gm^2+k^2l)}T}{2lm}\right)}{\sqrt{l(4gm^2+k^2l)}} \right)}{lm} \end{bmatrix}$$

O valor absoluto dos autovalores para A_{dis} são

$$\lambda_1 = \sqrt{e^{-\frac{kT}{m}} \sin \left[\left(\frac{\sqrt{4glm^2 - k^2l^2T}}{2lm} \right)^2 \right] + e^{-\frac{kT}{m}} \cos \left[\left(\frac{\sqrt{4glm^2 - k^2l^2T}}{2lm} \right)^2 \right]}$$

$$\lambda_2 = \sqrt{e^{-\frac{kT}{m}} \sin \left[\left(\frac{\sqrt{4glm^2 - k^2l^2T}}{2lm} \right)^2 \right] + e^{-\frac{kT}{m}} \cos \left[\left(\frac{\sqrt{4glm^2 - k^2l^2T}}{2lm} \right)^2 \right]}$$

Note que, ambos os autovalores tem magnitude menor que 1 se considerarmos que $g, l, m, k > 0$. Ou seja, o ponto de equilíbrio ($x_1 = 0, x_2 = 0$) é assintoticamente estável.

No caso da matriz A'_{dis} , os autovalores são

$$\lambda'_1 = e^{-\frac{2kT}{m}} \left| e^{\frac{3kT}{2m}} \sinh \left(\frac{\sqrt{4glm^2 + k^2l^2T}}{2lm} \right) - e^{\frac{3kT}{2m}} \cosh \left(\frac{\sqrt{4glm^2 + k^2l^2T}}{2lm} \right) \right|$$

$$\lambda'_2 = e^{-\frac{2kT}{m}} \left| e^{\frac{3kT}{2m}} \sinh \left(\frac{\sqrt{4glm^2 + k^2l^2T}}{2lm} \right) + e^{\frac{3kT}{2m}} \cosh \left(\frac{\sqrt{4glm^2 + k^2l^2T}}{2lm} \right) \right|$$

Neste caso pelo menos um dos autovalores tem magnitude maior que 1 (considerando $g, l, m, k > 0$), ou seja, o ponto de equilíbrio ($x_1 = \pi, x_2 = 0$) não é estável. \diamond

Exemplo 2: Considere o sistema escalar com origem em $\underline{w} = 0$

$$w_{n+1} = aw_n^3.$$

Linearizando em torno do ponto de equilíbrio, temos:

$$A = \frac{\partial f}{\partial w_n} \Big|_{w_n=\underline{w}} = 3aw_n^2 \Big|_{w_n=\underline{w}}.$$

Veja que a linearização falha na análise do sistema para estabilidade em torno do ponto de equilíbrio. O sistema poderia ser assintoticamente estável, estável ou instável, dependendo do valor de a . Se $-1 < a < 0$, então a origem é assintoticamente estável. Isso poderia ser observado considerando o segundo método de Lyapunov com a função de Lyapunov $V(w_n) = w_n^4$. \diamond

3.3 Segundo Método de Lyapunov

Nesta seção, são apresentados alguns conceitos sobre o segundo método de Lyapunov para os casos determinístico e estocástico, bem como as hipóteses e restrições consideradas neste trabalho.

3.3.1 Caso determinístico

Teorema 3. *Dado um sistema dinâmico **determinístico**, o ponto de equilíbrio $\underline{\mathbf{w}}$ é estável no sentido de Lyapunov se existe um função $V(\mathbf{w}_n)$ tal que [37][21]*

1. *Existe uma região fechada e limitada em torno do ponto de equilíbrio $\underline{\mathbf{w}}$;*
2. *$V(\mathbf{w}_n)$ é uma função contínua e positiva definida, com $V(\underline{\mathbf{w}}) = 0$; ;*
3. *$\Delta V(\mathbf{w}_n) = V(\mathbf{w}_{n+1}) - V(\mathbf{w}_n) \leq 0$;*
4. *Alternativamente, se $V(\cdot)$ tem uma função limitante superior e inferior, ou seja, $\alpha(\|\mathbf{w}_n\|) \leq V(\mathbf{w}_n) \leq \beta(\|\mathbf{w}_n\|)$, então tem-se estabilidade uniforme;*
5. *Por outro lado, se $\Delta V(\mathbf{w}_n) = V(\mathbf{w}_{n+1}) - V(\mathbf{w}_n) < 0$ para todo $\mathbf{w}_n \neq \underline{\mathbf{w}}$, então tem-se estabilidade assintótica;*
6. *Caso $\lim_{\|\mathbf{w}_n\| \rightarrow \infty} V(\mathbf{w}_n) = \infty$, ou seja, $V(\cdot)$ é radialmente ilimitada, então $\underline{\mathbf{w}}$ é globalmente estável. O caráter assintótico ou uniforme dependerá se as condições anteriores (4 ou 5) serão satisfeitas.*

A função V é chamada de candidata à função de Lyapunov quando a mesma ainda está sob investigação do ponto do vista do Teorema 3. Quando se prova pelo menos a estabilidade do sistema, então a função V passa a ser chamada de função de Lyapunov.

Nota 7. *Note que o Teorema 3 foi enunciado de maneira a contemplar os sistemas variantes no tempo.*

Nota 8. *Geralmente, a função V apresenta as propriedades $V(\cdot) > 0$ e $\Delta V(\cdot) \leq 0$, o que implica que $\lim_{n \rightarrow \infty} V(\cdot) = V_\infty$ existe. Contudo, não há nenhuma garantia de que $\Delta V(\cdot) \rightarrow 0$ quando $n \rightarrow \infty$ [56]. Isso estimula os estudos de estabilidade assintótica e regiões de convergência.*

Nota 9. *No teorema 3, a desigualdade $\alpha(\|\mathbf{w}_n\|) \leq V(\mathbf{w}_n)$ corresponde a chamar $V(\mathbf{w}_n)$ de função positiva definida [20].*

Nota 10. No teorema 3, a desigualdade $V(\mathbf{w}_n) \leq \beta(\|\mathbf{w}_n\|)$ é relevante, pois garante o comportamento de V em torno do ponto de equilíbrio [20]. Essa é uma exigência adicional nos sistemas não-autônomos⁵ para provar estabilidade uniforme⁶. Caso contrário, a função V fica vulnerável a picos dependentes do tempo onde o sistema foi inicializado, o que descaracteriza a propriedade de uniformidade.

A figura 3.6 ilustra o significado geométrico das notas 9 e 10. Note que função de Lyapunov está limitada inferiormente e superiormente, o que caracteriza as funções de Lyapunov como funções que não exibem comportamentos abruptos.

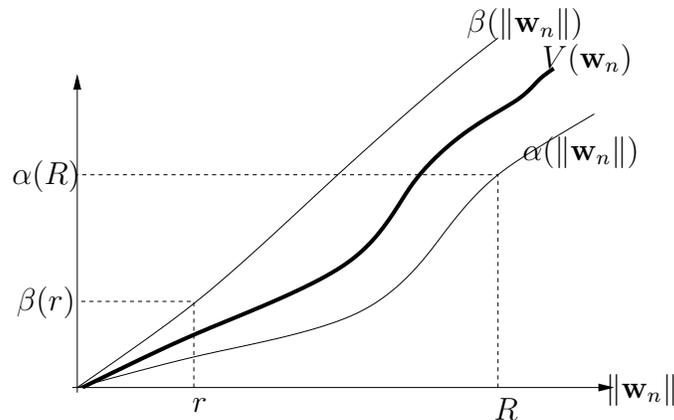


Fig. 3.6: Função positiva definida e decrescente.

Exemplo 3: Seja uma função $g(t)$ contínua e diferenciável que coincide com $e^{-t/2}$ exceto em alguns picos onde a função tem valor 1. A figura 3.7 representa $g^2(t)$ [20].

Se a função de Lyapunov é tal que $\Delta V = -x^2 < 0$, então temos uma das condições obedecidas, mas a origem (ou ponto de equilíbrio) nunca será assintoticamente estável.

Note que, quando $V(\cdot)$ é decrescente, então o caráter de estabilidade assintótica será possível de ser encontrado. Contudo, a figura 3.7 mostra um exemplo onde, mesmo $V(\cdot)$ localmente positiva definida, a energia do sistema não dissipa ao longo da trajetória. Por isso, temos os picos representados na figura. Consequentemente, para garantir a característica assintótica é requerido que $V(\cdot)$ seja decrescente.

◇

Exemplo 4: Considere um sistema determinístico ARMA de primeira ordem [37]

$$y_n = -ay_{n-1} + bu_{n-1}.$$

⁵Sistemas autônomos são aqueles que cujo o comportamento é invariante com relação ao tempo [22].

⁶No caso de sistemas autônomos, o conceito de estabilidade é mais intuitivo, uma vez que necessário apenas V positiva definida e $V(\mathbf{w}_{n+1}) - V(\mathbf{w}_n) < 0$. Contudo, o mesmo não se aplica em sistemas não-autônomos.

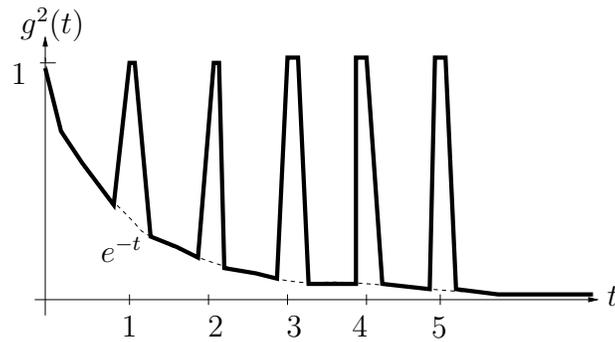


Fig. 3.7: Análise de função decrescente.

Esse sistema pode ser expresso em função de

$$\phi_{n-1}^T = [-y_{n-1}, u_{n-1}],$$

$$\theta_0^T = [a, b],$$

e a saída do sistema é dada por

$$y_n = \phi_{n-1}^T \theta_0.$$

Seja o algoritmo para estimar parâmetros [37]

$$\hat{\theta}_n = \hat{\theta}_{n-1} + \frac{\phi_{n-1}}{\phi_{n-1}^T \phi_{n-1}} (y_n - \phi_{n-1}^T \hat{\theta}_{n-1}).$$

Seja $\tilde{\theta}_n = \hat{\theta}_n - \theta_0$ o vetor de erro dos parâmetros $\tilde{\theta}_n$ pode ser escrito em função do erro de estimação como

$$\tilde{\theta}_n = \left(I - \frac{\phi_{n-1} \phi_{n-1}^T}{\phi_{n-1}^T \phi_{n-1}} \right) \tilde{\theta}_{n-1}.$$

Tomando como função candidata a Lyapunov:

$$V_n = \tilde{\theta}^T \tilde{\theta},$$

consequentemente,

$$V_{n+1} = V_n - \frac{\tilde{\theta}_n^T \phi_{n-1} \phi_{n-1}^T \tilde{\theta}_n}{\phi_{n-1}^T \phi_{n-1}}$$

Somando k termos

$$\begin{aligned}
 V_{n+k} &= V_n - \sum_{i=0}^{k-1} \frac{\tilde{\theta}_{n+i}^T \phi_{n+i} \phi_{n+i}^T \tilde{\theta}_{n+i}}{\phi_{n+i}^T \phi_{n+i}} \\
 &= V_n - \sum_{i=0}^{k-1} \tilde{\theta}_n^T \frac{\Psi_{n+i}^T \phi_{n+i} \phi_{n+i}^T \Psi_{n+i}}{\phi_{n+i}^T \phi_{n+i}} \tilde{\theta}_n \\
 &= V_n - \tilde{\theta}_n^T \left(\sum_{i=0}^{k-1} \frac{\Psi_{n+i}^T \phi_{n+i} \phi_{n+i}^T \Psi_{n+i}}{\phi_{n+i}^T \phi_{n+i}} \right) \tilde{\theta}_n \\
 &= V_n - cV_n = (1 - c)V_n, \quad c > 0.
 \end{aligned}$$

sendo Ψ a matriz de transição de estados e c uma constante positiva.

Para garantir estabilidade então é necessário que $c < 1$. Observando que $(1 - c) < e^{-c}$, conseqüentemente

$$V_{n+k} \leq e^{-c} V_n.$$

No caso em que temos N passos de comprimento k

$$V_{n+kN} \leq e^{-cN} V_0.$$

Ou seja, a grandeza $\tilde{\theta}$ tem convergência exponencial para zero. \diamond

Hipótese 5. *Os sistemas discutidos nos próximos capítulos deste trabalho são sistemas autônomos. Mais especificamente, não existe nenhuma dependência explícita da variável discreta temporal n nas outras variáveis que definem o sistema dinâmico, como \mathbf{w}_n [22][20].*

3.3.2 Caso estocástico

A teoria de Lyapunov para sistemas estocásticos pode ser apresentada em outras formas, nas quais, por exemplo, é considerado o comportamento do processo estocástico como um semi-martingale [57] ou mesmo como sistema markoviano [7]. No entanto, essas abordagens não abrangem todos os casos de sistemas dinâmicos. Neste trabalho, consideramos o tratamento abrangente introduzido em [6].

Uma outra possibilidade seria considerar o sistema estocástico como uma cadeia de Markov. Na verdade, muitos sistemas dinâmicos se enquadram nessa situação. Neste caso, é possível empregar os conceitos de recorrência e irredutibilidade. Essa abordagem é conhecida como critério Foster-Lyapunov [59], sendo utilizada em problemas de otimização para controle de produção e estoques [60], e também na análise de redes complexas (e.g. redes de comunicação) [59] ou até mesmo estrutura de protocolos de comunicação de dados [61].

No caso de sistemas markovianos discretos, pode-se enunciar:

Teorema 4. *Seja z_1, \dots um processo de Markov discreto nos parâmetros, sendo $V(\mathbf{z}_n) \geq 0$ e $Q_m = \{\mathbf{z}_n : V(\mathbf{z}_n) < m\}$. Em Q_m seja*

$$E\{V(\mathbf{z}_{n+1})|\mathbf{z}_n \in Q_m\} - V(\mathbf{z}_n) = -k(\mathbf{z}_n) \leq 0, .$$

Então,

$$P\left\{\sup_{0 \leq n < \infty} V(\mathbf{z}_n) \geq 0\right\} \leq \frac{V(\mathbf{z}_n)}{m}.$$

Existe uma variável aleatória v , $0 \leq v \leq m$, tal que $V(\mathbf{z}_n) \rightarrow v$ com probabilidade $\geq [1 - \frac{V(\mathbf{z}_n)}{m}]$. Também $k(\mathbf{z}_n) \rightarrow 0$ em Q_m com pelo menos probabilidade que \mathbf{z}_n está em Q_m , para todo $n < \infty$ [5].

Contudo, a abordagem deste trabalho se restringe ao caso mais simples onde não se levam em conta as possibilidades de sistema estável em probabilidade, o que é um dos casos apresentados em [6].

Em linhas gerais, o sistema estocástico estável deveria ser tal que $E\{V_n(\mathbf{z}_n)\}$ seja limitada [57].

O sistema estocástico poderia ser aproximado por uma representação (sistema) determinístico com ruído aditivo [22]. Contudo, essa aproximação não cobre os casos gerais, pois nem sempre a característica estocástica de um sistema pode ser simplesmente representada como um termo aditivo.

Teorema 5. *O ponto de equilíbrio de um sistema dinâmico **estocástico** é uniformemente estável em $\mathcal{L}_2(\Omega)$ com respeito a $\mathcal{A} = \{\mathbf{z}, \mathbf{z} \in \mathcal{L}_2(\Omega), \|\mathbf{z}\| \leq H\}$ se [6]:*

1. *A função V é contínua e positiva definida. V tem limitante superior e inferior α e β , sendo ambas funções de classe K . Ou seja, $\alpha(\|\mathbf{z}_n\|) \leq V(\mathbf{z}_n) \leq \beta(\|\mathbf{z}_n\|)$;*
2. *$E\{V(\mathbf{z}_{n+1}(n_0, \mathbf{z}_0, \omega))\} \leq E\{V(\mathbf{z}_n(n_0, \mathbf{z}_0, \omega))\}$ para $n_0 \in \mathbb{N}$, $n > n_0$ e $\|\mathbf{z}_{n_0}\| \geq H$.*

O teorema 5 assegura que um sistema é estável se a esperança da função de energia $V(\cdot)$ (o raciocínio da teoria de Lyapunov é a análise da energia do sistema) não aumenta ao longo do tempo [57].

3.4 Interpretação Geométrica

A teoria de Lyapunov para análise de estabilidade pode ser considerada à luz de uma interpretação baseada na evolução do uso da energia total de um sistema dinâmico. É possível que o sistema

dispenda energia ao longo do tempo e que essa mesma energia não seja repostada [57]. Nesse caso, o estado do sistema pode ser direcionado lentamente para algum ponto ou região em que tenderá a ficar encerrado indefinidamente, o que nos remete, no caso de Lyapunov, à noção de estabilidade.

Se considerarmos que V é uma função genérica de energia do sistema dinâmico, então a explanação acima implica

$$V(\mathbf{w}_n) \geq V(\mathbf{w}_{n+1}) \geq \dots V(\mathbf{w}_{n+N}).$$

Se for considerado um sistema estocástico, interessante, a abordagem intuitiva permanece, de certa forma, aplicável, desde que sejam consideradas todas as possíveis realizações [57]. Consequentemente, temos

$$E\{V(\mathbf{w}_n)\} \geq E\{V(\mathbf{w}_{n+1})\} \geq \dots E\{V(\mathbf{w}_{n+N})\}.$$

Do ponto de vista geométrico, se considerarmos a função energia do sistema representada juntamente com a trajetória, então é possível notar a perda de energia. Os sistemas que possuem essa característica apresentarão uma região de confinamento, de onde o sistema não consegue sair após adentrá-la.

A figura 3.8 apresenta a visão geométrica de um caso simples em que a trajetória do sistema é representada na superfície que descreve a função energia bem como em curvas de nível refletidas no plano. O sistema perde energia passando do nível de energia V_1 para V_3 . Uma analogia simples é o tanque de combustível de um carro. No começo de uma viagem, o tanque está cheio, ou seja, com nível de energia potencial V_1 . Contudo, ao longo do caminho, a energia potencial é consumida para movimentar o veículo. Isso faz com que o veículo, após alguns quilômetros, tenha se associado ao nível de energia V_2 . Ao término do percurso, o carro chega ao nível de energia V_3 . Obviamente, abastecer o tanque do carro significa adicionar energia no sistema. Geralmente estamos analisando o sistema dinâmico nos momentos em que o mesmo não sofre influência externa para acrescentar mais energia. Mas isso não impede o uso da teoria de Lyapunov se considerarmos que a energia do sistema pode ser um dos parâmetros do modelo. Apresentamos dois exemplos abaixo para exemplificar melhor um sistema estável do ponto de vista de Lyapunov.

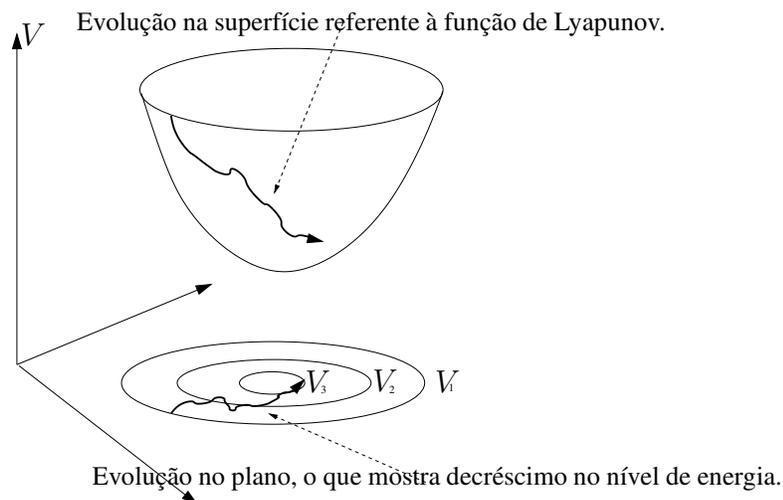


Fig. 3.8: Interpretação geométrica para teoria de Lyapunov.

Exemplo 5: Seja um sistema dinâmico com o modelo

$$w_{n+1} = aw_n, |a| < 1, n > 0.$$

Para qualquer ponto de inicialização w_0 , o sistema terá um valor de $|w_n|$ menor que o do passo anterior. Note que o sistema é estável devido à condição $|a| < 1$ (considere a função de Lyapunov $V = w_n^2 \implies \Delta V = (a^2 - 1)w_n^2$). \diamond

Exemplo 6: Seja um sistema escalar [57]

$$w_{n+1} = (a + \xi_n)w_n, n > 0.$$

onde ξ_n é i.i.d com

$$E\{\xi_n\} = 0, E\{\xi_n^2\} = \sigma^2.$$

Escolhendo a candidata à função de Lyapunov como

$$V(w_n) = w_n^2,$$

então

$$E\{V(w_{n+1})|w_n\} = (a^2 + \sigma^2)w_n^2,$$

Portanto, se $(a^2 + \sigma^2) < 1$, então

$$E\{V(w_{n+1})|w_n\} - V(w_n) = (a^2 + \sigma^2 - 1)w_n^2 < 0.$$

Consequentemente, a origem do sistema é assintoticamente estável. Este exemplo pode ser estendido para cobrir o caso vetorial. \diamond

Intuitivamente, um sistema estocástico estável deveria ser tal que $E\{V(\mathbf{w}_n)\}$ permanece limitada superiormente. Em termos da sequência $\{\mathbf{w}_n\}$, um sistema estocástico estável deveria ser tal que, para a maioria das amostras de $\{\mathbf{w}_n\}$, isto é, para um conjunto de amostras com probabilidade arbitrariamente próxima a 1, $\|\mathbf{w}_n\|$ deveria ser limitado [57].

Um sistema estocástico assintoticamente estável deve obedecer à condição $E\{V(\mathbf{w}_n)\}$ limitada para todo n e fazer com que $E\{V(\mathbf{w}_n)\}$ decresça monotonicamente para zero com probabilidade 1 [57].

3.5 Análise de Convergência

Os resultados obtidos via análise de Lyapunov também podem auxiliar a análise de convergência. Os teoremas e definições apresentados até o momento valem tanto para sistemas autônomos como para não-autônomos. Contudo, se considerado autônomos, é possível obter resultados mais fortes com relação ao caráter assintótico da estabilidade.

Definição 19. *O conjunto G em \mathbb{R}^n é um conjunto invariante para um sistema dinâmico se toda trajetória do sistema que começa de um ponto de G permanece em G para todo o tempo futuro [20].*

Qualquer ponto de equilíbrio é um conjunto invariante. O domínio de atração de um ponto de equilíbrio também é um conjunto invariante [20]. Para um sistema autônomo, qualquer trajetória no espaço de estados é um conjunto invariante.

Os resultados sobre conjuntos invariantes permitem estender os conceitos da teoria de Lyapunov para descrever convergência de sistemas dinâmicos em termos mais abrangentes que o de ponto de equilíbrio, como a convergência para um ciclo limite.

Comentário 1. *Existe uma extensão do princípio original de invariância de LaSalle para o caso discreto [62]. Basicamente, se as hipóteses expostas no Teorema 3 são válidas para um conjunto invariante, então o ponto estável será assintoticamente estável. Veja o corolário abaixo descrevendo o princípio da invariância para o caso discreto no tempo.*

Corolário 1. *Se uma função de Lyapunov tem derivadas parciais contínuas, limitadas inferiormente em \mathbb{C}^n e satisfazendo $\Delta V(\mathbf{w}_n) \leq 0$ ao longo de trajetórias para frente, então qualquer trajetó-*

ria limitada tende para o maior subconjunto invariante contido no lugar geométrico G dos pontos definidos por $\Delta V(\mathbf{w}_n) = 0$ [62].

Nota 11. Sob o princípio da invariância para sistemas discretos no tempo, existirá estabilidade assintótica mesmo quando $\Delta V(\mathbf{w}_n)$ não é estritamente menor que 0 [20].

Exemplo 7: Considere o sistema de duas dimensões

$$x_{n+1} = \frac{ay_n}{1+x_n^2} \quad y_{n+1} = \frac{bx_n}{1+y_n^2}.$$

Vamos definir em G o seguinte conjunto

$$F = \{x_n, y_n : \Delta V = 0, x_n, y_n \in G\}.$$

Então, considerando $V = x_n^2 + y_n^2$,

$$\Delta V = \left(\frac{b^2}{(1+y_n^2)^2} - 1 \right) x_n^2 + \left(\frac{a^2}{(1+x_n^2)^2} - 1 \right) y_n^2.$$

A análise de ΔV vai requerer diversos casos.

Caso 1: $a^2 < 1, b^2 < 1$

Para essa situação

$$\Delta V \leq (b^2 - 1)x_n^2 + (a^2 - 1)y_n^2 \leq 0$$

Veja que a função V é uma função de Lyapunov em \mathbb{R}^2 . Neste caso, $F = \{(0, 0)\}$ e toda solução é limitada. Toda solução se aproximará da origem quando $n \rightarrow \infty$. Este é um caso clássico de Lyapunov onde V e $-\Delta V$ são funções positivas definidas. Neste cenário, tem-se estabilidade assintótica.

Caso 2: $a^2 \leq 1, b^2 \leq 1$ e $a^2 + b^2 < 2$

Vamos assumir que $a^2 < 1$ e $b^2 = 1$. V ainda é uma função de Lyapunov em \mathbb{R}^2 , mas $-\Delta V$ não é positiva definida. Veja que

$$\Delta V \leq (a^2 - 1)y_n^2 \leq 0$$

Neste caso F corresponde ao eixo x_n , ou seja, todos os pontos onde $y_n = 0$. Mas quando $x_n \neq 0$ e $y_n = 0$ então

$$x_n \neq 0, y_n = 0 \implies x_{n+1} = 0, y_{n+1} = bx_n.$$

Ou seja, o ponto de equilíbrio é a origem. Neste caso, o ponto de equilíbrio terá estabilidade assintótica.

Caso 3: $a^2 = b^2 = 1$

V é uma função de Lyapunov e todas as soluções são limitadas. Neste caso, as possíveis soluções serão

$$\{(c, 0), (0, c), (-c, 0), (0, -c)\}$$

onde c representa o ponto de intersecção com o lugar geométrico $x_n^2 + y_n^2 = c^2$. Existem dois sub-casos para análise.

Caso 3-a: $ab = 1$

$$x_n = c, y_n = 0 \implies x_{n+1} = 0, y_{n+1} = bc \implies x_{n+2} = abc = c, y_{n+2} = 0.$$

Veja que, neste caso, a solução apresenta um caráter periódico (período 2). Consequentemente, qualquer solução ou vai para a origem ou vai para o movimento periódico.

Caso 3-b: $ab = -1$

$$x_n = c, y_n = 0 \implies x_{n+1} = 0, y_{n+1} = bc \implies x_{n+2} = abc = -c,$$

$$y_{n+2} = 0 \implies x_{n+3} = 0, y_{n+3} = -bc \implies x_{n+4} = -abc = c, y_{n+4} = 0.$$

Considerando $c \neq 0$, tem-se movimentos periódicos com período igual a 4. Como no caso 3-a, ou a solução é a origem ou o movimento periódico.

Caso 4: $a^2 > 1, b^2 > 1$

Seja o conjunto

$$B_\delta = \{x_n, y_n : x_n^2 + y_n^2 < \delta^2\},$$

então, para $x_n \in B_\delta$ e δ suficientemente pequeno

$$\Delta V \geq \left(\frac{b^2}{(1 + \delta^2)} - 1 \right) x_n^2 + \left(\frac{a^2}{(1 + \delta^2)} - 1 \right) y_n^2 \geq 0$$

Portanto, nenhuma solução que comece em B_δ conseguirá convergir para origem, pois a distância com relação à origem somente aumenta a medida que o sistema evolui no tempo. Consequentemente, as condições de Lyapunov não são obedecidas e o sistema é considerado instável.

Do ponto de vista do princípio da invariância, cada solução deve sair da região deli-

mitada por B_δ e, como não existe nenhuma solução que consiga avançar diretamente para a origem (exceto a própria origem), então não temos solução trivial quando $n \rightarrow \infty$.

◇

O tratamento de sistemas de equações a diferenças para o caso não-autônomo é descrito em [63]. Contudo, esse caso pode ser representado pela figura 3.9 onde a intenção é dar uma noção intuitiva de conjuntos invariantes. Dado o ponto $y \in G$, tomando a sequência $\mathbf{w}_k \rightarrow \mathbf{m}$ quando $k \rightarrow \infty$. A figura 3.9 representa tanto a evolução espacial quanto a temporal do sistema dinâmico. As expressões que limitam o sistema (lugar geométrico de y) de uma certa forma apresentam propriedades de tratamento invariante [63], apesar de nem sempre estarmos tratando um conjunto invariante para o caso não-autônomo [20].

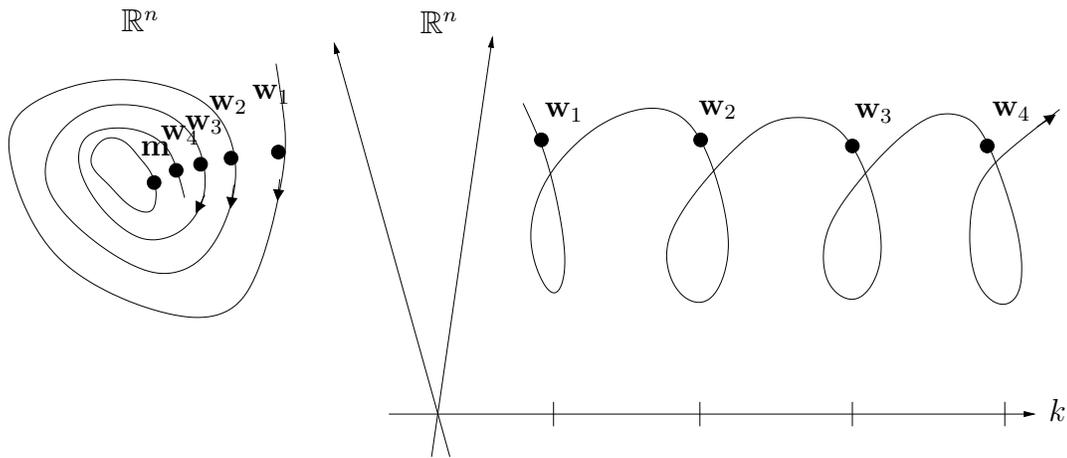


Fig. 3.9: Representação do princípio da invariância.

Nota 12. Assuma que exista uma função de Lyapunov $V(\mathbf{w}_n)$ definida em um conjunto D tal que $\Delta V(\mathbf{w}_n) \leq 0$ para qualquer $\mathbf{w}_n \in D$. Então as variáveis de estado (ou trajetórias do sistema) tendem ao maior conjunto invariante que contém a lugar geométrico definido por $\Delta V(\mathbf{w}_n) = 0$ para qualquer $n > 0$ com probabilidade 1⁷ [64].

