



Denis Gustavo Fantinato

**Equalização Não Supervisionada: Contribuições ao
Estudo do Critério do Módulo Constante
e de Métodos Baseados na Teoria da Informação**

Campinas
2013



UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Elétrica e de Computação

Denis Gustavo Fantinato

Equalização Não Supervisionada: Contribuições ao Estudo do Critério do Módulo Constante e de Métodos Baseados na Teoria da Informação

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica, na Área de Engenharia de Computação.

Orientador: Prof. Dr. Romis Ribeiro de Faissol Attux
Co-orientadora: Profa. Dra. Aline de Oliveira Neves Panazio

Este exemplar corresponde à versão final da dissertação defendida pelo aluno Denis Gustavo Fantinato, e orientada pelo Prof. Dr. Romis Ribeiro de Faissol Attux

Campinas
2013

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

F218e Fantinato, Denis Gustavo, 1985-
Equalização não supervisionada : contribuições ao estudo do critério do módulo constante e de métodos baseados na teoria da informação / Denis Gustavo Fantinato. – Campinas, SP : [s.n.], 2013.

Orientador: Romis Ribeiro de Faissol Attux.
Coorientador: Aline de Oliveira Neves Panazio.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Processamento de sinais. 2. Filtros adaptativos. 3. Teoria da informação. I. Attux, Romis Ribeiro de Faissol, 1978-. II. Neves Panazio, Aline de Oliveira. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Blind Equalization : contributions to the study of the constant modulus criterion and of the information-theoretic Learning methods

Palavras-chave em inglês:

Signal processing

Adaptive filters

Information theory

Área de concentração: Engenharia de Computação

Titulação: Mestre em Engenharia Elétrica

Banca examinadora:

Romis Ribeiro de Faissol Attux [Orientador]

Maria das Dores dos Santos Miranda

João Marcos Travassos Romano

Data de defesa: 01-11-2013

Programa de Pós-Graduação: Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE MESTRADO

Candidato: Denis Gustavo Fantinato

Data da Defesa: 1 de novembro de 2013

Título da Tese: "Equalização Não Supervisionada: Contribuições ao Estudo do Critério do Módulo Constante e de Métodos Baseados na Teoria da Informação"

Prof. Dr. Romis Ribeiro de Faissol Attux (Presidente): _____
Profa. Dra. Maria das Dores dos S. Miranda: _____
Prof. Dr. João Marcos Travassos Romano: _____

Resumo

Este trabalho parte da análise de duas abordagens distintas no âmbito do problema de equalização não supervisionada. Na primeira delas, revisita-se alguns trabalhos na literatura sobre a formulação polinomial do critério do módulo constante (CM) e suas aplicações, visando estender o alcance dessas contribuições em diversos aspectos. Propõe-se então: a investigação do limitante inferior do custo CM no domínio complexo; o uso de uma nova heurística de inicialização para o algoritmo do CM; e, por fim, uma técnica baseada no método dos multiplicadores de Lagrange para lidar com as restrições existentes. Na segunda abordagem, investigam-se três critérios cegos pertencentes ao ramo de aprendizagem baseada na teoria da informação ou ITL (do inglês *Information-Theoretic Learning*), uma linha que busca extrair as informações estatísticas presentes nos sinais de interesse em maior completude. Analisam-se diversos aspectos, como função custo, superfície de erro e desempenho de equalização, revelando as conexões entre essas três metodologias. De maneira complementar, explora-se também os elementos em comum à formulação clássica do critério CM.

Palavras-chave: Equalização Cega, Filtragem Adaptativa, Critério do Módulo Constante, Filtros Polinomiais, Inicialização do Algoritmo do Módulo Constante, Método dos Multiplicadores de Lagrange, Aprendizado Baseado na Teoria da Informação.

Abstract

This work presents an analysis for the unsupervised equalization problem that is twofold. In the first one, two studies in the literature concerning about the polynomial formulation of the constant modulus (CM) cost function and its applications are revisited, with the goal of extending the scope of these contributions. Based on this, we propose: the investigation of the CM lower bound in the complex domain; the use of a new initialization heuristic for CM algorithm; and, finally, a technique based on the method of Lagrange Multipliers to deal with the existing restrictions in coefficients. In the second analysis, we investigate three blind criteria belonging to the class of Information-Theoretic Learning (ITL), a branch that is characterized by the use of the statistics present in the signals of interest in a more extensive way. The analysis is carried in terms of their cost functions, error surfaces and equalization performance, revealing the connections between these three methodologies. In a complementary way, we also explore the elements that share similarities to the classic formulation of the CM criterion.

Key-words: Blind Equalization, Adaptive Filtering, Constant Modulus Criterion, Polynomial Filters, Initialization of the Constant Modulus Algorithm, Method of Lagrange Multipliers, Information Theoretic Learning.

Sumário

Introdução Geral	1
1 Fundamentos da Filtragem Adaptativa	5
1.1 Filtros Lineares Adaptativos	6
1.2 Equalização de Canais de Comunicação	7
1.3 Equalização Supervisionada	9
1.3.1 Filtragem de Wiener	10
1.3.2 Algoritmos Adaptativos	12
1.4 Equalização Não Supervisionada	17
1.4.1 Técnicas de Bussgang	18
1.4.2 O Critério do Módulo Constante	21
1.4.3 O Critério de Shalvi-Weinstein	23
1.5 Aprendizado Baseado na Teoria da Informação	24
1.5.1 Estimadores de Densidade Não Paramétricos	25
1.5.2 Os Critérios Baseados em Teoria da Informação	27
1.6 Conclusão	30
2 Formulação Polinomial do Critério do Módulo Constante	33
2.1 Formulação Polinomial	34
2.1.1 Formulação Com Restrições	34
2.1.2 Formulação Sem Restrições	39
2.2 Limitante Inferior para o Critério CM	40
2.2.1 Índice de Equalizabilidade	41
2.2.2 Resultados - Caso no Domínio dos Reais	43
2.2.3 Resultados - Caso no Domínio dos Complexos	49

2.3	Inicialização do Algoritmo do Módulo Constante	52
2.3.1	Heurísticas de Inicialização	53
2.3.2	Resultados	55
2.4	Conclusão	59
3	Técnica para Ajuste dos Parâmetros das Restrições Quadráticas	61
3.1	Formulação Polinomial do Critério CM Usando Restrições Quadráticas . . .	62
3.1.1	Equalizador de Primeira Ordem	62
3.1.2	Equalizador de Ordem N	65
3.2	Técnica de Obtenção dos Coeficientes de Lagrange	67
3.2.1	Identificação do Coeficiente Ótimo de Lagrange	67
3.2.2	Técnica Baseada no Método do Gradiente	77
3.2.3	Resultados	80
3.3	Conclusão	85
4	Critérios Cegos Baseados na Teoria da Informação	87
4.1	Critérios Selecionados	88
4.1.1	Entropia da Dispersão de Ordem p	88
4.1.2	Casamento de PDF's por Distância Quadrática	90
4.1.3	Casamento de PDF's por Correlação	92
4.2	Pontos de Contato Entre as Funções Custo	93
4.2.1	Conexões ao Critério CM	93
4.2.2	Relações entre os Critérios ITL	94
4.3	Análise das Superfícies de Erro	95
4.4	Análise Comparativa de Desempenho	98
4.4.1	Ajuste dos Parâmetros	99
4.4.2	Desempenho da Equalização	101
4.5	Conclusão	107
5	Conclusões e Perspectivas	109
	Bibliografia	113
A	Equação Normal Obtida por Diferenciação Complexa	117
B	Função Custo Polinomial Usando Restrições Quadráticas	121

*Aos meus pais,
Célia e Wanderlei.*

Agradecimentos

Em primeiro lugar, gostaria de agradecer a Deus, cujos desígnios compreendiam a concretização de mais esse importante passo em minha vida.

Agradeço também aos meus pais, Wanderlei e Célia, que esforçaram-se incansavelmente para poder me oferecer uma boa formação; por todo o apoio, amor, paciência e compreensão durante esse caminho que venho trilhando. Esse importante passo em minha carreira só foi possível graças a eles. À minha irmã, Debora, pelo companheirismo e ajuda durante esse período. Sou muito grato a Deus por tê-los em minha vida.

À Raíssa, pelo grande amor e carinho demonstrados em todos os momentos; pela insubstituível companhia, pela paciência, compreensão e grande incentivo para continuar neste caminho.

Ao Prof. Romis Attux, que, além de um ótimo orientador, é um grande amigo. Sou muito grato por todo o tempo que dedicou a mim, sempre demonstrando grande empenho e competência nos trabalhos conduzidos, desde o início da iniciação científica; e, também, por me instruir não apenas no âmbito acadêmico, mas também em diversos aspectos da minha vida.

À minha co-orientadora, Profa. Aline Neves, pelas valiosas discussões e contribuições que enriqueceram este trabalho.

Ao Prof. Ricardo Suyama, por ter acompanhado de perto este trabalho e oferecido, sempre que possível, ideias inspiradoras.

À Profa. Maria Miranda e ao Prof. João Marcos Romano, por terem aceito participar da avaliação deste trabalho, desde a etapa de qualificação, e pelas sábias contribuições.

Aos amigos a quem tive o privilégio de trabalhar conjuntamente e que, em grande medida, auxiliaram em diversos aspectos durante a redação dessa dissertação, Daniel Guerreiro, Filipe Fazanaro, Kenji Nose Filho, Levy Boccato, Rafael Ferrari e Rafael Krummenauer.

À todos os amigos do DSPPCom, pela sempre frutífera e agradável companhia.

Aos amigos da graduação (Turma EE06 - Noturno).

À Celi, Noêmia, Edson, Mazé, e a todos funcionários da FEEC, pelo imprescindível apoio.

À CAPES, pelo apoio financeiro.

Persiga um sonho, mas não deixe ele viver sozinho.
Descubra-se todos os dias, deixe-se levar pelas vontades, mas não enlouqueça por elas.
Procure, sempre procure o fim de uma história, seja ela qual for.
Dê um sorriso para quem esqueceu como se faz isso.
Acelere seus pensamentos, mas não permita que eles te consumam.
Olhe para o lado, alguém precisa de você.
Abasteça seu coração de fé, não a perca nunca.
Mergulhe de cabeça nos seus desejos e satisfaça-os.
Procure os seus caminhos, mas não magoe ninguém nessa procura.
Arrependa-se, volte atrás, peça perdão!
Não se acostume com o que não o faz feliz, revolte-se quando julgar necessário.
Alague seu coração de esperanças, mas não deixe que ele se afogue nelas.
Se achar que precisa voltar, volte!
Se perceber que precisa seguir, siga!
Se estiver tudo errado, comece novamente.
Se estiver tudo certo, continue.
Se sentir saudades, mate-a.
Se perder um amor, não se perca!
Se achá-lo, segure-o!

Fernando Pessoa

Lista de Figuras

1.1	Filtro Linear de Resposta ao Impulso Finita (FIR).	6
1.2	Diagrama de blocos para o sistema de comunicação.	8
2.1	Sistema no domínio de Volterra - Combinador quadrático e filtro polinomial.	36
2.2	Relação entre os coeficientes ótimos associados aos critérios CM, Wiener e LI.	43
2.3	Comparação entre os valores mínimos dos custos CM e Wiener e o limitante inferior $J_{LI}(\boldsymbol{\theta}_o)$ do custo CM para canais somente com polos.	44
2.4	Comparação entre os valores mínimos dos custos CM e Wiener e o limitante inferior $J_{LI}(\boldsymbol{\theta}_o)$ do custo CM para sinais com modulação BPSK.	46
2.5	Comparação entre os valores mínimos dos custos CM e Wiener e o limitante inferior do custo CM $J_{LI}(\boldsymbol{\theta}_o)$ para modulações 4- e 8-PAM.	48
2.6	Constelação do sinal $ y(n) ^2$ supondo modulação 4-PAM.	48
2.7	Comparação entre os valores mínimos dos custos CM e Wiener e o limitante inferior do custo CM $J_{LI}(\boldsymbol{\theta}_o)$ para para sinais com modulação 4-QAM. . . .	50
2.8	Comparação entre os valores mínimos dos custos CM e Wiener e o limitante inferior do custo CM $J_{LI}(\boldsymbol{\theta}_o)$ para para sinais com modulação 16-QAM. . . .	52
3.1	Análise de $G_s(\boldsymbol{\theta}_\lambda)$ e soluções para λ em um canal $H(z) = 1 + 0,6z^{-1}$	70
3.2	Relação entre os coeficientes no domínio original do equalizador e as correspondentes restrições.	74
3.3	Superfície e curvas de nível de $G_s(\boldsymbol{\theta}_\lambda)$ e $D_s(\boldsymbol{\theta}_\lambda)$ fazendo-se $\boldsymbol{\lambda}$ variável.	76
3.4	$J_G(\boldsymbol{\theta}_\lambda)$ para o canal $H(z) = 1 + 0,6z^{-1}$ e equalizadores de ordem $N = 1$ e $N = 2$	78
3.5	Convergência de $\lambda(n)$ à $\lambda_{(1)}$ para o canal $H(z) = 1 + 0,6z^{-1}$	80
3.6	Convergência dos parâmetros $\boldsymbol{\lambda}$ e dos coeficientes \mathbf{w}_λ associados, considerando 3 e 2 restrições.	83

3.7	Convergência dos parâmetros λ para o canal $H_1(z) = 1 + 0,989z^{-1}$	84
4.1	Curvas de nível das funções custo - canal $H(z) = 1 + 0,6z^{-1}$	96
4.2	Kernel Gaussiano para $\sigma = 5$ e $\sigma = 1$	100
4.3	Desempenho dos algoritmos em termos de IIS.	102
4.4	Desempenho dos algoritmos em termos de IIS para diferentes tipos de ruído.	104

Lista de Tabelas

2.1	Taxa de convergência global e valor médio do custo CM para canais somente com polos.	56
2.2	Taxa de convergência global e valor médio do custo CM para canais de primeira ordem.	57
2.3	Taxa de convergência global e valor médio do custo CM para canais de segunda ordem.	58
2.4	Taxa de convergência global e valor médio do custo CM para canais de quinta ordem.	58
2.5	Taxa de convergência global e valor médio do custo CM para canais de segunda ordem e modulação 4-PAM.	59
2.6	Taxa de convergência global e valor médio do custo CM para canais de segunda ordem e modulação 8-PAM.	59
3.1	Taxa de convergência para λ_o com equalizador original de ordem $N = 1$. . .	81
3.2	Valor do custo CM para \mathbf{w}_λ (com três e duas restrições) e para \mathbf{w}_{CMg}	82
3.3	Taxa de convergência para λ_o e valor médio do custo CM para equalizador de ordem $N = 2$ e canal de primeira ordem $H_1(z)$	84
3.4	Taxa de convergência para λ_o e valor médio do custo CM para equalizador de ordem $N = 2$ e canal de segunda ordem $H_2(z)$	85
4.1	Variância dos coeficientes \mathbf{w} após a convergência dos algoritmos.	101
4.2	Tamanhos de passo μ para o canal $H(z) = 1 + 0,6z^{-1}$	101