3.6 Sumário

A teoria de Lyapunov para o estudo de estabilidade de sistemas dinâmicos, abrangendo tanto o primeiro quanto o segundo método, foi apresentada neste capítulo.

⁷Para um ambiente estocástico, a origem é estável com probabilidade 1 quando existe $\delta > 0$ e $\epsilon > 0$ e ainda existe $\phi(\delta, \epsilon)$ tal que $\|\mathbf{u}\| \leq \phi(\delta, \epsilon)$ então $Probabilidade(\sup_n \|\mathbf{u}\| \leq \epsilon) \leq \delta$ [57].

No caso do primeiro método de Lyapunov, a exposição realizada mostrou que, em essência, seu *modus operandi* já é clássico na análise de algoritmos de equalização, embora nem sempre a potencial sinergia entre teoria de sistemas dinâmicos e processamento de sinais seja plenamente explorada. Por outro lado, o segundo método de Lyapunov se mostra uma ferramenta mais poderosa, pois permite tratar sistemas não-lineares determinísticos e estocásticos segundo um arcabouço conceitualmente similar. Outra vantagem muito importante do segundo método é que sua aplicação não exige nem a prévia obtenção dos pontos de equilíbrio do sistema dinâmico nem a adoção de simplificações de modelo (por exemplo, linearização). A busca pelo emprego do segundo método, como verá o leitor, será uma tônica ao longo dos próximos capítulos.

Como reflexão final, desejamos chamar a atenção para o fato de que, no caso de processamento digital de sinais, é bastante usado conceito de estabilidade BIBO (*Bounded-Input Bounded-Output*), que, em termos simples, baseia-se na idéia de que, dada uma sequência discreta como entrada de um sistema dinâmico caracterizada por uma limitação superior de amplitude das amostras, espera-se que uma limitação similar seja válida para a saída de um sistema estável. Interessantemente, a estabilidade de Lyapunov exige que as variáveis do sistema dinâmico fiquem confinadas em uma região ou contorno, o que apresenta pontos de contato que gostaríamos de ressaltar com o conceito de estabilidade BIBO, apesar de ambas as noções serem aplicadas em contextos distintos.

Capítulo 4

Análise da Estabilidade de Algoritmos Supervisionados Baseada no Segundo Método de Lyapunov

Neste capítulo, investigaremos a estabilidade no sentido de Lyapunov de dois algoritmos de filtragem supervisionada: o SDA e o LMS. Essas análises são as primeiras contribuições específicas deste trabalho.

A análise do SDA servirá ao propósito de introduzir a teoria de Lyapunov no escopo de análise de estabilidade de algoritmos de equalização adaptativa. Isso servirá de subsídio para análises mais complexas, como as do LMS e do CMA.

Vale ressaltar que as técnicas aplicadas neste capítulo podem ser estendidas para outros tipos de metodologia de filtragem, tanto para canais e equalizadores lineares quanto para o caso não-linear.

4.1 Análise do Algoritmo SDA

Apresentaremos agora a análise, do ponto de vista do segundo método de Lyapunov, do algoritmo SDA, exposto no capítulo 2.3.2. Tendo em vista a estrutura do algoritmo, trata-se de uma análise em cenário inerentemente determinístico.

A função custo do SDA um passo à frente pode ser escrita como

$$J_{SDA}(\hat{\mathbf{w}}_{n+1}) = \sigma_s^2 - \hat{\mathbf{w}}_{n+1}^T \mathbf{p} - \mathbf{p}^T \hat{\mathbf{w}}_{n+1} + \hat{\mathbf{w}}_{n+1}^T \mathbf{R} \hat{\mathbf{w}}_{n+1}. \quad (4.1)$$

Da definição da expressão de adaptação do SDA ($\hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n + \mu(\mathbf{p} - \mathbf{R}\hat{\mathbf{w}}_n)$) da equação (2.23),

pode-se escrever a função custo um passo à frente de uma nova maneira:

$$J_{SDA}(\hat{\mathbf{w}}_{n+1}) = \sigma_s^2 - (\hat{\mathbf{w}}_n + \mu(\mathbf{p} - \mathbf{R}\hat{\mathbf{w}}_n))^T \mathbf{p} - \mathbf{p}^T (\hat{\mathbf{w}}_n + \mu(\mathbf{p} - \mathbf{R}\hat{\mathbf{w}}_n)) + (\hat{\mathbf{w}}_n + \mu(\mathbf{p} - \mathbf{R}\hat{\mathbf{w}}_n))^T \mathbf{R} (\hat{\mathbf{w}}_n + \mu(\mathbf{p} - \mathbf{R}\hat{\mathbf{w}}_n)). \quad (4.2)$$

A função candidata a função Lyapunov será proposta conforme o trabalho de [18], ou seja, a função custo que gera o algoritmo será utilizada como base. Vale ressaltar que, no caso do SDA, temos uma função custo convexa [27][28], o que facilita a análise de Lyapunov. Os resultados obtidos em [18] sobre funções de Lyapunov convexas são aplicáveis aqui, o que significa que o algoritmo, se convergir, o fará para um único ponto. Isto se deve ao fato de o princípio da invariância - vide nota 11 (seção 3.5 na página 57) - ser aplicável neste contexto.

$$V_{SDA}(\hat{\mathbf{w}}_n) = J_{SDA}(\hat{\mathbf{w}}_n) - J_{SDA}(\underline{\mathbf{w}}) \geq 0, \quad (4.3)$$

sendo $V_{SDA}(\hat{\mathbf{w}}_n) = 0$ somente se $\hat{\mathbf{w}}_n = \hat{\mathbf{w}}_{n,min}$. Note que a função candidata a função de Lyapunov obedece as restrições com relação a V ser positiva definida e contínua. Isso é facilmente observado se considerarmos que J_{SDA} é uma função quadrática, portanto convexa. Portanto, conforme o teorema 3, a função ΔV_{SDA}

$$\begin{aligned} \Delta V_{SDA} &= J_{SDA}(\hat{\mathbf{w}}_{n+1}) - J_{SDA}(\hat{\mathbf{w}}_n) \\ &= -(\hat{\mathbf{w}}_n + \mu(\mathbf{p} - \mathbf{R}\hat{\mathbf{w}}_n))^T \mathbf{p} - \mathbf{p}^T (\hat{\mathbf{w}}_n + \mu(\mathbf{p} - \mathbf{R}\hat{\mathbf{w}}_n)) \\ &\quad + (\hat{\mathbf{w}}_n + \mu(\mathbf{p} - \mathbf{R}\hat{\mathbf{w}}_n))^T \mathbf{R} (\hat{\mathbf{w}}_n + \mu(\mathbf{p} - \mathbf{R}\hat{\mathbf{w}}_n)) \\ &\quad + \hat{\mathbf{w}}_n^T \mathbf{p} + \mathbf{p}^T \hat{\mathbf{w}}_n - \hat{\mathbf{w}}_n^T \mathbf{R} \hat{\mathbf{w}}_n. \end{aligned} \quad (4.4)$$

A expressão (4.4) pode ser simplificada para

$$\begin{aligned} \Delta V_{SDA} &= \mu^2 (-\mathbf{p}^T \mathbf{R}^T \mathbf{R} \hat{\mathbf{w}}_n - \hat{\mathbf{w}}_n^T \mathbf{R}^T \mathbf{R} \mathbf{p} + \hat{\mathbf{w}}_n^T \mathbf{R}^T \mathbf{R} \mathbf{R} \hat{\mathbf{w}}_n + \mathbf{p}^T \mathbf{R} \mathbf{p}) \\ &\quad + \mu (-2\mathbf{p}^T \mathbf{p} + 2\mathbf{p}^T \mathbf{R} \hat{\mathbf{w}}_n + 2\hat{\mathbf{w}}_n^T \mathbf{R} \mathbf{p} - 2\hat{\mathbf{w}}_n^T \mathbf{R}^T \mathbf{R} \hat{\mathbf{w}}_n) \\ &\stackrel{?}{\leq} 0. \end{aligned} \quad (4.5)$$

O próximo passo é analisar a inequação presente na expressão (4.5). O ponto de interrogação (?) foi incluído apenas para alertar o leitor que neste ponto ainda não sabemos se inequação realmente será válida. Veja que, $\Delta V_{SDA} = 0$ é um polinômio com relação a μ sem o termo constante. Consequentemente, para fazer valer a inequação (4.5), basta considerar os valores de μ tais que o segundo

membro da igualdade seja menor que um dado θ , conforme equação (4.6).

$$\mu \leq \theta(\mathbf{p}, \mathbf{R}, \hat{\mathbf{w}}_n) = \frac{+2\mathbf{p}^T \mathbf{p} - 2\mathbf{p}^T \mathbf{R} \hat{\mathbf{w}}_n - 2\hat{\mathbf{w}}_n^T \mathbf{R} \mathbf{p} + 2\hat{\mathbf{w}}_n^T \mathbf{R}^T \mathbf{R} \hat{\mathbf{w}}_n}{-\mathbf{p}^T \mathbf{R}^T \mathbf{R} \hat{\mathbf{w}}_n - \hat{\mathbf{w}}_n^T \mathbf{R}^T \mathbf{R} \mathbf{p} + \hat{\mathbf{w}}_n^T \mathbf{R}^T \mathbf{R} \mathbf{R} \hat{\mathbf{w}}_n + \mathbf{p}^T \mathbf{R} \mathbf{p}}. \quad (4.6)$$

Se o passo de adaptação μ é escolhido de acordo com a inequação (4.6), então o sistema determinístico será estável no sentido de Lyapunov pois $\Delta V_{SDA} \leq 0$. Por outro lado, μ deve obedecer à hipótese 2, ou seja, o numerador e denominador de equação (4.6) devem ter o mesmo sinal. Neste ponto, a candidata a função de Lyapunov V_{SDA} passa a se chamar função de Lyapunov uma vez que é possível extrair considerações sobre estabilidade. Vale lembrar que consideramos que $\mu > 0$.

A função de Lyapunov $V(\hat{\mathbf{w}}_n)$ pode ser limitada por $\kappa \|\hat{\mathbf{w}}_n - \underline{\mathbf{w}}\|^p \geq J_{CMA}(\hat{\mathbf{w}}_n) - J_{CMA}(\underline{\mathbf{w}}) \geq \vartheta \|\hat{\mathbf{w}}_n - \underline{\mathbf{w}}\|^p$ para $p \geq 1, p \in \mathbb{Z}$ e $0 < \kappa < \vartheta < \infty$ [18], o que prova que a mesma função é decrescente e localmente positiva definida.

4.1.1 Análise de resultados

A figura 4.1 apresenta $\theta(\mathbf{p}, \mathbf{R}, \hat{\mathbf{w}}_n)$ para o caso de um canal de fase máxima com dois coeficientes (com função de transferência $h(z) = 1 + 1,5z^{-1}$ e atraso $d = 2$). Conforme mencionado no capítulo 2, o sinal transmitido é bipolar. Neste caso, o canal é de fase não-mínima. Cada ponto da superfície foi obtido através da média entre os valores gerados pelos estados do canal mencionado com o sinal de entrada escolhido ¹. A solução de Wiener para o atraso $d = 2$ é $[-0,2706, 0,5864]^T$.

¹Um canal com 2 coeficientes alimentado com sinal bipolar produz 8 possíveis vetores de saída. As oito possíveis saídas são obtidas fazendo vetores entrada combinando as possibilidades para s_n , isto é, $+1$ e -1

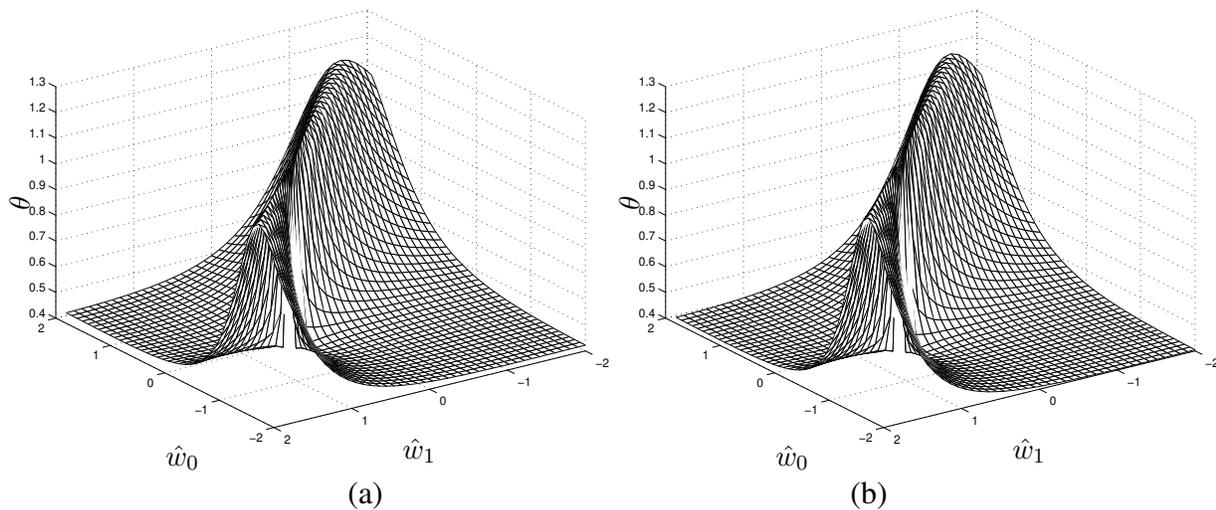


Fig. 4.1: (a) Superfície de θ para canal $h(z) = 1 + 1,5z^{-1}$ e $d = 2$, sem ruído aditivo. (b) Mesma simulação com ruído aditivo branco e gaussiano (AWGN) com SNR = 20 dB - 100 tentativas de Monte Carlo.

A superfície θ é um limitante superior para o passo de adaptação. Isso significa que o valor do passo de adaptação máximo é representado pela superfície da figura 4.1. Contudo, nada impede que o algoritmo utilize passos de adaptação menores que aqueles representados na figura. Existe uma relação direta entre passo de adaptação e velocidade de convergência, porém essa questão não é coberta aqui.

Para efeitos de equalização, devemos lembrar que consideramos um algoritmo de passo constante. Portanto, o projeto do equalizador deveria ser feito tomando o menor valor de θ disponível para evitar flutuações.

A primeira conclusão interessante é que existe uma dependência muito forte entre a escolha do ponto de inicialização do vetor de pesos e o valor máximo do passo de adaptação θ para que o algoritmo seja estável. Nos pontos próximos à solução de Wiener, é necessário um passo de adaptação menor, pois, do contrário, o algoritmo iria "saltar" o ponto de equilíbrio e nunca ficaria próximo o suficiente dele para ser considerado estável. Já nos pontos distantes do ponto de equilíbrio, é interessante ter um passo de adaptação maior. No entanto, não existe na superfície da figura 4.1 uma relação linear entre ponto de inicialização e passo de adaptação estável.

A dependência entre ponto de inicialização e passo de adaptação parece ser aceitável se for considerada a análise de algoritmos do tipo gradiente, uma vez que temos a razão entre os autovalores diferente de 1². Consequentemente, são esperados diferentes padrões de convergência associados a direções distintas determinadas pelos autovetores. Conforme mencionado em [27], os autovetores da

²Para o canal [1 1,5] os autovalores da matriz de autocorrelação são 1,75 e 4,75. Os autovetores são $[-0,70711 \ 0,70711]^T$ e $[0,70711 \ 0,70711]^T$.

matriz de autocorrelação definem os eixos principais da superfície de custo, como mostram as linhas mais destacadas da figura 4.2. A figura ainda representa as curvas de nível da função custo, bem como as curvas de nível relativas à superfície θ .

Outra observação é que a superfície θ não é plana. Portanto, esse é o motivo para a relação não-linear entre ponto de inicialização e passo de adaptação.

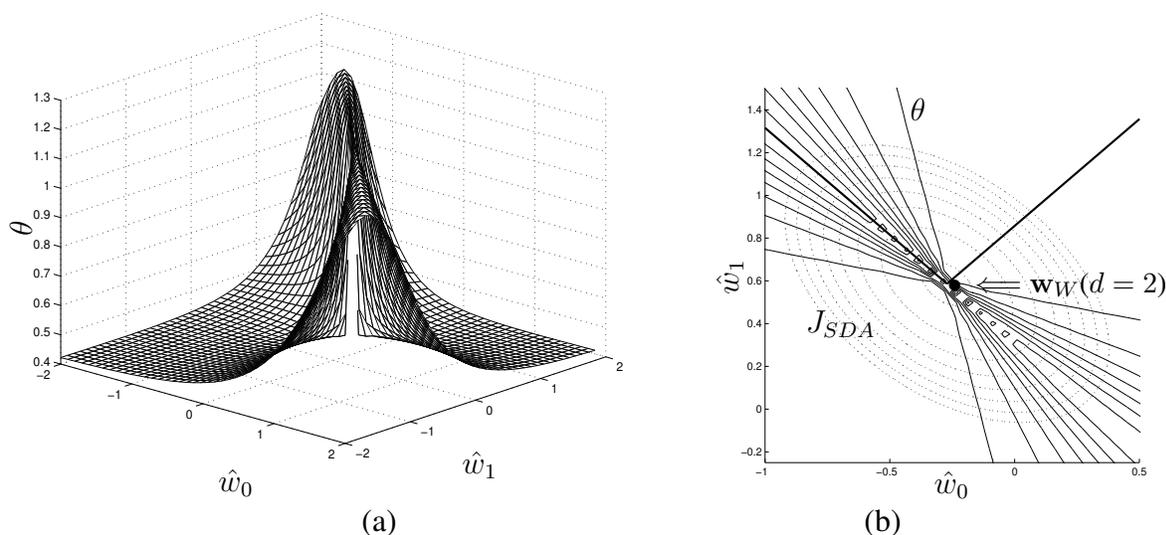


Fig. 4.2: (a) Superfície de θ para canal $h(z) = 1 + 1,5z^{-1}$ e $d = 2$, sem ruído aditivo. (b) Curva de nível da função custo (linha pontilhada) e θ , ou seja, o limitante superior do passo de adaptação (linha contínua).

Como a função de Lyapunov é baseada na função custo, nota-se que as curvas de nível da função custo e o limitante superior do passo de adaptação (o qual é derivada da função de Lyapunov) apresentam um ponto de mínimo comum. Esse mínimo comum é exatamente a solução de Wiener que oferece o menor valor da função custo. Nas regiões infinitesimalmente próximas do mínimo de Wiener é necessário um menor passo de adaptação para evitar que o algoritmo fique saltando em torno do ponto de equilíbrio.

É possível notar a dependência do formato da função custo e função θ através de outras simulações com outros canais, vide figuras 4.3, 4.4 e 4.5. Note que, novamente, os autovetores coordenam o formato das superfícies. A grande vantagem em termos de simplicidade de análise, no caso do SDA, é que o ponto de equilíbrio já é conhecido, uma vez que o mesmo coincide com a solução de Wiener.

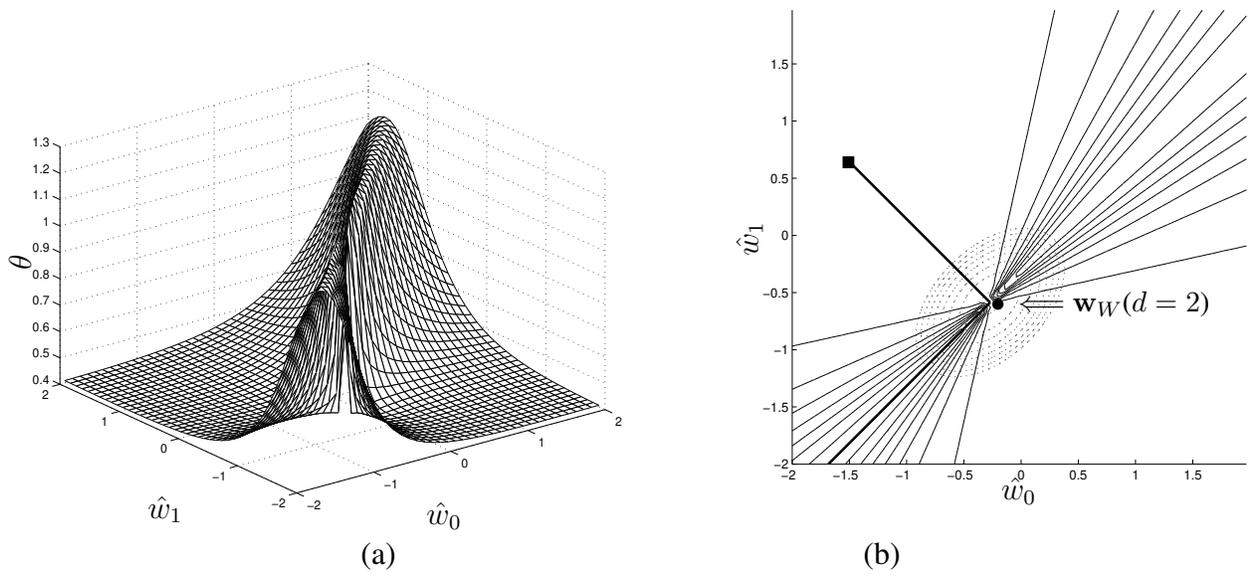


Fig. 4.3: (a) Superfície de θ para canal $h(z) = 1 - 1,5z^{-1}$ e $d = 2$, sem ruído aditivo. (b) Curva de nível da função custo (linha pontilhada) e θ , ou seja, o limitante superior do passo de adaptação (linha contínua).

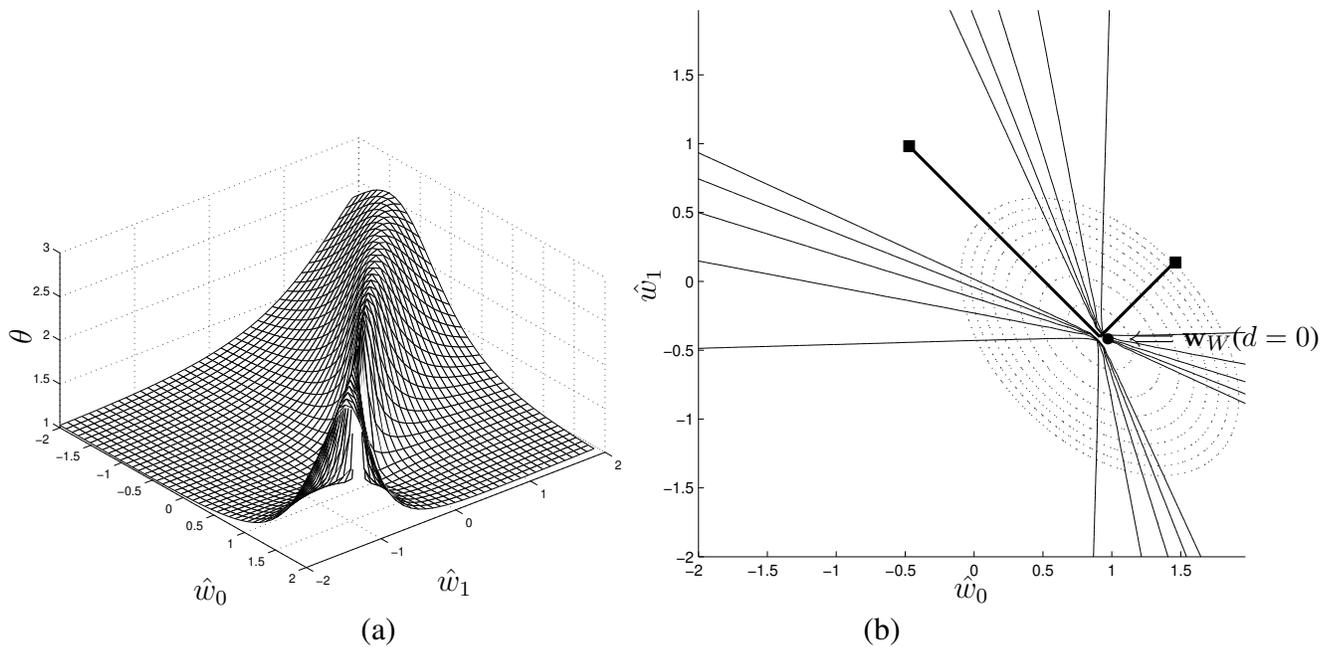


Fig. 4.4: (a) Superfície de θ para canal $h(z) = 1 + 0,6z^{-1}$ e $d = 0$, sem ruído aditivo. (b) Curva de nível da função custo (linha pontilhada) e θ , ou seja, o limitante superior do passo de adaptação (linha contínua).

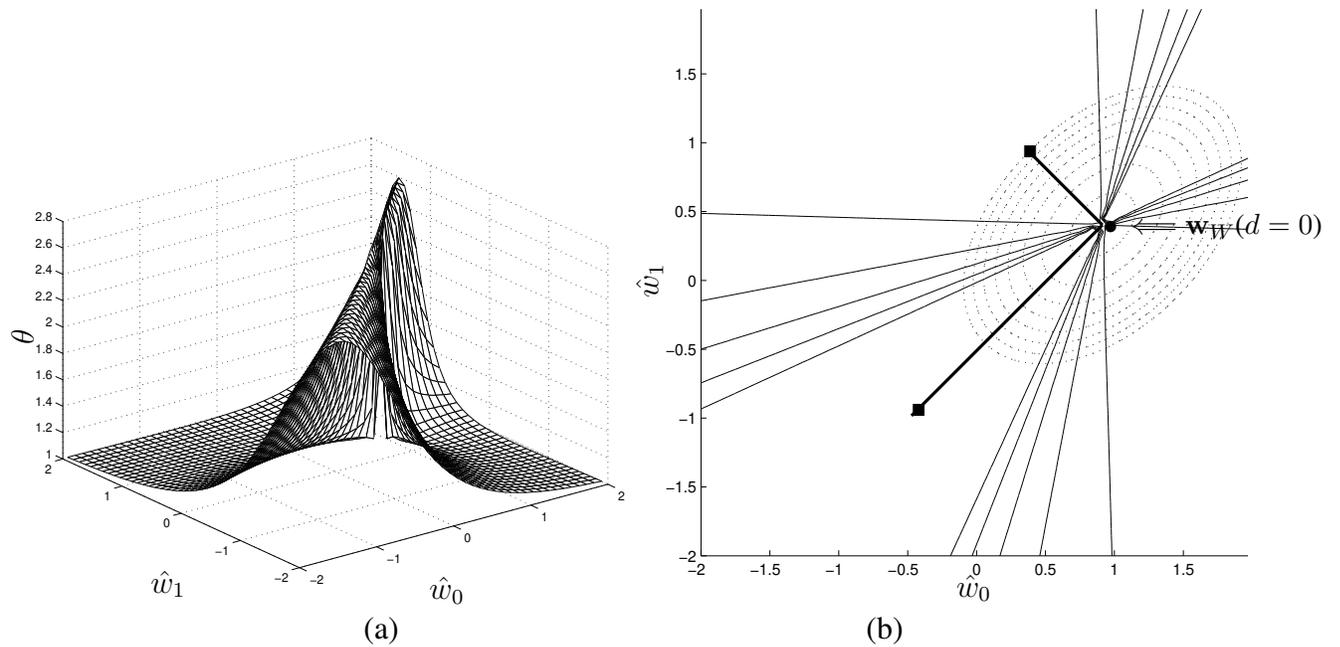


Fig. 4.5: (a) Superfície de θ para canal $h(z) = 1 - 0,6z^{-1}$ e $d = 0$, sem ruído aditivo. (b) Curva de nível da função custo (linha pontilhada) e θ , ou seja, o limitante superior do passo de adaptação (linha contínua).

A relação entre a matriz de autocorrelação \mathbf{R} e as curvas de nível da função de Lyapunov não surge por acaso. Se considerada a equação (2.23), que é a expressão de adaptação do algoritmo SDA, é possível invocar o conceito do primeiro método de estabilidade de Lyapunov e promover uma análise comum para sistemas lineares.

4.1.2 Extensão dos resultados

A expressão (4.6) pode ser escrita de maneira expandida como

$$\mu \leq \theta(\mathbf{p}, \mathbf{R}, \hat{\mathbf{w}}_n) = \frac{+2[(\mathbf{p} - \mathbf{R}\hat{\mathbf{w}}_n)^T(\mathbf{p} - \mathbf{R}\hat{\mathbf{w}}_n)]}{-\mathbf{p}^T \mathbf{R}^T \mathbf{R} \hat{\mathbf{w}}_n - \hat{\mathbf{w}}_n^T \mathbf{R}^T \mathbf{R} \mathbf{p} + \hat{\mathbf{w}}_n^T \mathbf{R}^T \mathbf{R} \mathbf{R} \hat{\mathbf{w}}_n + \mathbf{p}^T \mathbf{R} \mathbf{p}}. \quad (4.7)$$

A análise dos termos do denominador permite encontrar uma forma mais sucinta

$$-\mathbf{p}^T \mathbf{R}^T \mathbf{R} \hat{\mathbf{w}}_n - \hat{\mathbf{w}}_n^T \mathbf{R}^T \mathbf{R} \mathbf{p} + \hat{\mathbf{w}}_n^T \mathbf{R}^T \mathbf{R} \mathbf{R} \hat{\mathbf{w}}_n + \mathbf{p}^T \mathbf{R} \mathbf{p} = (\mathbf{p} - \mathbf{R}\hat{\mathbf{w}}_n)^T \mathbf{R} (\mathbf{p} - \mathbf{R}\hat{\mathbf{w}}_n). \quad (4.8)$$

Lembrando que, no caso do SDA, $\mathbf{p} - \mathbf{R}\mathbf{w}_n = \mathbf{e}$, conseqüentemente

$$\mu \leq \theta(\mathbf{p}, \mathbf{R}, \hat{\mathbf{w}}_n) = \frac{2\mathbf{e}^T \mathbf{e}}{\mathbf{e}^T \mathbf{R} \mathbf{e}} = 2 \left(\frac{\mathbf{e}^T \mathbf{R} \mathbf{e}}{\mathbf{e}^T \mathbf{e}} \right)^{-1}. \quad (4.9)$$

Isto nos dá uma expressão em função do erro e . Contudo, o extremo superior da expressão acima pode ser escrito em função dos autovalores da matriz de autocorrelação - de acordo com a idéia de quociente de Rayleigh - conforme [28, eq 8.21]

$$\mu \leq \theta(\mathbf{p}, \mathbf{R}, \hat{\mathbf{w}}_n) = \frac{2}{\lambda_{max}}. \quad (4.10)$$

Esse resultado é exatamente a expressão clássica do limite do passo de adaptação para o SDA, conforme exposto no capítulo 2. É importante dizer que, ao contrário do que foi feito aqui, as análises baseadas no método de Lyapunov disponíveis na literatura não lançam nenhuma conclusão com relação ao passo de adaptação [65][66][67].