Lista de Acrônimos e Notação

BGR	Benveniste-Goursat-Ruget
BPSK	<i>Binary Phase-Shift Keying</i> (Modulação por Deslocamento Binário de Fase)
CM	Módulo Constante
CMA	Algoritmo do Módulo Constante
DD	Algoritmo de decisão direta
EOS	Estatísticas de Ordem Superior
EQM	Erro Quadrático Médio
FIR	<i>Finite Impulse Response</i> (Resposta ao Impulso Finita)
HTC	Heurística dos Termos Cruzados
HTQ	Heurística dos Termos Quadráticos
ICA	<i>Independent Component Analysis</i> (Análise de Componentes Independentes)
<i>i.i.d.</i>	independente e identicamente distribuído
IIR	<i>Infinite Impulse Response</i> (Resposta ao Impulso Infinita)
IIS	Interferência Intersimbólica
ITL	<i>Information-Theoretic Learning</i> (Aprendizado Baseado na Teoria da Informação)
LI	Limitante Inferior
LMS	Algoritmo <i>Least-Mean-Square</i>
LS	<i>Least-Squares</i> (Mínimos Quadrados)
MQD	<i>Modified Quadratic Distance</i> (Distância Quadrática Modificada)

MSQD	<i>Modified SQD</i> (SQD Modificado)
PAM	<i>Pulse Amplitude Modulation</i> (Modulação por Amplitude de Pulso)
PDF	<i>Probability Density Function</i> (Função Densidade de Probabilidade)
QAM	<i>Quadrature Amplitude Modulation</i> (Modulação por Amplitude em Quadratura)
QD	<i>Quadratic Distance</i> (Distância Quadrática)
RD	<i>Rényi's entropy of the p-order Dispersion</i> (Entropia de Rényi da dispersão de ordem p)
RLS	Algoritmo <i>Recursive Least-Squares</i>
SNR	<i>Signal-to-Noise Ratio</i> (Relação Sinal-Ruído)
SQD	<i>Stochastic gradient algorithm for QD</i> (Algoritmo do Gradiente Estocástico para QD)
SRD	<i>Stochastic gradient algorithm for RD</i> (Algoritmo do Gradiente Estocástico para RD)
SW	Shalvi-Weinstein
VA	Variável Aleatória
ZF	<i>Zero-Forcing</i>

a	simbologia de um escalar
\mathbf{a}	simbologia de um vetor
\mathbf{A}	simbologia de uma matriz
\mathbf{I}	matriz identidade
$(\cdot)^*$	complexo conjugado
$(\cdot)^T$	transposição de um vetor ou matriz
$(\cdot)^H$	transposição hermitiana de um vetor ou matriz
$E[\cdot]$	operador esperança
$f(\cdot), g(\cdot)$	função
$\nabla f(\cdot)$	vetor gradiente de $f(\cdot)$
$(\hat{\cdot})$	estimativa de um escalar, vetor ou matriz
$\ \mathbf{a}\ $	norma euclidiana do vetor \mathbf{a}
\mathcal{R}	conjunto dos números reais
\mathcal{C}	conjunto dos números complexos
$\log(\cdot)$	função logarítmica
$p(\cdot)$	função densidade de probabilidade

Introdução Geral

Na história da humanidade, a comunicação desempenhou um papel fundamental para o desenvolvimento do que chamamos de sociedade. No século XIX, em especial, algumas marcantes invenções, como o telégrafo e o telefone, abriram caminho para uma intensa transformação no mundo. Nos dias atuais, os sistemas de comunicações digitais puderam propiciar um aumento na capacidade de transmissão e uma maior confiabilidade, permitindo o uso de tecnologias com grande fluxo de informações, como dados, imagens e áudio, o que revolucionou o modo de viver do ser humano de hoje. Mesmo assim, tais sistemas ainda precisam lidar constantemente com os desafios inerentes à comunicação.

De maneira sucinta, o objetivo de um sistema de comunicação é o de assegurar o fluxo de informação entre interlocutores através de um meio de transmissão ou canal. Contudo, a presença do canal pode introduzir distorções na informação transmitida, prejudicando a correta interpretação, por parte de um receptor, da mensagem ou sinal enviado. Nesse sentido, o emprego de um filtro denominado *equalizador*, capaz de mitigar, parcial ou completamente, os efeitos causados pela ação do canal é uma solução de grande relevância [Haykin, 1996]. Em aplicações práticas nas quais o canal não é conhecido ou, de alguma forma, é variante no tempo, torna-se pertinente que a equalização seja realizada de maneira *adaptativa*, isto é, que os parâmetros da estrutura do equalizador sejam passíveis de um ajuste iterativo segundo algum critério de otimização pré-estabelecido.

Do ponto de vista do receptor, a disponibilidade ou a ausência de um sinal de referência (ou, alternativamente, sequência de treinamento) caracteriza o tipo de metodologia de equalização, originando a distinção entre abordagens *supervisionadas* e *não supervisionadas*, respectivamente. Em ambos os casos, a adequada escolha do critério de otimização associado à metodologia empregada constitui um ponto chave no processo de equalização, pois lhe compete – através da definição de uma função custo – quantificar, em termos estatísticos, o desempenho associado à operação do dispositivo de filtragem.

No caso supervisionado, emerge como abordagem paradigmática aquela baseada no *critério de Wiener* [Haykin, 1996; Romano et al., 2010]. Este baseia-se na minimização da função custo definida pelo erro quadrático médio (EQM) entre a saída atual do equalizador e o sinal de treinamento. Sua importância se deve, em grande medida, à elegante e intuitiva formulação advinda do uso de estatísticas de ordem dois, a qual, ademais, é bastante adequada do ponto de vista de distorções lineares. A partir da formulação de Wiener e da noção associada de mínimos quadrados, surgiram os algoritmos LMS e RLS, ferramentas que podem ser consideradas canônicas.

Por outro lado, metodologias não supervisionadas ou cegas, por não utilizarem um sinal de referência durante o ajuste do equalizador, não exigem do transmissor o envio de um sinal já conhecido, o que é bastante conveniente, pois permite que a taxa transmitida de informação útil seja maior que aquela em relação ao caso supervisionado. Contudo, devido a essa ausência da sequência de treinamento, torna-se necessário ao equalizador recorrer à outros tipos de informações a respeito do sinal transmitido, que devem ser conhecidos *a priori* no processo de filtragem. Particularmente, uma abordagem que se mostrou bastante frutífera foi aquela baseada na informação trazida pelas estruturas estatísticas dos sinais transmitido e recebido. Nessa linha, os primeiros critérios não supervisionados propostos na literatura são aqueles que pertencem à família de técnicas de Bussgang [Haykin, 1996], dentre as quais podemos citar como exemplos os algoritmos de decisão direta (DD) [Lucky, 1965], de Sato [Sato, 1975] e do módulo constante (CMA, do inglês *Constant Modulus Algorithm*) [Godard, 1980]. Este último, em particular, tem sido amplamente estudado em vários contextos teóricos e práticos ao longo das três últimas décadas [Ding et al., 1991; Johnson Jr. et al., 1998], e constituirá um dos focos deste trabalho.

A grande motivação para os estudos do critério do módulo constante (ou, simplesmente, CM) pode ser atribuída à necessidade de uma maior fundamentação teórica que contribuisse para que a aplicação desse algoritmo fosse mais robusta e confiável. Contudo, esses estudos foram conduzidos de forma lenta e gradual, o que se justifica à luz, por exemplo, da dificuldade de se avaliar sua função custo que é não quadrática e que, por conseguinte, possui múltiplos mínimos. Mesmo assim, os resultados alcançados se revelaram muito interessantes, sendo possível estabelecer, num âmbito comparativo, conexões com o critério de Shalvi-Weinstein [Ding et al., 1991; Shalvi e Weinstein, 1993; Regalia, 1999] e também com o critério de Wiener [Zeng et al., 1998, 1999; Schniter e Johnson Jr., 2000]. Em especial, este último caso contribuiu para a proposição de uma formulação polinomial do critério CM [Sousa Júnior, 2011; Suyama, 2003], capaz de explicitar algumas estatísticas de maior

ordem presentes no custo CM.

Um dos objetivos deste trabalho é analisar a formulação polinomial do critério CM e suas aplicações, visando estender, sempre que possível, o alcance dessas contribuições. Serão avaliados o limitante inferior do custo CM, a heurística de inicialização para o CMA e o uso das restrições existentes através do método dos multiplicadores de Lagrange.

Além disso, mantendo o escopo não supervisionado, o segundo objetivo desse trabalho consiste em investigar alguns critérios cegos pertencentes ao recente ramo de aprendizagem baseada na teoria da informação ou ITL (do inglês *Information-Theoretic Learning*). Dessa linha, que busca extrair as informações estatísticas presentes nos sinais de interesse de maneira mais extensiva, participarão três critérios em uma análise comparativa, que, em termos de função custo, superfície de erro e desempenho de equalização, buscará estabelecer algumas conexões entre essas metodologias, e, de maneira complementar, explicitar alguns elementos “herdados” da formulação clássica do critério CM.

Estrutura do Trabalho

Esta dissertação está organizada da seguinte maneira:

- Capítulo 1: é feita uma recapitulação do problema de equalização de canais de comunicação, tanto em um âmbito supervisionado como não supervisionado. Também são apresentados alguns conceitos do ramo de aprendizado baseado na teoria da informação.
- Capítulo 2: resgata-se o conceito da formulação polinomial do critério CM, introduzindo, como contribuição, sua derivação para caso inserido no domínio dos complexos. Em seguida, é avaliado o limitante inferior do custo CM para uma ampla gama de cenários, inclusive para modulações complexas e de módulo não constante (ou multinível). Por fim, é proposta uma nova heurística de inicialização do CMA, cuja análise compreenderá comparações com outras metodologias existentes, como a center-spike.
- Capítulo 3: a formulação polinomial do critério CM usando restrições quadráticas através do método dos multiplicadores de Lagrange é apresentada. Em vista da complexidade do problema, propõe-se o uso de uma técnica que facilita a obtenção dos valores desejados para esses multiplicadores.
- Capítulo 4: inicialmente, são apresentados três critérios encontrados na literatura e pertencentes à classe de ITL. Em seguida, tais critérios, juntamente com o CM, são

comparados em termos de função custo, superfície de erro e desempenho de equalização.

- Capítulo 5: são apresentadas as conclusões gerais sobre o trabalho e as perspectivas futuras.

Fundamentos da Filtragem Adaptativa

Os filtros adaptativos foram foco de expressiva atenção como objeto de estudo ao longo de quase um século, o que contribuiu para lhes conferir o título de ferramenta fundamental e também sua consolidação como parte importante na área de processamento estatístico de sinais.

Em situações onde se exige o processamento de sinais resultantes de uma operação em cenários cujas propriedades estatísticas são desconhecidas, o emprego de filtros adaptativos surge como solução bastante atrativa ao problema, principalmente por prover, na maioria dos casos, um aprimoramento do desempenho sobre o uso de filtros fixos projetados por métodos convencionais. Além disso, os filtros adaptativos possuem amplo leque de aplicações em áreas como as de comunicações, controle, radar, sonar e sísmica [Romano et al., 2010].

No presente trabalho, concentraremos nossos esforços no problema específico de equalização de canais de comunicação, sob a conjectura de que os modelos de canal e filtro são discretos no tempo, sendo as amostras do sinal recebido uniformemente amostradas. Consideraremos ainda o caso genérico de séries temporais no domínio complexo, visto que, em muitas situações práticas, sinais de informação podem nele ser expressos de maneira vantajosa.

Este capítulo apresenta brevemente os fundamentos da filtragem adaptativa aplicada ao problema de equalização de canais, estabelecendo, em linhas gerais, o arcabouço teórico que figura como pano de fundo para o restante da dissertação. Serão explicitadas ainda as questões que atuam como motivação ao estudo mais aprofundado sobre os critérios envolvidos.

1.1 Filtros Lineares Adaptativos

O termo *filtro* está associado à estrutura de um dispositivo físico (hardware) ou de um software computacional capaz de desempenhar a extração de informação sobre uma quantidade de interesse usando dados medidos até o tempo presente [Haykin, 1996]. É dito linear se sua relação entrada-saída $f(\cdot)$ obedece ao princípio da superposição, ilustrado a seguir:

$$y = f(ax_1 + bx_2) = af(x_1) + bf(x_2). \quad (1.1)$$

O dispositivo linear emblemático que porta tal característica é o filtro de resposta ao impulso finita (FIR, do inglês *Finite Impulse Response*), em que um leque finito de amostras presentes e passadas de um sinal são ponderadas a fim de se obterem dados relevantes. Sua estrutura é ilustrada na Fig. 1.1.

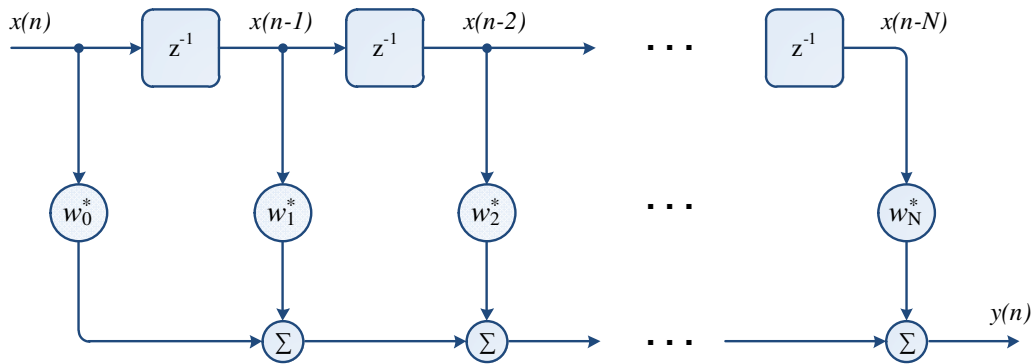


Figura 1.1: Filtro Linear de Resposta ao Impulso Finita (FIR).

Matematicamente, podemos expressar a saída de um filtro FIR de ordem N – dotado, portanto, de $N+1$ coeficientes – usando notação vetorial, da seguinte forma:

$$y(n) = \begin{bmatrix} w_0^* & w_1^* & \cdots & w_N^* \end{bmatrix} \begin{bmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-N) \end{bmatrix} \quad (1.2)$$

$$= \mathbf{w}^H \mathbf{x}(n),$$

sendo \mathbf{w} o vetor de coeficientes do filtro, $\mathbf{x}(n)$ seu vetor de entrada e $y(n)$ o sinal de saída; o sobrescrito H denota transposição *Hermitiana* (ou seja, conjugação complexa combinada à transposição), e o asterisco $*$ denota conjugação. Comumente, encontramos na literatura a descrição de um filtro linear por meio de sua função de transferência, que, neste caso, fica:

$$W(z) = w_0^* + w_1^* z^{-1} + \cdots + w_N^* z^{-N}. \quad (1.3)$$

Em muitas situações, os sistemas em que os filtros estão inseridos possuem características variáveis com o tempo ou são, até mesmo, desconhecidos. Torna-se imprescindível nestes casos que os coeficientes do filtro \mathbf{w} sejam passíveis de ajuste, ou, em outras palavras, que um *filtro adaptativo* seja empregado. Nas próximas seções veremos como proceder com esse ajuste.

Em termos mais gerais, a escolha de um dispositivo de filtragem não precisa, necessariamente, limitar-se a estruturas do tipo FIR: podem ser empregados também filtros de distintas características, como recursão, não linearidade ou a combinação destas. Por não serem foco deste trabalho, destacamos apenas alguns deles, como, por exemplo, os filtros com resposta ao impulso infinita (IIR, do inglês *Infinite Impulse Response*), cujas funções de transferência tem a seguinte forma:

$$W(z) = \frac{B(z)}{1 - A(z)} = \frac{b_0^* + b_1^*z^{-1} + \dots + b_N^*z^{-N}}{1 - a_1^*z^{-1} - a_2^*z^{-2} - \dots - a_M^*z^{-M}}, \quad (1.4)$$

em que N é a ordem da parte direta (ou *forward*) do filtro IIR – i.e., com N zeros em sua função de transferência – e M é a ordem da parte recorrente (ou dotada de *feedback*). Outras possibilidades são os filtros não lineares polinomiais [Mathews e Sicuranza, 2000] e as redes neurais [Haykin, 1998].

Em suma, independentemente da estrutura, vigora nos filtros, em geral, como característica chave, sua capacidade intrínseca de extrair diversos tipos de informações, conferindo-lhes uma ampla gama de aplicações. Em especial, trataremos aqui do problema de equalização de canais de comunicação.

1.2 Equalização de Canais de Comunicação

A aplicação dos filtros ao problema de equalização de canais pode ser mais bem compreendida no contexto de modelos de *sistemas de comunicação*. Basicamente, em um sistema de comunicação, o fluxo de informação entre os interlocutores se dá através de um meio de transmissão ou canal que, por sua vez, pode introduzir distorções na informação transmitida, prejudicando a correta interpretação, por parte de um receptor, da mensagem ou sinal enviado. A fim de mitigar, parcial ou completamente, os efeitos causados pela ação do canal, adota-se, tipicamente, um procedimento de *equalização*, por meio da utilização de filtros lineares ou não lineares, que também podem receber o nome de *equalizadores*. A Fig. 1.2 ilustra o diagrama de blocos para o sistema de comunicação [Haykin, 1996; Romano et al., 2010].

Embora o canal possa representar diferentes entidades, como um cabo coaxial, a atmosfera ou uma fibra ótica, geralmente este é modelado matematicamente como um filtro do tipo FIR ou IIR (por meio de sua função de transferência $H(z)$) que atua sobre o sinal

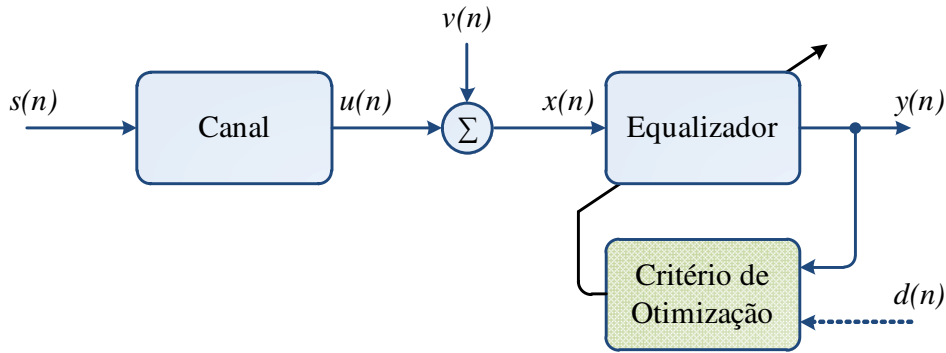


Figura 1.2: Diagrama de blocos para o sistema de comunicação.

transmitido $s(n)$. Não obstante, o sinal resultante $u(n)$ ainda se encontra sujeito a uma possível adição de ruído $v(n)$ – que pode ser descrito segundo uma distribuição, como a Gaussiana, a impulsiva, a Laplaciana, etc. – gerado, por exemplo, no *front end* do receptor ou como interferência de outras fontes. Ao equalizador cabe, portanto, da versão degradada $x(n)$, gerar uma versão $y(n)$ tão fiel quanto possível do sinal original transmitido $s(n)$.

À luz do que foi descrito, destacamos como principais metas associadas ao emprego de um equalizador o combate ao efeito dispersivo dos canais, fenômeno denominado *interferência intersimbólica* (IIS), e à atenuação de ruído.

De um ponto de vista teórico, admite-se que o sinal equalizado $y(n)$ forneça, idealmente, uma versão de $s(n)$ sujeito a variações de escala a e atraso k :

$$y(n) = as(n - k). \quad (1.5)$$

Se este resultado for alcançado, terá havido de fato a completa equalização, e o filtro projetado será uma inversa perfeita do canal.

Este conceito pode ser melhor entendido se considerarmos a resposta combinada canal+equalizador $c(n) = \sum_{i=-\infty}^{\infty} w_i^* h^*(n - i)$, que nada mais é do que a convolução entre o canal $h(n)$ e o vetor de coeficientes \mathbf{w} do filtro (onde foi considerado hipoteticamente um comprimento infinito tanto para a parte causal como não causal). Supondo que não há ruído aditivo $v(n)$, podemos escrever $y(n)$ como

$$\begin{aligned} y(n) &= \mathbf{w}^H \mathbf{x}(n) = \sum_{i=-\infty}^{\infty} w_i^* x(n - i) \\ &= \sum_{i=-\infty}^{\infty} w_i^* \left(\sum_{j=-\infty}^{\infty} h^*(j) s(n - i - j) \right), \end{aligned} \quad (1.6)$$

fazendo $j = l - i$, temos

$$\begin{aligned}
 y(n) &= \sum_{i=-\infty}^{\infty} w_i^* \left(\sum_{l=-\infty}^{\infty} h^*(l-i)s(n-l) \right) \\
 &= \sum_{l=-\infty}^{\infty} \left(\sum_{i=-\infty}^{\infty} w_i^* h^*(l-i) \right) s(n-l) \\
 &= \sum_{l=-\infty}^{\infty} c(l)s(n-l)
 \end{aligned} \tag{1.7}$$

para que se alcance plenamente a solução almejada em (1.5), $\mathbf{c} = \{c(n)\}$ deverá ser um vetor com apenas um elemento não nulo igual a a , isto é,

$$\mathbf{c} = \left[0 \quad \dots \quad 0 \quad a \quad 0 \quad \dots \quad 0 \right]. \tag{1.8}$$

Nessa situação, temos a *condição Zero-Forcing* (ZF) [Haykin, 1996], cujo nome deriva do fato do equalizador “forçar” \mathbf{c} a ser zero em todos os instantes exceto naquele de interesse.

Vale ressaltar que nem sempre a condição ZF poderá ser satisfeita. Na maioria dos casos, os equalizadores empregados apresentam limitações estruturais que inviabilizam sua atuação como inversa perfeita do canal. Contudo, os esforços podem ser direcionados para mitigar a IIS tanto quanto possível.

De qualquer forma, o sucesso da equalização dependerá exclusivamente do correto ajuste dos parâmetros do filtro. Apesar da condição ZF ser um indicativo deste ajuste desejado, na prática, o não conhecimento do canal somado, muitas vezes, à impossibilidade de se alcançar a condição ZF, inviabiliza o uso direto desta abordagem. Desta forma, tipicamente, esta tarefa é desempenhada através do auxílio de um *critério de otimização* que, por sua vez, baseia-se na definição de uma função custo, isto é, uma métrica que busca quantificar o desempenho desejado do equalizador, tornando possível lançar mão de uma abordagem estatística para melhor caracterizar o problema. Em face dessa questão, o problema de equalização suscita duas principais vertentes: equalização supervisionada e equalização não supervisionada.

1.3 Equalização Supervisionada

O processo de equalização supervisionada tem como elemento distintivo a mandatória presença de uma sequência de referência $d(n)$ (vide Fig. 1.2) utilizada para guiar o ajuste dos coeficientes do filtro. De maneira mais específica, nessa abordagem, o processo de adaptação do equalizador ocorre durante um certo período, chamado período de treinamento, no qual a sequência $d(n)$ - conhecida *a priori* pelo receptor - é transmitida. O sinal recebido $x(n)$

portando os efeitos inerentes do canal é então comparado com a sequência de referência $d(n)$ no receptor, cuja informação pode ser utilizada na adaptação do filtro.

Seguindo esta linha de raciocínio, emerge nesse âmbito um dos maiores pilares da teoria de filtragem supervisionada, o paradigmático *critério de Wiener* [Haykin, 1996; Romano et al., 2010].

1.3.1 Filtragem de Wiener

O emblemático critério de Wiener tem origem nos trabalhos independentes de Kolmogorov e Wiener, de 1941 e 1942, respectivamente [Haykin, 1996]. Baseia-se na minimização de uma função custo, o erro quadrático médio (EQM) entre a saída atual do equalizador $y(n)$ e o sinal de treinamento $d(n)$:

$$J_{Wiener}(\mathbf{w}) = E[|e(n)|^2] = E[|d(n) - y(n)|^2], \quad (1.9)$$

sendo $E[\cdot]$ o operador esperança estatística e $d(n)$ o sinal de referência, que, em equalização, pode ser considerado igual a $s(n-k)$, o sinal transmitido com atraso k constante.

Considerando os sinais envolvidos no sistema como processos estocásticos estacionários, é possível definir as medidas de segunda ordem que relacionam-se diretamente com a função custo (1.9). Seja o vetor de entrada do filtro FIR de ordem N dado por $\mathbf{x}(n) = [x(n) \ x(n-1) \ \dots \ x(n-N)]$, a *matriz de autocorrelação* \mathbf{R}_x pode ser escrita como

$$\mathbf{R}_x = E[\mathbf{x}(n)\mathbf{x}^H(n)] = \begin{bmatrix} r(0) & r(1) & \dots & r(N) \\ r(-1) & r(0) & \dots & r(N-1) \\ \vdots & \vdots & \ddots & \vdots \\ r(-N) & r(-N+1) & \dots & r(0) \end{bmatrix}, \quad (1.10)$$

em que $r(k) = E[x(n)x^*(n-k)]$ é a função de autocorrelação. Ainda, podemos definir também o *vetor de correlação cruzada* \mathbf{p}_x da seguinte maneira:

$$\mathbf{p}_x = E[\mathbf{x}(n)d^*(n)]. \quad (1.11)$$

A partir dessas definições, considerando $|e(n)|^2 = e(n)e^*(n)$, é possível expandir o custo

EQM dado por (1.9):

$$\begin{aligned}
J_{Wiener}(\mathbf{w}) &= E [|e(n)|^2] = E [e(n)e^*(n)] \\
&= E [(d(n) - \mathbf{w}^H \mathbf{x}(n)) (d^*(n) - \mathbf{x}(n)^H \mathbf{w})] \\
&= E [d(n)d^*(n)] - \mathbf{w}^H E [\mathbf{x}(n)d^*(n)] - E [d(n)\mathbf{x}^H(n)] \mathbf{w} \\
&\quad + \mathbf{w}^H E [\mathbf{x}(n)\mathbf{x}^H(n)] \mathbf{w} \\
&= \sigma_d^2 - \mathbf{p}_x^H \mathbf{w} - \mathbf{w}^H \mathbf{p}_x + \mathbf{w}^H \mathbf{R}_x \mathbf{w},
\end{aligned} \tag{1.12}$$

sendo σ_d^2 a variância de $d(n)$. A Eq. (1.12) revela que o custo $J_{Wiener}(\mathbf{w})$ é precisamente uma função de segunda ordem em relação ao vetor de coeficientes \mathbf{w} . Dada a típica positividade de \mathbf{R}_x , isto implica que existe um único ponto de mínimo para o paraboloide formado por $J_{Wiener}(\mathbf{w})$.

A relação quadrática possui a vantagem de conduzir a expressões relativamente simples no processo de cálculo de suas derivadas. De fato, a solução que minimiza (1.9) pode ser obtida analiticamente através da derivação de (1.12) em relação ao vetor complexo \mathbf{w} e igualando-se o resultado a zero [Haykin, 1996]. A esta solução atribui-se o nome *solução de Wiener*, que pode ser definida como (Apêndice A):

$$\mathbf{w}_o = \mathbf{R}_x^{-1} \mathbf{p}_x, \tag{1.13}$$

em que \mathbf{w}_o é o vetor de coeficientes ótimo no sentido de menor EQM.

Substituindo (1.13) em (1.12), chegamos a uma relação ainda mais simples:

$$\begin{aligned}
J_{Wiener}(\mathbf{w}_o) &= \sigma_d^2 - \mathbf{p}_x^H \mathbf{w}_o - \mathbf{w}_o^H \mathbf{p}_x + \mathbf{w}_o^H \mathbf{R}_x \mathbf{w}_o \\
&= \sigma_d^2 - \mathbf{p}_x^H \mathbf{R}_x^{-1} \mathbf{p}_x,
\end{aligned} \tag{1.14}$$

sendo $J_{Wiener}(\mathbf{w}_o)$ o mínimo custo EQM ou mínimo custo Wiener.

Vale ressaltar que a solução de Wiener (1.13) depende também do atraso k considerado na definição $d(n) = s(n-k)$, o que pode conduzir a soluções com diferentes valores residuais de EQM. A determinação *a priori* do atraso k que conduz à solução com menor custo mínimo em (1.14) ainda é um problema em aberto dentro deste ramo de pesquisa, e contribuirá como motivação à abordagem seguida neste trabalho.

A simplicidade que envolve a expressão da Eq. (1.13) é, sem dúvida, bastante atrativa. Contudo, principalmente quando são empregados equalizadores com muitos coeficientes, é subjacente à estimação do vetor de correlação cruzada \mathbf{p}_x e da matriz de correlação \mathbf{R}_x um considerável custo computacional, sobretudo para o cálculo da inversa \mathbf{R}_x^{-1} . Uma alternativa menos custosa para o problema pode ser estabelecida através do uso dos algoritmos adaptativos.

1.3.2 Algoritmos Adaptativos

Os algoritmos adaptativos se caracterizam por desempenharem um processo de otimização de forma iterativa. Além de apresentarem um custo computacional reduzido quando comparado, por exemplo, ao cálculo da solução de Wiener (1.13), estes algoritmos carregam também elementos que permitem a operação em cenários variantes. Nesse sentido, existem diversas metodologias de caráter iterativo que são capazes de produzir estimativas do filtro ótimo \mathbf{w}_o , como os algoritmos do gradiente determinístico, *Least-Mean-Square* (LMS) e *Recursive Least-Squares* (RLS) [Romano et al., 2010].

Algoritmo do Gradiente Determinístico

O ajuste iterativo dos coeficientes \mathbf{w} do filtro conferem-lhe um “estado” (de valores associados a \mathbf{w}) que pode diferir em cada instante de tempo: por isso, doravante o denominaremos $\mathbf{w}(n)$. Nesta abordagem, permite-se que seja feito um ajuste gradual em busca da solução desejada.