4.2 Análise do Algoritmo LMS Utilizando Abordagem de Lyapunov Determinística

Algumas definições serão importantes para a análise da estabilidade do LMS. Primeiramente, definamos o erro entre o vetor ótimo e o vetor de parâmetros estimados,

$$\tilde{\mathbf{w}}_n = \hat{\mathbf{w}}_n - \mathbf{w}_W. \quad (4.11)$$

A partir de (4.11), podemos reescrever a expressão de adaptação do algoritmo LMS (2.31) como

$$\tilde{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n + \mu e_n \mathbf{x}_n - \mathbf{w}_W = \tilde{\mathbf{w}}_n + \mu e_n \mathbf{x}_n. \quad (4.12)$$

Denotando por v_n o erro associado à solução de Wiener, i.e., $\mathbf{x}_n^T \mathbf{w}_W + v_n = s_{n-d}$ podemos reescrever o sinal de erro e_n como

$$\begin{aligned} e_n &= s_{n-d} - \mathbf{x}_n^T \hat{\mathbf{w}}_n = \mathbf{x}_n^T \mathbf{w}_W + v_n - \mathbf{x}_n^T \hat{\mathbf{w}}_n \\ &= v_n - \mathbf{x}_n^T \tilde{\mathbf{w}}_n. \end{aligned} \quad (4.13)$$

Note ainda que a representação do erro em (4.13) se assemelha à forma obtida no problema de identificação de sistemas, onde o erro depende da distância entre o vetor de coeficientes da planta e do modelo ($\tilde{\mathbf{w}}$), e possivelmente de um ruído de medida (v_n) descorrelacionado do sinal de entrada

[13]. Conseqüentemente, substituindo (4.13) em (4.12), temos

$$\tilde{\mathbf{w}}_{n+1} = \tilde{\mathbf{w}}_n + \mu v_n \mathbf{x}_n - \mu \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n. \quad (4.14)$$

A hipótese adicional para essa análise é descrita como:

Hipótese 6. *O sinais envolvidos no problema de filtragem são sinais reais.*

Seja a candidata a função de Lyapunov ³

$$V_{LMS}(\tilde{\mathbf{w}}_n) = \tilde{\mathbf{w}}_n^T \tilde{\mathbf{w}}_n, \quad (4.15)$$

Nota 13. *A abordagem determinística é comumente considerada na literatura, pois seria uma forma particular do caso estocástico e pode ser considerada como a limitante para o pior caso. Alguns trabalhos que utilizam uma função de Lyapunov determinística mesmo em problemas estocásticos são [68][24][23].*

Nota 14. *Na construção da equação (4.15) é considerado que todos os termos da equação (2.31) são determinísticos. Conseqüentemente, a teoria de Lyapunov para sistemas determinísticos é aplicável.*

A diferença temporal da candidata a função de Lyapunov é expressa como (vide teorema 3)

$$\Delta V = V_{LMS}(\tilde{\mathbf{w}}_{n+1}) - V_{LMS}(\tilde{\mathbf{w}}_n) = \tilde{\mathbf{w}}_{n+1}^T \tilde{\mathbf{w}}_{n+1} - \tilde{\mathbf{w}}_n^T \tilde{\mathbf{w}}_n. \quad (4.16)$$

Na análise de (4.16), é preciso considerar a decomposição de $\tilde{\mathbf{w}}_{n+1}^T \tilde{\mathbf{w}}_{n+1}$. Da equação (4.14), temos

$$\begin{aligned} \tilde{\mathbf{w}}_{n+1}^T \tilde{\mathbf{w}}_{n+1} &= \tilde{\mathbf{w}}_n^T \tilde{\mathbf{w}}_n + \mu \tilde{\mathbf{w}}_n^T \mathbf{x}_n v_n - \mu \tilde{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n + \mu \mathbf{x}_n^T \tilde{\mathbf{w}}_n v_n \\ &\quad + \mu^2 \mathbf{x}_n^T \mathbf{x}_n v_n^2 - \mu^2 v_n \mathbf{x}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n - \mu \tilde{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n \\ &\quad - \mu^2 v_n \mathbf{x}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n + \mu^2 (\mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n)^T (\mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n). \end{aligned} \quad (4.17)$$

considerando a hipótese 6 (i.e $\mathbf{x}_n^T \tilde{\mathbf{w}}_n = \tilde{\mathbf{w}}_n^T \mathbf{x}_n$),

$$\begin{aligned} \tilde{\mathbf{w}}_{n+1}^T \tilde{\mathbf{w}}_{n+1} &= \tilde{\mathbf{w}}_n^T \tilde{\mathbf{w}}_n + 2\mu \tilde{\mathbf{w}}_n^T \mathbf{x}_n v_n - 2\mu \tilde{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n \\ &\quad + \mu^2 \mathbf{x}_n^T \mathbf{x}_n v_n^2 - 2\mu^2 v_n \mathbf{x}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n + \mu^2 (\mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n)^T (\mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n). \end{aligned} \quad (4.18)$$

³Veja que a condição 2 do teorema 3 é obedecida para esta função.

Portanto a expressão (4.16) pode ser reescrita como

$$\begin{aligned} \Delta V = & +2\mu\tilde{\mathbf{w}}_n^T \mathbf{x}_n v_n - 2\mu\tilde{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n \\ & + \mu^2 \mathbf{x}_n^T \mathbf{x}_n v_n^2 - 2\mu^2 v_n \mathbf{x}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n + \mu^2 (\mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n)^T (\mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n). \end{aligned} \quad (4.19)$$

Para provar estabilidade em torno do ponto de equilíbrio devemos ter $\Delta V \leq 0$, ou seja, da expressão (4.19), a seguinte desigualdade precisa ser verdadeira

$$\mu [\mathbf{x}_n^T \mathbf{x}_n v_n^2 - 2v_n \mathbf{x}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n + (\mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n)^T (\mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n)] \leq -2\tilde{\mathbf{w}}_n^T \mathbf{x}_n v_n + 2\tilde{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n. \quad (4.20)$$

Consequentemente, o passo de adaptação deve ser limitado por

$$\mu < \frac{2\tilde{\mathbf{w}}_n^T \mathbf{x}_n (\mathbf{x}_n^T \tilde{\mathbf{w}}_n - v_n)}{\mathbf{x}_n^T \mathbf{x}_n (\mathbf{x}_n^T \tilde{\mathbf{w}}_n - v_n)^T (\mathbf{x}_n^T \tilde{\mathbf{w}}_n - v_n)} = \theta. \quad (4.21)$$

Em casos próximos a condições de equalização/identificação perfeitas, podemos considerar v_n como sendo valor desprezível. Nesse caso, a expressão (4.21) para um limitante superior do passo de adaptação fica ainda mais simplificada, tornando-se uma regra prática do tipo:

$$\mu < \frac{2}{tr\{\mathbf{x}_n \mathbf{x}_n^T\}}. \quad (4.22)$$

Nota-se que essa expressão simplificada é similar à obtida através da análise clássica do LMS [28] vista no capítulo 2, o que indica a existência de coerência entre a abordagem de Lyapunov e a metodologia dependente da idéia de convergência no sentido do erro quadrático.

Na figura 4.6, apresentamos o resultado correspondente à equação (4.21) no contexto do problema de equalização de um canal composto por dois elementos com um filtro com dois coeficientes. Cada ponto da superfície foi obtido através da média de várias realizações do sinal de entrada considerado o mesmo vetor de pesos. Para explorar o limitante superior θ para cada ponto de inicialização, talvez o correto seria tomar como referência o menor valor de θ obtido em uma realização específica. Contudo, para um contexto geral de equalização, é necessário inicializar o algoritmo com $\mu_{max} < \min(\theta)$ para garantir estabilidade em todo o conjunto de estados. Portanto, tomar várias realizações para obter θ como uma referência para o passo de adaptação não nos parece danoso. Veja que a superfície apresenta uma forma convexa nas proximidades da solução de Wiener. Quanto mais próximo ao mínimo, menor deve ser o valor do passo de adaptação para que se tenha um passo minimizante no que se refere ao critério de Lyapunov. Tomando a equação (4.21), notamos uma dependência com $\tilde{\mathbf{w}}_n$, ou seja, o valor do limite para passo de adaptação depende da localização do vetor de pesos.

Note que a obtenção dessa superfície significa que conseguimos obter um guia para escolha do passo de adaptação para o problema abordado, o que era a nossa meta fundamental.

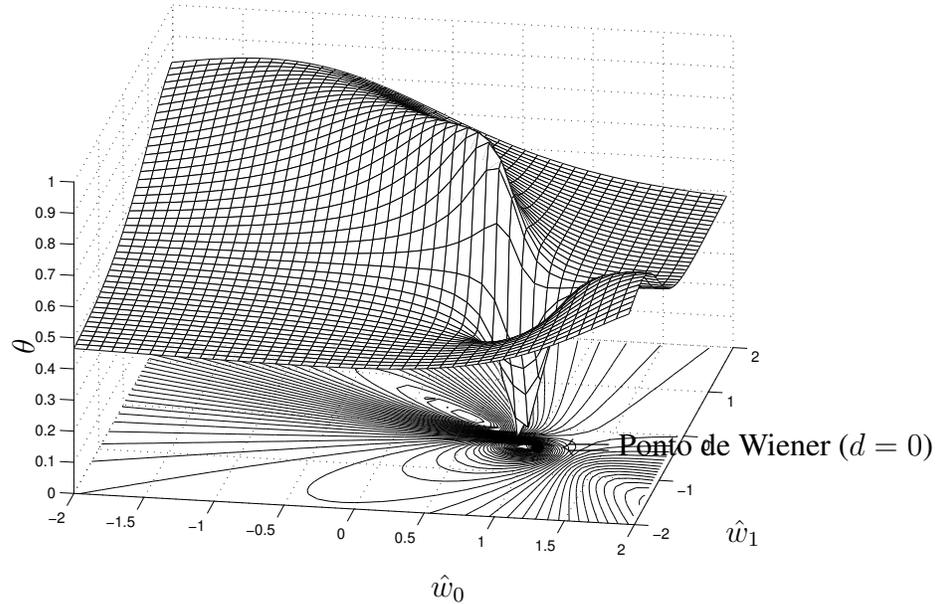


Fig. 4.6: Superfície que simboliza o limite para o passo de adaptação para equação (4.21). Canal $h(z) = 1 + 0,6z^{-1}$ e $d = 0$.

4.3 Análise do Algoritmo LMS Utilizando Abordagem de Lyapunov Estocástica

No caso estocástico, é necessário considerar uma versão da candidata à função de Lyapunov que utiliza o operador esperança, ou seja

$$V_{LMS}(\tilde{\mathbf{w}}_n) = E\{\tilde{\mathbf{w}}_n^T \tilde{\mathbf{w}}_n\}. \quad (4.23)$$

Consequentemente,

$$\begin{aligned} \Delta V &= V_{LMS}(\tilde{\mathbf{w}}_{n+1}) - V_{LMS}(\tilde{\mathbf{w}}_n) \\ &= E\{\tilde{\mathbf{w}}_{n+1}^T \tilde{\mathbf{w}}_{n+1}\} - E\{\tilde{\mathbf{w}}_n^T \tilde{\mathbf{w}}_n\}. \end{aligned} \quad (4.24)$$

Utilizando a expressão (4.18), obtemos (vide teorema 5)

$$\begin{aligned}
 \Delta V &= V_{LMS}(\tilde{\mathbf{w}}_{n+1}) - V_{LMS}(\tilde{\mathbf{w}}_n) \\
 &= E\{\tilde{\mathbf{w}}_n^T \tilde{\mathbf{w}}_n\} + E\{2\mu \tilde{\mathbf{w}}_n^T \mathbf{x}_n v_n\} - E\{2\mu \tilde{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n\} \\
 &\quad + E\{\mu^2 \mathbf{x}_n^T \mathbf{x}_n v_n^T v_n\} - E\{2\mu^2 v_n \mathbf{x}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n\} \\
 &\quad + E\{\mu^2 (\mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n)^T (\mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n)\} - E\{\tilde{\mathbf{w}}_n^T \tilde{\mathbf{w}}_n\}.
 \end{aligned} \tag{4.25}$$

Considerando a equação (4.25) podemos estabelecer as condições para que $\Delta V < 0$

$$\mu \leq \frac{2E\{\tilde{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n\} - 2E\{\tilde{\mathbf{w}}_n^T \mathbf{x}_n v_n\}}{E\{\mathbf{x}_n^T \mathbf{x}_n v_n^T v_n\} - E\{2v_n \mathbf{x}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n\} + E\{(\mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n)^T (\mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n)\}} = \theta. \tag{4.26}$$

Para casos que o valor de v_n (erro residual) é desprezível em comparação com os outros termos, a equação (4.26) pode ser reduzida a

$$\mu \leq \frac{2E\{\tilde{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n\}}{E\{(\mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n)^T (\mathbf{x}_n \mathbf{x}_n^T \tilde{\mathbf{w}}_n)\}} = \theta. \tag{4.27}$$

A figura 4.7 apresenta uma exemplo de superfície de nível θ da equação (4.26). As condições para esta simulação são as mesmas da figura 4.6.

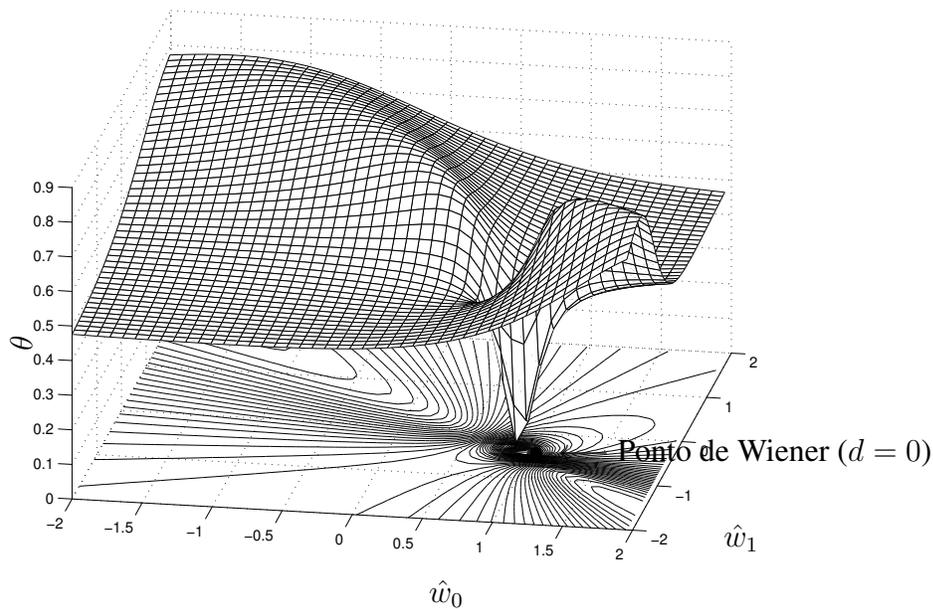


Fig. 4.7: Canal $h(z) = 1 + 0,6z^{-1}$ com atraso $d = 0$. Superfície que simboliza o passo de adaptação para equação (4.26).

A superfície referente à equação (4.27) é representada na figura 4.8.

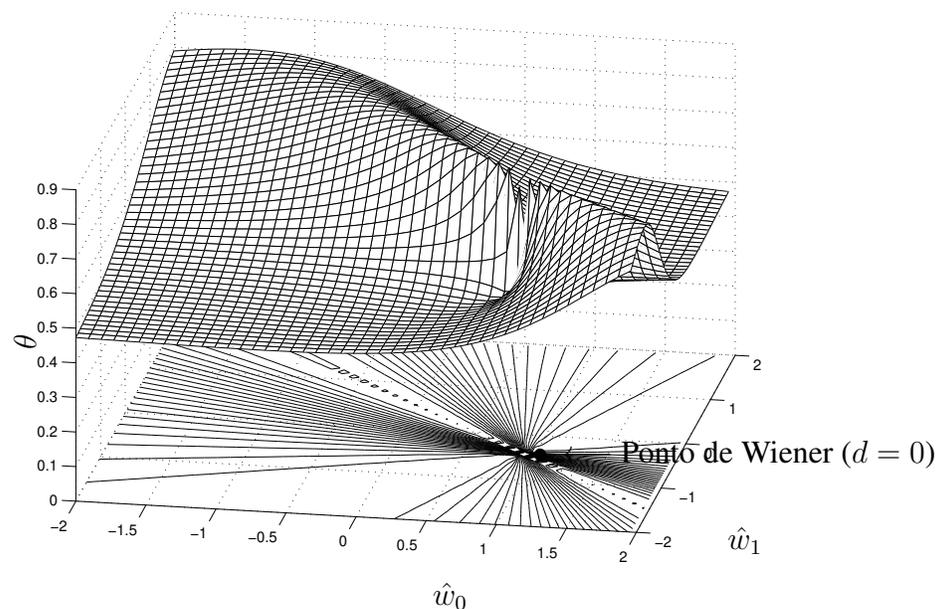


Fig. 4.8: Canal $h(z) = 1 + 0,6z^{-1}$ com atraso $d = 0$. Superfície que simboliza o passo de adaptação para equação (4.27).

É possível perceber que a figura 4.7 apresenta uma notável semelhança em sua forma geral com a figura 4.6, o que revela uma conexão entre a abordagem estocástica e a abordagem determinística quando diversas realizações são levadas em conta. Ainda mais interessante que isso, no entanto, é refletir sobre o limitante obtido através da expressão mostrada na figura 4.7. Antes de mais nada, é preciso ter em mente que a figura foi gerada para um caso de equalização em que a solução de Wiener não era capaz de obter um erro nulo. Isso faz com que seja esperada a existência de flutuações do vetor de pesos em torno da solução de Wiener, o que nos conduz à noção fundamental de *misadjustment* [28]. A existência dessas flutuações, de certo modo, justifica o "pico invertido" que se verifica na vizinhança do ótimo, pois, para, em termos simples, obter uma tendência de confinamento do vetor numa região cada vez menor em torno da solução de Wiener, passa a ser necessário empregar passos de adaptação cada vez mais reduzidos. Quando o vetor é igual ao ótimo, a única forma de mantê-lo nessa posição é zerar o passo para evitar flutuações residuais. Essa linha de raciocínio abre perspectivas interessantes em termos de análise de *misadjustment* vs. passo.

Para dar um pouco mais de suporte a essas idéias, mostramos, na figura 4.8, uma curva em que o erro residual da solução de Wiener é considerado nulo: nesse caso, não se verifica a existência do

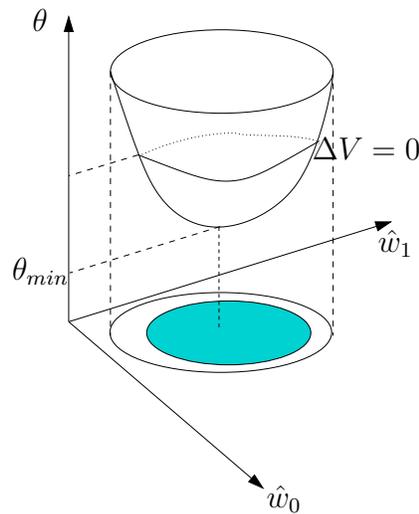


Fig. 4.9: Superfície representando o limitante superior do passo de adaptação e o ponto onde $\Delta V = 0$.

”pico invertido” que está disponível na figura 4.7, pois mesmo um passo não-nulo é incapaz de perturbar o equilíbrio associado ao ótimo. Esse seria o caso típico em problemas ideais de identificação, por exemplo. Contudo, ao ignorar o erro residual da solução de Wiener, estamos desprezando um fator relevante para determinar o valor preciso do passo de adaptação. Isso significa que desconsideramos a necessidade de um passo de adaptação menor em torno do mínimo de Wiener pois o ”pico invertido” da figura 4.7 desaparece. Ou seja, se utilizarmos a figura 4.8 como referência para o passo adaptação, nunca alcançaremos os valores ótimos do passo de adaptação em torno do ponto de equilíbrio.

Outro ponto interessante é que as curvas obtidas parecem estar razoavelmente limitadas por um valor máximo, fato que pode indicar, a partir de uma análise mais rigorosa, a viabilidade de se estabelecerem resultados relacionando valores do passo de adaptação que levam, em algum sentido estocástico, à convergência para o ótimo a partir de regiões iniciais diversas.

4.4 Análise de Convergência para LMS

4.4.1 Região de Convergência

No caso da teoria de Lyapunov, a região de convergência é dada por

$$\Delta V = 0, \quad (4.28)$$

ou seja, o lugar geométrico onde a função de Lyapunov passaria de uma função que ”perde energia” para uma situação onde a função ”acrescenta energia” no sistema dinâmico.

Portanto, para o cálculo do *misadjustment*, podemos tomar o pior caso e considerar que

$$J_{ex}(\infty) = J(n) |_{n \rightarrow \infty} - J_{min} = J(n) |_{\Delta V=0} - J_{min}, \quad (4.29)$$

A superfície limite para estabilidade é dada por

$$\Delta V = 0. \quad (4.30)$$

Contudo, a condição para o que o LMS determinístico seja estável é (vide equação (4.21))

$$\mu \leq \theta \implies \Delta V \leq 0. \quad (4.31)$$

Portanto, tomando

$$\mu = \theta \implies \Delta V = 0. \quad (4.32)$$

temos a região de convergência. Consequentemente, a região de convergência pode ser escrita como função de μ . Vide como exemplo para interpretação geométrica a figura 4.9.

Vejamos alguns exemplos que ilustram essa idéia para dois tipos diferentes de sinal transmitido.

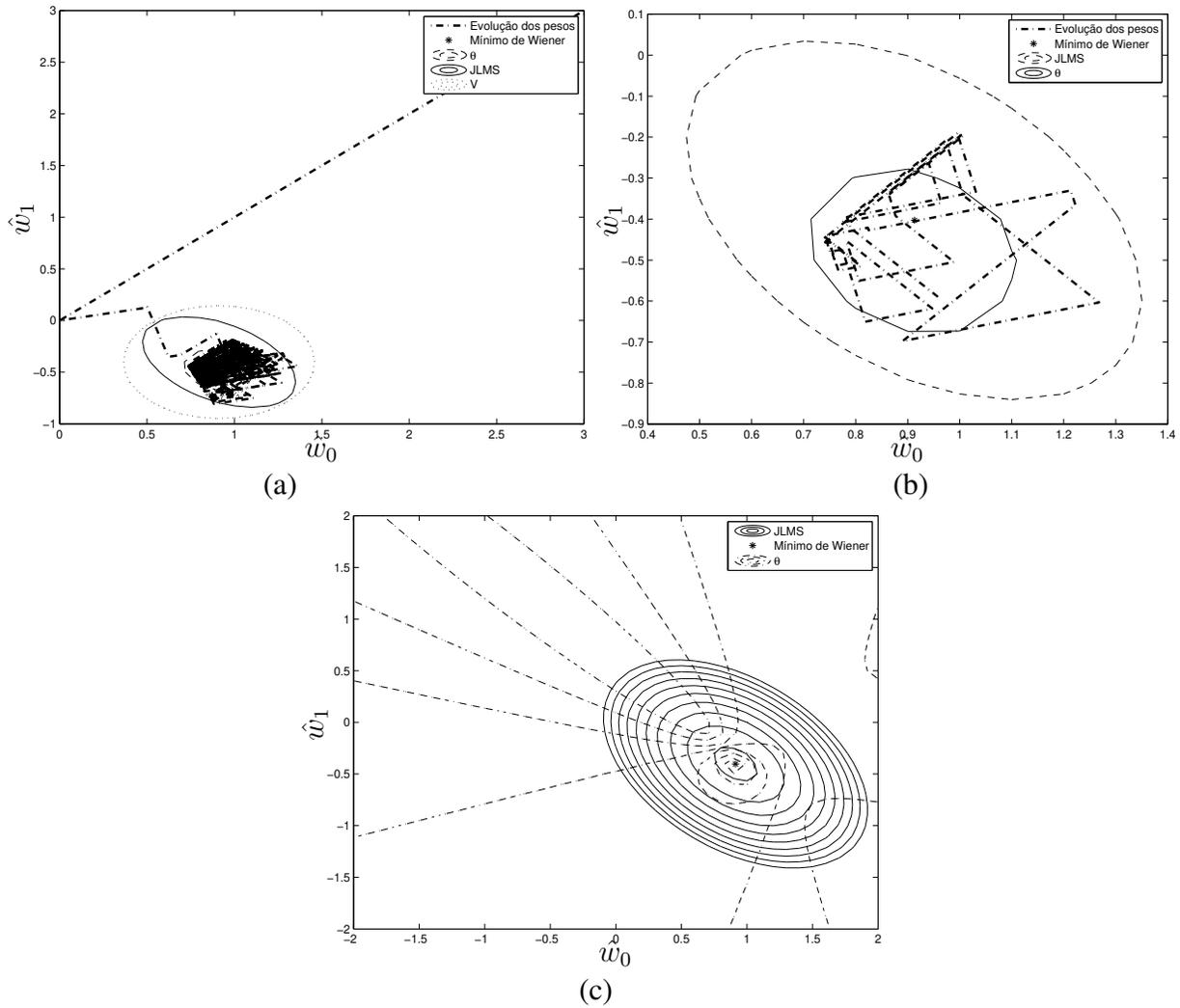


Fig. 4.10: (a) Evolução dos pesos do algoritmo LMS com análise de Lyapunov determinística. Canal $h(z) = 1 + 0,6z^{-1}$ com sinal s_n binário, $\mu = 0,3$. (b) Últimos passos da evolução dos pesos. (c) Curvas de nível de θ e J_{LMS} .

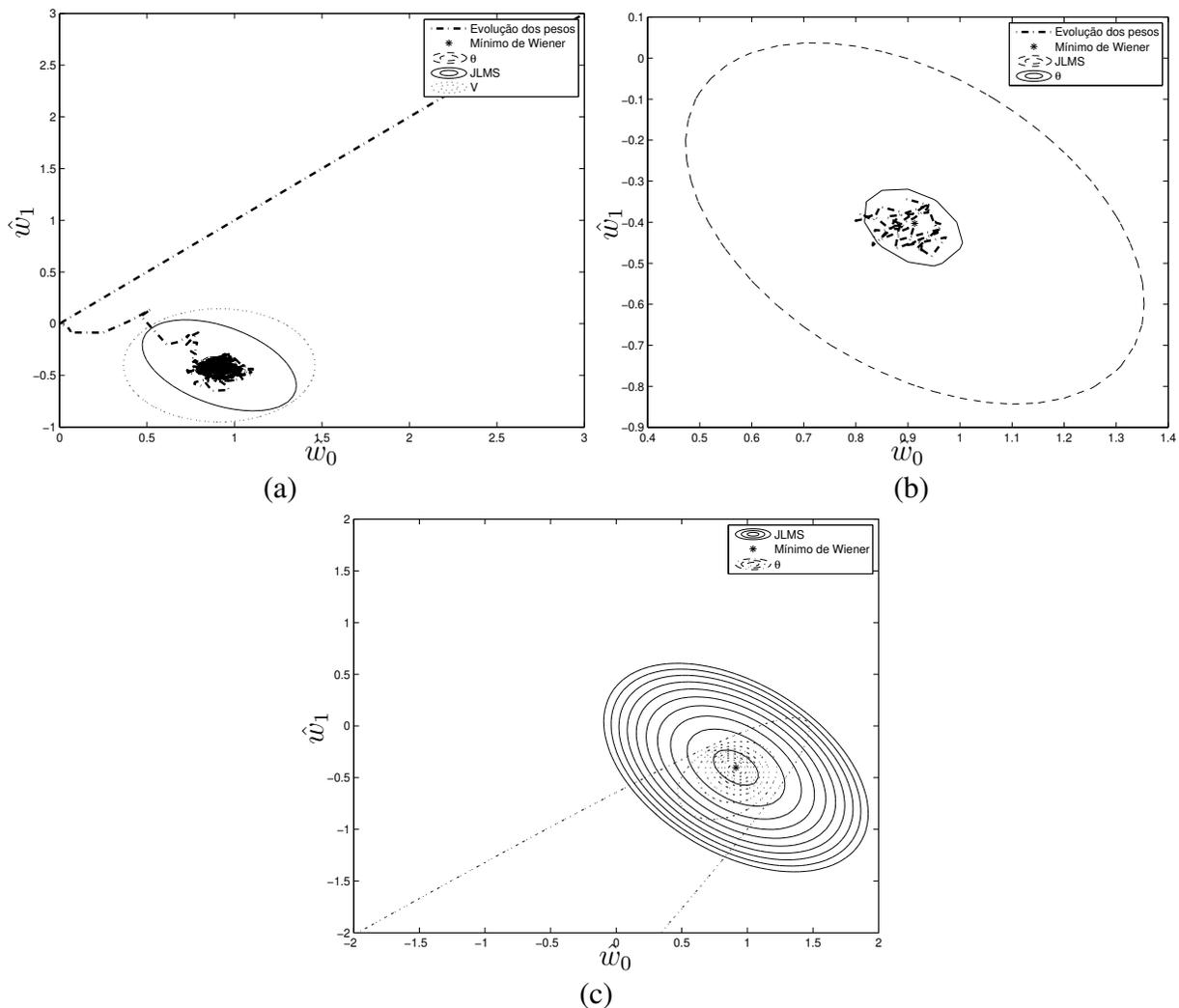


Fig. 4.11: (a) Evolução dos pesos do algoritmo LMS com análise de Lyapunov determinística. Canal $h(z) = 1 + 0,6z^{-1}$ com sinal $-1 + 2 * rand(1, N)$, $\mu = 0,3$. (b) Últimos passos da evolução dos pesos. (c) Curvas de nível de θ e J_{LMS} .

Nas figuras 4.10 e 4.11, o algoritmo evolui para a região definida por θ e permanece aproximadamente restrita a ela. As curvas de nível foram produzidas considerando o passo de adaptação μ , o que mostra claramente que o resultado é um fruto da análise de Lyapunov construída, ou seja, a metodologia proposta permite que seja estimado o alcance do efeito de "deriva" ao qual se associa o *misadjustment*.

No caso específico da figura 4.10, para a qual o sinal transmitido é uma excitação binária, a convergência tem um padrão mais complexo, já que, quando o algoritmo evolui para pontos próximos do mínimo, o valor da excitação é suficientemente grande para causar uma perturbação significativa. Neste sentido, o algoritmo produz como resultado passos para os quais os pesos de filtro não conver-

gem perfeitamente para o mínimo, mas gravitam em torno dele. Isto é, na verdade, uma consequência de se considerarem apenas algoritmos com passo de adaptação fixo.

Os efeitos indesejados do passo de adaptação fixo e sinal binário são menos severos no caso apresentado na figura 4.11.

4.4.2 Análise de convergência do LMS utilizando abordagem de Lyapunov determinística

Consideremos a expressão que garante a estabilidade do algoritmo LMS, ou seja, a equação (4.21):

$$\mu < \frac{2\tilde{\mathbf{w}}_n^T \mathbf{x}_n (\mathbf{x}_n^T \tilde{\mathbf{w}}_n - v_n)}{\mathbf{x}_n^T \mathbf{x}_n (\mathbf{x}_n^T \tilde{\mathbf{w}}_n - v_n)^T (\mathbf{x}_n^T \tilde{\mathbf{w}}_n - v_n)} = \theta. \quad (4.33)$$

Tomando $\mu = \theta$, então é possível determinar a região de convergência como

$$\mu = \frac{2\tilde{\mathbf{w}}_n^T \mathbf{x}_n (\mathbf{x}_n^T \tilde{\mathbf{w}}_n - v_n)}{\mathbf{x}_n^T \mathbf{x}_n (\mathbf{x}_n^T \tilde{\mathbf{w}}_n - v_n)^T (\mathbf{x}_n^T \tilde{\mathbf{w}}_n - v_n)}. \quad (4.34)$$

ou seja, a região de convergência para um dado passo μ pode ser obtida através da formulação

$$\mu \mathbf{x}_n^T \mathbf{x}_n (\mathbf{x}_n^T \tilde{\mathbf{w}}_n - v_n)^T (\mathbf{x}_n^T \tilde{\mathbf{w}}_n - v_n) - 2\tilde{\mathbf{w}}_n^T \mathbf{x}_n (\mathbf{x}_n^T \tilde{\mathbf{w}}_n - v_n) = 0. \quad (4.35)$$

Levando em conta o caso de um equalizador com dois coeficientes, que é favorável à ilustração de desempenho por meio de gráficos, podemos definir:

$$\tilde{\mathbf{w}}_n = [\tilde{w}_0 \quad \tilde{w}_1], \quad (4.36)$$

$$\mathbf{x}_n = [x_n \quad x_{n-1}], \quad (4.37)$$

Portanto, expandindo a expressão (4.35) e agrupando os termos em \tilde{w}_0 e \tilde{w}_1 , temos:

$$\begin{aligned} & (\mu x_{n-1}^4 + \mu x_n^2 x_{n-1}^2 - 2x_{n-1}^2) \tilde{w}_1^2 + (\mu x_n^2 x_{n-1}^2 + \mu x_n^4 - 2x_n^2) \tilde{w}_0^2 \\ & + (2\mu x_n x_{n-1}^3 + 2\mu x_n^3 x_{n-1} - 4x_n x_{n-1}) \tilde{w}_0 \tilde{w}_1 + (-2\mu v_n x_{n-1}^3 - 2\mu v_n x_n^2 x_{n-1} + 2v_n x_{n-1}) \tilde{w}_1 \\ & + (-2\mu v_n x_n x_{n-1}^2 - 2\mu v_n x_n^3 + 2v_n x_n) \tilde{w}_0 + (\mu v_n^2 x_{n-1}^2 + \mu v_n^2 x_n^2) = 0. \end{aligned} \quad (4.38)$$

A expressão (4.38) pode ainda ser reescrita para se obter uma representação mais simples.

$$a\tilde{w}_1^2 + b\tilde{w}_0^2 + c\tilde{w}_0\tilde{w}_1 + d\tilde{w}_1 + e\tilde{w}_0 + f = 0. \quad (4.39)$$

Notamos que a equação anterior representa, como lugar geométrico, uma seção cônica ($a\tilde{w}_0^2 + b\tilde{w}_1^2 + c\tilde{w}_0\tilde{w}_1 + d\tilde{w}_1 + e\tilde{w}_0 + f = 0$). Em linhas gerais a expressão $a\tilde{w}_0^2 + b\tilde{w}_1^2 + c\tilde{w}_0\tilde{w}_1 + d\tilde{w}_1 + e\tilde{w}_0 + f = 0$ pode representar uma elipse, parábola ou hipérbole. O critério para determinar qual é a cônica em questão depende do sinal de $c^2 - 4ab = \chi$. Se $\chi > 0$, então a cônica é uma hipérbole. Caso $\chi < 0$, trata-se de uma elipse. Em último caso, quando $\chi = 0$, tem-se uma parábola. Observando as superfícies de θ (vide como exemplo a figura 4.6), é possível observar a existência de elipses e hipérbolas como lugares geométricos disponíveis nas curvas de nível de θ .

Para o equalizador de dois coeficientes, do ponto de vista da análise de Lyapunov Determinística, os parâmetros que caracterizam a seção cônica no sistema de coordenadas $\tilde{w}_0 \times \tilde{w}_1$ são (vide equação (4.38)):

$$\begin{aligned} a &= \mu x_1^4 + (\mu x_0^2 - 2)x_1^2 \\ b &= \mu x_0^2 x_1^2 + \mu x_0^4 - 2x_0^2 \\ c &= 2\mu x_0 x_1^3 + (2\mu x_0^3 - 4x_0)x_1 \\ d &= -2\mu v_n x_1^3 + (2v_n - 2\mu v_n x_0^2)x_1 \\ e &= -2\mu v_n x_0 x_1^2 - 2\mu v_n x_0^3 + 2v_n x_0 \\ f &= \mu v_n^2 x_1^2 + \mu v_n^2 x_0^2. \end{aligned} \quad (4.40)$$

É bom salientar que os resultados obtidos até então são referentes ao sistemas de coordenadas $\tilde{w}_0 \times \tilde{w}_1$. Consequentemente, é possível transladar os resultados para o sistema de coordenadas original $w_0 \times w_1$.

As coordenadas dos pontos referentes ao lugar geométrico definido por $a\tilde{w}_0^2 + b\tilde{w}_1^2 + c\tilde{w}_0\tilde{w}_1 + d\tilde{w}_0 + e\tilde{w}_1 + f = 0$ são definidas, de maneira vetorial, como:

$$\tilde{\mathbf{w}}_{n,conv} = [\tilde{w}_0 \quad \tilde{w}_1] , \quad (4.41)$$

ou no caso do sistema nas coordenadas originais, isto é, sem a translação para o mínimo de Wiener

$$\hat{\mathbf{w}}_{n,conv} = [\hat{w}_0 \quad \hat{w}_1] . \quad (4.42)$$

Conseqüentemente, uma estimativa para o *misadjustment* pode ser escrita como

$$M = \frac{J(n) |_{\mathbf{w}_{n,conv}} - J_{min}}{J_{min}}. \quad (4.43)$$

Utilizando a expressão do custo para o LMS (equação (2.26)), o valor do *misadjustment* pode ser expresso como

$$M = \max \left(\frac{\sigma_s^2 - \hat{\mathbf{w}}_{n,conv}^T \mathbf{p}_n - \mathbf{p}_n^T \hat{\mathbf{w}}_{n,conv} + \hat{\mathbf{w}}_{n,conv}^T \mathbf{R}_n \hat{\mathbf{w}}_{n,conv} - J_{min}}{J_{min}} \right), \quad (4.44)$$

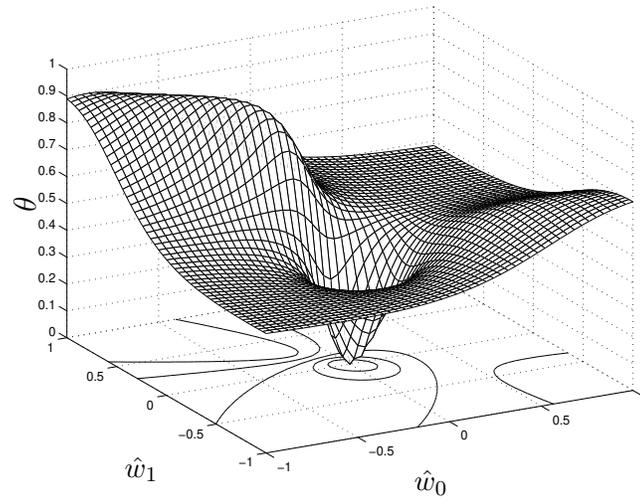
A expressão para o *misadjustment* (equação (4.44)) tem algumas propriedades clássicas. Uma delas é a relação direta entre M e μ . Contudo, no caso da expressão (4.44), é possível considerar uma sensível redução no esforço para se determinar o *misadjustment*, uma vez que não é necessário calcular os autovalores da matriz de autocorrelação. Isso pode ser uma vantagem no sentido de permitir a análise de sistemas invariantes no tempo e/ou não-lineares.

Ao utilizar a equação (4.44) devemos notar que necessitamos da variável $\hat{\mathbf{w}}_{n,conv}$. Ou seja, precisamos de informações do lugar geométrico que limita a região de convergência. Os pontos do lugar geométrico são dados pelas expressões (4.39) e (4.40).

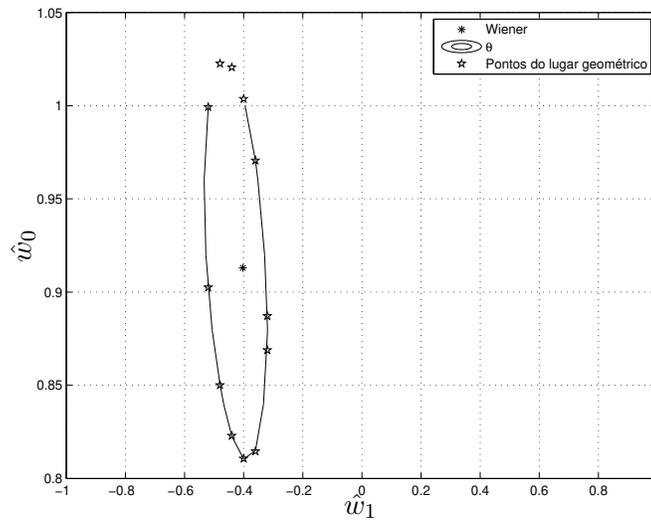
É interessante notar que o cálculo do *misadjustment* disponível na literatura não considera a região de convergência. Contudo, o *misadjustment* serve como métrica para saber a eficiência do algoritmo em estado de regime. Na nossa abordagem (equação (4.44)), essa situação ainda perdura mas com o diferencial que utilizamos as informações referentes ao lugar geométrico para o cálculo de M . Essa diferença é algo relevante, pois dá uma nova perspectiva para o conceito de *misadjustment* além de tratar informações adicionais como a localização onde o vetor de pesos ficará confinado.

Infelizmente, essa análise não tem relação com velocidade de convergência. Portanto, não conseguimos estabelecer qual seria a região de convergência após certo número de iterações.

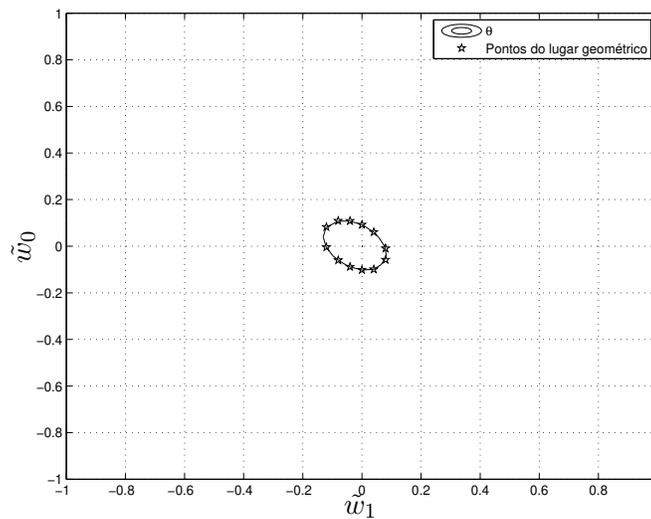
As figuras 4.12, 4.13 e 4.14 apresentam casos de simulação para mostrar a relação entre a região de convergência, o passo de adaptação μ e a função limitante superior para o passo de adaptação θ obtida por intermédio da análise de Lyapunov. Vale observar que as curvas de nível foram obtidas através de cortes feitos na superfície de θ bem como calculando os pontos do lugar geométrico através da expressão (4.38). Para obter figuras mais confiáveis e com uma representação que ficasse mais próxima do caso real, foi considerada no cálculo dos parâmetros da equação (4.40) a média, dado que, em um canal com 2 coeficientes excitado com sinal bipolar, temos oito possibilidades de sinal de saída e, em um canal com 3 coeficientes, o número de estados possíveis é 16.



(a)

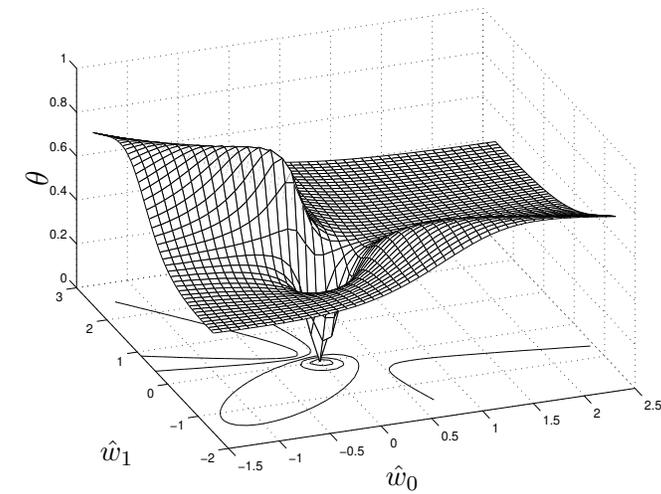


(b)

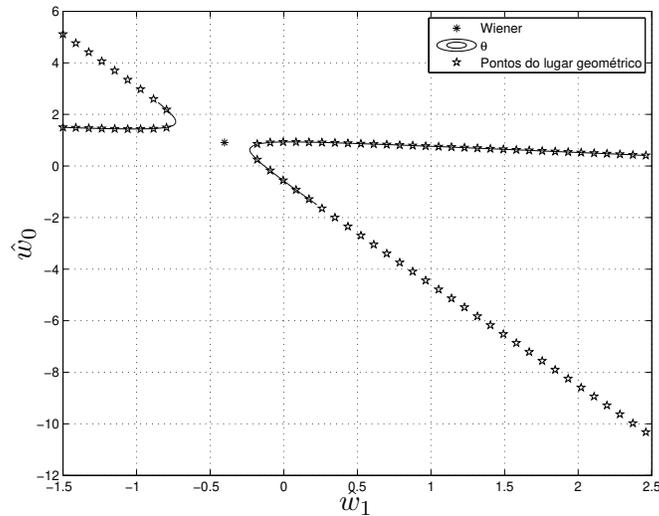


(c)

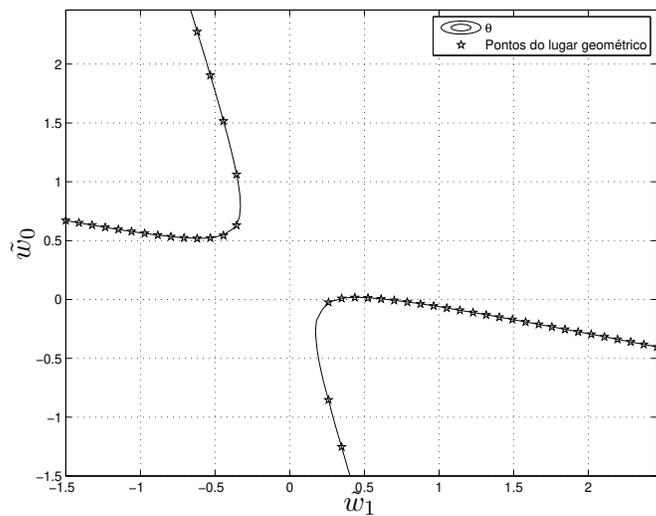
Fig. 4.12: Canal $h(z) = 1 + 0,6z^{-1}$ com atraso $d = 0$ e $\mu = 0,15$. (a) Superfície θ e respectivas curvas de nível. (b) Curvas de nível e lugar geométrico no sistema de coordenadas $\hat{w}_0 \times \hat{w}_1$. (c) Curvas de nível e lugar geométrico no sistema de coordenadas $\tilde{w}_0 \times \tilde{w}_1$.



(a)

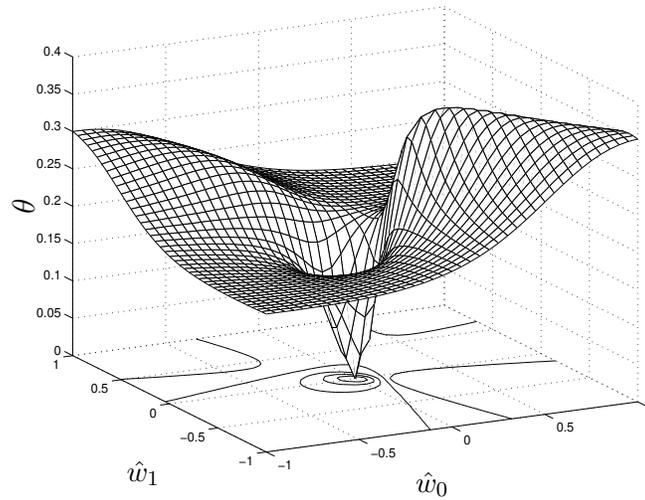


(b)

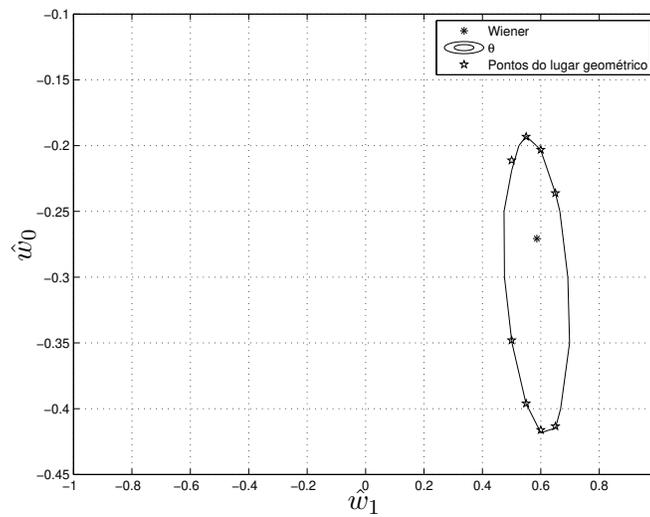


(c)

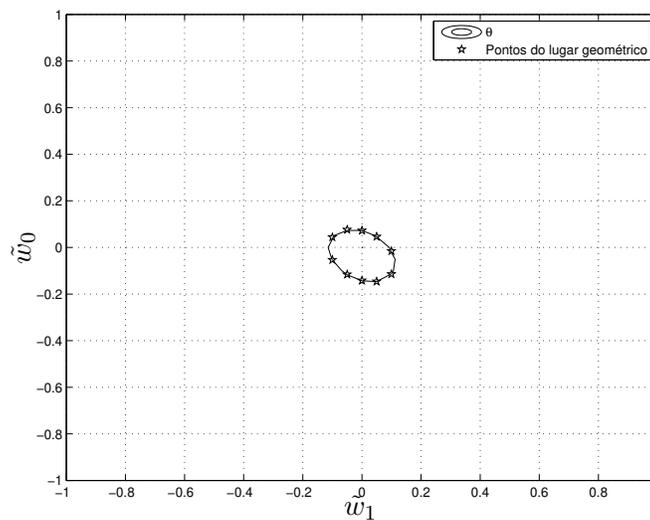
Fig. 4.13: Canal $h(z) = 1 + 0,6z^{-1}$ com atraso $d = 0$, $\mu = 0,6$. (a) Superfície θ e respectivas curvas de nível. (b) Curvas de nível e lugar geométrico no sistema de coordenadas $\hat{w}_0 \times \hat{w}_1$. (c) Curvas de nível e lugar geométrico no sistema de coordenadas $\tilde{w}_0 \times \tilde{w}_1$.



(a)

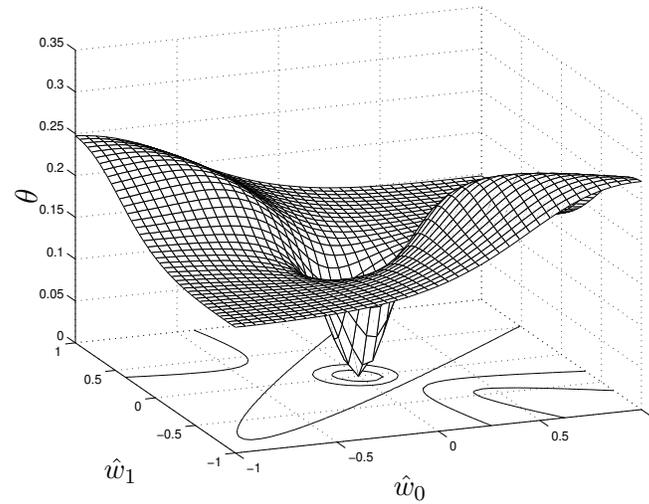


(b)

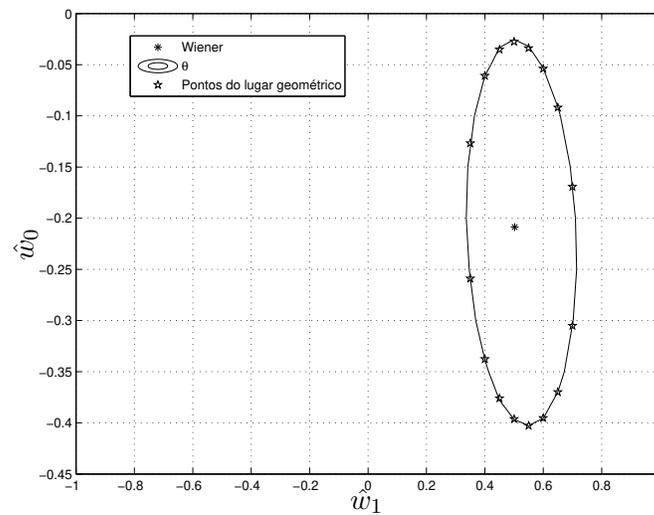


(c)

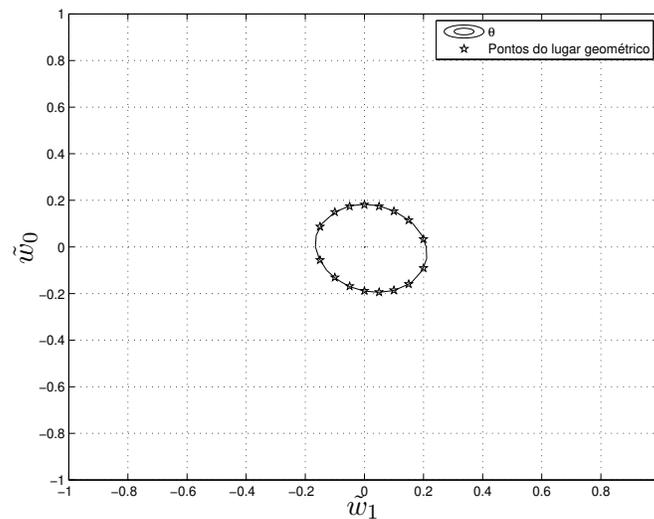
Fig. 4.14: Canal $h(z) = 1 + 1,5z^{-1}$ com atraso $d = 2$, $\mu = 0,1$. (a) Superfície θ e respectivas curvas de nível. (b) Curvas de nível e lugar geométrico no sistema de coordenadas $\hat{w}_0 \times \hat{w}_1$. (c) Curvas de nível e lugar geométrico no sistema de coordenadas $\tilde{w}_0 \times \tilde{w}_1$.



(a)



(b)



(c)

Fig. 4.15: Canal $h(z) = 1 + 0,6z^{-1} + 1,5z^{-2}$ com atraso $d = 3$, $\mu = 0,1$. (a) Superfície θ e respectivas curvas de nível. (b) Curvas de nível e lugar geométrico no sistema de coordenadas $\hat{w}_0 \times \hat{w}_1$. (c) Curvas de nível e lugar geométrico no sistema de coordenadas $\tilde{w}_0 \times \tilde{w}_1$.

A diferença entre os gráficos da figura 4.12 e 4.13 é o valor do passo de adaptação, o que mostra claramente a relação entre passo e lugar geométrico de convergência. Para passos de adaptação pequenos, o lugar geométrico tem o formato de elipse, o que mostra que, se a curva de nível em questão tem um formato elíptico, então a superfície apresenta uma formação convexa, o que leva a situação de mínimo local único.

Nos casos onde o valor do passo de adaptação é mais elevado (vide figura 4.13), a forma da região de convergência pode ser hiperbólica ou parabólica. No entanto, a região de convergência acompanha esse formato. Isso significa que para esse valor de passo de adaptação μ não oferece uma região de convergência fechada como a elipse. Conseqüentemente, a tendência é que o algoritmo não leve os elementos do vetor de pesos a ficarem confinados em uma região fechada. Isso é uma situação que leva a instabilidade. Vide como exemplo um caso de simulação da figura 4.16 para um passo de adaptação grande o suficiente para forma uma região de convergência com formato hiperbólico.

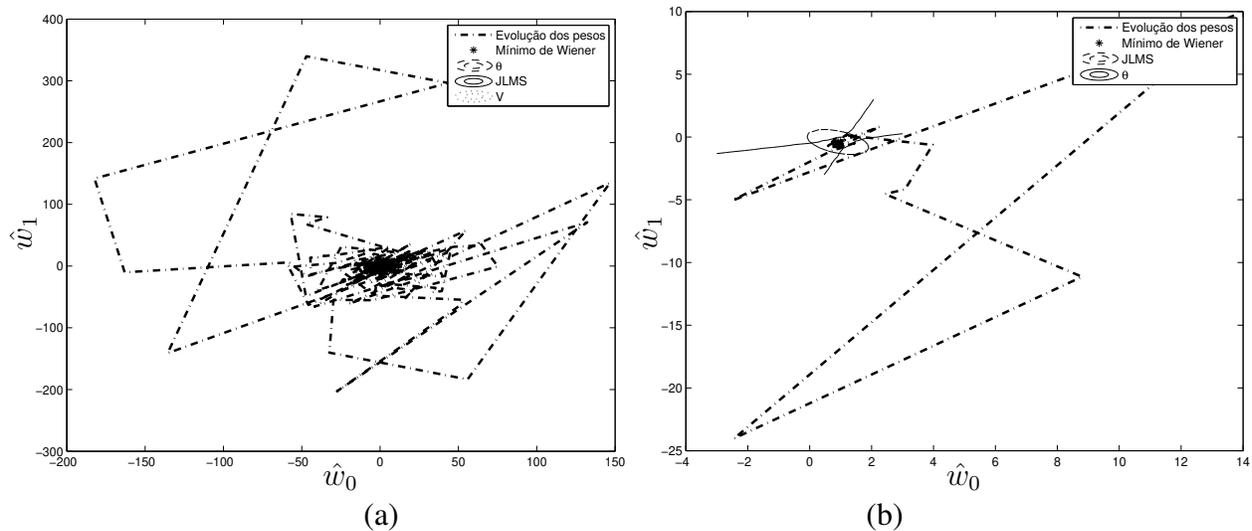


Fig. 4.16: Canal $h(z) = 1 + 0,6z^{-1}$ com atraso $d = 0$, $\mu = 1, 2$. (a) Evolução dos pesos do algoritmo LMS. (b) Últimos 50 passos da evolução de pesos.

As figuras 4.14 e 4.15 apresentam as mesmas situações de contorno da região de convergência, mas para um canal de fase não-mínima e um canal com três coeficientes. Em ambos os casos, a região de convergência tem um formato elíptico pois foi escolhido um passo de adaptação suficientemente pequeno a fim de que a análise ficasse próxima ao ponto de equilíbrio. Porém é possível notar nas figuras 4.14-a e 4.15-a que, para passos de adaptação maiores o formato da região de convergência pode ser parabólico ou hiperbólico.

Todas as figuras apresentaram um versão da região de convergência no sistema de coordenadas $\tilde{w}_0 \times \tilde{w}_1$ e $\hat{w}_0 \times \hat{w}_1$. Isso é relevante para apontar que o mínimo de Wiener é localizado no região central do pico invertido que temos nas figuras 4.12-a a 4.13-a nas proximidades do ponto de equilíbrio. Veja

que o ponto de Wiener não está necessariamente no centro dos elipsóides que são formados a partir da superfície θ . Contudo, o mais importante é notar que o mínimo de Wiener está dentro da região de convergência. Tal observação não é nova [28], mas, no escopo deste trabalho, ela é corroborada com a ajuda da teoria de Lyapunov e da análise da região de convergência.

A análise de convergência promovida nesta seção pode ser aplicada a outros tipos de algoritmo, inclusive algoritmos não-supervisionados. Isso é possível pois a análise toma em consideração a expressão $\Delta V = 0$. No caso de algoritmos estocásticos teremos, como dificuldade adicional, de calcular e considerar o valor das esperanças que possivelmente existirão na função de Lyapunov. Vide, como exemplo, a função de Lyapunov considerando a formulação estocástica para o algoritmo LMS disponível na equação (4.23).

4.5 Sumário

Uma das características mais marcantes do ferramental de Lyapunov é que a formulação é praticamente a mesma para todos os algoritmos ou técnicas. Isso estabelece um contraste, no caso supervisionado, com a abordagem clássica, a qual geralmente emprega aproximações ou hipóteses ad hoc como a teoria da independência. Neste capítulo, a aplicação da teoria de Lyapunov na análise de algoritmos se mostrou uma ferramenta importante tanto em termos de definição da região de convergência quanto da obtenção de um limitante superior para o passo de adaptação com potencial de aplicação prática. Os resultados apresentados no capítulo podem ainda ser estendidos para englobar estabilidade global se for considerado que o espaço de estados e a função de Lyapunov são radialmente ilimitados.

Em termos práticos, podemos resumir os resultados desse capítulo da seguinte forma:

- A estabilidade do algoritmo não é garantida para $\mu > \theta$.
- No início da operação do algoritmo, deve ser considerado que o valor de $\mu < \theta$ é um parâmetro que depende do ponto de inicialização. Ou seja, nas próximas iterações ou passos, o valor de μ pode não mais ser necessariamente o mais adequado. Isso significa que, em uma nova região definida pelo vetor dos pesos \hat{w}_n , o algoritmo pode necessitar de ajustes (mais especificamente, de redução) do passo de adaptação para facilitar a convergência.
- Cada canal apresenta sua própria faixa de valores de μ para garantir a estabilidade.
- O conceito de estabilidade no sentido de Lyapunov não significa convergência, o que significa que, a rigor, a análise de convergência deve ser considerada como um problema à parte.

Contudo, por intermédio da técnica de Lyapunov, não deixa de ser possível levantar algumas conclusões sobre a convergência, conforme indicado na análise adicional exposta na seção 4.4.

Tendo realizado um estudo representativo no campo de algoritmo supervisionados, trataremos, em seguida, da análise de Lyapunov no contexto de um representante importante da classe de técnicas de equalização não-supervisionadas: o algoritmo CMA.

Capítulo 5

Análise da Estabilidade de Algoritmos Não-Supervisionados Baseada no Segundo Método de Lyapunov

Neste capítulo, aplicaremos o segundo método de Lyapunov ao mais estudado dos algoritmos da classe de Bussgang, o CMA. Primeiramente, trataremos de um caso simples, caracterizado por um equalizador com um único coeficiente, para, em seguida, abordarmos o caso multidimensional.

5.1 Análise do algoritmo CMA de um coeficiente

A análise do CMA é significativamente mais complexa que a do LMS devido à presença de termos de ordem superior na função de módulo constante, o que leva, em geral, a um sistema de equações não-lineares na condição de gradiente nulo e a um comportamento dinâmico que transcende o usual binômio divergência / convergência [69].

Esta maior complexidade influencia diretamente a elaboração de candidatas a função de Lyapunov. Mesmo em uma análise à maneira do primeiro método de Lyapunov, as características do CMA são suficientes para dificultar o desenvolvimento e o entendimento.