Com base nisso, o algoritmo do gradiente determinístico busca aplicar sucessivas correções sobre $\mathbf{w}(n)$ no sentido contrário ao do gradiente da função custo, isto é, na direção de sua maior declividade local (devido a isso, esse algoritmo também recebe o nome de *steepest descent* [Haykin, 1996; Romano et al., 2010]) com o intuito de alcançar o ponto de mínimo, onde os coeficientes do filtro assumem seu valor ótimo \mathbf{w}_o . Dessa forma, dada a função custo de Wiener com coeficientes atuais $\mathbf{w}(n)$, a atualização dos parâmetros do equalizador para o próximo estado $\mathbf{w}(n+1)$ pode ser expressa como:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \nabla J_{Wiener}(\mathbf{w}(n)), \quad (1.15)$$

onde $\nabla J_{Wiener}(\mathbf{w}(n))$ é o gradiente de $J_{Wiener}(\mathbf{w}(n))$ e μ é o chamado *passo de adaptação*. Alternativamente, μ pode ser denominado também *tamanho de passo*, pois controla a amplitude da correção incremental aplicada aos coeficientes do filtro, assumindo apenas valores reais e positivos, podendo ser fixo ou variável com o tempo. Uma escolha inadequada de seu valor pode provocar a divergência do algoritmo, ou seja, fazer com que este não alcance o ponto de mínimo [Haykin, 1996].

Tomando o gradiente de $J_{Wiener}(\mathbf{w}(n))$ com respeito aos coeficientes do filtro $\mathbf{w}(n)$ (os detalhes da derivação complexa podem ser conferidos no Apêndice A), chegamos à seguinte relação:

$$\nabla J_{Wiener}(\mathbf{w}(n)) = -2\mathbf{p}_x + 2\mathbf{R}_x \mathbf{w}(n). \quad (1.16)$$

Em suma, as Eqs. (1.15) e (1.16) definem o algoritmo do gradiente determinístico, cuja operação não depende de nenhuma inversão de matriz. Todavia, persiste a demanda por estimativas da matriz de correlação \mathbf{R}_x e do vetor de correlação cruzada \mathbf{p}_x , o que implica

em considerável processamento computacional. Como veremos a seguir, um passo adiante nesse problema virá do emprego de aproximações estocásticas.

Algoritmo LMS

O algoritmo *Least-Mean-Square* ou, simplesmente, LMS, é uma ferramenta amplamente usada em filtragem adaptativa linear, especialmente por sua combinação de robustez e simplicidade. Baseia-se na aproximação estocástica inicialmente proposta por [Robbins e Monro, 1951], cuja essência jaz na substituição das esperanças matemáticas presentes nas medidas de correlação por estimativas instantâneas baseadas em amostras do sinal de entrada $x(n)$ e do sinal desejado $d(n)$. Assim, as estimativas da matriz de correlação \mathbf{R}_x e do vetor de correlação cruzada \mathbf{p}_x ficam:

$$\begin{aligned}\mathbf{R}_x &\longrightarrow \hat{\mathbf{R}}_x = \mathbf{x}(n)\mathbf{x}^H(n) \\ \mathbf{p}_x &\longrightarrow \hat{\mathbf{p}}_x = \mathbf{x}(n)d^*(n).\end{aligned}\tag{1.17}$$

De posse dessas aproximações, é possível obter um novo tipo de algoritmo baseado nas relações do gradiente determinístico. Nesse caso, teremos, portanto, como estimativa instantânea do vetor gradiente

$$\hat{\nabla} J_{Wiener}(\mathbf{w}(n)) = -2\mathbf{x}(n)d^*(n) + 2\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{w}(n).\tag{1.18}$$

Substituindo o gradiente por sua estimativa instantânea (1.18) em (1.15), é possível obter uma relação recursiva para atualização dos coeficientes do filtro:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu (d^*(n) - \mathbf{x}^H(n)\mathbf{w}(n)) \mathbf{x}(n),\tag{1.19}$$

onde o fator multiplicativo 2 foi incluído na definição do passo de adaptação μ .

A forma complexa do algoritmo LMS foi originalmente proposta em [Widrow et al., 1975] e também é conhecida pelo nome de *algoritmo do gradiente estocástico*. Uma análise mais detalhada do seu comportamento revelou que as estimativas instantâneas de \mathbf{R}_x e \mathbf{p}_x portam variâncias relativamente altas, o que, a princípio, parece impactar de forma bastante negativa sobre o seu desempenho. O LMS, contudo, dada sua natureza recursiva, realiza, de certa maneira, uma média destas estimativas durante o processo iterativo, de forma que sua convergência estará ligada, majoritariamente, apenas à correta escolha do passo de adaptação μ [Macchi, 1996].

Vale ressaltar também que, por se tratar de uma aproximação estocástica baseada em valores instantâneos, o algoritmo não fixa sua convergência ao ponto de mínimo, mas oscila os parâmetros em torno desta solução (no caso, a de Wiener) - o chamado efeito de *misadjustment* [Haykin, 1996]. Isto implica em um valor residual EQM ligeiramente superior

àquele obtido pela solução analítica. Neste sentido, a aproximação do gradiente também cria, relativamente ao passo, um compromisso entre velocidade de convergência e flutuação em regime permanente.

Algoritmo RLS

A abordagem seguida na dedução do algoritmo do gradiente determinístico e LMS baseia-se na formulação de Wiener e no arcabouço teórico fundamentado em médias estatísticas. No entanto, esta não é a única abordagem possível a ser seguida na derivação dos algoritmos adaptativos.

O algoritmo *Recursive Least-Squares* (RLS), de maneira alternativa, busca a minimização do erro quadrático médio relativo às amostras de que se dispõe, i.e., orienta-se exclusivamente pelos dados disponíveis. Apesar de essa abordagem estabelecer relações com o caso estatístico, devem estar bem claras essas diferenças conceituais.

A base do método é uma versão recursiva da abordagem de mínimos quadrados (LS, do inglês *Least-Squares*), a qual permite uma atualização iterativa da solução ótima. Também é usado o lema de inversão de matrizes [Golub e Van Loan, 1996], que evita a necessidade de realizar a inversão da versão amostral da matriz de correlação. Sob algumas hipóteses de análise, é possível afirmar que o RLS possui uma taxa de convergência tipicamente de uma ordem de grandeza mais rápida que a do algoritmo LMS simples e, ademais, não experimenta o fenômeno de *misadjustment* [Haykin, 1996]. Entretanto, o aprimoramento no desempenho é conquistado à custa de um aumento na complexidade computacional.

À luz do que foi exposto, torna-se necessário a proposição de uma nova função custo, baseada na soma dos quadrados de um sinal de erro:

$$J_{LS}(\mathbf{w}) = \sum_{i=1}^L \lambda^{L-i} |e(i)|^2, \quad (1.20)$$

onde $e(n) = d(n) - y(n)$, L é o número de amostras disponível e λ é uma constante positiva menor ou igual a um, que desempenha o papel de fator de esquecimento, pois tende a reduzir o efeito de amostras antigas. Para $\lambda = 1$, não há esquecimento, i.e., todas as amostras são igualmente ponderadas, e, complementarmente, $\lambda = 0$ implica num caso sem memória, onde é utilizada apenas a amostra atual.

De maneira análoga à exposição do critério de Wiener, é possível definir, neste novo âmbito, as matrizes de correlação e o vetor de correlação cruzada, agora denominados, respectivamente, $\Phi(n)$ e $\pi(n)$:

$$\Phi(n) = \sum_{i=1}^L \lambda^{L-i} \mathbf{x}(i) \mathbf{x}^H(i); \quad (1.21)$$

$$\boldsymbol{\pi}(n) = \sum_{i=1}^L \lambda^{L-i} \mathbf{x}(i) d^*(i). \quad (1.22)$$

A partir dessas relações, podemos expandir (1.20), chegando a uma expressão muito similar à Eq. (1.12) no âmbito da filtragem de Wiener:

$$\begin{aligned} J_{LS}(\mathbf{w}) &= \sum_{i=1}^L \lambda^{L-i} e(i) e^*(i) \\ &= \sum_{i=1}^L \lambda^{L-i} (d(i) - \mathbf{w}^H \mathbf{x}(i)) (d^*(i) - \mathbf{x}(i)^H \mathbf{w}) \\ &= \left(\sum_{i=1}^L \lambda^{L-i} d^2(i) \right) - \boldsymbol{\pi}^H(n) \mathbf{w} - \mathbf{w}^H \boldsymbol{\pi}(n) + \mathbf{w}^H \boldsymbol{\Phi}(n) \mathbf{w}. \end{aligned} \quad (1.23)$$

Dadas as semelhanças com a abordagem de Wiener, não é de se espantar que o único ponto de mínimo da Eq. (1.23) tenha forma similar à da solução de Wiener (1.13). Usando a condição do gradiente nulo, chegamos em

$$\mathbf{w}_{LS} = \boldsymbol{\Phi}^{-1}(n) \boldsymbol{\pi}(n). \quad (1.24)$$

A solução dada por (1.24) tem um custo computacional relativamente alto, da mesma maneira que seu par sob a ótica de Wiener. Nesse sentido, torna-se oportuno representar o problema através de um algoritmo adaptativo, beneficiando-se de maior simplicidade computacional. Para isso, o primeiro passo é expressar $\boldsymbol{\Phi}(n)$ e $\boldsymbol{\pi}(n)$ de maneira recursiva.

É possível reescrever a Eq. (1.21) como

$$\begin{aligned} \boldsymbol{\Phi}(n) &= \sum_{i=1}^L \lambda^{L-i} \mathbf{x}(i) \mathbf{x}^H(i) = \lambda \left(\sum_{i=1}^{L-1} \lambda^{L-1-i} \mathbf{x}(i) \mathbf{x}^H(i) \right) + \mathbf{x}(n) \mathbf{x}^H(n) \\ &= \lambda \boldsymbol{\Phi}(n-1) + \mathbf{x}(n) \mathbf{x}^H(n), \end{aligned} \quad (1.25)$$

onde o produto matricial $\mathbf{x}(n) \mathbf{x}^H(n)$ atua como um termo de correção.

Similarmente, para o vetor de correlação cruzada $\boldsymbol{\pi}(n)$ dado em (1.22), temos

$$\begin{aligned} \boldsymbol{\pi}(n) &= \sum_{i=1}^L \lambda^{L-i} \mathbf{x}(i) d^*(i) = \lambda \left(\sum_{i=1}^{L-1} \lambda^{L-1-i} \mathbf{x}(i) d^*(i) \right) + \mathbf{x}(n) d^*(n) \\ &= \lambda \boldsymbol{\pi}(n-1) + \mathbf{x}(n) d^*(n). \end{aligned} \quad (1.26)$$

Para obter os coeficientes do filtro \mathbf{w}_{LS} de acordo com (1.24), resta ainda determinar a inversa da matriz de correlação $\boldsymbol{\Phi}(n)$. No entanto, este custoso processo pode ser evitado se lançarmos mão de um sofisticado resultado matemático: o lema de inversão de matrizes

[Golub e Van Loan, 1996], enunciado a seguir.

Considerando duas matrizes $M \times M$ definidas positivas \mathbf{A} e \mathbf{B} , vale a seguinte relação:

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^H, \quad (1.27)$$

onde \mathbf{D} é uma matriz $K \times K$ definida positiva e \mathbf{C} uma matriz $M \times K$.

O lema de inversão de matrizes afirma que a inversa de \mathbf{A} pode ser expressa como:

$$\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B}\mathbf{C}(\mathbf{D} + \mathbf{C}^H\mathbf{B}\mathbf{C})^{-1}\mathbf{C}^H\mathbf{B}. \quad (1.28)$$

Como se assume que a matriz de correlação $\Phi(n)$ é definida positiva e, portanto, não singular, o lema de inversão de matrizes é aplicável a este contexto. Comparando as Eqs. (1.25) e (1.27), é possível estabelecer a seguinte associação:

$$\begin{aligned} \mathbf{A} &= \Phi(n); \\ \mathbf{B}^{-1} &= \lambda\Phi(n-1); \\ \mathbf{C} &= \mathbf{x}(n); \\ \mathbf{D} &= 1. \end{aligned} \quad (1.29)$$

Usando estas definições no lema de inversão de matrizes (1.28), obtemos a seguinte equação recursiva para a inversa da matriz de correlação:

$$\Phi^{-1}(n) = \lambda^{-1}\Phi^{-1}(n-1) - \frac{\lambda^{-2}\Phi^{-1}(n-1)\mathbf{x}(n)\mathbf{x}^H(n)\Phi^{-1}(n-1)}{1 + \lambda^{-1}\mathbf{x}^H(n)\Phi^{-1}(n-1)\mathbf{x}(n)}. \quad (1.30)$$

Este cálculo iterativo, somado à algumas definições de variáveis auxiliares, finalmente nos conduz ao algoritmo RLS, mostrado no Alg. 1.1 [Haykin, 1996].

Quando comparado ao LMS, o algoritmo RLS se destaca por apresentar algumas peculiaridades. Em primeiro lugar, sua taxa de convergência não depende do condicionamento da matriz de correlação (i.e., a distribuição dos autovalores associados à matriz de correlação não causa impactos sobre a convergência), permitindo-lhe convergir de maneira mais rápida que o algoritmo LMS em diversos casos. Além disso, quando utiliza-se $\lambda = 1$ – se o cenário não muda suas características com o tempo, isto é, é estacionário –, o RLS possui a vantagem de ser mais robusto ao efeito de *misadjustment*, reduzindo cada vez mais o EQM residual à medida que o número de amostras disponíveis aumenta. Tal comportamento é justificado pelo fato de a média temporal sobre um grande número de amostras se aproximar do valor ideal provido pelo operador esperança estatística.

Algoritmo 1.1 Algoritmo RLS

Inicializar o algoritmo escolhendo:

$$\begin{aligned}\mathbf{P}(0) &= \delta^{-1}\mathbf{I}; \\ \mathbf{w}(0) &= \mathbf{0}.\end{aligned}$$

onde δ é um valor constante pequeno e positivo e \mathbf{I} é a matriz identidade.

A cada amostra recebida, atualizar os coeficientes do filtro de acordo com as seguintes equações:

$$\begin{aligned}\mathbf{k}(n) &= \frac{\lambda^{-1}\mathbf{P}(n-1)\mathbf{x}(n)}{1 + \lambda^{-1}\mathbf{x}^H(n)\mathbf{P}(n-1)\mathbf{x}(n)}; \\ \alpha(n) &= d(n) - \mathbf{w}^H(n-1)\mathbf{x}(n); \\ \mathbf{w}(n) &= \mathbf{w}(n-1) + \mathbf{k}(n)\alpha^*(n); \\ \mathbf{P}(n) &= \lambda^{-1}\mathbf{P}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{x}^H(n)\mathbf{P}(n-1).\end{aligned}$$

1.4 Equalização Não Supervisionada

Conforme visto na Seção 1.3, o emprego de uma sequência de referência como auxílio no processo de adaptação dos parâmetros do filtro equalizador caracteriza o que chamamos de modo supervisionado. Os desdobramentos dessa abordagem, como o emblemático critério de Wiener e os algoritmos LMS e RLS, não puderam, entretanto, compensar a necessidade do aumento das taxas de transmissão em muitas aplicações práticas. O peso dessa limitação recaiu principalmente sobre o período de treinamento necessário nesses cenários, visto que não há transmissão efetiva de informação durante este intervalo de tempo, já que a sequência transmitida já é conhecida pelo receptor. Essa linha de raciocínio serviu como motivação para a busca de novas técnicas capazes de permitir um efetivo processo de equalização de canais sem depender, portanto, do sinal de referência. A esta nova vertente atribuiu-se o nome de *equalização não supervisionada* ou *cega*.

Uma das opções encontradas capazes de reduzir a IIS foi a de recorrer exclusivamente à informações estatísticas gerais acerca da mensagem que se deseja recuperar. Embora existam outros tipos de abordagens que não fazem uso direto dessas estatísticas, como, por exemplo, o uso dos filtros de erro de predição ou a aplicação de restrições (e.g., o filtro de mínima variância com restrições lineares) [Romano et al., 2010], o foco desse trabalho se restringirá ao primeiro caso.

As técnicas não supervisionadas constituíram um tema de pesquisa abordado com bastante frequência durante as últimas três décadas. Elas possuem um amplo leque de aplicações, como telecomunicações, processamento de áudio e fala, radares e sonares, sinais e imagens biomédicas e geofísica [Romano et al., 2010].

Apesar de ser uma área já consolidada em termos teóricos, os primeiros resultados

positivos foram decorrentes de algumas tentativas de caráter mais intuitivo, direcionadas principalmente à casos envolvendo sinais com modulação PAM (do inglês *Pulse Amplitude Modulation*) ou QAM (do inglês *Quadrature Amplitude Modulation*). Pouco mais tarde, foi mostrado em [Godfrey e Rocca, 1981; Bellini e Rocca, 1986] que os algoritmos dessa linha baseada em resultados empíricos eram pertencentes à classe das chamadas técnicas de Bussgang.

1.4.1 Técnicas de Bussgang

As abordagens não supervisionadas conhecidas como algoritmos (ou técnicas) de Bussgang se caracterizam por adotarem um estimador não linear sem memória do sinal desejado numa expressão do tipo LMS. A não linearidade desse estimador faz com que os algoritmos se baseiem implicitamente em estatísticas de ordem superior (EOS). Por outro lado, esse caráter não linear torna-se também responsável pela existência de mínimos locais (e outros pontos de equilíbrio, como um máximo e pontos de sela), o que dá origem a um cenário dinâmico bastante rico.

A seguir, apresentaremos três abordagens que, embora tenham motivações conceituais diferentes, enquadram-se naquilo que chamamos de algoritmos de Bussgang.

Algoritmo de Decisão Direta

O algoritmo de decisão direta (DD) teve sua origem [Lucky, 1965] da necessidade de soluções práticas e simples que fossem capazes de melhorar o desempenho de sistemas de comunicação digital, procedendo o ajuste do equalizador diretamente do sinal recebido, sem a necessidade de interromper a transmissão de informação pela sequência de treinamento.

O critério DD pode ser visto como um “critério de Wiener modificado”, cujo sinal desejado é substituído por uma estimativa obtida pelo dispositivo decisor (não linear). A função custo associada ao critério DD pode ser definida como:

$$J_{DD}(\mathbf{w}) = E [|dec(y(n-k)) - y(n)|^2], \quad (1.31)$$

em que $dec(\cdot)$ é a função de decisão e k um atraso arbitrário. Calculando o gradiente dessa função com respeito aos parâmetros do equalizador \mathbf{w} e aplicando a aproximação estocástica usual, obtém-se:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu (dec^*(y(n-k)) - \mathbf{x}^H(n)\mathbf{w}(n)) \mathbf{x}(n). \quad (1.32)$$

Caso o sinal transmitido possua modulação BPSK (do inglês *Binary Phase-Shift Keying*),

a função decisão pode ser definida como:

$$dec(y(n-k)) = sgn(y(n-k)) = \begin{cases} +1, & \text{se } y(n-k) \geq 0 \\ -1, & \text{se } y(n-k) < 0 \end{cases}. \quad (1.33)$$

É importante ressaltar que o potencial de equalização associado à técnica é limitado pela natureza da não linearidade escolhida. O decisor só proverá estimativas fidedignas de algum atraso do sinal transmitido se a condição inicial apresentada ao algoritmo permitir um grau de redução de IIS suficientemente adequado. Isso explica por que é bastante comum utilizar o algoritmo DD em conjunto com uma abordagem supervisionada, tipicamente o LMS. Enquanto se tem acesso, por exemplo, a um cabeçalho de transmissão, usa-se a técnica supervisionada para “abrir o diagrama de olho”. Uma vez que isso tenha acontecido, chaveia-se para o algoritmo DD, que converge para a solução desejada, que, idealmente, será igual à solução de Wiener [Macchi e Eweda, 1984].

Algoritmo de Sato

Com o objetivo de alcançar um grau de robustez maior que o associado ao algoritmo DD, Sato propôs o primeiro sistema recursivo capaz de operar em condições de olho fechado [Sato, 1975]. Foi concebido inicialmente para ser utilizado em sistemas M-PAM (*Pulse Amplitude Modulation M-ário*), i.e., sistemas com símbolos puramente reais. Considerando, portanto, que o sistema e os sinais estejam contidos no domínio dos reais \mathcal{R} , o algoritmo de Sato baseia-se na minimização da seguinte função custo não convexa:

$$J_{Sato}(\mathbf{w}) = E [(\hat{s}(n) - y(n))^2], \quad (1.34)$$

sendo $\hat{s}(n)$ uma estimativa de $s(n)$ obtida através de uma função não linear de memória nula:

$$\hat{s}(n) = \gamma sgn(y(n-k)), \quad (1.35)$$

em que k é um atraso arbitrário e γ uma constante que pode ser definida como:

$$\gamma = \frac{E [s^2(n)]}{E [|s(n)|]}. \quad (1.36)$$

Na Eq. (1.35), o emprego da função sinal $sgn(\cdot)$ funciona como um aproximador grosseiro do operador esperança estatística. Seu algoritmo adaptativo pode ser expresso pela seguinte equação:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu (\gamma sgn(y(n-k)) - \mathbf{x}^T(n)\mathbf{w}(n)) \mathbf{x}(n). \quad (1.37)$$

As considerações acerca da convergência do algoritmo DD são, em grande medida, vá-

lidas também para o algoritmo de Sato. Essa similaridade surge, por exemplo, do fato de que, para sinais do tipo BPSK binário, o algoritmo de Sato coincide com o de decisão direta. Assim, ambos possuem susceptibilidade a uma convergência para mínimos locais. Essa constatação se mantém mesmo no caso em que o equalizador é duplamente infinito e a modulação é PAM [Ding et al., 1993].

Algoritmos de Godard

As técnicas de equalização cega propostas por Godard [Godard, 1980] se baseiam num critério em que se quantifica a dispersão do sinal equalizado relativamente a um valor pré-determinado sem que seja considerada a fase, de forma a assegurar a independência do processo de desconvolução em relação à estimação de parâmetros de uma portadora. Esta abordagem permite que modulações complexas do tipo QAM sejam contempladas. A função custo, também chamada “dispersão de ordem p ”, assume a forma:

$$J_{Godard}(\mathbf{w}, p) = E [(|y(n)|^p - R_p)^2], \quad (1.38)$$

em que p é um inteiro positivo e R_p é uma constante que controla o ganho das soluções obtidas, dada por:

$$R_p = \frac{E [|s(n)|^{2p}]}{E [|s(n)|^p]}. \quad (1.39)$$

Calculando o gradiente e lançando mão da clássica aproximação estocástica, Godard obteve a expressão que posteriormente recebeu seu nome:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu y^*(n) |y(n)|^{p-2} (|y(n)|^p - R_p) \mathbf{x}(n). \quad (1.40)$$

Dentre a família de algoritmos apresentada, Godard concentrou sua análise principalmente aos casos com $p = 1$ e $p = 2$. No primeiro caso, o algoritmo resultante mostrou-se como uma versão modificada do algoritmo de Sato para sinais complexos. Já no segundo, para $p = 2$, observou que o tratamento matemático tornou-se mais conveniente e, surpreendentemente, com resultados mais satisfatórios no sentido de que, notada a presença de mínimos locais indesejados em ambas as formulações, o caso com $p = 2$ era mais robusto e com convergência mais rápida nos casos simulados [Godard, 1980]. De fato, foi constatado posteriormente que este último caso apresentava menor número de mínimos locais indesejáveis, o que contribuiu para sua rápida aceitação em diversas aplicações práticas. Mesmo assim, por possuir uma função custo não convexa, a adequada inicialização dos coeficientes \mathbf{w} do filtro seria de fundamental importância para assegurar a convergência correta do algoritmo.

Para certas modulações, pode-se interpretar que o algoritmo de Godard penaliza desvios

da saída do equalizador $y(n)$ com relação a um módulo constante, o que contribuiu para que a função custo para $p = 2$ viesse a receber o nome de critério do módulo constante (CM) [Treichler e Agee, 1983]. Por aliar simplicidade, robustez e eficácia, o critério CM e o algoritmo dele derivado, chamado de algoritmo do módulo constante (CMA, do inglês *Constant Modulus Algorithm*), tornou-se um dos maiores temas de pesquisa dentro do problema de equalização não supervisionada. Entretanto, como uma consequência da complexidade da função custo CM, de certa forma, permanecem ainda questões centrais de âmbito teórico e prático associadas à formulação CM que exigem maiores investigações.

A seguir, dedicaremos especial atenção à este interessante critério, apresentando alguns detalhes sobre seus aspectos que servem como motivação para as análises conduzidas neste trabalho.

1.4.2 O Critério do Módulo Constante

Como mencionado anteriormente, o caso específico para $p = 2$ na família de critérios de Godard, Eq. (1.38), dá origem à função custo do critério CM:

$$J_{CM}(\mathbf{w}) = E \left[(|y(n)|^2 - R_2)^2 \right], \quad (1.41)$$

em que

$$R_2 = \frac{E [|s(n)|^4]}{E [|s(n)|^2]}. \quad (1.42)$$

O algoritmo resultante dessa formulação, o célebre CMA, pode ser expresso como:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu y^*(n) (|y(n)|^2 - R_2) \mathbf{x}(n). \quad (1.43)$$

Apesar de considerado o algoritmo mais robusto dentre os algoritmos de Bussgang, o CMA não é capaz de recuperar a fase da portadora. Esse comportamento é justificado pelo fato de sua função custo basear-se somente na amplitude do sinal recebido. Mesmo assim, sua convergência para os mínimos do custo CM tende a aproximar, em grande medida, às soluções do critério de Wiener [Suyama, 2003]. Mostra-se, portanto, oportuno analisarmos em mais detalhes a questão da convergência do CMA, um assunto de grande interesse teórico até os dias atuais.

Uma hipótese muito útil porém pouco realista foi utilizada inicialmente nas análises de convergência das técnicas não supervisionadas: *comprimento duplamente infinito do equalizador*. Sob essa hipótese, lida-se com equalizadores que possuem resposta ao impulso infinita tanto na parte causal quanto na anticausal, estendendo-se de $n = -\infty$ a $n = +\infty$. O valor dessa hipótese, que explica a sua utilização por muitos anos na análise do CMA, é que ela permite que se estudem os mínimos do algoritmo em termos da resposta combinada

canal+equalizador, que está num domínio em que sabemos a posição dos bons mínimos, que sempre serão soluções ZF [Foschini, 1985]. Em outras palavras, nesse domínio, é possível realizar inversão perfeita de um canal FIR.

Já em seu artigo original [Godard, 1980], Godard mostrou que, para o caso $p = 2$, se $E[|s(n)|^4] < 2E[|s(n)|^2]^2$ (condição se que relaciona à ideia de subgaussianidade), os mínimos globais da função custo corresponderão a soluções de equalização perfeita.

Uma análise mais detalhada para o caso do equalizador duplamente infinito foi feita por Foschini [Foschini, 1985], concluindo que o algoritmo de Godard não possui mínimos locais indesejáveis. Ele também empreendeu uma importante análise da matriz hessiana, na qual demonstrou a existência de três possíveis tipos de pontos de equilíbrio do CMA no domínio da resposta combinada \mathbf{c} :

1. Um máximo local, correspondente ao vetor nulo;
2. Mínimos globais, já detectados por Godard, correspondentes a soluções com apenas um elemento não nulo;
3. Pontos de sela, correspondentes a soluções com mais de um elemento não-nulo.

Embora a análise de Foschini tenha sido um marco, ele frisava que ainda mais relevante seria uma análise detalhada no domínio dos parâmetros do equalizador, no qual ocorre a adaptação. Isso, no entanto, não é simples devido à complexidade da superfície de custo do critério CM.

Em 1991, Ding e outros [Ding et al., 1991] constataram a convergência do CMA para mínimos locais, ou seja, mínimos subótimos. Eles também mostraram a deficiência do método de análise baseado na resposta combinada, que só equivale à análise no domínio dos parâmetros do filtro no caso em que a matriz de convolução tem um espaço nulo trivial. Em suma, demonstraram a existência de uma nova classe de pontos de equilíbrio do CMA: a dos mínimos locais. Isso causou um certo descrédito no que diz respeito ao critério CM, levando a crer que este era inferior a técnicas como a de Wiener. No entanto, tal ideia é questionável, pois a análise de [Ding et al., 1991] se deu no contexto de um modelo de canal IIR (de natureza, portanto, distinta da de um modelo FIR), e, em alguns casos, mesmo soluções obtidas com o critério de Wiener não são capazes de levar a uma condição de olho aberto (se o atraso k for considerado um grau de liberdade) [Neves et al., 2006]. Talvez se possa dizer que, atualmente, há uma certa linha de análise que conjectura serem os mínimos do CMA, no caso de canal e equalizador FIR, soluções “próximas” às soluções de Wiener para os melhores atrasos [Zeng et al., 1998, 1999; Schniter e Johnson Jr., 2000; Suyama, 2003]. Essa ideia recebeu interpretações também à luz das relações entre o critério CM e o critério de Shalvi-Weinstein (apresentado adiante) e entre a formulação de Shalvi-Weinstein e a ideia de maximização de não gaussianidade, que é bastante relevante no contexto de

separação cega de fontes.

Outra importante constatação feita por Li e Ding em 1995 [Li e Ding, 1995] estabelece a dependência entre o desempenho do CMA e a inicialização de seus parâmetros. Os autores mostram ainda que a inicialização pelo método *center-spike* [Foschini, 1985] reduz a probabilidade de convergência para mínimos locais. Neste mesmo trabalho, tais resultados são estendidos para o critério de Shalvi-Weinstein, reforçando a conexão entre as duas abordagens.

1.4.3 O Critério de Shalvi-Weinstein

O critério de Shalvi-Weinstein (SW) [Shalvi e Weinstein, 1990] não se enquadra nos moldes das técnicas de Bussgang, pois tem por ponto de partida o teorema homônimo, cuja validade estabelece as premissas para a equalização não supervisionada. O elemento fundamental dessa abordagem envolve o cálculo da grandeza denominada *curtose*, capaz de extrair EOS a respeito do sinal de interesse. A curtose $K(\cdot)$ associada a $y(n)$ pode ser definida como:

$$K(y(n)) = E [|y(n)|^4] - 2E^2 [|y(n)|^2] - |E [y^2(n)]|^2. \quad (1.44)$$

Em 1990, aliada à ideia de cumulantes, tal medida serviu de ensejo para Shalvi e Weinstein proporem o que pode ser entendido como um refinamento do teorema de Benveniste-Goursat-Ruget (BGR) [Benveniste et al., 1980], alcançando uma expressão que caracteriza a condição de perfeita equalização a partir de uma quantidade de informação menos restritiva acerca dos sinais envolvidos. Enquanto o teorema BGR atesta uma equalização perfeita através da equivalência entre as densidades de probabilidade do sinal transmitido e do sinal recebido, o teorema SW afirma que a mesma condição pode ser observada igualando-se apenas as estatísticas de segunda ordem dos sinais e alguma outra EOS [Shalvi e Weinstein, 1990].