A expressão adaptativa para o algoritmo CMA de um coeficiente no contexto de um canal trivial, correspondente a um ganho, é dada por:

$$\hat{w}_{n+1} = \hat{w}_n + k^2 \mu \hat{w}_n (1 - k^2 \hat{w}_n^2) A^4. \quad (5.1)$$

A expressão (5.1) é a equação (2.37) reescrita considerando o canal trivial com ganho k , sinal binário

com magnitude A e filtro com um coeficiente.

Os pontos de equilíbrio desta expressão podem ser encontrados considerando a condição $\hat{w}_{n+1} = \hat{w}_n$, o que significa que a equação (5.1) pode ser reescrita como

$$k^2 \mu \hat{w}_n (1 - k^2 \hat{w}_n^2) A^4 = 0. \quad (5.2)$$

Consequentemente, as soluções possíveis para este caso são dadas por

$$\begin{aligned} \overline{\hat{w}_{n,0}} &= 0, \\ \overline{\hat{w}_{n,1}} &= -\frac{1}{k}, \\ \overline{\hat{w}_{n,2}} &= \frac{1}{k}. \end{aligned} \quad (5.3)$$

Considerando $\hat{w}_{n+1} = F(\hat{w}_n)$, a condição de estabilidade em torno do ponto de equilíbrio - perceba que se trata de uma abordagem de primeiro método de Lyapunov - é dada por $|\frac{dF(\hat{w}_n)}{d\hat{w}_n}| < 1$ [69]. A derivada vale

$$\frac{dF(\hat{w}_n)}{d\hat{w}_n} = 1 + k^2 A^4 - 3\mu k^4 A^4 \hat{w}_n^2, \quad (5.4)$$

$$\left. \frac{dF(\hat{w}_n)}{d\hat{w}_n} \right|_{\hat{w}_n = \overline{\hat{w}_{n,1}}} = \left. \frac{dF(\hat{w}_n)}{d\hat{w}_n} \right|_{\hat{w}_n = \overline{\hat{w}_{n,2}}} = 1 - 2\mu k^2 A^4. \quad (5.5)$$

Isto leva à seguinte faixa de passo de adaptação para a qual a estabilidade é garantida:

$$0 < \mu < \frac{1}{k^2 A^4}. \quad (5.6)$$

5.1.1 Formulação para análise de estabilidade

No cenário do CMA com um coeficiente, a função custo J_{CMA} pode ser representada como

$$J_{CMA}(\hat{w}_n) = (k^2 \hat{w}_n^2 A^2 - A^2)^2. \quad (5.7)$$

Utilizando uma candidata a função de Lyapunov baseada na expressão de custo [18], temos:

$$V_{CMA} = J_{CMA}(\hat{w}_n) - J_{CMA}(\hat{w}_{n,min}). \quad (5.8)$$

Portanto, temos

$$\begin{aligned} \Delta V_{CMA} &= J_{CMA}(\hat{w}_{n+1}) - J_{CMA}(\hat{w}_{n,min}) - (J_{CMA}(\hat{w}_n) - J_{CMA}(\hat{w}_{n,min})) \\ &= \left(k^2 A^2 (k^2 \mu \hat{w}_n (1 - k^2 \hat{w}_n^2) A^4 + \hat{w}_n)^2 - A^2 \right)^2 - (k^2 \hat{w}_n^2 A^2 - A^2)^2. \end{aligned} \quad (5.9)$$

A expressão (5.9) pode ser reescrita em função do passo de adaptação μ

$$\begin{aligned} \Delta V_{CMA} &= \mu^4 (k^{20} \hat{w}_n^{12} - 4 k^{18} \hat{w}_n^{10} + 6 k^{16} \hat{w}_n^8 - 4 k^{14} \hat{w}_n^6 + k^{12} \hat{w}_n^4) A^{20} \\ &\quad + \mu^3 (-4 k^{16} \hat{w}_n^{10} + 12 k^{14} \hat{w}_n^8 - 12 k^{12} \hat{w}_n^6 + 4 k^{10} \hat{w}_n^4) A^{16} \\ &\quad + \mu^2 (6 k^{12} \hat{w}_n^8 - 14 k^{10} \hat{w}_n^6 + 10 k^8 \hat{w}_n^4 - 2 k^6 \hat{w}_n^2) A^{12} \\ &\quad + \mu (-4 k^8 \hat{w}_n^6 + 8 k^6 \hat{w}_n^4 - 4 k^4 \hat{w}_n^2) A^8, \end{aligned} \quad (5.10)$$

ou seja, ΔV_{CMA} é um polinômio de quarta ordem cuja as raízes são ¹

$$\begin{aligned} \mu_0 &= 0, \\ \mu_1 &= -\frac{\sqrt{2 - k^2 \hat{w}_n^2} - k \hat{w}_n}{(k^5 \hat{w}_n^3 - k^3 \hat{w}_n) A^4}, \\ \mu_2 &= \frac{\sqrt{2 - k^2 \hat{w}_n^2} + k \hat{w}_n}{(k^5 \hat{w}_n^3 - k^3 \hat{w}_n) A^4}, \\ \mu_3 &= \frac{2}{k^4 \hat{w}_n^2 A^4 - k^2 A^4}. \end{aligned} \quad (5.11)$$

A figura 5.1 apresenta um caso particular da expressão (5.10) para ilustrar o comportamento do sistema dinâmico.

Veja que ΔV_{CMA} é negativo entre a raiz μ_0 e a próxima raiz positiva μ_i (onde $i = 1, 2$ ou 3), que é a faixa para passo de adaptação que leva o algoritmo a ser estável no sentido de Lyapunov. Alterar

¹Como temos uma raiz em $\mu = 0$, teremos pelo menos uma raiz real e outras duas raízes que podem ser um par complexo ou duas raízes reais.

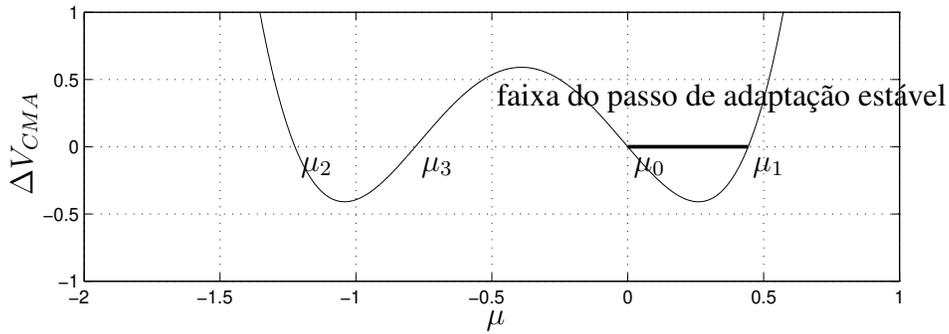


Fig. 5.1: Esboço da equação (5.10) ($k = 2, A = 1$ e $\hat{w}_n = 0, 3$).

os parâmetros (k, A e \hat{w}_n) significa que o comportamento e o intervalo onde $\Delta V_{CMA} < 0$ serão afetados. Consequentemente, a raiz μ_i que limitará superiormente a faixa do passo de adaptação para o algoritmo estável pode mudar dependendo do cenário do valor dos parâmetros envolvidos. Contudo, ΔV_{CMA} é uma função convexa para valores positivos de μ , o que garantirá a condição de Lyapunov ($\Delta V_{CMA} < 0$) para estabilidade.

Seguem outros exemplos variando os parâmetros do sistema dinâmico nas figuras 5.2, 5.3 e 5.4

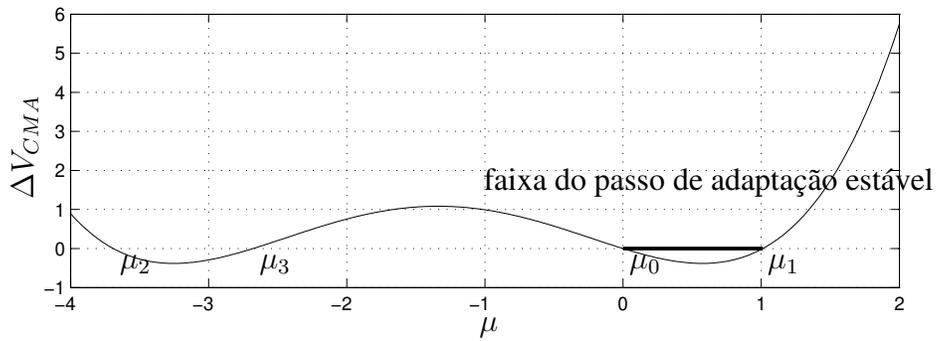


Fig. 5.2: Esboço da equação (5.10) ($k = 1, 0, A = 1, 1$ e $\hat{w}_n = 0, 7$).

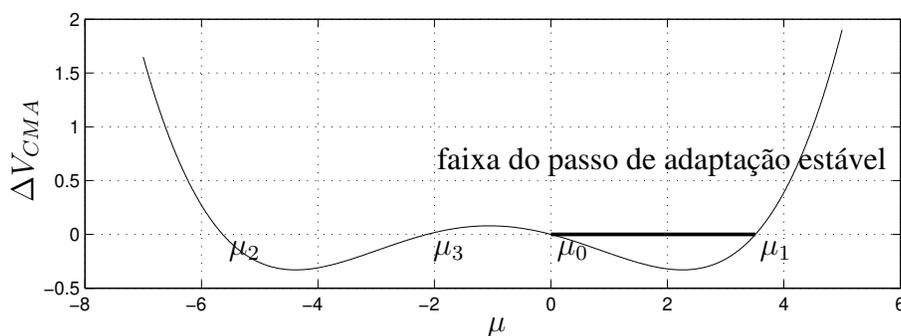


Fig. 5.3: Esboço da equação (5.10) ($k = 1, 6$, $A = 0, 8$ e $\hat{w}_n = 0, 2$).

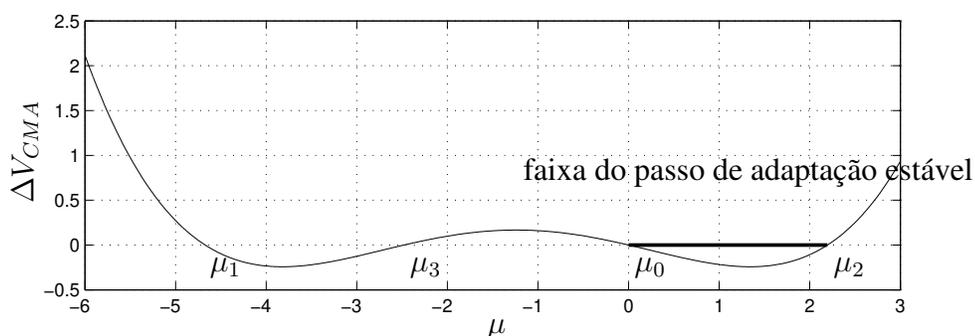


Fig. 5.4: Esboço da equação (5.10) ($k = 1, 6$, $A = 0, 8$ e $\hat{w}_n = -0, 3$).

Na figura 5.4, o intervalo dos valores de passo de adaptação para o algoritmo CMA estável está entre μ_0 e μ_2 , o que difere das figuras 5.1, 5.2 e 5.3, que têm o intervalo $\mu_0 - \mu_1$. As figuras apresentam uma visão das condições de estabilidade do algoritmo no sentido de Lyapunov. Até esse momento, nenhuma menção foi feita em relação ao ponto de equilíbrio ou mesmo ao valor máximo do passo de adaptação.

Considerando as possíveis soluções para o CMA de um coeficiente (vide equação (5.3)), é necessário analisar o comportamento do sistema em torno dos pontos de equilíbrio ($\overline{\hat{w}_{n,1}} = -\frac{1}{k}$ e $\overline{\hat{w}_{n,2}} = \frac{1}{k}$) para determinar o valor máximo passo de adaptação.

Ponto de equilíbrio $\overline{\hat{w}_{n,1}} = -\frac{1}{k}$

Para este caso, os limites para as raízes apresentadas na equação (5.11) podem ser escritos conforme abaixo.

$$\begin{aligned}
 \lim_{\hat{w}_n \rightarrow -\frac{1}{k}} \mu_1 &= \lim_{\hat{w}_n \rightarrow -\frac{1}{k}} -\frac{\sqrt{2 - k^2 \hat{w}_n^2} - k \hat{w}_n}{(k^5 \hat{w}_n^3 - k^3 \hat{w}_n) A^4} \rightarrow -\infty, \\
 \lim_{\hat{w}_n \rightarrow -\frac{1}{k}} \mu_2 &= \lim_{\hat{w}_n \rightarrow -\frac{1}{k}} \frac{\sqrt{2 - k^2 \hat{w}_n^2} + k \hat{w}_n}{(k^5 \hat{w}_n^3 - k^3 \hat{w}_n) A^4} = \frac{0}{0}, \\
 \lim_{\hat{w}_n \rightarrow -\frac{1}{k}} \mu_3 &= \lim_{\hat{w}_n \rightarrow -\frac{1}{k}} \frac{2}{k^4 \hat{w}_n^2 A^4 - k^2 A^4} \rightarrow \infty.
 \end{aligned} \tag{5.12}$$

É importante salientar que apenas uma das raízes (μ_2) tem o limite tal que permita uma continuação imediata da análise. Informações adicionais sobre a raiz μ_2 podem ser obtidas utilizando a regra de L'Hospital

$$\begin{aligned}
 \lim_{\hat{w}_n \rightarrow -\frac{1}{k}} \mu_2 &= \lim_{\hat{w}_n \rightarrow -\frac{1}{k}} \frac{\sqrt{2 - k^2 \hat{w}_n^2} + k \hat{w}_n}{(k^5 \hat{w}_n^3 - k^3 \hat{w}_n) A^4}, \\
 &\stackrel{\text{L'Hospital}}{=} \frac{\lim_{\hat{w}_n \rightarrow -\frac{1}{k}} -k^2 (2 - k^2 \hat{w}_n^2)^{-1/2} \hat{w}_n + k}{\lim_{\hat{w}_n \rightarrow -\frac{1}{k}} (k^5 3 \hat{w}_n^2 - k^3) A^4} \\
 &= \frac{2k}{2k^3 A^4}, \\
 &= \frac{1}{k^2 A^4}.
 \end{aligned} \tag{5.13}$$

Concluindo, em torno do ponto de equilíbrio $\overline{\hat{w}_{n,1}} = -\frac{1}{k}$ o valor máximo do passo de adaptação para um sistema dinâmico estável é dado pela raiz μ_2 . O valor desse limitante do passo de adaptação é dado por $\frac{1}{k^2 A^4}$.

Ponto de equilíbrio $\overline{\hat{w}_{n,2}} = \frac{1}{k}$

Repetindo a análise para o ponto de equilíbrio $\overline{\hat{w}_{n,2}}$, vemos que as raízes da equação (5.11) apresentam os limites conforme abaixo

$$\begin{aligned}
\lim_{\hat{w}_n \rightarrow \frac{1}{k}} \mu_1 &= \lim_{\hat{w}_n \rightarrow \frac{1}{k}} - \frac{\sqrt{2 - k^2 \hat{w}_n^2} - k \hat{w}_n}{(k^5 \hat{w}_n^3 - k^3 \hat{w}_n) A^4} = \frac{0}{0}, \\
\lim_{\hat{w}_n \rightarrow \frac{1}{k}} \mu_2 &= \lim_{\hat{w}_n \rightarrow \frac{1}{k}} \frac{\sqrt{2 - k^2 \hat{w}_n^2} + k \hat{w}_n}{(k^5 \hat{w}_n^3 - k^3 \hat{w}_n) A^4} = \frac{2}{0} \rightarrow \infty, \\
\lim_{\hat{w}_n \rightarrow \frac{1}{k}} \mu_3 &= \lim_{\hat{w}_n \rightarrow \frac{1}{k}} \frac{2}{k^4 \hat{w}_n^2 A^4 - k^2 A^4} \rightarrow \infty.
\end{aligned} \tag{5.14}$$

Para μ_1 , aplicaremos a regra de L'Hospital novamente:

$$\begin{aligned}
\lim_{\hat{w}_n \rightarrow \frac{1}{k}} \mu_1 &= \lim_{\hat{w}_n \rightarrow \frac{1}{k}} - \frac{\sqrt{2 - k^2 \hat{w}_n^2} - k \hat{w}_n}{(k^5 \hat{w}_n^3 - k^3 \hat{w}_n) A^4} \\
&\stackrel{\text{L'Hospital}}{=} - \frac{\lim_{\hat{w}_n \rightarrow \frac{1}{k}} -k^2 (2 - k^2 \hat{w}_n^2)^{-1/2} \hat{w}_n - k}{\lim_{\hat{w}_n \rightarrow \frac{1}{k}} (k^5 3 \hat{w}_n^2 - k^3) A^4} \\
&= - \frac{-2k}{2k^3 A^4}, \\
&= \frac{1}{k^2 A^4}.
\end{aligned} \tag{5.15}$$

O resultado $\mu_{max} = \frac{1}{k^2 A^4}$ (que é o mesmo para ambos pontos de equilíbrio $\overline{\hat{w}_{n,1}}$ e $\overline{\hat{w}_{n,2}}$) é exatamente o passo de adaptação obtido em [69]. Contudo, perceba que, neste trabalho, isto foi apenas um caso particular de uma abordagem mais genérica baseada no segundo método de Lyapunov, o que significa uma nova perspectiva que, no entanto, é coerente com resultados obtidos segundo vias mais tradicionais, como os de [69].

Análise das raízes da equação (5.11)

Na seção anterior, tivemos a oportunidade de observar que o máximo passo de adaptação para o algoritmo CMA de um coeficiente é determinado por μ_1 ou μ_2 . Isto dependerá de qual ponto de equilíbrio será considerado para análise.

De acordo com a equação (5.12), para o ponto de equilíbrio $\overline{\hat{w}_{n,1}} = -\frac{1}{k}$, o máximo passo de adaptação será determinado por μ_2 . Por outro lado, quando o ponto de equilíbrio é $\overline{\hat{w}_{n,2}} = \frac{1}{k}$, o máximo passo de adaptação que leva à estabilidade será limitado por μ_1 .

É importante dizer que apenas uma das raízes μ_i , ($i = 1, 2, 3$) será o limite do passo de adaptação para um dado valor do vetor de pesos \hat{w}_n . As outras possibilidades levarão o sistema a divergir e,

consequentemente, não são relevantes para a determinação do limite de passo.

Isto mostra que existirá um passo de adaptação válido, finito e real em torno do ponto de equilíbrio independentemente de onde este se localize. É importante frisar que o segundo método de Lyapunov permite investigações analíticas sem restrições sobre a localização do ponto de equilíbrio, o que pode ser particularmente interessante para um algoritmo multimodal como o CMA.

Ilustramos na figura 5.5 um exemplo do que acabamos de descrever. Nesta simulação, foi considerado $k = 0,5$ e $A = 3$, o que leva aos pontos de equilíbrio $\hat{w}_n = 2$ e $\hat{w}_n = -2$. A representação gráfica requer três curvas (uma para cada valor de $\mu_i, (i = 1, 2, 3)$) como função de \hat{w}_n .

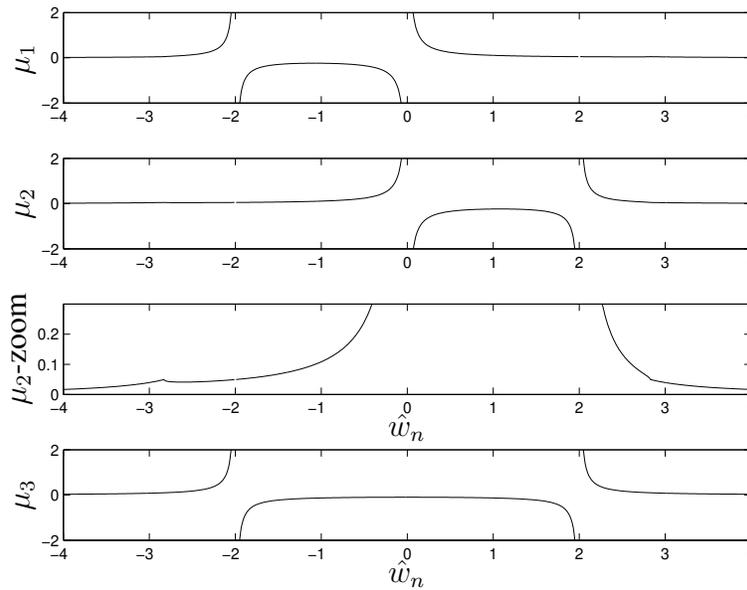


Fig. 5.5: Análise das raízes ($k = 0,5, A = 3 \implies \overline{\hat{w}_{n,2}} = -2$ e $\overline{\hat{w}_{n,2}} = 2$).

Para o ponto de equilíbrio $\hat{w}_n = -2$, μ_1 e μ_3 divergirão. Contudo, μ_2 convergirá para um valor finito e real. Consequentemente, μ_2 é um limite para passo de adaptação válido para pontos em torno de $\hat{w}_n = -2$. A raiz μ_0 é sempre zero.

Quando o ponto de equilíbrio é $\hat{w}_n = 2$, os passos de adaptação μ_2 and μ_3 terão valores infinitos. O limite do passo de adaptação válido para $\hat{w}_n = 2$ é determinado por μ_1 , que é um número real e finito.

A próxima figura descreve outro cenário como maneira de exemplificar que para sistema dinâmicos diferentes sempre existirá um limite de passo de adaptação para cada ponto de equilíbrio.

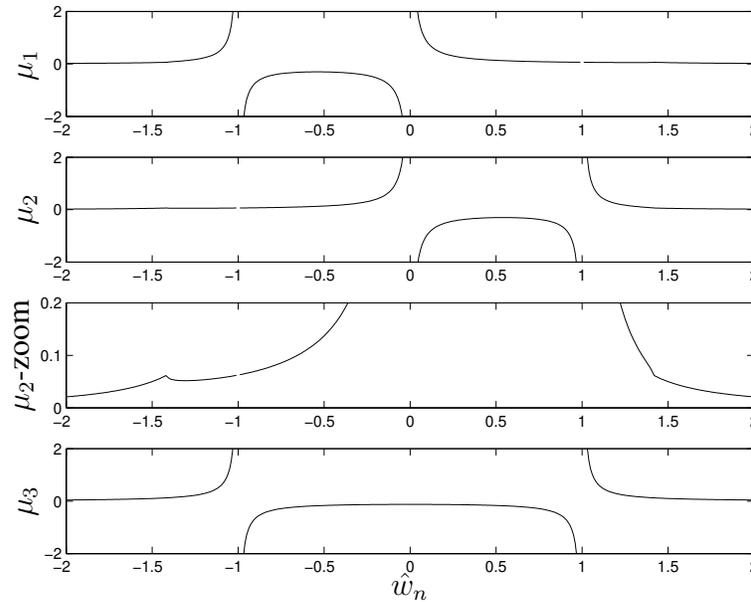


Fig. 5.6: Análise das raízes ($k = 1, A = 2 \implies \overline{\hat{w}_{n,2}} = -0,5$ e $\overline{\hat{w}_{n,2}} = 0,5$).

Na próxima seção, será analisado o algoritmo CMA com N coeficientes, ou seja, um caso mais genérico e com maior representatividade prática.

5.2 Análise do algoritmo CMA de N coeficientes

A versão estocástica do CMA para um equalizador de N coeficientes é dada por [28]

$$\hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n + \mu e_n \mathbf{x}_n, \quad (5.16)$$

onde $e_n = y_n(R_2 - |y_n|^2)$, a saída do equalizador é $y_n = \hat{\mathbf{w}}_n^T \mathbf{x}_n$ e $R_2 = \frac{E\{|\mathbf{s}_n|^4\}}{E\{|\mathbf{s}_n|^2\}}$.

A candidata a função de Lyapunov é baseada na função custo, conforme a proposta de [18], portanto

$$V_{CMA}(\hat{\mathbf{w}}_n) = J_{CMA}(\hat{\mathbf{w}}_n) - J_{CMA}(\hat{\mathbf{w}}_{n,min}). \quad (5.17)$$

Consequentemente,

$$\begin{aligned} \Delta V_{CMA} &= E\{J_{CMA}(\hat{\mathbf{w}}_{n+1})\} - J_{CMA}(\hat{\mathbf{w}}_n) \\ &= E\{|\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} + \mu y_n R_2 \mathbf{x}_n^T \mathbf{x}_{n+1} - \mu y_n^3 \mathbf{x}_n^T \mathbf{x}_{n+1}|^4 \\ &\quad - 2|\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} + \mu y_n R_2 \mathbf{x}_n^T \mathbf{x}_{n+1} - \mu y_n^3 \mathbf{x}_n^T \mathbf{x}_{n+1}|^2 R_2 - y_n^4 + 2y_n^2 R_2\}. \end{aligned} \quad (5.18)$$

Hipótese 7. Considerando que o sistema tem apenas variáveis reais, o que significa $|\cdot|^2 = (\cdot)^2$.

A equação (5.18) pode ser reescrita como

$$\Delta V_{CMA} = E\{[(\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} + \mu \mathbf{x}_n^T \mathbf{x}_{n+1} y_n (R_2 - y_n^2))^2 - R_2]^2\} - E\{(y_n^2 - R_2)^2\}. \quad (5.19)$$

Ou, tomando a versão expandida,

$$\begin{aligned} \Delta V_{CMA} &= E\{J_{CMA}(\hat{\mathbf{w}}_{n+1})\} - J_{CMA}(\hat{\mathbf{w}}_n) \\ &= E\{[(\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} + \mu \mathbf{x}_n^T \mathbf{x}_{n+1} y_n (R_2 - y_n^2))^2 - R_2]^2\} - E\{(y_n^2 - R_2)^2\} \\ &= \\ &\mu^4 E\{-4(\mathbf{x}_n^T \mathbf{x}_{n+1})^4 y_n^6 R_2^3 + (\mathbf{x}_n^T \mathbf{x}_{n+1})^4 y_n^{12} + (\mathbf{x}_n^T \mathbf{x}_{n+1})^4 y_n^4 R_2^4 \\ &\quad - 4(\mathbf{x}_n^T \mathbf{x}_{n+1})^4 y_n^{10} R_2 + 6(\mathbf{x}_n^T \mathbf{x}_{n+1})^4 y_n^8 R_2^2\} + \\ &\mu^3 E\{-12\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} (\mathbf{x}_n^T \mathbf{x}_{n+1})^3 y_n^5 R_2^2 - 4\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} (\mathbf{x}_n^T \mathbf{x}_{n+1})^3 y_n^9 \\ &\quad + 4\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} (\mathbf{x}_n^T \mathbf{x}_{n+1})^3 y_n^3 R_2^3 + 12\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} (\mathbf{x}_n^T \mathbf{x}_{n+1})^3 y_n^7 R_2\} + \\ &\mu^2 E\{-2(\mathbf{x}_n^T \mathbf{x}_{n+1})^2 y_n^6 R_2 + 6(\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1})^2 (\mathbf{x}_n^T \mathbf{x}_{n+1})^2 y_n^2 R_2^2 \\ &\quad - 12(\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1})^2 (\mathbf{x}_n^T \mathbf{x}_{n+1})^2 y_n^4 R_2 + 6(\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1})^2 (\mathbf{x}_n^T \mathbf{x}_{n+1})^2 y_n^6 \\ &\quad - 2(\mathbf{x}_n^T \mathbf{x}_{n+1})^2 y_n^2 R_2^3 + 4(\mathbf{x}_n^T \mathbf{x}_{n+1})^2 y_n^4 R_2^2\} + \\ &\mu E\{-4(\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1})^3 (\mathbf{x}_n^T \mathbf{x}_{n+1}) y_n^3 + 4\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} (\mathbf{x}_n^T \mathbf{x}_{n+1}) y_n^3 R_2 \\ &\quad + 4(\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1})^3 (\mathbf{x}_n^T \mathbf{x}_{n+1}) y_n R_2 - 4\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} (\mathbf{x}_n^T \mathbf{x}_{n+1}) y_n R_2^2\}. \end{aligned} \quad (5.20)$$

A expressão (5.20) é um polinômio de quarta ordem se consideramos que a variável é μ .

De acordo com o teorema 5, a estabilidade será garantida para μ quando $A\mu^4 + B\mu^3 + C\mu^2 + D\mu < 0$, ou seja, para os valores de μ que $\Delta V_{CMA} < 0$.

Hipótese 8. O algoritmo CMA não tem nenhuma representação analítica que seja simples para os pontos de mínimo. Neste sentido, para avaliar a estabilidade, o cenário de inversão perfeita é considerado. Isso significa que o vetor de pesos será equivalente à solução de Wiener, a qual, em

termos da condição zero-forcing, é dada por (constelação BPSK)

$$\mathbf{w}^* = (H^T H)^{-1} H \mathbf{Z}_F, \quad (5.21)$$

onde $\mathbf{Z}_F = \left[\cdots \ 1 \ 0 \ 0 \ \cdots \right]^T$ e H é a matriz de convolução do canal.

O elemento do vetor \mathbf{Z}_F que é igual a 1 é localizado na posição associada a fonte e respectivo atraso a ser recuperado².

Esta abordagem e hipótese é considerada em outros estudos, como [16], onde a solução zero-forcing é tida como a solução de inversão do canal.

Para facilitar a análise do algoritmo, vamos remover o operador esperança $E\{\cdot\}$ da equação (5.20). Consequentemente, estaremos promovendo uma avaliação instantânea. A equação (5.20) pode ser reescrita como

$$\begin{aligned} \Delta V_{CMA} = & \\ & \mu^4 \left[-4(\mathbf{x}_n^T \mathbf{x}_{n+1})^4 y_n^6 R_2^3 + (\mathbf{x}_n^T \mathbf{x}_{n+1})^4 y_n^{12} + (\mathbf{x}_n^T \mathbf{x}_{n+1})^4 y_n^4 R_2^4 \right. \\ & \quad \left. - 4(\mathbf{x}_n^T \mathbf{x}_{n+1})^4 y_n^{10} R_2 + 6(\mathbf{x}_n^T \mathbf{x}_{n+1})^4 y_n^8 R_2^2 \right] + \\ & \mu^3 \left[-12 \hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} (\mathbf{x}_n^T \mathbf{x}_{n+1})^3 y_n^5 R_2^2 - 4 \hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} (\mathbf{x}_n^T \mathbf{x}_{n+1})^3 y_n^9 \right. \\ & \quad \left. + 4 \hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} (\mathbf{x}_n^T \mathbf{x}_{n+1})^3 y_n^3 R_2^3 + 12 \hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} (\mathbf{x}_n^T \mathbf{x}_{n+1})^3 y_n^7 R_2 \right] + \\ & \mu^2 \left[-2(\mathbf{x}_n^T \mathbf{x}_{n+1})^2 y_n^6 R_2 + 6(\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1})^2 (\mathbf{x}_n^T \mathbf{x}_{n+1})^2 y_n^2 R_2^2 \right. \\ & \quad \left. - 12(\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1})^2 (\mathbf{x}_n^T \mathbf{x}_{n+1})^2 y_n^4 R_2 + 6(\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1})^2 (\mathbf{x}_n^T \mathbf{x}_{n+1})^2 y_n^6 \right. \\ & \quad \left. - 2(\mathbf{x}_n^T \mathbf{x}_{n+1})^2 y_n^2 R_2^3 + 4(\mathbf{x}_n^T \mathbf{x}_{n+1})^2 y_n^4 R_2^2 \right] + \\ & \mu \left[-4(\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1})^3 (\mathbf{x}_n^T \mathbf{x}_{n+1}) y_n^3 + 4 \hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} (\mathbf{x}_n^T \mathbf{x}_{n+1}) y_n^3 R_2 \right. \\ & \quad \left. + 4(\hat{\mathbf{w}}_n^T \mathbf{x}_{n+1})^3 (\mathbf{x}_n^T \mathbf{x}_{n+1}) y_n R_2 - 4 \hat{\mathbf{w}}_n^T \mathbf{x}_{n+1} (\mathbf{x}_n^T \mathbf{x}_{n+1}) y_n R_2^2 \right]. \end{aligned} \quad (5.22)$$

Nota 15. Dada as condições do sinal transmitido s_n especificadas no capítulo 2, $R_2 = 1$.

²Por exemplo, para atraso zero, $\mathbf{Z}_F = [1 \ 0 \ 0 \ 0 \ \cdots]^T$, para atraso 1, $\mathbf{Z}_F = [0 \ 1 \ 0 \ 0 \ \cdots]^T$.

A expressão (5.22) tem as raízes

$$\begin{aligned}
 \mu_0 &= 0, \\
 \mu_1 &= -\frac{\sqrt{2 - \hat{\mathbf{w}}_n^T \mathbf{x}_{n-1} \mathbf{x}_{n-1}^T \hat{\mathbf{w}}_n + \hat{\mathbf{w}}_n^T \mathbf{x}_{n-1}}}{\hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1} - \hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1} \hat{\mathbf{w}}_n^T \hat{\mathbf{w}}_n}, \\
 \mu_2 &= \frac{\sqrt{2 - \hat{\mathbf{w}}_n^T \mathbf{x}_{n-1} \mathbf{x}_{n-1}^T \hat{\mathbf{w}}_n - \hat{\mathbf{w}}_n^T \mathbf{x}_{n-1}}}{\hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1} - \hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1} \hat{\mathbf{w}}_n^T \hat{\mathbf{w}}_n}, \\
 \mu_3 &= \frac{2}{\hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_n \mathbf{x}_n^T \hat{\mathbf{w}}_n - \mathbf{x}_n^T \mathbf{x}_n}.
 \end{aligned} \tag{5.23}$$

O passo de adaptação para o equalizador CMA com N coeficientes pode ser estudado na mesma linha de raciocínio que aquela descrita na seção 5.1.

Hipótese 9. *O sinal recebido, ou seja, a entrada do equalizador \mathbf{x}_n , pode ser escrito como Hs_n , sendo H a matriz de convolução. Isto implica que o ruído aditivo b_n é desprezível ou mesmo considerado.*

Sob a hipótese de inversão perfeita, podemos assumir que

$$\lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}} \hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \hat{\mathbf{w}}_n = 1. \tag{5.24}$$

Veja que $\hat{\mathbf{w}}_n^T \mathbf{x}_{n-1}$ pode ser 1 ou -1 para a solução *zero-forcing*. Isso nos dá duas possibilidades para análise que serão discutidas em detalhes a seguir.

Caso A - análise do CMA de N coeficientes quando $\hat{\mathbf{w}}_n^T \mathbf{x}_{n-1} = 1$

Portanto, considerando $\hat{\mathbf{w}}_n^T \mathbf{x}_{n-1} = 1$, podemos calcular o limite para as raízes da equação (5.23)

$$\begin{aligned}
 \lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}} \mu_1 &= \lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}^*} -\frac{\sqrt{2 - \hat{\mathbf{w}}_n^T \mathbf{x}_{n-1} \mathbf{x}_{n-1}^T \hat{\mathbf{w}}_n + \hat{\mathbf{w}}_n^T \mathbf{x}_{n-1}}}{\hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1} - \hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1} \hat{\mathbf{w}}_n^T \hat{\mathbf{w}}_n} \rightarrow \infty, \\
 \lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}} \mu_2 &= \lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}^*} \frac{1 - \hat{\mathbf{w}}_n^T \mathbf{x}_{n-1}}{\hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1} (1 - \hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \hat{\mathbf{w}}_n)} = \frac{0}{0}, \\
 \lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}} \mu_3 &= \lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}^*} \frac{2}{\hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_n \mathbf{x}_n^T \hat{\mathbf{w}}_n - \mathbf{x}_n^T \mathbf{x}_n} \rightarrow \infty.
 \end{aligned} \tag{5.25}$$

Seja a variável auxiliar $g_n = \hat{\mathbf{w}}_n^T \mathbf{x}_n$, portanto:

$$\lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}} \mu_2 = \lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}} \frac{1 - \hat{\mathbf{w}}_n^T \mathbf{x}_{n-1}}{g_n \mathbf{x}_n^T \mathbf{x}_{n-1} (1 - g_n^2)}. \quad (5.26)$$

multiplicando o numerador e o denominador por $\hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1}$

$$\lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}} \mu_2 = \lim_{g_n \rightarrow 1} \frac{g_n \mathbf{x}_n^T \mathbf{x}_{n-1} - g_n^2 \mathbf{x}_{n-1} \mathbf{x}_{n-1}}{g_n^2 \mathbf{x}_{n-1}^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1} (1 - g_n^2)} = \lim_{g_n \rightarrow 1} \frac{\mathbf{x}_n^T \mathbf{x}_{n-1} - g_n \mathbf{x}_{n-1} \mathbf{x}_{n-1}}{g_n \mathbf{x}_{n-1}^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1} (1 - g_n^2)}. \quad (5.27)$$

Aplicando a regra de L'Hospital para μ_2 , mais especificamente na variável g_n

$$\lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}} \mu_2 = \lim_{g_n \rightarrow 1} \mu_2 = \frac{-\mathbf{x}_{n-1}^T \mathbf{x}_{n-1}}{-2\mathbf{x}_{n-1}^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1}} = \frac{\mathbf{x}_{n-1}^T \mathbf{x}_{n-1}}{2\mathbf{x}_{n-1}^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1}}. \quad (5.28)$$

Caso B - análise do CMA de N coeficientes quando $\hat{\mathbf{w}}_n^T \mathbf{x}_{n-1} = -1$

Se considerarmos $\hat{\mathbf{w}}_n^T \mathbf{x}_{n-1} = -1$ e repetirmos o cálculo do limite conforme a equação (5.25), temos

$$\begin{aligned} \lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}} \mu_1 &= \lim_{\hat{\mathbf{w}}_n \rightarrow \mathbf{w}^*} - \frac{1 + \hat{\mathbf{w}}_n^T \mathbf{x}_{n-1}}{\hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1} - \hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1} \hat{\mathbf{w}}_n^T \hat{\mathbf{w}}_n} = \frac{0}{0}, \\ \lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}} \mu_2 &= \lim_{\hat{\mathbf{w}}_n \rightarrow \mathbf{w}^*} \frac{1 - \hat{\mathbf{w}}_n^T \mathbf{x}_{n-1}}{\hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1} (1 - \hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \hat{\mathbf{w}}_n)} \rightarrow \infty, \\ \lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}} \mu_3 &= \lim_{\hat{\mathbf{w}}_n \rightarrow \mathbf{w}^*} \frac{2}{\hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_n \mathbf{x}_n^T \hat{\mathbf{w}}_n - \mathbf{x}_n^T \mathbf{x}_n} \rightarrow \infty. \end{aligned} \quad (5.29)$$

Podemos aplicar a regra de L'Hospital para μ_1 (mesma idéia empregada na equação (5.27))

$$\lim_{\hat{\mathbf{w}}_n \rightarrow \underline{\mathbf{w}}} \mu_1 = \frac{\mathbf{x}_{n-1}^T \mathbf{x}_{n-1}}{2\mathbf{x}_{n-1}^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1}}. \quad (5.30)$$

Apesar das mudanças de condições para esse caso, chegamos nos mesmo resultados com relação ao valor do limite ($\mu_{max} = \frac{\mathbf{x}_{n-1}^T \mathbf{x}_{n-1}}{2\mathbf{x}_{n-1}^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1}}$).

As equações (5.28) e (5.30) apresentam apenas uma realização da expressão do passo máximo de adaptação para o CMA. Para alcançar uma exatidão maior no cálculo do limite máximo para o passo de adaptação, propomos considerar várias realizações, se possível, um número elevado. Neste

sentido, o valor máximo do passo de adaptação para o equalizador CMA de N coeficientes é

$$\mu_{max} = \frac{E\{\mathbf{x}_{n-1}^T \mathbf{x}_{n-1}\}}{2E\{\mathbf{x}_{n-1}^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{x}_{n-1}\}}. \tag{5.31}$$

Simulações

Na figura 5.7, variamos o passo de adaptação μ e representamos a norma do vetor de pesos do equalizador $\hat{\mathbf{w}}_n$ para as últimas 50 passos antes do algoritmos se tornar instável. O objetivo deste esboço é obter informação de qual é o maior passo de adaptação existente mas que permita o algoritmo trabalhar em situação de estabilidade. Tal representação evoca os diagramas de bifurcação usados na análise de sistemas dinâmicos não-lineares. Essa representação pode ser encontrada em [69] para o CMA de um coeficiente.

Um histograma com valores máximos do passo de adaptação apresentados na figura 5.7 foi apresentado na figura 5.8. O valor do passo máximo de adaptação foi calculado considerando que, ao variar o passo de adaptação μ , o algoritmo tende a divergir para qualquer inicialização mesmo que próxima ao ponto de Wiener.

Da expressão (5.31), o valor teórico do passo de adaptação é 0,1051, o que mostra uma concordância razoável entre o resultado algébrico e simulações.

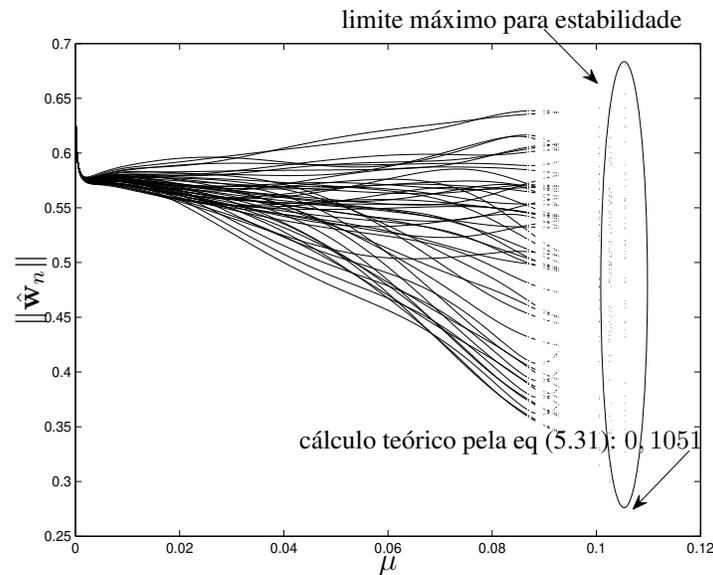


Fig. 5.7: Norma das últimas 50 amostras do vetor de pesos $\hat{\mathbf{w}}_n$ para o canal $h(z) = 1 + 1,5z^{-1}$.

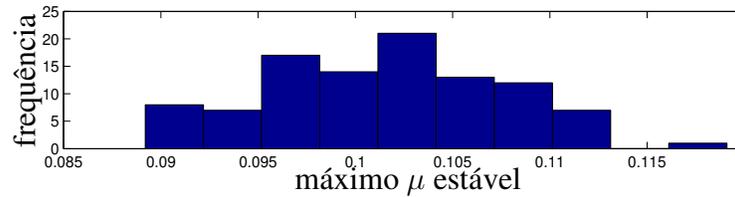


Fig. 5.8: Histograma para 100 realizações com o canal $h(z) = 1 + 1,5z^{-1}$.

Para o canal de fase mínima, o valor máximo do passo de adaptação foi obtido por intermédio de simulações e considerando o valor de μ para qual a norma $\|\hat{w}_n\|$ começa a divergir. Isto pode ser encontrado na figura 5.9.

O histograma com o resultado de 100 simulações independentes da figura 5.9 está esboçado na figura 5.10.

Neste caso, o valor do limitante teórico é 0,2578, que mostra consonância entre resultados práticos e o limite teórico.

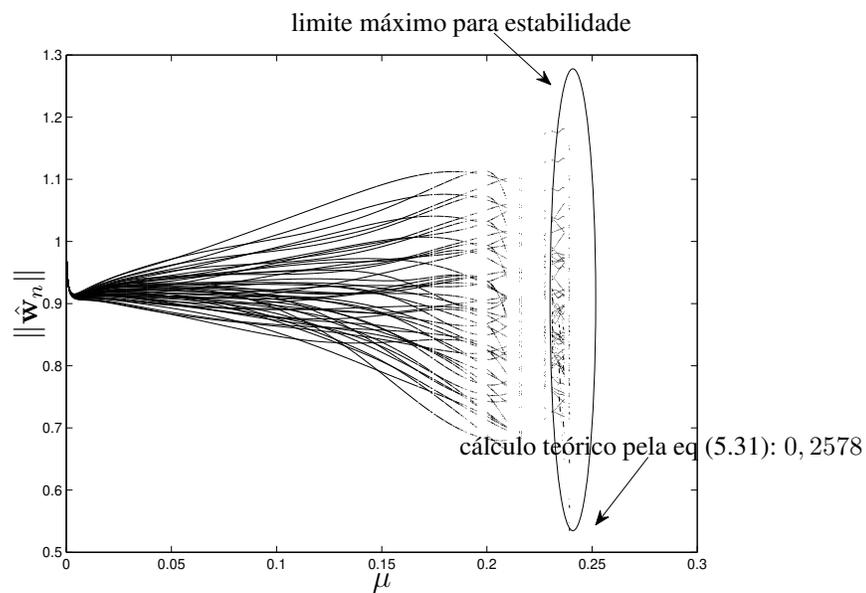


Fig. 5.9: Norma das últimas 50 amostras do vetor de pesos \hat{w}_n para o canal $h(z) = 1 + 0,6z^{-1}$.

A discussão e resultados desta seção revelam que o segundo método de Lyapunov é capaz de produzir resultados sólidos mesmo no caso de um algoritmo não-supervisionado com mais de um parâmetro em um contexto estocástico.

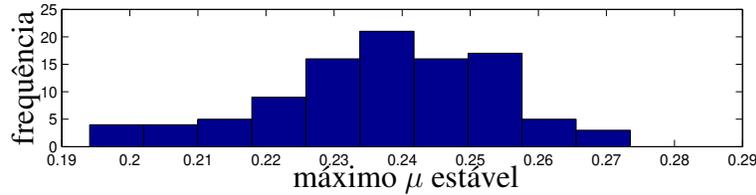


Fig. 5.10: Histograma para 100 realizações com o canal $h(z) = 1 + 0,6z^{-1}, .$

No caso de aplicações práticas sugerimos utilizar a expressão simplificada representada pelo equação (5.31). Contudo, devemos salientar que da mesma forma que na análise do algoritmo CMA com apenas um coeficiente, o equalizador para o caso multidimensional apresenta quatro raízes, as quais estão descritas na equação (5.23).

Observando as equações (5.28) e (5.30), notamos que para cada caso temos sempre uma raiz para determinar o limite do passo de adaptação considerando como requisito a estabilidade do algoritmo. Vale lembrar que, da mesma forma que o algoritmo de um coeficiente (vide equações (5.13) e (5.15)), o limitante para o passo de adaptação é determinado ora por uma das raízes e ora por outra.

É possível utilizar a equação (5.23) diretamente para delimitar o passo de adaptação. Contudo, obteremos três valores possíveis de limitante para o passo de adaptação. O limitante válido deve ser o menor valor positivo.

A próxima seção resumirá os resultados obtidos e discorrerá brevemente sobre a relevância da análise de Lyapunov.

5.3 Sumário

Neste capítulo, apresentamos a análise de duas versões do algoritmo CMA: com um coeficiente e com N coeficientes. Em ambos os casos, foi possível obter um limitante algébrico para o passo de adaptação no sentido de que não haja instabilidade.

Um aspecto interessante foi o fato de que os resultados para o algoritmo de um coeficiente, obtidos de acordo com o segundo método de Lyapunov, corroboraram os obtidos em [69], em que foi usada um abordagem conceitualmente similar à do primeiro método de Lyapunov. Essa similaridade foi um indício de coerência por parte da análise empreendida.

Os resultados para o caso multidimensional, por outro lado, trouxeram novas perspectivas que complementam abordagens já existentes na literatura, como [16]. Note que a formulação geral para determinar o limitante superior do algoritmo CMA é dada pela equação (5.23). Essa expressão poderia até mesmo permitir análises com auxílio de figuras, como foi o caso do algoritmo LMS (vide figura 4.7). Contudo, devido ao fato de que temos um formulação derivada de um polinômio de quarta

ordem (equação (5.22)), as superfícies que conseguimos traçar são um tanto irregulares ou mesmo dotadas de variações bruscas. Portanto, optamos por fazer a análise em torno de um possível ponto de equilíbrio ideal (solução *zero-forcing*).

As observações sobre estabilidade para o caso supervisionado, expostas na seção 4.5, são análogas às aplicáveis ao caso não-supervisionado:

- A estabilidade do algoritmo não é garantida para $\mu > \mu_{max}$.
- O valor de passo de adaptação $\mu < \mu_{max}$ é dependente do local de inicialização, o que pode ser observado na expressão (5.23).
- Para cada canal, temos uma faixa de μ estável que é particular.
- A análise proposta neste capítulo aponta apenas resultados sobre estabilidade. A questão da convergência não foi abordada. Contudo, o lugar geométrico definido por $\Delta V_{CMA} = 0$ determina o esboço da região de atração do algoritmo.
- O uso da condição *zero-forcing* foi apenas um subterfúgio para analisar o limite do algoritmo. Em linhas gerais, o que é relevante é saber o valor de $\lim_{\hat{\mathbf{w}}_n \rightarrow \mathbf{w}^*} \hat{\mathbf{w}}_n^T \mathbf{x}_n \mathbf{x}_n^T \hat{\mathbf{w}}_n$ para permitir avançar na análise do limite para o passo de adaptação. Veja, que neste caso não foi necessário calcular a solução algébrica do ponto de equilíbrio \mathbf{w}^* .

Capítulo 6

Inicialização do Algoritmo CMA

Nos capítulos anteriores, tivemos como foco a temática de estabilidade para algoritmos de equalização supervisionada e não-supervisionada. No entanto, no contexto de técnicas não-supervisionadas, outro problema correlato tem um papel fundamental no que diz respeito a desempenho: o problema da inicialização dessas técnicas.

Quando se lida com algoritmos de Bussgang, um primeiro aspecto relevante é que não é possível adotar uma inicialização sobre a origem, opção comum no âmbito de abordagens supervisionadas como o LMS. Ademais, uma escolha inadequada das condições iniciais pode levar à convergência para soluções sub-ótimas e mesmo à indesejada atração por pontos de sela [70]. Portanto, torna-se importante a definição de uma metodologia de inicialização que seja eficiente e, ao mesmo tempo, requiera um mínimo de informação *a priori* sobre as condições de operação.

A metodologia que, atualmente, pode ser considerada padrão é a *center-spike*, proposta por Foschini em 1985 [47]. Nessa metodologia, o "centro de massa" do vetor de coeficientes é colocado em sua região central, o que tende a valorizar soluções com atrasos intermediários, potencialmente seguras tanto para canais de fase mínima quanto para canais de fase não-mínima. Entretanto, mostraremos que, por meio de uma formulação polinomial do critério CM, é possível chegar a uma proposta que faz uso de informação contida no sinal recebido para prover as condições iniciais, sendo, portanto, potencialmente mais eficiente que a idéia de Foschini. Mostraremos ainda que essa formulação leva à dedução de um limitante para a função custo CM que, além de ser uma ferramenta teórica interessante, tem o valor de um "índice de equalizabilidade" do canal em questão.

6.1 Nova formulação para o algoritmo CMA

Seja a função custo do algoritmo CMA definida no capítulo 5:

$$J_{CMA}(y_n) = E\{|y_n|^2 - R_2\}^2. \quad (6.1)$$

Hipótese 10. Assumimos, sem perda de generalidade, que o sinal transmitido é tal que $R_2 = \frac{E\{|\mathbf{x}_n|^4\}}{E\{|\mathbf{x}_n|^2\}} = 1$.

Com a hipótese 10, a expressão (6.1) pode ser reescrita como:

$$J_{CMA}(y_n) = E\{|y_n|^2 - 1\}^2. \quad (6.2)$$

A equação (6.2) pode ser interpretada como uma função custo do critério de Wiener (MSE) com $s_{n-d} = 1$ e um equalizador cujo sinal de saída é

$$\nu_n = y_n^2, \quad (6.3)$$

o que nos daria um sinal de erro:

$$e_n = s_{n-d} - \nu_n = 1 - y_n^2. \quad (6.4)$$

O equalizador com sinal de saída da equação (6.3) pode ser interpretado como um filtro polinomial com restrições. O equalizador é polinomial é considerado com restrições pois a expressão para o sinal de saída é composta de multiplicações cruzadas e potências quadráticas dos termos do vetor de entrada do equalizador, ou seja, $\nu_n = \sum_{i,j,k,l}^N z_{i,j,k,l} x_{i,j} x_{k,l}$.

A restrição, neste caso, é a arquitetura do filtro e a dependência (note os termos com multiplicação) de seus parâmetros com respeito aos parâmetros do equalizador. Vide a equação (6.5) para maiores detalhes sobre os termos ou condições de restrição do filtro.

$$\begin{aligned} \nu_n = (\mathbf{w}_n \mathbf{x}_n^T)^2 &= \left(\underbrace{[w_0 \ w_1 \ \cdots \ w_N]}_{\text{pesos}} \underbrace{\begin{bmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_{n-N} \end{bmatrix}}_{\text{entradas}} \right)^2 \\ &= \left(\underbrace{w_0 x_n}_{\text{restrição nos termos multiplicativos}} + w_1 x_{n-1} \underbrace{+}_{\text{restrição de arquitetura}} \cdots + w_N x_{n-N} \right)^2 \end{aligned} \quad (6.5)$$

Contudo, podemos definir um equalizador sem considerar a restrição de arquitetura (ou mesmo multiplicativas), o qual é denotado por $\nu_n(u)$. O equalizador sem restrições é linear nos parâmetros e possui como entrada os mesmos sinais presentes no filtro com restrições, ou seja, elementos de segundo ordem compostos a partir de \mathbf{x}_n . A idéia proposta para o equalizador está esboçada na figura 6.1

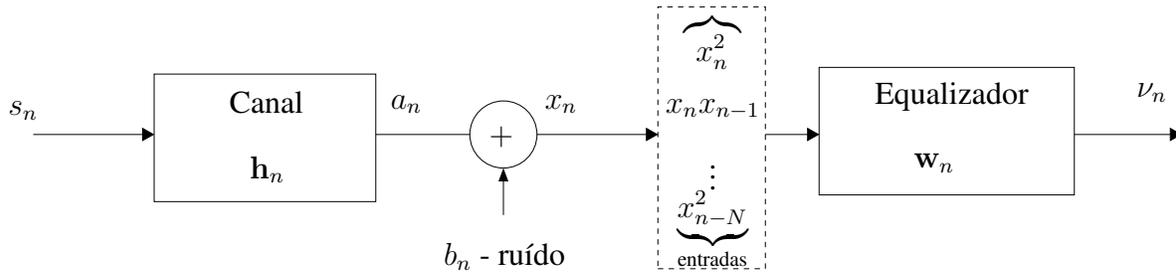


Fig. 6.1: Esboço do equalizador adaptativo polinomial.

Note que, na figura 6.1, o vetor de entrada para o equalizador ($[x_n^2 \ x_n x_{n-1} \ \dots \ x_{n-N}^2]$) é estruturado com os termos de \mathbf{x}_n . Portanto, temos uma estrutura de equalização diferente do que foi aplicado ao se definir o problema de equalização no capítulo 2.

Na próxima seção descreveremos as idéias do equalizador polinomial tomando um canal e equalizador com apenas dois coeficientes. Esse exercício ajudará no entendimento das particularidades da técnica proposta.

6.2 Canal com 2 coeficientes

Sejam os parâmetros de um equalizador com dois coeficientes:

$$\mathbf{w}_n = [w_0 \ w_1] , \quad (6.6)$$

e seja também o vetor de entrada do equalizador, que também é formado por dois coeficientes:

$$\mathbf{x}_n = [x_n \ x_{n-1}] . \quad (6.7)$$

O sinal de saída do equalizador é dado por:

$$y_n^2 = \nu_n = (\mathbf{w}_n \mathbf{x}_n^T)^2 = w_0^2 x_n^2 + 2w_0 w_1 x_n x_{n-1} + w_1^2 x_{n-1}^2 . \quad (6.8)$$

Portanto, podemos escrever a saída do equalizador em notação vetorial

$$y_n^2 = \nu_n = \overbrace{[w_0^2 \quad w_0 w_1 \quad w_1^2]}^{\text{pesos}} \overbrace{\begin{bmatrix} x_n^2 \\ 2x_n x_{n-1} \\ x_{n-1}^2 \end{bmatrix}}^{\text{entradas}}. \quad (6.9)$$

Veja que, até este momento, temos restrições na arquitetura (posição e termos multiplicativos específicos) do vetor de pesos. Ou seja, o vetor de pesos é formado por combinações fixas utilizando w_0 e w_1 . Essa restrição é ainda marcada pelo fato que temos posições fixas para os elementos w_i bem como as combinações cruzadas. Ignorando, por ora, a regra de formação do vetor de pesos, podemos escrever o sinal de saída do equalizador como

$$\nu_{n(u)} = [a \quad b \quad c] \begin{bmatrix} x_n^2 \\ 2x_n x_{n-1} \\ x_{n-1}^2 \end{bmatrix}. \quad (6.10)$$

Assim, é possível aplicar o critério de Wiener (que é unimodal) para determinar uma estimativa para o vetor de pesos do sistema sem restrições. A matriz de autocorrelação pode ser escrita como

$$\mathbf{R}_{(u)} = E \left\{ \begin{bmatrix} x_n^2 \\ 2x_n x_{n-1} \\ x_{n-1}^2 \end{bmatrix} \begin{bmatrix} x_n^2 & 2x_n x_{n-1} & x_{n-1}^2 \end{bmatrix} \right\}, \quad (6.11)$$

e o vetor de correlação cruzada é dado por:

$$\mathbf{p}_{(u)} = E \left\{ \begin{bmatrix} x_n^2 \\ 2x_n x_{n-1} \\ x_{n-1}^2 \end{bmatrix} \right\}. \quad (6.12)$$

A solução de Wiener é escrita como:

$$\mathbf{w}_{n(u)} = \mathbf{R}_{(u)}^{-1} \mathbf{p}_{(u)} = \begin{bmatrix} w_{0,n(u)} & w_{1,n(u)} & w_{2,n(u)} \end{bmatrix}^T. \quad (6.13)$$

Observamos que o EQM (Erro Quadrático Médio) associado à solução de Wiener obtida, ou seja, o EQM associado ao filtro polinomial sem restrições, é menor ou igual ao EQM associado ao filtro polinomial com restrições. Uma vez que o EQM do filtro com restrições é idêntico ao custo CM, o erro associado à solução de Wiener calculada em (6.13) é um limitante inferior da função custo que dá origem ao CMA. Esse limitante, além de ser relevante para o estudo do desempenho do algoritmo do módulo constante, serve ainda como um "índice de equalizabilidade" de um canal, uma vez que

sabemos que o custo CM está limitado a faixa que vai de 0 a 1. Por exemplo, para um dado canal e para um certo comprimento de equalizador, se tivermos o limitante inferior na ordem de, digamos, 0,1, podemos esperar que a situação seja, do ponto de vista de equalização, mais favorável que a encontrada para um canal que dê origem a um EQM irrestrito de 0,4. Um ponto interessante é que esse índice é obtido sem que se disponha de um sinal de referência e sem que se fixe *a priori* um atraso de equalização.

Combinando as equações (6.9) e (6.13), podemos estabelecer uma equivalência entre o equalizador com restrições (\mathbf{w}_n) e o sem restrições ($\mathbf{w}_{n(u)}$) em termos dos coeficientes do equalizador. Em outras palavras, cada termo do vetor da equação (6.13) possui uma correspondência com um dos elementos do vetor de pesos da expressão (6.9). Note ainda que o vetor de pesos da equação (6.9) é formado pela combinação de w_0 e w_1 . Portanto, propomos uma estimativa dos elementos que formam o vetor de pesos da equação (6.9) conforme a expressão:

$$\mathbf{w}_n = \begin{bmatrix} w_{0,n} \\ w_{1,n} \end{bmatrix} = \begin{bmatrix} \sqrt{w_{0,n(u)}} \\ \text{sign}(w_{1,n(u)})\sqrt{w_{2,n(u)}} \end{bmatrix}. \quad (6.14)$$

A estimativa construída a partir da equação (6.14) pode servir como condição inicial para um equalizador que utiliza como expressão de adaptação o algoritmo CMA¹. Nesse caso, buscamos aproveitar uma relação potencialmente estreita entre as soluções restrita e irrestrita para escolher favoravelmente uma região inicial do espaço de busca.

Na próxima seção vamos generalizar esses resultados para o caso onde o equalizador possui M coeficientes.

6.3 Equalizador com M coeficientes

Considerando o equalizador com M coeficientes, é possível escrever o vetor de pesos e vetor de entrada da seguinte maneira:

$$\mathbf{w}_n = [w_0 \quad w_1 \quad \cdots \quad w_{M-1}] , \quad (6.15)$$

$$\mathbf{x}_n = [x_n \quad x_{n-1} \quad \cdots \quad x_{M-1}] . \quad (6.16)$$

Portanto, o sinal de saída do equalizador é escrito como

$$y_n^2 = \nu_n = (\mathbf{x}_n \mathbf{w}_n^T)^2 = (w_0 x_n + w_1 x_{n-1} + \cdots + w_{M-1} x_{M-1})^2. \quad (6.17)$$

¹Esse método foi desenvolvido para o caso de sinais binários, vide capítulo 2. A análise para outros tipos de constelação necessita ser analisada.