Com base nestas informações, foi possível definir o critério SW como um problema de maximização sujeito a uma restrição, isto é:

$$\max_{\mathbf{w}} J_{SW}(\mathbf{w}) = \max_{\mathbf{w}} |K(y(n))|; \quad (1.45)$$

sujeito a:

$$E [|y(n)|^2] = E [|s(n)|^2]. \quad (1.46)$$

Em suma, o critério pode ser explicado pelo fato de que, considerando-se a transmissão do sinal $s(n)$ independente e identicamente distribuído (i.i.d.) através de um canal linear, se for satisfeita a condição de igualdade das estatísticas de ordem dois entre $y(n)$ e $s(n)$,

a curtose do sinal $y(n)$ – filtrado por um equalizador linear – será sempre menor ou igual àquela do sinal transmitido. É intuitivo esperar então que a maximização da curtose de $y(n)$ conduzirá idealmente ao valor da curtose de $s(n)$, situação na qual o filtro equalizador será capaz de inverter perfeitamente o canal.

Como mencionado anteriormente, o critério SW estabelece diversos pontos em comum quando comparado com o critério CM. As abordagens de Li e Ding [Li e Ding, 1995] e, posteriormente, de Regalia [Regalia, 1999] revelaram correspondências entre as funções custo dos referidos critérios, principalmente no tocante aos pontos críticos associados a elas. Torna-se portanto evidente a grande faceta escondida nas entrelinhas do critério CM: ele permite o uso implícito de EOS, sendo capaz de resgatar estatísticas de ordem quatro através de uma estrutura consideravelmente mais simples.

Sabe-se que o CMA apresenta um desempenho satisfatório – do ponto de vista indireto da maximização da curtose [Regalia, 1999] – quando o sinal transmitido é composto de amostras independentes. Sob tais condições, o potencial de recuperação do sinal transmitido por essas técnicas se relaciona, estruturalmente, a um problema de extração de fontes com mais sinais que sensores [Haykin, 2000]. Em contrapartida, para sinais fora do contexto clássico, como aqueles dotados de algum tipo de dependência estatística ou com distribuições esparsas ou uniforme, o desempenho dessas técnicas se revela insatisfatório, levando a funções custo sujeitas a modificações nocivas à obtenção de soluções de inversão do canal.

Nesse âmbito, em uma abordagem mais recente, aliada ao aumento do poder computacional, emergiu a possibilidade de se estender a análise das EOS através do emprego direto das densidades de probabilidade dos sinais envolvidos nos critérios de otimização. Dessa forma, espera-se poder extrair as informações estatísticas de interesse de maneira mais extensiva, sobretudo, a partir do uso dos conceitos e medidas trazidos da área de teoria da informação, como entropia e informação mútua [Shannon, 1948; Cover e Thomas, 1991], conforme veremos na próxima seção.

1.5 Aprendizado Baseado na Teoria da Informação

O estudo dos critérios de filtragem adaptativa e algoritmos que visam extrair em maior completude as estatísticas contidas nos sinais de interesse constitui a motivação da área de pesquisa conhecida como aprendizado baseado em teoria da informação ou, simplesmente, ITL (do inglês *Information-Theoretic Learning*).

Mesmo com o aumento do poder computacional, o efetivo uso dos critérios baseados em ITL apenas se tornou possível com o uso de estimadores não paramétricos como os métodos de kernel associados ao janelamento de Parzen [Principe, 2010] e sua consequente aplicação

ao problema de estimação de funções densidade de probabilidade (PDF's, do inglês *Probability Density Functions*). Apesar de possuírem maior complexidade em relação aos critérios clássicos, a possibilidade de lidar diretamente com a PDF de uma variável aleatória e indiretamente com todos os momentos estatísticos associados a ela, principalmente àqueles de ordem maior que dois – i.e., as EOS –, abre um horizonte repleto de perspectivas interessantes quanto à aplicação em cenários onde os critérios mais clássicos conhecidamente apresentam fraquezas.

1.5.1 Estimadores de Densidade Não Paramétricos

Idealmente, a obtenção de uma PDF associada a um sinal exigiria uma sequência infinita de amostras para se conseguir uma representação fidedigna aos moldes de sua definição teórica. Existem, porém, maneiras de se estimar a PDF lançando mão de um conjunto de amostras finito. Uma delas é a estimação não paramétrica de Parzen, cuja aplicação se vincula à estrutura de métodos de kernel [Principe, 2010].

Pode-se dizer que os métodos de kernel possuem grande utilidade em virtude do horizonte que abrem quanto à possibilidade de resolução de problemas não lineares no próprio espaço dos dados de entrada, que chamaremos de S . Basicamente, tais métodos realizam, de forma implícita, operações de produto interno $\langle \cdot, \cdot \rangle$ num espaço de características dimensionais potencialmente maiores, que, sob a ótica do espaço de entrada S , representam operações não lineares sobre os dados. Além disso, essas operações podem ser totalmente efetuadas no próprio espaço S , permitindo-nos obter versões não lineares de operações lineares (expressa em termos de produtos internos) [Principe, 2010].

Para entendermos melhor esta abordagem, podemos supor um mapeamento não linear Φ sobre os dados \mathbf{x}_i , que levam do espaço S para o espaço H :

$$\begin{aligned} \Phi : S &\rightarrow H \\ \mathbf{x} &\rightarrow \Phi(\mathbf{x}). \end{aligned} \tag{1.47}$$

Associada a este mapeamento Φ , é possível provar que existe uma função kernel bivariada $\kappa(\mathbf{x}_i, \mathbf{x}_j)$, definida positiva e simétrica (condições de Mercer [Vapnik, 1995]) que é capaz de estabelecer a seguinte relação:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle. \tag{1.48}$$

Em palavras, a Eq. (1.48) diz que o uso da função kernel $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ sobre os dados de entrada em S corresponde à um produto interno dos dados mapeados – de maneira não linear – no espaço H (o chamado “truque do kernel” [Principe, 2010]). Essa elegante ideia nos permite, através do emprego das funções kernel, lidar implicitamente com as transformações não

lineares Φ – isto é, o uso do kernel não demanda o conhecimento de Φ . Entretanto, o tipo de mapeamento não linear Φ estará diretamente ligado à função kernel escolhida [Aronszajn, 1943].

Embora o contexto original em que os métodos de kernel foram introduzidos não contemple uma interpretação estatística em relação aos dados, a abordagem caiu como uma luva para o trabalho de Parzen realizado em 1962 [Parzen, 1962], cujo o objetivo era o de prover estimativas de PDF's associadas a um conjunto de dados. A dedução decorre de um ponto de vista estatístico acerca dos sinais através de sua abordagem como processos estocásticos. Seguindo esta linha de raciocínio, podemos associar a um determinado sinal em certo instante, digamos $x(n)$, uma variável aleatória (VA) contínua denominada X . Supondo que se dispõe de L amostras *i.i.d.* $\{x_1, \dots, x_L\}$ dessa VA, a estimativa Parzen [Parzen, 1962] da PDF associada a X - denotada por $p_X(x)$ - é dada por:

$$\hat{p}_X(x) = \frac{1}{L} \sum_{i=1}^L \kappa_\sigma \left(\frac{x - x_i}{\sigma} \right), \quad (1.49)$$

em que $\kappa(\cdot)$ é uma função kernel arbitrária e σ é o tamanho do kernel.

Conforme vimos anteriormente, a função kernel $\kappa(\cdot)$ a ser utilizada deve respeitar as condições de Mercer [Vapnik, 1995]. Contudo, para que esta abranja as definições associadas às distribuições dos dados, deverá observar também às três seguintes restrições:

$$\begin{aligned} (i) \quad & \kappa(x) \geq 0; \\ (ii) \quad & \int_R \kappa(x) dx = 1; \\ (iii) \quad & \lim_{x \rightarrow \infty} |x\kappa(x)| = 0. \end{aligned} \quad (1.50)$$

Assim, os kernels empregados são simétricos e normalizados, tipicamente com posição do pico associado ao valor da amostra e, preferencialmente, que sejam contínuos e diferenciáveis caso se queira usar métodos baseados no gradiente da função.

Um dos kernels mais utilizados na literatura é, sem dúvida, o Gaussiano $G_\sigma(\cdot)$ [Principe, 2010]. Tal kernel pode ser definido como:

$$G_\sigma(x_i - x_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left(\frac{-|x_i - x_j|^2}{2\sigma^2} \right). \quad (1.51)$$

Medidas alternativas para estimação de PDF's (e.g, as baseadas em histogramas) [Chapeau-Blondeau e Rousseau, 2009] podem ser também bastante eficientes, mas, por outro lado, podem demandar um número maior de amostras para a estimação e possuir maior sensibilidade ao ajuste de parâmetros. Aliado a isso, estes métodos não podem ser empregados em algoritmos que fazem uso do gradiente da função custo, já que, muitas vezes, as PDF's

estimadas de acordo com essa linha não fornecem derivadas contínuas. Este fator, apesar de não ser um grande limitador, contribuiu para a não consideração destes estimadores neste trabalho, limitado apenas à abordagens que propiciam a derivação de técnicas baseadas no método do gradiente.

1.5.2 Os Critérios Baseados em Teoria da Informação

Em termos simples, as formulações de ITL buscam o ideal de extrair e tratar a informação de maneira tão eficiente quanto possível, no sentido de levar em conta a estrutura das EOS associadas aos sinais de interesse, o que pode ser crucial quando, por exemplo, lida-se com sinais e ruídos associados à distribuições não Gaussianas e, também, estruturas não lineares e recorrentes. Tais aspectos permitem, em tese, que se alcance uma formulação de maior generalidade que aquela baseada apenas em estatísticas de ordem dois [Erdogmus e Principe, 2006].

No âmbito da filtragem não supervisionada, abordagem esta que será seguida neste trabalho, o vínculo com a teoria de informação já é bastante bem estabelecido. Um marco nesse sentido se deu no desenvolvimento da análise de componentes independentes (ICA, do inglês *Independent Component Analysis*) e de aplicações relacionadas à separação cega de fontes, onde se utilizou, por exemplo, a ideia de quantificar independência com auxílio de grandezas como a informação mútua e também a adoção de conceitos como o de transferência ou fluxo de informação num sistema [Hyvärinen et al., 2001; Héroult et al., 1985].

Na realização de ICA, podem ser empregadas diferentes abordagens baseadas na teoria da informação, dentre as quais merecem destaque aquelas baseadas na informação mútua, um eficiente meio para quantificar o grau de independência entre variáveis aleatórias, e na negentropia [Suyama et al., 2006]. No entanto, quando se lida com o problema de filtragem no domínio do tempo, a aplicação das grandezas supracitadas ainda é relativamente recente.

Como vimos na seção anterior, a estimação das PDF's das VA's associadas aos sinais de interesse abre caminho para a aplicação direta dos conceitos e medidas trazidos do campo da teoria da informação. Um ponto destacado nesse contexto é o recurso a definições de entropia alternativas à de Shannon, especialmente a de Alfred Rényi [Rényi, 1976]. Observa-se uma predileção pelo caso particular que leva à entropia quadrática, já que ela se presta mais à derivação de algoritmos baseados em diferenciação [Fisher, 1997]. Exemplos dessas novas formulações serão mostrados a seguir.

Entropia Quadrática de Rényi

Uma proposta bastante interessante surgida nos últimos anos corresponde ao uso da entropia quadrática de Rényi para gerar uma nova versão do critério CM [Santamaría

et al., 2002a]. A escolha da entropia quadrática se justifica em termos de simplicidade de estimação e derivação de algoritmos baseados em gradiente. Em termos mais genéricos, emprega-se a entropia de Rényi de ordem α , dada por:

$$H_\alpha(y) = \frac{1}{1-\alpha} \log \left(\int p_Y^\alpha(y) dy \right). \quad (1.52)$$

Considerando agora a função custo CM, Eq. (1.41), torna-se possível, em vez de penalizar desvios quadráticos da condição de módulo constante, a penalização sobre a entropia dos desvios. Com a utilização da informação adicional extraída pela entropia, espera-se uma melhora no desempenho, principalmente em termos de velocidade de convergência. Dessa maneira, a função custo assume a forma:

$$J_{H_\alpha}(\mathbf{w}) = H_\alpha(|y(n)|^2 - R_2) = H_\alpha(|y(n)|^2), \quad (1.53)$$

onde a constante R_2 definida em (1.42) é desprezada na última igualdade pelo fato da entropia não depender da média de um sinal.

Considerando a entropia quadrática de Rényi – i.e., fazendo $\alpha = 2$ – e kernels Gaussianos, Eq. (1.51), podemos usar as relações (1.52) e (1.49) em (1.53) para obter a estimativa:

$$\begin{aligned} \hat{J}_{H_2}(\mathbf{w}) &= \hat{H}_2(|y(n)|^2 = \zeta) \\ &= -\log \int_{-\infty}^{\infty} \left(\frac{1}{L} \sum_{i=0}^{L-1} G_\sigma(\zeta - |y(n-i)|^2) \right)^2 d\zeta \\ &= -\log \frac{1}{L^2} \int_{-\infty}^{\infty} \left(\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} G_\sigma(\zeta - |y(n-j)|^2) G_\sigma(\zeta - |y(n-i)|^2) \right) d\zeta \quad (1.54) \\ &= -\log \frac{1}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \int_{-\infty}^{\infty} G_\sigma(\zeta - |y(n-j)|^2) G_\sigma(\zeta - |y(n-i)|^2) d\zeta \\ &= -\log \left(\frac{1}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} G_{\sigma\sqrt{2}}(|y(n-j)|^2 - |y(n-i)|^2) \right). \end{aligned}$$

Este resultado é obtido graças à observação de que a seguinte relação, decorrente do uso de kernels Gaussianos, é válida:

$$\int_{-\infty}^{\infty} G_\sigma(\zeta - |y(n-j)|^2) G_\sigma(\zeta - |y(n-i)|^2) d\zeta = G_{\sigma\sqrt{2}}(|y(n-j)|^2 - |y(n-i)|^2). \quad (1.55)$$

Para outras funções kernel não se pode afirmar o mesmo.

Existem também algumas considerações feitas sobre o critério definido em (1.54) que

tentam tornar esta função custo mais simples: dado que a função logaritmo $\log(\cdot)$ é uma função monotônica, esta pode ser desconsiderada, pois os pontos de mínimo em (1.54) ainda se mantêm; além disso, o sinal negativo pode ser igualmente descartado, com a ressalva de que o problema agora torna-se uma maximização. Ao resultado destas simplificações dá-se o nome de *estimador do potencial de informação quadrático* $V_2(\cdot)$ [Santamaría et al., 2002a], com a nova função custo:

$$\begin{aligned} \hat{J}_{V_2}(\mathbf{w}) &= V_2(|y(n)|^2 = \zeta) \\ &= \frac{1}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} G_{\sigma\sqrt{2}}(|y(n-j)|^2 - |y(n-i)|^2). \end{aligned} \quad (1.56)$$

O critério definido em (1.56) será melhor estudado no Cap. 4, onde será também comparado a outros critérios de interesse.