Expandindo a expressão (6.17), temos

$$\begin{aligned} \nu_n = & w_{M-1}^2 x_{n-M+1}^2 + 2w_{M-1} x_{n-M+1} \sum_{i=0}^{M-2} w_i x_{n-i} + w_{M-2} x_{n-M+2}^2 \\ & + 2w_{M-2} x_{n-M+2} \sum_{i=0}^{M-3} w_i x_{n-i} + \cdots + w_0^2 x_n^2. \end{aligned} \quad (6.18)$$

Veja que a expressão (6.18) tem termos com w_i e x_n . Vamos escrever a matriz \mathbf{A} para facilitar a análise dos termos com x_n que formam o vetor de entradas do equalizador

$$\mathbf{A} = \begin{bmatrix} x_n \\ x_{n-1} \\ \dots \\ x_{n-M+1} \end{bmatrix} \begin{bmatrix} x_n & x_{n-1} & \dots & x_{n-M+1} \end{bmatrix} = \begin{bmatrix} x_n^2 & x_n x_{n-1} & \dots & x_n x_{n-M+1} \\ x_{n-1} x_n & x_{n-1}^2 & \dots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ x_{n-M+1} x_n & \dots & \dots & x_{n-M+1}^2 \end{bmatrix}_{M \times M}, \quad (6.19)$$

Reorganizando a matriz \mathbf{A} para formar o vetor de entradas temos o vetor \mathbf{A}' , temos:

$$\mathbf{A}' = \begin{bmatrix} x_n^2 \\ x_{n-1}^2 \\ \vdots \\ x_{n-M+1}^2 \\ x_n \sum_{i=1}^{M-1} x_{n-i} \\ x_{n-1} \sum_{i=2}^{M-1} x_{n-i} \\ \vdots \end{bmatrix}_{\sum_{i=0}^M (M-i) \times 1}, \quad (6.20)$$

A transição entre a matriz \mathbf{A} e o vetor \mathbf{A}' é esboçada na figura 6.2.

Na figura 6.2, o termo x_n^2 da matriz \mathbf{A} é posicionado no primeiro elemento do vetor \mathbf{A}' , x_{n-1}^2 ocupa o segundo do elemento de \mathbf{A}' , e o processo é finalizado utilizando os termos da diagonal de \mathbf{A} . Após preencher o vetor \mathbf{A}' com a matriz diagonal de \mathbf{A} , retorna-se na primeira linha da matriz \mathbf{A} e utiliza-se os termos $x_n x_{n-1}$, $x_n x_{n-2}$ até $x_n x_{n-M+1}$. Ao término da análise da primeira linha de \mathbf{A} , é considerada a segunda linha. Contudo, o processo recomeça com $x_{n-1} x_{n-2}$. Novamente, os termos do lado direito da diagonal são utilizados para preencher o vetor \mathbf{A}' . Resumidamente, os termos da matriz triangular superior de \mathbf{A} são utilizados para formar o vetor \mathbf{A}' .

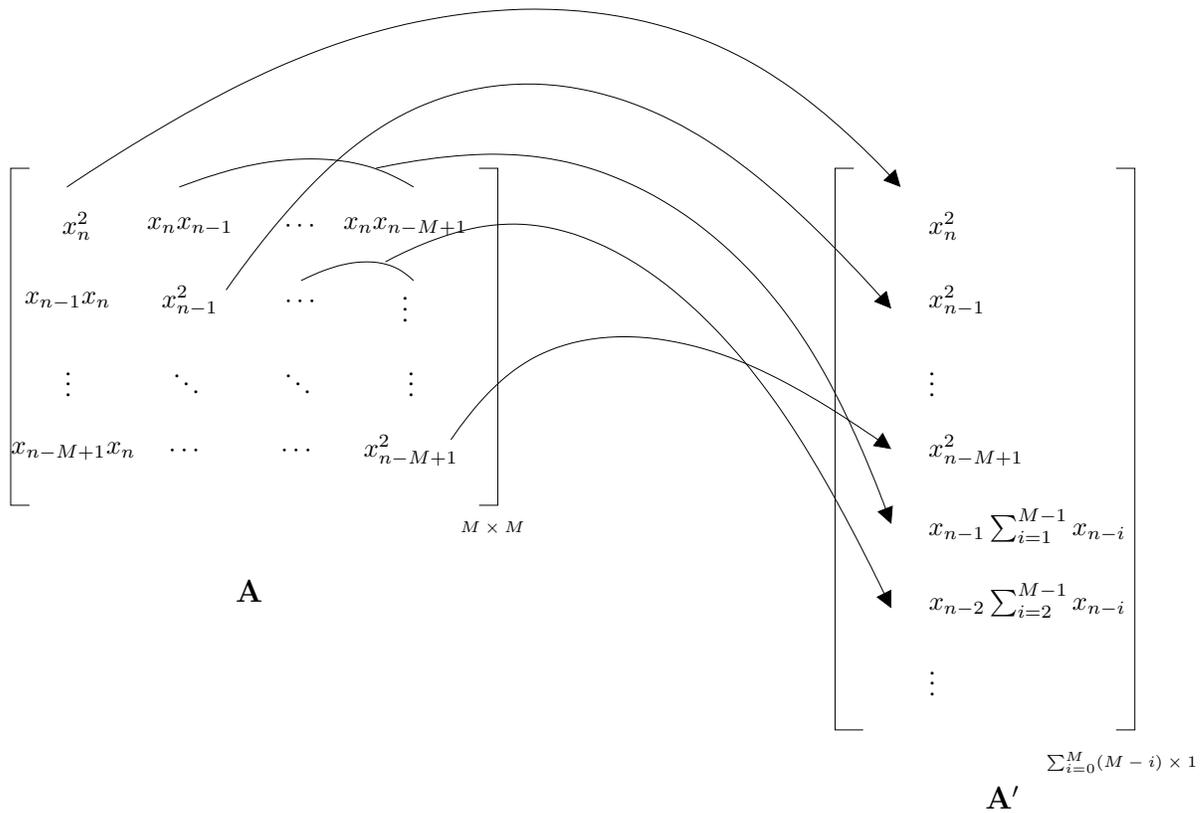


Fig. 6.2: Formação do vetor de entrada para inicialização do algoritmo CMA, o que corresponde a transição entre as equações (6.19) e (6.20).

Analisemos agora o arranjo do vetor de pesos do equalizador adequado à análise de inicialização. Considerando a equação (6.18), o vetor de pesos pode ser escrito como

$$\mathbf{B}' = 2 \left[w_0^2 \quad w_1^2 \quad \cdots \quad w_{M-1}^2 \quad w_1 \sum_{i=M-1}^{M-1} w_{M+1-i} \quad w_2 \sum_{i=M-2}^{M-1} w_{M+1-i} \quad \cdots \right]_{1 \times \sum_{i=0}^{M-1} (M-i)}, \quad (6.21)$$

Portanto, o sinal de saída do equalizador pode ser escrito como

$$y_n^2 = \nu_n \approx \mathbf{B}' \mathbf{A}'. \quad (6.22)$$

Tornando o problema de inicialização uma tarefa sem restrições, então podemos reescrever a equação (6.18) como

$$\nu_{n,(u)} = 2 \left[a \quad b \quad c \quad d \quad e \quad \cdots \right]_{1 \times \sum_{i=0}^{M-1} (M-i)} \mathbf{A}'. \quad (6.23)$$

Da mesma forma que na seção 6.2, a abordagem de Wiener pode ser utilizada para escrevermos

uma solução com base no vetor de pesos \mathbf{A}' . Portanto, a matriz de autocorrelação é dada por

$$\mathbf{R}_{(u)} = E \{ \mathbf{A}' \mathbf{A}'^T \} . \quad (6.24)$$

Como o sinal desejado é considerado unitário ($s_{n-d} = 1$) então o vetor de correlação cruzada é escrito como

$$\mathbf{p}_{(u)} = E \{ \mathbf{A}' \} . \quad (6.25)$$

A solução de Wiener é dada por (tomando $\sum_{i=0}^M (M-i) = P$):

$$\mathbf{w}_{n(u)} = \left[w_{0,n(u)} \quad w_{1,n(u)} \quad w_{2,n(u)} \quad \cdots \quad w_{P,n(u)} \right]^T = \mathbf{R}_{(u)}^{-1} \mathbf{p}_{(u)} , \quad (6.26)$$

O vetor da equação (6.26) pode ser reorganizado de maneira semelhante a matriz da equação (6.19):

$$\mathbf{C} = \begin{bmatrix} w_{0,n(u)} & w_{M,n(u)} & w_{M+1,n(u)} & \cdots & w_{2M-2,n(u)} \\ w_{M,n(u)} & w_{1,n(u)} & w_{2M-1,n(u)} & \cdots & w_{3M-3,n(u)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ w_{2M-2,n(u)} & w_{3M-3,n(u)} & \cdots & w_{P,n(u)} & w_{M-1,n(u)} \end{bmatrix}_{M \times M} , \quad (6.27)$$

Selecionando a linha da matriz \mathbf{C} que tem o elemento com maior valor numérico, então restringimos a matriz \mathbf{C} a apenas um vetor \mathbf{C}'

$$\mathbf{C}' = \mathbf{C}_i = \left[w_{M+i,n(u)} \quad \cdots \quad w_{i-1,n(u)} \quad \cdots \quad w_{(i+1)M-(i+1),n(u)} \right]_{1 \times M} , \quad (6.28)$$

onde $w_{i-1,n(u)}$ é o elemento de maior valor da diagonal da matriz \mathbf{C} .

Tomando as equações (6.21) e (6.28), podemos escrever a expressão abaixo como método para calcular os coeficientes do vetor de pesos para inicialização

$$\mathbf{w}_n = \begin{bmatrix} w_{0,n} \\ w_{1,n} \\ \vdots \\ w_{M-1,n} \end{bmatrix} = \begin{bmatrix} \text{sign}(\mathbf{C}_{i,1}) \sqrt{w_{0,n(u)}} \\ \text{sign}(\mathbf{C}_{i,2}) \sqrt{w_{1,n(u)}} \\ \text{sign}(\mathbf{C}_{i,3}) \sqrt{w_{2,n(u)}} \\ \vdots \\ \sqrt{w_{i-1,n(u)}} \\ \vdots \\ \text{sign}(\mathbf{C}_{i,M}) \sqrt{w_{M-1,n(u)}} \end{bmatrix}_{1 \times M} = \begin{bmatrix} \text{sign}(w_{M+i,n(u)}) \sqrt{w_{0,n(u)}} \\ \text{sign}(w_{M+i+1,n(u)}) \sqrt{w_{1,n(u)}} \\ \text{sign}(w_{M+i+2,n(u)}) \sqrt{w_{2,n(u)}} \\ \vdots \\ \sqrt{w_{i-1,n(u)}} \\ \vdots \\ \text{sign}(w_{2M+i-1,n(u)}) \sqrt{w_{M-1,n(u)}} \end{bmatrix}_{1 \times M} . \quad (6.29)$$

Note que o termo que tem o maior valor numérico da diagonal da matriz \mathbf{C} não tem a multiplicação

pelo função $sign(\cdot)$.

Conforme exposto anteriormente, se considerarmos o vetor de pesos apresentado na equação (6.26), é possível obter o mínimo EQM a ele associado:

$$J_{Wiener}(\mathbf{w}_{n(u)}) = 1 - \mathbf{w}_{n(u)}^T \mathbf{p}_{(u)} - \mathbf{p}_{(u)}^T \mathbf{w}_{n(u)} + \mathbf{w}_{n(u)}^T \mathbf{R}_{(u)} \mathbf{w}_{n(u)}. \quad (6.30)$$

Analisando a equação (6.30), podemos salientar que:

- o valor da função custo depende exclusivamente da informação contida no vetor de entradas (\mathbf{x}_n);
- temos uma superfície de custo convexa, o que é diferente da superfície de custo original do algoritmo CMA;
- a expressão (6.30) corresponde a um limitante inferior para a função custo do algoritmo CMA.

Na próxima seção apresentaremos simulações para verificar a eficácia do método proposto.

6.4 Simulações

Nesta seção, além de avaliarmos a solidez do limitante inferior, vamos analisar a performance da heurística de inicialização tendo por base uma comparação com duas outras técnicas de inicialização:

- inicialização conforme a técnica *center spike*;
- inicialização dos coeficientes do canal randomicamente com valores entre +1 e -1.

6.4.1 Caso A - Canal com dois coeficientes

Consideremos um canal com expressão do tipo

$$\mathbf{h}_n = \begin{bmatrix} 1 & \alpha \end{bmatrix}, \quad (6.31)$$

onde α é um valor positivo. Para alterar o canal basta modificar o valor de α (lembrando que sempre há uma subsequente normalização do vetor de coeficientes). Considerando um passo de adaptação fixo $\mu = 0,0005$ e 10.000 iterações, o algoritmo CMA foi executado utilizando a nova heurística de inicialização proposta na equação (6.14).

A figura 6.3 apresenta os valores de custo obtidos por três caminhos distintos:

- utilizando o limitante inferior deduzido ($J_{heu,min} = 1 - \mathbf{p}_{(u)}^T \mathbf{w}_{n(u)}$);
- o custo do CMA (equação (6.2), ou seja, $J_{CMA} = E\{(y_n^2 - 1)^2\}$, sendo $\mathbf{w}_{n=0} = \mathbf{w}_{Wiener}$) inicializando o algoritmo no mínimo de Wiener e considerando o vetor de saída do equalizador após todas as iterações; e
- o custo mínimo da função de Wiener (equação (2.20), ou seja, $J_{Wiener,min}(\mathbf{w}_n) = 1 - \mathbf{p}^T \mathbf{R}^{-1} \mathbf{p} = 1 - \mathbf{p}^T \mathbf{w}_{Wiener}$).

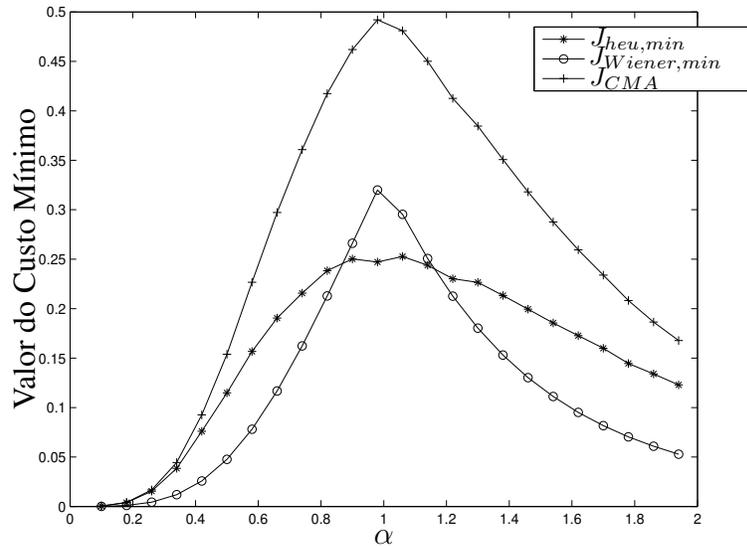


Fig. 6.3: Valor da função custo para simulação com algoritmo de 2 coeficientes.

Através da figura 6.3 é possível notar que o valor do custo associado ao limitante inferior é, de fato, menor que a estimativa do mínimo custo do CMA. O custo do limitante inferior (e dos demais métodos) aumenta em torno do ponto $\alpha = 1$, ou seja, na transição entre canal de fase mínima e fase não-mínima, que corresponde a zero sobre o círculo unitário. Essa região se associa a uma situação de notável dificuldade para equalização, o que gera uma maior discrepância entre a solução restrita e a solução irrestrita.

Para analisar a eficácia da heurística de inicialização, é bastante relevante verificar o número de realizações que apresentam convergência global. Neste caso, a referência de comportamento ideal é obtida por meio do algoritmo CMA inicializado com a melhor solução de Wiener. Considera-se que o algoritmo converge quando a norma da diferença entre o vetor de pesos resultante e o vetor de pesos de referência - levando em conta a ambiguidade de sinal inerente a algoritmos de Bussgang - é menor

Tab. 6.1: Porcentagem de realizações que apresentam convergência global (%) utilizando a inicialização com a nova heurística, *center spike* e inicialização aleatória.

Nova Heurística	Inicialização <i>Center Spike</i>	Inicialização Aleatória
95,83	45,83	25,00

que 0, 1. Conforme dito anteriormente, outras duas técnicas também foram analisadas: a inicialização *center spike* e uma inicialização aleatória em que os pesos do equalizador foram uniformemente escolhidos na faixa que vai de +1 a -1.

A tabela 6.1 apresenta a porcentagem de realizações com convergência, ou seja, o percentual de vezes que o algoritmos convergiu, para cada um dos métodos utilizados.

Percebe-se que os resultados obtidos com a nova heurística foram significativamente melhores que os atingidos pelas técnicas *center spike* e inicialização randômica do ponto de vista de convergência. Isso atesta que a solução irrestrita foi capaz de trazer informações significativas para a escolha adequada da região do espaço de busca em que haveria a inicialização.

6.4.2 Caso B - Canal com três coeficientes

Quando se utilizam canais com três coeficientes ($[1 \ \alpha \ \beta]$), ainda é possível encontrar uma representação gráfica para a superfície das funções custo. Neste sentido, elaboramos a simulação para representar a situação do limitante no mesmo contexto exposto na seção 6.4.1.

O resultado da simulação é sumarizado pela superfície da figura 6.4. Note que, mais uma vez, como esperado, o limitante inferior se mostrou sólido e foi capaz de acompanhar o perfil do menor erro quadrático médio atingível para o canal, o que confirma a sua relevância como "índice de equalizabilidade".

A porcentagem de realizações que apresentam convergência global para os três métodos possui uma ordem numérica análoga a do caso do canal de dois coeficientes, conforme mostra a tabela 6.2. Mais uma vez, a heurística sobressaiu em termos de desempenho. Vale ressaltar que, nas simulações desta seção, foram considerados um passo de adaptação fixo $\mu = 0,0007$ e 30.000 interações para análise da convergência.

Tab. 6.2: Porcentagem de realizações que apresentam convergência global (%) para canal com 3 coeficientes utilizando a inicialização com a nova heurística, *center spike* e inicialização randômica.

Nova Heurística	Inicialização <i>Center Spike</i>	Inicialização Aleatória
72,32	10,73	17,99

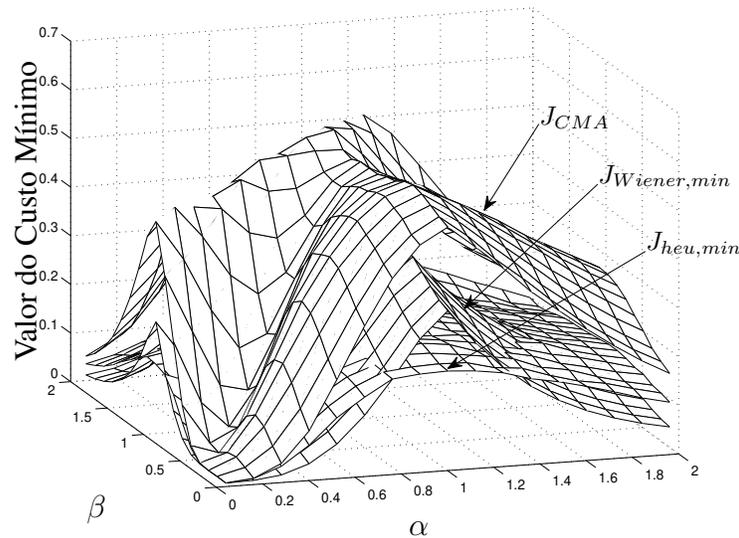


Fig. 6.4: Valor da função custo para simulação com algoritmo de 3 coeficientes.

6.4.3 Caso C - Canais com comprimento entre 2 e 5 coeficientes

Consideraremos agora canais de comunicação com comprimento aleatoriamente escolhido entre 2 e 5 coeficientes para realizar uma simulação ainda mais abrangente.

Na figura 6.5, apresentamos o mínimo custo para alguns canais de comunicação cujo comprimento e parâmetros foram escolhidos também aleatoriamente. Novamente, o custo associado ao limitante é inferior aos valores mínimos para o CMA e para o critério de Wiener. Na tabela 6.3, apresentamos os valores de porcentagem de realizações com convergência global para as três propostas.

A simulação foi realizada com passo de adaptação fixo igual a 0,0003. Considerou-se a ocorrência de convergência quando a diferença, em módulo, entre a solução de Wiener e o último vetor de pesos atingido pelo CMA foi menor que 0,1. As simulações para verificar a convergência contaram com 100 realizações de 30.000 iterações cada.

Tab. 6.3: Porcentagem de realizações com convergência global (%) utilizando a inicialização com a nova heurística, *center spike* e inicialização randômica para o diversos comprimentos de canal.

Nova Heurística	Inicialização <i>Center Spike</i>	Inicialização Aleatória
9,00	1,00	4,00

A figura 6.5 apresenta dez casos de cálculo de custo mínimo para canais diferentes comprimentos. A tabela 6.3 resume os resultados sobre a convergência utilizando três técnicas de inicialização.

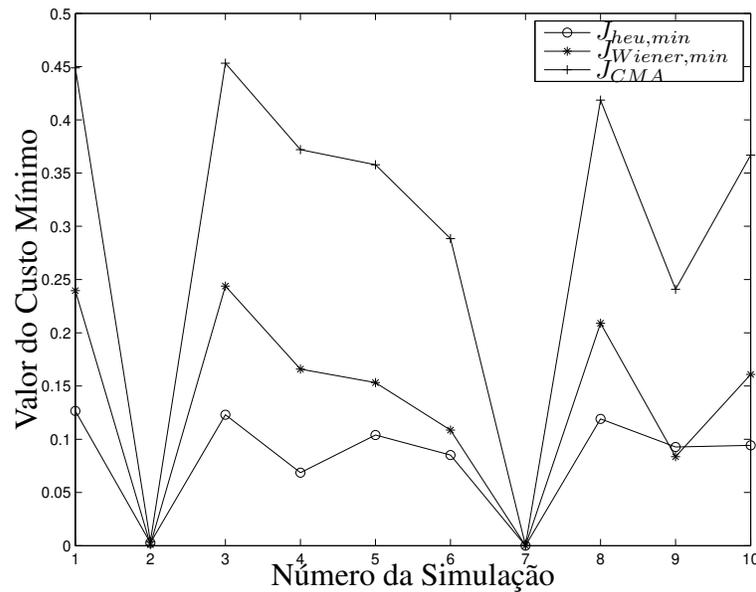


Fig. 6.5: Valor da função custo para simulação com algoritmo de 2 a 5 coeficientes.

6.5 Sumário

Neste capítulo, apresentamos um limitante inferior para a função custo CM e uma nova proposta para inicialização do CMA. As duas contribuições partem da idéia de estabelecer uma formulação do critério CM correspondente a uma versão do critério de Wiener para um filtro polinomial com restrições. Para o limitante inferior, isso é suficiente, mas, do ponto de vista da heurística de inicialização, faz-se necessário um passo adicional de relaxamento das restrições para que a solução de Wiener possa ser obtida em forma fechada.

A proposta de inicialização do CMA foi desenvolvida para equalizadores com dois coeficientes, e depois estendida para o caso de várias variáveis. Vale ressaltar que a proposta de inicialização faz relativamente poucas restrições sobre o sinal transmitido e não toma por base diretamente elementos do modelo de canal.

Os resultados de simulação indicam que tanto o limitante quanto a heurística são sólidos e podem ser, dessa forma, úteis para o efetivo emprego de equalizadores cegos. Vale ressaltar que a metodologia levou a um desempenho melhor que o da clássica abordagem *center spike* e que o de uma abordagem aleatória para diferentes modelos de canal.

Capítulo 7

Conclusão

Este trabalho apresenta resultados e contribuições oriundas da aplicação da teoria de Lyapunov ao estudo de estabilidade dos algoritmos de equalização adaptativa SDA, LMS e CMA. No caso do algoritmo LMS, a análise de estabilidade é estendida para discutir questões relacionadas à convergência. O algoritmo CMA conta ainda com uma proposta de um limitante inferior para a função custo CM, ao qual se associa a derivação de uma heurística de inicialização para o algoritmo do módulo constante.

No caso dos algoritmos de equalização supervisionada SDA e LMS, chegamos a limitantes bastante informativos para o passo de adaptação, os quais, sob certas hipóteses, corroboravam resultados já existentes na literatura. Concluímos disso que a metodologia expande análises e ferramentas aplicadas anteriormente ao problema de equalização mantendo, não obstante, uma significativa coerência com resultados obtidos segundo abordagens mais clássicas.

O segundo método de Lyapunov foi aplicado ao estudo do algoritmo LMS tanto na versão determinística (i.e. SDA) quanto na estocástica. A análise estocástica permitiu um estudo efetivo, sem alterações estruturais do problema investigado nem adoção de hipóteses restritivas quanto às estimativas instantâneas. Deve-se explicitar que o estudo do LMS, do ponto de vista de Lyapunov, não é algo novo, como atestam trabalhos importantes como [13], [24] e [23]. Contudo, a contribuição desta tese apresenta como diferencial a análise estocástica, sem o uso da teoria da independência e sem levantar hipóteses ou restrições com relação ao comportamento do canal de comunicação ou do sinal transmitido.

No caso específico do LMS, a análise realizada resultou também em uma investigação sobre o *misadjustment*. Por intermédio dessa análise, determinou-se o lugar geométrico para onde o algoritmo tenderá a convergir, que é, em geral, dotado de formato cônico. Vale ressaltar que essa análise aprofunda os resultados clássicos sobre o conceito de *misadjustment* e insinua uma mudança de paradigma interessante no que diz respeito ao conceito de convergência do algoritmo. Em vez de

considerar como convergência bem-sucedida a estabilização em apenas um ponto, é possível pensar em convergência para uma região (mesmo que pequena) em torno do ponto de equilíbrio. Vale ressaltar que não encontramos precedentes para a análise do *misadjustment* com recursos de geometria analítica.

Passando ao caso de algoritmos não-supervisionados, realizou-se também uma análise do algoritmo CMA do ponto de vista de Lyapunov. Em paralelo com o desenvolvimento desta tese, foi publicado um dos poucos trabalhos relacionados à investigação de estabilidade desse método utilizando análise estocástica [16]. Contudo, não existe sobreposição entre as técnicas utilizadas em [16] e o desenvolvimento feito no capítulo 5.

A investigação de estabilidade para o CMA foi corroborada com a análise da versão do CMA de um coeficiente. Para este caso específico, a análise de Lyapunov promoveu o mesmo resultado já obtido em [69] quando o algoritmo se aproxima do ponto de equilíbrio. Esse resultado deu subsídios e motivação para a análise genérica do algoritmo CMA com vários parâmetros.

Na abordagem para o algoritmo CMA, foi possível notar que existem três possibilidades para o limitante do passo de adaptação. Contudo, apenas uma das possibilidades representa o passo de adaptação máximo a ser utilizado. O mais interessante nesta análise é que chegamos a uma situação onde os limitantes para o passo de adaptação são representados por raízes de um polinômio de quarta ordem (uma das raízes é o valor 0 e as demais são limitantes). Portanto, existem escolhas de passo de adaptação que podem estar fora da faixa onde o algoritmo será estável. Novamente, hipóteses como a teoria da independência foram evitadas.

Em resumo, no estudo dos algoritmos SDA, LMS e CMA, alcançamos um objetivo comum: determinar um limitante superior para o passo de adaptação para o algoritmo apresentar estabilidade do ponto de vista de Lyapunov.

Durante as discussões que levaram ao desenvolvimento deste trabalho, percebemos que o ponto de inicialização é tão importante quanto o passo de adaptação. Assim, no capítulo 6, obtivemos um limitante inferior para a função custo CM que permite avaliar o desempenho atingível pelo equalizador para um determinado canal de comunicação. Em seguida, derivamos uma heurística de inicialização para o algoritmo do módulo constante cuja meta é aumentar tanto quanto possível a porcentagem de realizações que resultam em convergência global do método. Os resultados obtidos confirmaram o potencial dessa linha de trabalho como um todo.

Conseqüentemente, fechamos um ciclo para tratamento de algoritmos para equalização do ponto de vista de estabilidade pois consideramos, em alguma medida, as questões de:

- passo de adaptação
- ponto de inicialização

- representação matemática para a região de convergência.

7.1 Perspectivas para trabalhos futuros

A relação entre o processo de aprendizado de um equalizador adaptativo e a idéia de sistema dinâmico permitiu o uso da teoria de Lyapunov para investigar questões ligadas à estabilidade. Portanto, uma possibilidade para trabalhos futuros é considerar outras técnicas pertencentes ao escopo de sistemas não-lineares para investigação das propriedades dos equalizadores adaptativos. Citamos alguns casos promissores:

1. estender os resultados para o problema de separação de fontes ou para cenário de existência de múltiplos receptores;
2. considerar estabilidade no sentido entrada-saída e fazer uso da estabilidade \mathcal{L} ou mesmo de metodologias para sistemas robustos;
3. considerar o equalizador adaptativo como sistema markoviano, o que abre espaço para refinar os teoremas de Lyapunov conforme é proposto em [5];
4. propor novas arquiteturas fazendo uso de equalizadores com modelos mistos (parte do modelo é um equalizador FIR e outra parte é um composto por uma rede neural ou outro mecanismo proveniente de técnicas de *machine learning*);
5. considerar equalizadores baseados em redes neurais (ou redes wavelets) conforme já verificamos em [71] (ou no apêndice A); esta técnica mostra-se viável e de grande potencial, especialmente no sentido de introduzir novos algoritmos de treinamento e/ou novas arquiteturas

Referências Bibliográficas

- [1] M. Yelderman, B. Widrow, J. M. Cioffi, E. Hesler, e J. A. Leddy. ECG enhancement by adaptive cancellation of electrosurgical interference. *IEEE Transactions on Biomedical Engineering*, BME-30(7):392–398, jul. 1983.
- [2] P. Parks. Liapunov redesign of model reference adaptive control systems. *IEEE Transactions on Automatic Control*, 11 (3):362–367, jul. 1966.
- [3] A. C. Faleiros e T. Yoneyama. *Teoria Matemática de Sistemas*. Arte & Ciência, São Paulo, 2002.
- [4] C. Sousa Jr, E. M. Hemerly, e R. K. H. Galvão. Adaptive control of mobile robot using wavelet networks. *IEEE Trans. on Systems, Man and Cybernetics - Part B: Cybernetics*, 32(4):493–504, ago. 2002.
- [5] H. Kushner. *Stochastic Stability and Control*. Academic Press, New York, 1967.
- [6] T. Morozan. Stability of stochastic discrete systems. *Journal of Mathematical Analysis and Applications*, 23:1–9, 1968.
- [7] R. Z. Has'minskii. *Stochastic Stability of Differential Equations*. Sijthoff & Noordhoff, Maryland, 1980.
- [8] G. C. Goodwin e R. L. Payne. *Dynamic System Identification: Experiment Design and Data Analysis*. Academic Press, New York, NY, USA, 1977.
- [9] K. S. Narendra e A. M. Annaswamy. A new adaptive law for robust adaptation without persistent excitation. *IEEE Transactions on Automatic Control*, AC-32(2):134–145, fev. 1987.
- [10] R. Fierro e F. L. Lewis. Control of a nonholonomic mobile robots: Backstepping kinematics into dynamics. *Journal of Robotic Systems*, 14(3):149–163, 1997.