Casamento de PDF's por Distância Quadrática

Uma alternativa interessante a ser adotada como critério consiste em tentar forçar o casamento entre as PDF's associadas ao sinal de saída do equalizador com a do sinal transmitido (de constelação conhecida) [Lázaro et al., 2005]. Mais especificamente, as PDF's consideradas utilizam as VA's associadas ao módulo dos símbolos elevados à potência q , isto é, $|y(n)|^q$ e $|s(n)|^q$. Na prática, os melhores resultados são obtidos para $q = 2$, contudo, o método admite uma formulação geral.

O casamento entre as distribuições de interesse pode ser traduzido como um critério de equalização estabelecendo-se a distância quadrática (QD, do inglês *Quadratic Distance*) entre as PDF's como função custo, que é dada por:

$$J_{QD}(\mathbf{w}) = \int_{-\infty}^{\infty} (p_{Y^q}(z) - p_{S^q}(z))^2 dz, \quad (1.57)$$

onde Y^q e S^q são as VA's associadas aos sinais $|y(n)|^q$ e $|s(n)|^q$, respectivamente.

Expandindo (1.57), chegamos a

$$\begin{aligned} J_{QD}(\mathbf{w}) &= \int_{-\infty}^{\infty} (p_{Y^q}^2(z) - 2p_{Y^q}(z)p_{S^q}(z) + p_{S^q}^2(z)) dz \\ &= \int_{-\infty}^{\infty} p_{Y^q}^2(z) dz - 2 \int_{-\infty}^{\infty} p_{Y^q}(z)p_{S^q}(z) dz + \int_{-\infty}^{\infty} p_{S^q}^2(z) dz. \end{aligned} \quad (1.58)$$

De modo análogo ao caso da entropia quadrática de Rényi, se considerarmos $q = 2$ e kernels Gaussianos (1.51), é possível chegar a uma estimativa da função custo, nos moldes de (1.54):

$$\begin{aligned}
\hat{J}_{QD}(\mathbf{w}) &= \frac{1}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} G_{\sigma\sqrt{2}}(|y(n-j)|^2 - |y(n-i)|^2) \\
&\quad - \frac{2}{LN_s} \sum_{i=1}^{N_s} \sum_{j=0}^{L-1} G_{\sigma\sqrt{2}}(|y(n-j)|^2 - |s_i|^2) \\
&\quad + \frac{1}{N_s^2} \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} G_{\sigma\sqrt{2}}(|s_j|^2 - |s_i|^2),
\end{aligned} \tag{1.59}$$

onde N_s é o número de símbolos complexos definidos na constelação da correspondente modulação e s_i os valores complexos associados ao i -ésimo símbolo. Note que foi necessário o uso da relação dada por (1.55) na derivação de $\hat{J}_{QD}(\mathbf{w})$.

Como foi visto, por fazerem uso dos métodos de kernel, os critérios ITL demandam um custo computacional mais elevado em relação à abordagem clássica (como o CMA). Assim, no processo de adaptação dos coeficientes do equalizador, empregam-se geralmente algoritmos baseados no método do gradiente como tentativa de amenizar a carga computacional.

Os trabalhos conduzidos em [Santamaría et al., 2002a] e [Lázaro et al., 2005] revelam pontos positivos quando da abordagem dos critérios ITL cegos em cenários comparativos que envolvem os métodos clássicos como o CMA e o algoritmo de Shalvi-Weinstein. Assim como para os demais critérios ITL propostos que compõem o cenário atual, estes destacam-se principalmente por apresentarem maior velocidade de convergência e robustez a ruídos não Gaussianos. No entanto, até o momento, não houve uma análise no sentido de reunir e comparar sistematicamente os critérios ITL não supervisionados. Neste trabalho, um esforço será feito nesse sentido envolvendo os critérios ITL aqui apresentados, buscando estabelecer, sempre que possível, conexões ao critério CM. A análise mais detalhada dos critérios considerados e a sua comparação sob diversos prismas como função custo, superfície de erro e desempenho, serão apresentadas no Cap. 4.

1.6 Conclusão

Neste capítulo, foram apresentados alguns conceitos e fundamentos teóricos utilizados para a equalização de canais, bem como os principais critérios empregados para desempenhar tal tarefa, de forma a estabelecer as bases necessárias para o desenvolvimento dos próximos capítulos.

Mais detalhadamente, foram apresentados os conceitos fundamentais dos filtros FIR e os aspectos básicos do problema de equalização de canais, seguido de uma revisão dos critérios clássicos baseados nas vertentes supervisionadas (como o de Wiener) e cegas (como o CM). Pontuamos ainda alguns aspectos de convergência do algoritmo CM, de onde surgem certos pontos de contato aos critérios de Wiener e de Shalvi-Weinstein. De maneira bastante

interessante, essas relações motivam a abordagem que será seguida nos Caps. 2 e 3, que, conforme iremos expor, nos permitirá explorar as estatísticas de ordem superior presentes no critério CM de maneira mais explícita.

Finalmente, apresentamos uma breve noção da nova geração de critérios cegos com aprendizado baseado na teoria da informação, que serão discutidos em maiores detalhes no Cap. 4.

Formulação Polinomial do Critério do Módulo Constante

O critério de módulo constante tem recebido significativa atenção por parte da comunidade científica ao longo de quase três décadas. Parte desse interesse teve como motivação o aprofundamento dos estudos de Godard envolvendo os pontos de mínimo associados ao custo CM e a convergência do algoritmo [Foschini, 1985; Ding et al., 1991; Li e Ding, 1995]. Como consequência dessas análises, ressaltamos, dentre os principais resultados, as conexões com o critério de Shalvi-Weinstein [Ding et al., 1991; Shalvi e Weinstein, 1993; Regalia, 1999] e também com o critério de Wiener. No primeiro caso, estabeleceu-se um ponto de contato de grande importância do ponto de vista de unificação de abordagens cegas e também do estabelecimento de pontos de contato com a teoria de separação cega de fontes. Numa visão mais detalhista, tais relações permitiram que se enxergasse com maior clareza uma equivalência entre os critérios SW e CM, revelando um efetivo uso de momentos de até quarta ordem pelo critério CM, mesmo que apenas implicitamente. Já no segundo caso, em particular, as conexões com o critério de Wiener levantaram uma conjectura importante que permeia este trabalho: espera-se que haja certa proximidade entre os mínimos do critério CM e as melhores soluções de Wiener (levando-se em conta diferentes atrasos k). Tal hipótese é fundamentada principalmente em trabalhos como [Schniter e Johnson Jr., 2000; Zeng et al., 1998, 1999], que, além de compararem os mínimos desses critérios, também teceram comparações aos valores associados a cada custo. Pouco mais tarde, uma análise abrangendo cenários mais genéricos foi apresentada em [Suyama, 2003], o que elevou, de maneira mais rigorosa, o nível de proximidade entre os dois critérios.

Seguindo essa linha de raciocínio, abordaremos neste capítulo a formulação polinomial do critério CM, que mostrou ser capaz de estreitar ainda mais as conexões entre os critérios Wiener e CM. Tal formulação, inicialmente reportada em 2011 [Sousa Júnior, 2011], também revelou ser bastante esclarecedora e capaz de exprimir, de maneira mais explícita, a presença

das estatísticas de ordem superior subjacentes à formulação CM.

Apesar de propiciar uma análise teórica bastante intrigante que explora os laços entre os critérios supervisionados e não supervisionados, a formulação polinomial também conduz a duas derivações de caráter mais prático de nosso interesse: o limitante inferior do custo CM – interpretado como índice de equalizabilidade cego – e a heurística de inicialização do CMA, ambas introduzidas em [Sousa Júnior, 2011]. Entretanto, notamos que este trabalho inicial ainda admite algumas extensões e novos desdobramentos. Nesse sentido, focaremos grande parte de nossos esforços nessas duas abordagens, de forma que as contribuições apresentadas neste capítulo atuem como complementação à estas propostas originais, visando torná-las mais abrangentes, conforme veremos adiante.

A primeira extensão que buscaremos abordar será a definição da formulação polinomial – e, conseqüentemente, do limitante inferior – considerando o caso em que os sinais e sistemas inserem-se no domínio dos complexos (os esforços anteriores compreendiam apenas o domínio dos reais [Sousa Júnior, 2011]). Em seguida, conduziremos uma análise mais extensiva do limitante inferior como índice de equalizabilidade em diferentes cenários de simulação, inclusive em situações em que o sinal transmitido possui modulação multinível (como 8-PAM e 16-QAM). Por fim, introduziremos uma nova heurística de inicialização do CMA capaz de elevar ainda mais a taxa de convergência global deste algoritmo.

2.1 Formulação Polinomial

Como vimos anteriormente, mesmo fazendo uso implícito de estatísticas de quarta ordem, o critério CM detém uma função custo relativamente simples – reproduzida novamente a seguir:

$$J_{CM}(\mathbf{w}) = E \left[(|y(n)|^2 - R_2)^2 \right], \quad (2.1)$$

sendo $R_2 = E [|s(n)|^4] / E [|s(n)|^2]$.

De maneira bastante interessante, visando um estreitamento das relações entre os critérios CM e de Wiener, podemos exprimir a função custo CM mostrada em (2.1) como uma expressão do tipo EQM, nos moldes daquela de Wiener (1.9): torna-se necessário, entretanto, que sejam feitas algumas considerações com respeito ao termo quadrático do sinal equalizado $y(n)$. Mostraremos a seguir como é feita essa dedução, resgatando as ideias apresentadas em [Sousa Júnior, 2011] que deram origem à formulação polinomial do critério CM; porém, iremos supor, adicionalmente, um cenário contido no domínio complexo.

2.1.1 Formulação Com Restrições

Supondo que o sinal transmitido seja composto por amostras de norma unitária, sem perda de generalidade, é possível interpretar a Eq. (2.1) como uma função de EQM entre um

sinal desejado $d(n) = R_2 = 1$ e o sinal de saída $v_r(n) = |y(n)|^2$ – que carrega intrinsecamente uma relação polinomial entre $x(n)$ e \mathbf{w} –, assumindo a forma:

$$J_{CM}(\mathbf{w}) = E [(v_r(n) - 1)^2]. \quad (2.2)$$

Nesta relação, a função custo CM não sofre qualquer tipo de prejuízo ou variação, porém, em contrapartida, implica em mudanças marcantes sobre o sinal filtrado, agora chamado de $v_r(n)$.

Para ilustrar essa ideia, será discutido inicialmente um caso relativamente simples. Consideramos um equalizador FIR no sistema original de ordem $N = 1$ e coeficientes \mathbf{w} , cuja saída $y(n)$ é dada por

$$\begin{aligned} y(n) &= \begin{bmatrix} w_0^* & w_1^* \end{bmatrix} \begin{bmatrix} x(n) \\ x(n-1) \end{bmatrix} \\ &= \mathbf{w}^H \mathbf{x}(n), \end{aligned} \quad (2.3)$$

onde assumimos um cenário composto por sistemas e parâmetros contidos no domínio complexo. Para este caso, podemos escrever a expansão $v_r(n) = |y(n)|^2 = y(n)y^*(n)$, em notação vetorial, como

$$\begin{aligned} v_r(n) &= y(n)y^*(n) = \mathbf{w}^H \mathbf{x}(n) \mathbf{x}^H(n) \mathbf{w} \\ &= \begin{bmatrix} |w_0|^2 & w_0^* w_1 & w_0 w_1^* & |w_1|^2 \end{bmatrix} \begin{bmatrix} |x(n)|^2 \\ x(n)x^*(n-1) \\ x^*(n)x(n-1) \\ |x(n-1)|^2 \end{bmatrix} \\ &= \boldsymbol{\theta}_r^H \boldsymbol{\xi}(n), \end{aligned} \quad (2.4)$$

sendo $\boldsymbol{\theta}_r$ o vetor de parâmetros do filtro polinomial [Mathews e Sicuranza, 2000], cujos valores associam-se àqueles do filtro original \mathbf{w} , e $\boldsymbol{\xi}(n)$ o vetor de entrada no domínio de Volterra [Mathews e Sicuranza, 2000]. Mais especificamente, tratando-se do filtro polinomial $\boldsymbol{\theta}_r$, podemos dizer que os valores dos seus coeficientes estarão restritos à algumas relações em função dos coeficientes de \mathbf{w} . Por exemplo, o primeiro coeficiente de $\boldsymbol{\theta}_r$ deverá respeitar a relação $|w_0|^2$, portanto, deverá assumir apenas valores positivos, e, além disso, uma vez determinado (ou fixado) o valor de w_0 , teremos que outros coeficientes de $\boldsymbol{\theta}_r$ dependerão desse valor, como nos casos do segundo e terceiro coeficientes de $\boldsymbol{\theta}_r$, pois estão associados à $w_0^* w_1$ e $w_0 w_1^*$, respectivamente. Assim, torna-se evidente a restrição imposta sobre $\boldsymbol{\theta}_r$ em relação aos valores de \mathbf{w} – o subscrito ‘ r ’ refere-se a esta restrição.

Se supormos ainda que o filtro polinomial $\boldsymbol{\theta}_r$ realiza ponderações no vetor de entrada $\boldsymbol{\xi}(n)$, podemos inferir, a partir da Eq. (2.4), que $\boldsymbol{\theta}_r$ corresponde à uma estrutura bastante similar a um filtro FIR – note a semelhança entre (2.4) e (2.3). Assim, podemos utilizar

um novo sistema de filtragem baseado em (2.4), dotado de um filtro polinomial θ_r e um vetor de entrada $\xi(n)$ (que engendra combinações quadráticas de $\mathbf{x}(n)$), capaz de gerar em sua saída diretamente o sinal $v_r(n)$.

Como é possível observar, o número de pesos do filtro polinomial nesta formulação é superior àquele do filtro original \mathbf{w} . Ademais, faz-se importante notar que, ao abrirmos mão da estrutura de filtragem original, alguns prejuízos são sofridos: não é mais possível reconstruir $y(n)$ de maneira fidedigna a partir de $v_r(n)$. Isto inviabiliza, por exemplo, o uso direto desta nova estrutura como um equalizador. Torna-se conveniente, portanto, face a essas distinções, escrever (2.2) explicitamente em função dos coeficientes do filtro polinomial θ_r , isto é, no domínio de Volterra:

$$J_{CMpoli}(\theta_r) = E [(v_r(n) - 1)^2]. \quad (2.5)$$

Vale ressaltar que (2.5), apesar de envolver uma estrutura de filtragem polinomial, é capaz de representar fielmente os valores do custo CM em (2.1), desde que sejam satisfeitas as restrições que vinculam θ_r a \mathbf{w} .

Considerando-se, agora, um equalizador FIR genérico de ordem N no sistema original, é possível conduzir uma análise similar à do caso de ordem unitária para se obter a representação equivalente para a formulação polinomial nesse caso mais abrangente. Convém, no entanto, distinguir mais claramente os processos envolvidos nessa abordagem: basicamente, o novo sistema no domínio de Volterra pode ser dividido em um combinador quadrático e um filtro polinomial, como ilustrado na Fig. (2.1).