- [11] R. W. Lucky. Automatic equalization for digital communication. *Bell System Technical Journal*, 44:547–588, 1965.
- [12] D. N. Godard. Self-recovering equalization and carrier tracking in two dimensional data communication systems. *IEEE Transactions on Communications*, COM-28(11):1867–1875, nov. 1980.
- [13] V. H. Nascimento e A. Sayed. An unbiased and cost-effective leaky-LMS filter. *Asilomar Conference on Signals, Systems and Computers, Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2:1078–1082, out. 1996.
- [14] B. Widrow, J.M. McCool, M.G. Larimore, e Jr. Johnson. Stationary and nonstationary learning characteristics of the LMS adaptive filter. *Proceedings of the IEEE*, 64(8):1151–1162, ago. 1976.
- [15] O. Dabeer e E. Masry. Convergence analysis of the constant modulus algorithm. *IEEE Transactions on Information Theory*, 49(6):1447–1464, jun. 2003.
- [16] V. H. Nascimento e M. T. M. Silva. Stochastic stability analysis for the constant-modulus algorithm. *IEEE Transactions on Signal Processing*, 56(10):4984–4989, out. 2008.
- [17] R. R. F. Attux C. C. Cavalcante F. J. V. Von Zuben L. N. Castro e J. M. T. Romano R. Ferrari, C. M. Panazio. Unsupervised channel equalization using fuzzy-prediction filters. *In Proceedings of IEEE Workshop on Neural Networks for Signal Processing (NNSP2003)*, 1:869–878.
- [18] D. Kundur e D. Hatzinakos. On the use of Lyapunov criteria to analyze the convergence of blind deconvolution algorithms. *IEEE Transactions on Signal Processing*, 46(11):2918–2925, nov. 1998.
- [19] O. Diene e A. Bhaya. Adaptive filtering algorithms designed using control Liapunov functions. *IEEE Signal Processing Letters*, 13(4):224–227, 2006.
- [20] J.-J. E. Slotine e W. Li. *Applied Nonlinear Control*. Prentice Hall, New Jersey, 1991.
- [21] M. Vidyasagar. *Nonlinear System Analysis*. Prentice-Hall Inc, New Jersey, segunda edição, 1993.
- [22] H. K. Khalil. *Nonlinear Systems*. Prentice Hall, New Jersey, segunda edição, 1996.
- [23] R. R. Bitmead e B. D. O Anderson. Lyapunov techniques for the exponential stability of linear difference equations with random coefficients. *IEEE Transactions on Automatic Control*, AC-25 (4):782–787, ago. 1980.

- [24] V. Solo. The stability of LMS. *IEEE Transactions on Signal Processing*, 45 (12):3017–3026, dez. 1997.
- [25] V. H. Nascimento e A. Sayed. Unbiased and stable leakage-based adaptive filters. *IEEE Transactions on Signal Processing*, 47 (12):3261–3276, dez. 1999.
- [26] D. A. Cartes, L. R. Ray, e R. D. Collier. Lyapunov tuning of the leaky LMS algorithm for single-source, single-point noise cancellation. *Mechanical Systems and Signal Processing*, 17 (5):925–944, 2003.
- [27] B. Widrow e S. D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, Upper Saddle River, 1985.
- [28] S. Haykin. *Adaptive Filter Theory*. Prentice-Hall, Upper Saddle River, terceira edição, 1996.
- [29] H. J. Kushner. *Approximation and weak convergence methods for random processes, with applications to stochastic systems theory*. MIT Press Series in Signal Processing, Optimization, and Control, 6. MIT Press, Cambridge, MA, 1984.
- [30] S. Meyn e R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, New York, NY, USA, 2005.
- [31] V. H. Nascimento e A. H. Sayed. On the learning mechanism of adaptive filters. *IEEE Transactions on Signal Processing*, 48(6):1609–1625, jun. 2000.
- [32] V. Krishnan. *Nonlinear Filtering and Smoothing: An Introduction to Martingales, Stochastic Integrals and Estimation*. Dover Publications, Inc., New York, NY, USA, 2005.
- [33] V. H. Nascimento e J. C. M. Bermudez. Probability of divergence for the least-mean fourth algorithm. *IEEE Transactions on Signal Processing*, 54(4):1376–1385, abr. 2006.
- [34] S. Abrar e A. K. Nandi. An adaptive constant modulus blind equalization algorithm and its stochastic stability analysis. *Signal Processing Letters*, 17(1):55–58, jan. 2010.
- [35] K. Dogancay. Blind compensation of nonlinear distortion for bandlimited signals. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 52(9):1872–1882, set. 2005.
- [36] A. V. Oppenheim, R. W. Schaffer, e J. R. Buck. *Discrete-Time Signal Processing*. Prentice-Hall, Upper Saddle River, segunda edição, 1999.
- [37] G. C. Goodwin e K. S. Sin. *Adaptive Filtering, Prediction and Control*. Prentice-Hall, Upper Saddle River, 1984.

- [38] A. O. Neves, R. R. de F. Attux, R. Suyama, M. D. Miranda, e J. M. T. Romano. Sobre critérios para equalização não-supervisionada. *Revista Controle & Automação*, 17(3):278–299, ago. 2006.
- [39] S. V. Vaseghi. *Advanced Digital Signal Processing and Noise Reduction*. John Wiley & Sons Ltd, New York, segunda edição, 2000.
- [40] L. Bottou. *On-line learning and stochastic approximations*. Cambridge University Press, New York, NY, USA, 1998.
- [41] S. Jaggi e A. B. Martinez. Upper and lower bounds of the misadjustment in the LMS algorithm. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(1):164–166, jan. 1990.
- [42] H. J. Butterweck. A new interpretation of the misadjustment in adaptive filtering. In *ICASSP '96: Proceedings of the Acoustics, Speech, and Signal Processing, 1996. on Conference Proceedings., 1996 IEEE International Conference*, páginas 1641–1643, Washington, DC, USA, 1996. IEEE Computer Society.
- [43] P. A. Regalia. On the equivalence between the Godard and Shalvi-Weinstein schemes of blind equalization. *Signal Process.*, 73(1-2):185–190, 1999.
- [44] J. R. Treichler e B. G. Agee. A new approach to multipath correction of constant modulus signals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-31(2):459–473, abr. 1983.
- [45] Z. Ding, A. Kennedy, B. D. O. Anderson, e C. R. Johnson. Ill-convergence of Godard blind equalizers in data communication systems. *IEEE Transactions on Communications*, 39(9):1313–1327, set. 1991.
- [46] C. R. Johnson Jr., P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, e R.A Casas. Blind equalization using the constant modulus criterion: A review. *Proceeding of The IEEE*, 86(10):1927–1950, 1998.
- [47] G. J. Foschini. Equalizing without altering or detecting data. *Bell System Technical Journal*, 64:1885–1911, 1985.
- [48] S. Haykin. *Neural Networks and Learning Machines*. Prentice Hall, Cambridge, MA, terceira edição, 2009.

- [49] B. Fisher e N. Bershad. The complex LMS adaptive algorithm—transient weight mean and covariance with applications to the ALE. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 31(1):34 – 44, fev. 1983.
- [50] E. Kaszkurewicz e A. Bhaya. *Matrix Diagonal Stability in Systems and Computation*. Birkhäuser, Boston, 2000.
- [51] B. Noble e J. W. Daniel. *Álgebra Linear Aplicada*. Rio de Janeiro, RJ, segunda edição, 1986.
- [52] R. A. Horn e C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1990.
- [53] R. Bellman. *Stability Theory of Differential Equations*. Dover, Mineola, New York, 2008.
- [54] K. Ogata. *Discrete-time control systems*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, segunda edição, 1987.
- [55] A. Iggidr e M. Bensoubaya. New results on the stability of discrete-time systems and applications to control problems. *Journal of Mathematical Analysis and Applications*, 219:392–414, 1998.
- [56] P. A. Ioannou e J. Sun. *Robust Adaptive Control*. Prentice-Hall, New Jersey, 1996.
- [57] M. Aoki. *Optimization of Stochastic Systems*. Academic Press, New York, 1967.
- [58] C.-T. Chen. *Linear System Theory and Design*. Oxford University Press, Inc., New York, NY, USA, 1984.
- [59] S. Meyn. *Control Techniques for Complex Networks*. Cambridge University Press, New York, NY, USA, 2007.
- [60] E. F. de Arruda. *Paradigma de Programação Dinâmica Discreta Em Problemas Estocásticos de Investimento e Produção*. Doutorado, Universidade Estadual de Campinas - UNICAMP, Campinas - SP - BR, 2006.
- [61] P. Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, New York, NY, USA, 1998.
- [62] V. Sundarapandian. An invariance principle for discrete-time nonlinear systems. *Applied Mathematics Letters*, 16:85–91, 2003.
- [63] J. P. LaSalle. *The Stability of Dynamical Systems*. SIAM, Philadelphia, 1976.

- [64] P. Florchinger. A stochastic Jurdjevic-Quinn theorem. *SIAM J Control Optim*, 41(1):83–88, 2003.
- [65] V. Rajaraman e H. Wertz. On stability and "steepest descent". *IEEE Transactions on Automatic Control*, 8(1):61–62, jan. 1963.
- [66] H. H. Dam, S. Nordholm, S. Y. Low, e A. Cantoni. Blind signal separation using steepest descent method. *IEEE Transactions on Signal Processing*, 55(8):4198–4207, ago. 2007.
- [67] M. Torii e M. T. Hagan. Stability of steepest descent with momentum for quadratic functions. *IEEE Transactions on Neural Networks*, 13(3):752–756, maio 2002.
- [68] P. I. Hubscher, J. C. M. Bermudez, e V. H. Nascimento. A mean-square stability analysis of the least mean fourth adaptive algorithm. *IEEE Transactions on Signal Processing*, 55(8):4018–4028, ago. 2007.
- [69] R. R. F. Attux e J. M. T. Romano. Chaotic phenomena in adaptive blind equalisers. *IEE Proceedings on Vision, Image and Signal Processing*, 150(6):360–364, dez. 2003.
- [70] R. R. F. Attux. *Sobre Dinâmica Caótica e Convergência em Algoritmos de Equalização Auto-didata*. Msc, Universidade Estadual de Campinas - UNICAMP, Campinas - SP - BR, 2001.
- [71] C. Sousa Jr, R. R. F. Attux, e J. M. T. Romano. Esquemas para filtragem adaptativa baseados em redes wavelet. In *Anais do XXII Simpósio Brasileiro de Telecomunicações (SBrT'05)*.
- [72] O. Rioul e P. Duhamel. Fast algorithms for discrete and continuous wavelet transforms. *IEEE Transactions on Information Theory*, 38(2):569–586, mar. 1998.
- [73] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.
- [74] J. Zhang, G. G. Walter, Y. Miao, e W. N. W. Lee. Wavelet neural networks for function learning. *IEEE Transactions on Signal Processing*, 43(6):1485–1497, jun. 1995.
- [75] B. Dorizzi e J. C. M. Mota C. C. Cavalcante, J. R. M. Filho. A neural predictor for blind equalization of digital communication systems: Is it plausible? *Proceedings of the 2000 IEEE Signal Processing Society Workshop*, 2:736–745.

Apêndice A

Apêndice 1 - Resultados do uso de redes wavelets em equalização

São apresentados neste apêndice os resultados obtidos com o uso de redes wavelets no problema de equalização adaptativa. Esses resultados foram reportados em [71] e são descritos aqui como registro de parte das pesquisas realizadas durante o período do trabalho de doutorado.

A.1 Redes Wavelet

A representação natural da transformada wavelet considera uma versão transladada e dilatada de uma função contínua ϕ , conhecida também como wavelet mãe:

$$\phi_{a,b}(t) = \frac{1}{\sqrt{a}} \phi\left(\frac{t-b}{a}\right), a \in \mathbb{R}^*, b \in \mathbb{R}. \quad (\text{A.1})$$

Contudo, a representação no domínio de funções contínuas é pouco apropriada para problemas envolvendo sistemas digitais. Sendo assim, é preferível adotar uma parametrização na qual as grandezas de tempo e escala (b e a) são discretizadas [72]

$$\phi_{j,k}(t) = 2^{-j/2} \phi\left(\frac{t-2^j k}{2^j}\right), j, k \in \mathbb{Z}. \quad (\text{A.2})$$

Uma dada função $f \in L^2(\mathbb{R})$ pode ser caracterizada ou reconstruída por uma superposição de elementos de funções wavelet como [73]

$$f = \sum_{j,k} \langle f, \phi_{j,k} \rangle \tilde{\phi}_{j,k}. \quad (\text{A.3})$$

Uma amostragem mais precisa dos parâmetros da wavelet pode ser obtida considerando

$$\phi_{m,k}(t) = 2^{-m/2M} \phi\left(\frac{t - 2^{m/M}k}{2^{m/M}}\right), m = 0, \dots, M - 1. \quad (\text{A.4})$$

onde $M \in \mathbb{Z}$ [72]. Para reduzir o número de operações por função wavelet, limitaremos o parâmetro k a apenas alguns valores.

No entanto, para garantir que a composição de wavelets seja capaz de reconstruir a função f , o conjunto $\Phi = \{\phi_{m,k}, m, k \in \mathbb{Z}\}$ deve constituir um *frame* para $L^2(\mathbb{R})$ [73], ou seja, devem existir $A > 0$ e $B < \infty$ tais que

$$A\|f\|^2 \leq \sum_{m,k} |\langle f, \phi_{m,k} \rangle|^2 \leq B\|f\|^2$$

para todo $f \in L^2(\mathbb{R})$. Maiores detalhes sobre as condições suficientes para formação de um *frame* podem ser encontrados em [73].

Sendo assim, será considerada aqui a seguinte parametrização para a rede wavelet

$$y_n = \sum_{m=0}^{M-1} \sum_{k=-K, k \neq 0}^K w_{m,k} \phi_{m,k}(x_n), \quad (\text{A.5})$$

onde $w_{m,k}$ representa os pesos da camada de saída para este nó da rede. A figura A.1 ilustra alguns nós da rede wavelet:

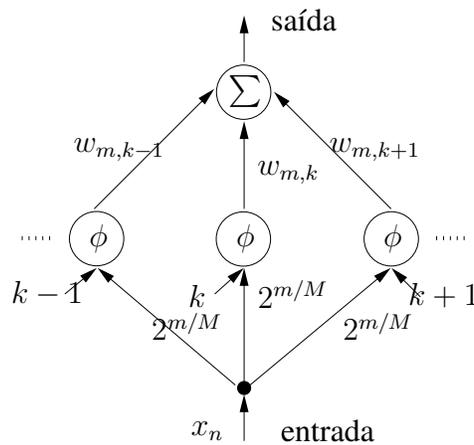


Fig. A.1: Representação de um nó da rede wavelet considerando apenas uma única entrada.

Quando a rede wavelet é alimentada por mais de uma entrada, torna-se necessário repetir a arquitetura da figura A.1 para cada uma das entradas.

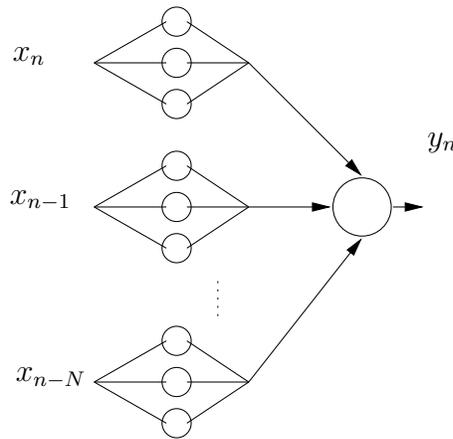


Fig. A.2: Representação da rede wavelet para múltiplas entradas.

Entretanto, do ponto de vista da parametrização, basta adicionar mais um somatório à expressão da saída da rede (equação (A.5)). Nas redes wavelet, vale também a representação por matrizes. Assim, a saída da rede wavelet pode ser reescrita como

$$y_n = \sum_{i=1}^N \sum_{m=0}^{M-1} \sum_{k=-K, k \neq 0}^K w_{i,m,k} \phi_{m,k}(x_{i,n}) = \underbrace{\mathbf{W}_n}_{1 \times M.N.2K} \underbrace{\Phi_n}_{M.N.2K \times 1}. \quad (\text{A.6})$$

onde N é o número de entradas para rede wavelet.

A parametrização descrita na equação (A.2) assemelha-se a um equalizador linear e também à representação de redes do tipo RBF. Existem alguns pontos em comum entre a rede wavelet e a RBF [74]:

- 1 Wavelets podem ser consideradas como versões deslocadas e dilatadas de uma dada função de ativação (que obedeça às condições de admissibilidade);
- 2 Valem as propriedades de aproximação universal e aproximação L^2 tanto para redes RBF quanto para redes wavelets;
- 3 Algoritmos para treinamento de redes neurais (por exemplo, o algoritmo LMS) podem ser utilizados para redes wavelets.

Neste apêndice, utilizaremos o algoritmo LMS normalizado para adaptar os pesos da camada de

saída da rede wavelet. A expressão iterativa dessa técnica é [28]:

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \mu \frac{\mathbf{z}_n e_n^*}{\|\mathbf{z}_n\|^2}, \quad (\text{A.7})$$

onde μ é o passo de adaptação, \mathbf{z}_n o vetor de entradas para o algoritmo e e_n o sinal de erro. O vetor \mathbf{z}_n e o escalar e_n serão devidamente explicitados nas próximas seções.

A.2 Adaptação Estrutural

Adotamos a estratégia de adaptação estrutural proposta em [4], ou seja, os nós da camada escondida serão divididos em três grupos conforme a evolução dos coeficientes da camada de saída:

$$\begin{aligned} \mathbb{L}^+ : \quad & \forall j \quad \text{se} \quad \|\mathbf{W}_j\|_n - \|\mathbf{W}_j\|_{n-1} > \mu_1 \\ \mathbb{L}^o : \quad & \forall j \quad \text{se} \quad -\mu_2 \leq \|\mathbf{W}_j\|_n - \|\mathbf{W}_j\|_{n-1} \leq \mu_1 \\ \mathbb{L}^- : \quad & \forall j \quad \text{se} \quad \|\mathbf{W}_j\|_n - \|\mathbf{W}_j\|_{n-1} < -\mu_2 \end{aligned} \quad (\text{A.8})$$

onde \mathbf{W}_j é o j -ésimo coeficiente da camada de saída.

As leis de adaptação da rede são:

- Selecione um nó do conjunto \mathbb{L}^- para ser removido a cada iteração no instante de tempo n ;
- Adicione um nó na rede se houver pelo menos um ponto em \mathbb{L}^+ ;

O conjunto \mathbb{L}^o não afeta a adaptação estrutural e não é permitido excluir ou adicionar mais de um nó por iteração. Quando um nó é introduzido na rede, seu peso e seus coeficientes para dilatação e translação são inicializados aleatoriamente.

A.3 Equalização com Rede Wavelet

No problema de equalização supervisionada, a rede wavelet terá por objetivo recuperar com a maior fidelidade possível uma versão atrasada s_{n-d} do sinal transmitido. O treinamento da rede terá por base uma medida de erro entre o sinal desejado e a saída y_n , como mostra a figura A.3.

No caso não-supervisionado, projetamos a rede wavelet para que ela faça o papel de um preditor num esquema em que o equalizador é o filtro de erro de predição resultante. Essa proposta foi fundamentada nos arcabouço teórico desenvolvido em [75] e [17]. Assumimos que a parte determinística

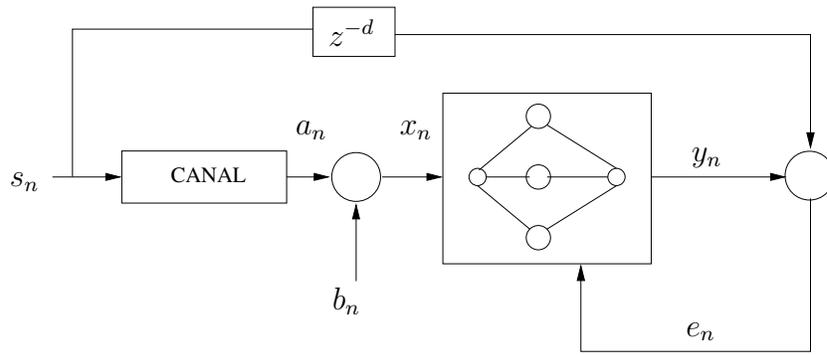


Fig. A.3: Rede wavelet para equalização supervisionada.

do canal pode ser modelada por um filtro FIR:

$$F(z) = \sum_{i=0}^{N_h-1} h_i z^i, \quad (\text{A.9})$$

onde N_h é o número de coeficientes do canal h .

Dessa forma, o sinal de entrada do equalizador (que é o filtro de erro de predição) pode ser representado como

$$\mathbf{x}_n = h_0 s_n + h_1 s_{n-1} + \dots + h_{N_h-1} s_{n-N_h+1} + b_n, \quad (\text{A.10})$$

onde b_n é o processo estocástico que representa o ruído aditivo. Dando o nome de $Predictor(\cdot)$ ao operador que representa o mapeamento efetuado pelo preditor, chegamos à seguinte forma para o erro de predição

$$\begin{aligned} e_n^{pred} &= x_n - Predictor(\mathbf{x}_{n-1}) \\ &= h_0 s_n + h_1 s_{n-1} + \dots + h_{N_h-1} s_{n-N_h+1} + b_n - \underbrace{Predictor(\mathbf{x}_{n-1})}_{\hat{x}_n} \\ &= h_0 s_n + h_1 s_{n-1} + \dots + h_{N_h-1} s_{n-N_h+1} + b_n - \underbrace{Predictor(h_0 s_{n-1} + \dots + h_{N_h-1} s_{n-N_h+2} + b_{n-1})}_{\hat{x}_n} \end{aligned} \quad (\text{A.11})$$

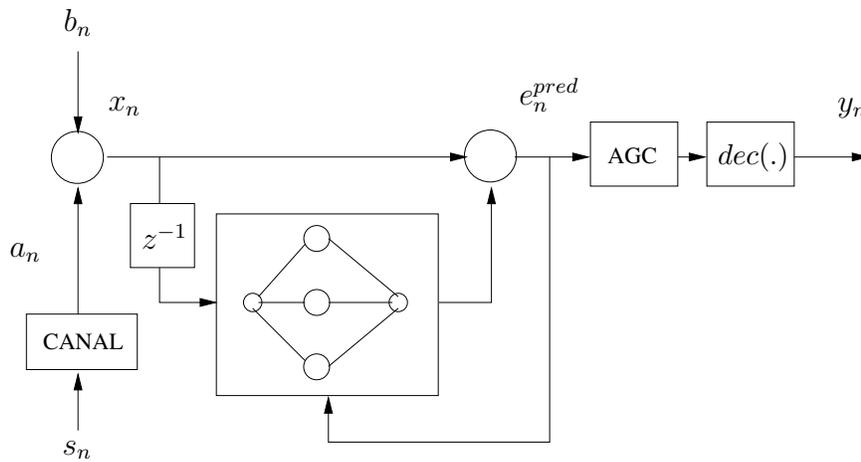


Fig. A.4: Rede wavelet para equalização cega.

No caso ideal, o preditor da equação (A.11) cancelará os termos redundantes como $(h_1 s_{n-1} + \dots + h_{N_h-1} s_{n-N_h+1})$ e também parte de b_n . Sendo assim, o erro de predição se resumirá a

$$e_n^{pred} = h_0 s_n + b'_n, \quad (\text{A.12})$$

onde b'_n representa o ruído residual. Portanto, foi recuperado o sinal transmitido, com atraso zero, a menos de um ganho, que pode ser compensado por um mecanismo do tipo AGC (Automatic Gain Control), conforme proposto por [75]. A saída do AGC é submetida a ação de um decisor $dec(\cdot)$, o qual, neste caso, é a função sinal.

Nos primeiros trabalhos que fizeram uso dessa proposta, foram utilizadas redes do tipo MLP e um filtro fuzzy [17][75]. A proposta apresentada aqui contém, portanto, a primeira aplicação de uma estrutura baseada em wavelets a esse problema particular.

A.4 Resultados de Simulação

Para o caso supervisionado, as configurações de canal, ruído, da rede wavelet e dos algoritmos de treinamento estão na primeira linha da Tabela A.1. Escolhemos um canal de fase mínima e o atraso de equalização que produz o problema de equalização mais complexo, o qual, aliás, não é passível de solução linear.

Os resultados para o caso supervisionado são apresentados nas figuras A.5-A.7. A figura A.5 mostra o sinal de saída do equalizador, ou seja, y_n . O sinal de erro instantâneo e_n é exibido na

figura A.6, onde chamamos a atenção para a redução acentuada do erro após os primeiros instantes da simulação.

A figura A.8 retrata o sinal e_n^{pred} para o caso não-supervisionado. O objetivo do preditor é tornar e_n^{pred} o mais próximo possível de s_n (pois $h_0 = 1$, vide a segunda linha da Tabela A.1). Tal meta é alcançada, conforme atesta o comportamento de "abertura do olho" durante a simulação, reportado na figura A.8, e os erros de decisão retratados na figura A.9. Portanto, concluímos que a rede wavelet cumpriu a tarefa de maneira adequada.

A figura A.8 retrata o caso não-supervisionado. Ambas figuras são divididas em três gráficos: a saída do equalizador, o erro entre sinal desejado e saída e uma estimativa do erro quadrático médio. Os três gráficos mostram que a rede consegue obter um sinal de saída próximo ao sinal desejado, sendo a diferença entre eles devida predominantemente ao ruído residual.

As figuras A.7 e A.10 mostram o número de neurônios (ou nós) da camada oculta da rede para cada uma das situações, o que nos permite ter uma boa noção do papel do mecanismo de adaptação estrutural. Observamos que, na fase inicial do processo adaptativo, há uma gradual inserção de neurônios até que, por fim, atinge-se uma configuração adequada.

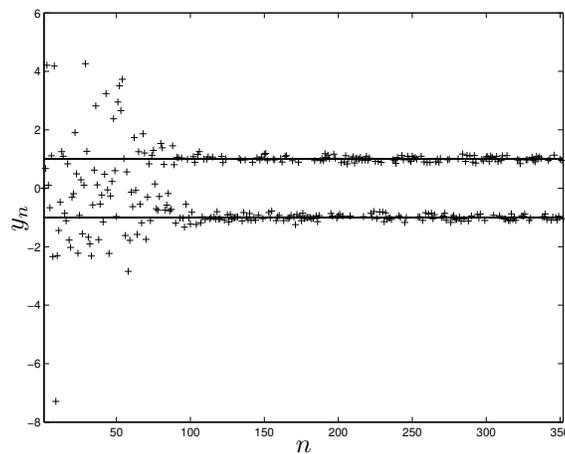


Fig. A.5: Rede wavelet com adaptação estrutural para o caso supervisionado.

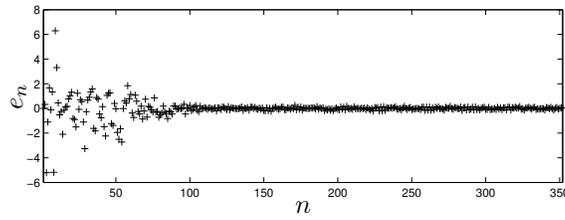


Fig. A.6: Erro instantâneo da simulação da figura A.5.

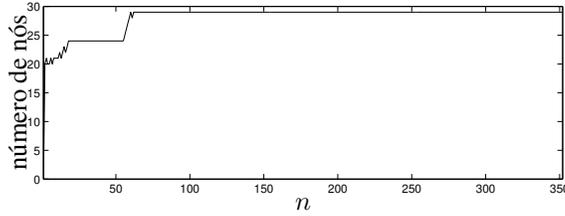


Fig. A.7: Evolução do número de nós da simulação da figura A.5.

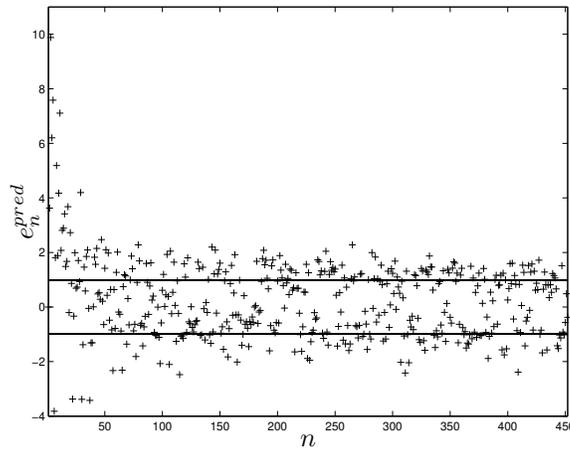


Fig. A.8: Rede wavelet com adaptação estrutural para o sistema não-supervisionado.

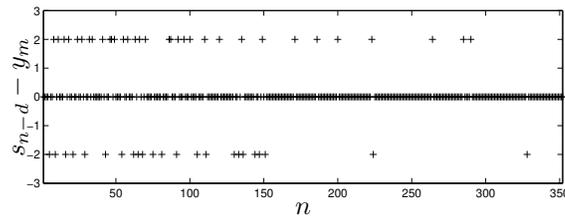


Fig. A.9: Diferença entre s_{n-d} e y_n para simulação da figura A.8 (caso não-supervisionado).

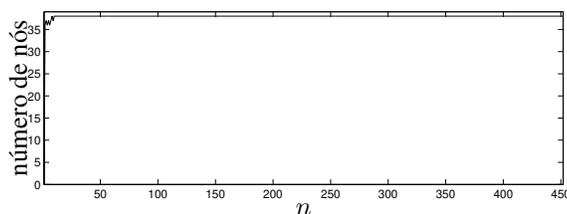


Fig. A.10: Evolução do número de nós da simulação da figura A.8.

Tab. A.1: Parâmetros das redes wavelet utilizadas nas simulações.

figuras	M	N	b_n	K	Canal	μ	μ_1	μ_1	Ganho AGC	d
A.5,A.6 e A.7	4	5	20 dB	2	$h(z) = 1 + 0,6z^{-1}$	0,7	0,15	0,15	0,009	2
A.8,A.9 e A.10	3	3	21 dB	3	$h(z) = 1 + 0,8z^{-1} + 0,4z^{-2}$	0,3	0,2	0,2	0,009	0

Nota 16. Em todas as simulações com wavelet foi considerado $\phi(t) = (1 - t^2) \exp(-t^2/2)$.

A.5 Sumário

Os resultados mostram que a proposta apresentada neste apêndice teve um bom desempenho tanto num contexto supervisionado (em que o treinamento teve por base o critério de Wiener) quanto num contexto não-supervisionado (baseado no critério de erro de predição), o que nos permite tirar conclusões otimistas acerca da aplicabilidade de redes wavelets a uma ampla gama de problemas de equalização. Dentre as características positivas da rede wavelet, desejamos destacar as seguintes:

1. Capacidade de aproximação universal, o que possibilita a aplicação em uma gama ampla de canais, incluindo os não-lineares e variantes no tempo;
2. A rede wavelet pode ser construída com diferentes funções (incluindo Gaussianas), o que pode ser encarado como uma generalização dos outros tipos de rede, por exemplo RBF e MLP.
3. A rede wavelet com adaptação estrutural permite tratar situações nas quais a dinâmica do sistema muda bruscamente durante a equalização. Isto inspira a investigação desta ferramenta para sinais com estatísticas não estacionárias.

Contudo, restam vários pontos para investigação como, por exemplo, elaborar uma estratégia para determinar os parâmetros iniciais da rede (M , N e K) e os ganhos do critério de adaptação estrutural (μ_1 e μ_2).