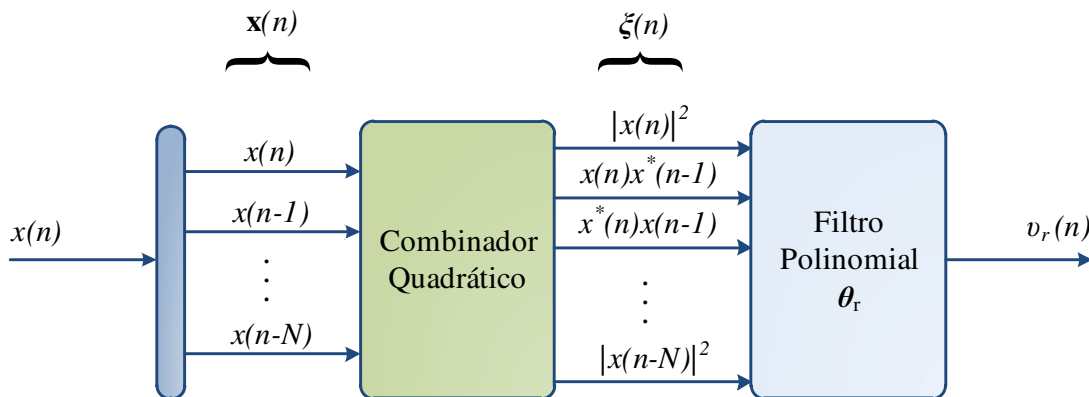


Figura 2.1: Sistema no domínio de Volterra - Combinador quadrático e filtro polinomial.

Ao combinador quadrático cabe a tarefa de gerar, a partir da entrada original $\mathbf{x}(n)$, o novo vetor de entrada $\xi(n)$ no domínio de Volterra. Já para o filtro polinomial θ_r , como vimos anteriormente, resta fazer a ponderação dos elementos que compõem o vetor de entrada, respeitando as restrições expressas pelas relações em \mathbf{w} .

Uma das abordagens possíveis para concretizar o modo de operação do combinador qua-

drático é aquela no qual usam-se os elementos da matriz resultante do produto $\mathbf{x}(n)\mathbf{x}^H(n)$ para compor o vetor de entrada $\boldsymbol{\xi}(n)$:

$$\mathbf{x}(n)\mathbf{x}^H(n) = \begin{bmatrix} |x(n)|^2 & x(n)x^*(n-1) & \cdots & x(n)x^*(n-N) \\ x^*(n)x(n-1) & |x(n-1)|^2 & \cdots & x(n-1)x^*(n-N) \\ \vdots & \vdots & \ddots & \vdots \\ x^*(n)x(n-N) & \cdots & \cdots & |x(n-N)|^2 \end{bmatrix}. \quad (2.6)$$

Se tomarmos, portanto, os elementos dessa matriz, linha a linha, e os concatenarmos, o vetor transposto resultante é o próprio $\boldsymbol{\xi}(n)$:

$$\boldsymbol{\xi}(n) = \begin{bmatrix} |x(n)|^2 \\ x(n)x^*(n-1) \\ \vdots \\ |x(n-N)|^2 \end{bmatrix}. \quad (2.7)$$

Composto o vetor de entrada $\boldsymbol{\xi}(n)$, a próxima etapa consiste no seu processamento pelo filtro polinomial. Este, por sua vez, deve assumir valores complexos limitados pelas relações de \mathbf{w} . A obtenção dessas restrições, que delimitam o conjunto de soluções atingíveis de $\boldsymbol{\theta}_r$, pode se dar, de modo análogo ao procedimento do combinador quadrático, pelo produto $\mathbf{w}\mathbf{w}^H$ (assumindo valores genéricos para \mathbf{w}):

$$\mathbf{w}\mathbf{w}^H = \begin{bmatrix} |w_0|^2 & w_0w_1^* & \cdots & w_0w_N^* \\ w_0^*w_1 & |w_1|^2 & \cdots & w_1w_N^* \\ \vdots & \vdots & \ddots & \vdots \\ w_0^*w_N & \cdots & \cdots & |w_N|^2 \end{bmatrix}. \quad (2.8)$$

Tomando os elementos dessa matriz, linha a linha, compomos o vetor de restrição sobre $\boldsymbol{\theta}_r$:

$$\boldsymbol{\theta}_r = \begin{bmatrix} |w_0|^2 \\ w_0w_1^* \\ \vdots \\ |w_N|^2 \end{bmatrix}. \quad (2.9)$$

Finalmente, com base nas relações (2.7) e (2.9), podemos expressar a saída do filtro

polinomial $v_r(n)$ para o caso genérico do equalizador de ordem N no domínio original:

$$\begin{aligned}
v_r(n) &= y(n)y^*(n) \\
&= \begin{bmatrix} |w_0|^2 & w_0^*w_1 & \cdots & |w_N|^2 \end{bmatrix} \begin{bmatrix} |x(n)|^2 \\ x(n)x^*(n-1) \\ \vdots \\ |x(n-N)|^2 \end{bmatrix} \\
&= \boldsymbol{\theta}_r^H \boldsymbol{\xi}(n).
\end{aligned} \tag{2.10}$$

Nesta formulação, o filtro polinomial $\boldsymbol{\theta}_r$ terá ordem M (isto é, terá $M + 1$ coeficientes), ordem esta que é sempre maior que a do filtro original \mathbf{w} . De fato, tem-se que $M = (N + 1)^2 - 1 = N(N + 2)$, em decorrência do uso das combinações quadráticas dos elementos de \mathbf{w} e $\mathbf{x}(n)$.

Parâmetros no Domínio dos Reais

Se limitarmos o escopo de análise para sinais e sistemas pertencentes ao conjuntos dos reais \mathcal{R} , como feito em [Sousa Júnior, 2011], é possível obter relações mais simples para $\boldsymbol{\theta}_r$ e $\boldsymbol{\xi}(n)$. A expansão de $v_r(n) = y^2(n)$, neste caso, para um equalizador de ordem N no contexto original, fica

$$\begin{aligned}
v_r(n) &= y^2(n) \\
&= \begin{bmatrix} w_0^2 & 2w_0w_1 & \cdots & w_N^2 \end{bmatrix} \begin{bmatrix} x^2(n) \\ x(n)x(n-1) \\ \vdots \\ x^2(n-N) \end{bmatrix} \\
&= \boldsymbol{\theta}_r^T \boldsymbol{\xi}(n).
\end{aligned} \tag{2.11}$$

Em relação ao caso com valores complexos, mostrado em (2.10), as dimensões de $\boldsymbol{\theta}_r$ e $\boldsymbol{\xi}(n)$ em (2.11) são menores: a ordem do filtro polinomial será $M = N(N + 3)/2$. Além disso, as relações de $\boldsymbol{\theta}_r$ com \mathbf{w} são menos restritivas devido à própria natureza dos reais.

As implicações sobre o combinador quadrático (vide Fig. 2.1) recaem basicamente na escolha adequada dos elementos da matriz resultante do produto $\mathbf{x}(n)\mathbf{x}^T(n) \in \mathcal{R}$ – como na Eq. (2.6) – para compor o vetor de entrada $\boldsymbol{\xi}(n)$. Visto que, no caso real, $\mathbf{x}(n)\mathbf{x}^T(n)$ resulta numa matriz simétrica, basta considerar os elementos contidos em sua parte triangular superior (ou, equivalentemente, em sua parte triangular inferior).

Quanto ao filtro polinomial, de maneira análoga, o vetor que guarda as restrições de \mathbf{w} sobre $\boldsymbol{\theta}_r$ pode ser obtido fazendo-se $\mathbf{w}\mathbf{w}^T$ (para valores quaisquer de \mathbf{w}) – de maneira

similar à exposta em (2.8) – e selecionando os elementos da parte triangular superior da matriz resultante (ou triangular inferior, pois esta também é simétrica). No entanto, para os elementos fora da diagonal, é necessário fazer a multiplicação por 2.

Conforme vimos até o momento, a Eq. (2.5) define uma formulação polinomial do critério CM tanto para sinais com valores complexos como para sinais com valores reais. Existem, porém, restrições acerca dos coeficientes do filtro polinomial $\boldsymbol{\theta}_r$ que, se convenientemente relaxadas, delineiam uma nova formulação, como veremos a seguir.

2.1.2 Formulação Sem Restrições

Uma abordagem interessante pode ser obtida da relaxação das restrições sobre os valores de \mathbf{w} em $\boldsymbol{\theta}_r$, de forma que os coeficientes do filtro polinomial tornem-se plenamente “livres” – eles serão denotados adiante apenas como $\boldsymbol{\theta}$. Ao permitir que os parâmetros do filtro no domínio de Volterra assumam quaisquer valores, abre-se a possibilidade de $\boldsymbol{\theta}$ ser definido como:

$$\boldsymbol{\theta}^H = \left[\theta_0^* \quad \theta_1^* \quad \dots \quad \theta_{N(N+2)}^* \right] \in \mathcal{C}; \quad (2.12)$$

$$\boldsymbol{\theta}^T = \left[\theta_0 \quad \theta_1 \quad \dots \quad \theta_{N(N+3)/2} \right] \in \mathcal{R}, \quad (2.13)$$

que são os casos correspondentes aos parâmetros no domínio complexo e real, respectivamente, não mais restritos aos valores presentes em \mathbf{w} – considerando um equalizador de ordem N .

Este novo sistema irrestrito em nada diverge daquele apresentado na Fig. 2.1, exceto pelo fato de não existirem mais os vínculos a \mathbf{w} como em (2.9), sendo possível expressar a saída do filtro polinomial sem restrições como

$$v(n) = \boldsymbol{\theta}^H \boldsymbol{\xi}(n). \quad (2.14)$$

Isto indica que uma nova função custo do tipo EQM também pode ser deduzida. Antes disso, porém, faz-se conveniente definirmos a matriz de autocorrelação \mathbf{R}_ξ e o vetor de correlação-cruzada \mathbf{p}_ξ (que tem como sinal desejado $d(n) = R_2 = 1$) no domínio de Volterra:

$$\mathbf{R}_\xi = E \left[\boldsymbol{\xi}(n) \boldsymbol{\xi}^H(n) \right]; \quad (2.15)$$

$$\mathbf{p}_\xi = E \left[\boldsymbol{\xi}(n) d^*(n) \right] = E \left[\boldsymbol{\xi}(n) \right]. \quad (2.16)$$

Com base nessas definições, a formulação polinomial irrestrita leva à seguinte função

custo:

$$J_{LI}(\boldsymbol{\theta}) = E [|v(n) - 1|^2]. \quad (2.17)$$

Fazendo a expansão de seus termos, teremos:

$$\begin{aligned} J_{LI}(\boldsymbol{\theta}) &= E [|v(n) - 1|^2] \\ &= E [(\boldsymbol{\theta}^H \boldsymbol{\xi}(n) - 1) (\boldsymbol{\xi}^H(n) \boldsymbol{\theta} - 1)] \\ &= \boldsymbol{\theta}^H E [\boldsymbol{\xi}(n) \boldsymbol{\xi}^H(n)] \boldsymbol{\theta} - \boldsymbol{\theta}^H E [\boldsymbol{\xi}(n)] - E [\boldsymbol{\xi}^H(n)] \boldsymbol{\theta} + 1 \\ &= \boldsymbol{\theta}^H \mathbf{R}_\xi \boldsymbol{\theta} - \boldsymbol{\theta}^H \mathbf{p}_\xi - \mathbf{p}_\xi^H \boldsymbol{\theta} + 1. \end{aligned} \quad (2.18)$$

Sem os laços que restringiam as soluções para o filtro polinomial em relação a \mathbf{w} , o novo custo $J_{LI}(\boldsymbol{\theta})$ apresenta um comportamento que não segue estritamente aquele do custo CM (2.1), mas é bastante informativo, conforme veremos a seguir.

2.2 Limitante Inferior para o Critério CM

A dedução da função custo polinomial irrestrita $J_{LI}(\boldsymbol{\theta})$, Eq. (2.17), distancia-se ligeiramente do custo original CM (2.1) ao permitir que se obtenha uma relaxação sobre (2.5) por parte dos coeficientes do filtro polinomial. Essa maior liberdade, por sua vez, indica que uma minimização sobre $J_{LI}(\boldsymbol{\theta})$ levaria a um valor de EQM inferior ou igual àquele obtido pela formulação com restrições (2.5), permitindo a interpretação de $J_{LI}(\boldsymbol{\theta})$ como um limitante inferior (LI) do custo CM.

Particularmente, $J_{LI}(\boldsymbol{\theta})$ guarda algumas peculiaridades bastante interessantes. Se dispostas as Eqs. (2.18) e (1.12) lado a lado, suas similaridades tornam-se bastante evidentes, revelando a existência de importantes laços à teoria de Wiener. De maneira mais explícita, podemos dizer que $J_{LI}(\boldsymbol{\theta})$ corresponde a um problema não linear de filtragem de Wiener, visto que $\boldsymbol{\xi}(n)$ pertence ao domínio de Volterra mas a dependência em relação aos parâmetros livres $\boldsymbol{\theta}$ se mantém linear. Logo, podemos inferir que (2.17) corresponde a um paraboloide, cujo mínimo pode ser obtido na forma de uma solução fechada do tipo Wiener. Derivando $J_{LI}(\boldsymbol{\theta})$ em relação a $\boldsymbol{\theta}$ e igualando o resultado a zero, chegamos ao seguinte ponto de mínimo:

$$\boldsymbol{\theta}_o = \mathbf{R}_\xi^{-1} \mathbf{p}_\xi. \quad (2.19)$$

É importante notar que a solução para $\boldsymbol{\theta}_o$ depende das estatísticas de segunda ordem do sinal recebido $x(n)$ graças a \mathbf{p}_ξ e das de quarta ordem em decorrência da estrutura de \mathbf{R}_ξ . Além disso, tomando $d(n) = R_2 = 1$, o vetor de correlação cruzada \mathbf{p}_ξ mostra-se imune a variações no atraso k do sinal de interesse – efeito não observado na abordagem clássica de Wiener.

O valor mínimo do custo sem restrições, $J_{LI}(\boldsymbol{\theta}_o)$, pode ser obtido diretamente pela

substituição de (2.19) em (2.18):

$$\begin{aligned} J_{LI}(\boldsymbol{\theta}_o) &= \boldsymbol{\theta}_o^H \mathbf{R}_\xi \boldsymbol{\theta}_o - \boldsymbol{\theta}_o^H \mathbf{p}_\xi - \mathbf{p}_\xi^H \boldsymbol{\theta}_o + 1 \\ &= 1 - \mathbf{p}_\xi^H \mathbf{R}_\xi^{-1} \mathbf{p}_\xi, \end{aligned} \quad (2.20)$$

que caracteriza o limitante inferior do custo CM, visto que

$$J_{CM}(\mathbf{w}) \geq J_{LI}(\boldsymbol{\theta}_o). \quad (2.21)$$

Vale frisar que o limitante inferior $J_{LI}(\boldsymbol{\theta}_o)$ é completamente definido num âmbito não supervisionado, dependendo apenas das estatísticas presentes em \mathbf{R}_ξ e \mathbf{p}_ξ . Ainda, por ser oriundo do critério CM, espera-se que algumas propriedades deste ainda estejam preservadas. É possível conjecturar, portanto, como feito em [Sousa Júnior, 2011], que o valor $J_{LI}(\boldsymbol{\theta}_o)$ seja um reflexo do desempenho passível de ser alcançado por meio de um equalizador FIR tradicional, como veremos na seção a seguir.

2.2.1 Índice de Equalizabilidade

A função custo CM (2.1), por definição, busca fornecer uma medida de desempenho alcançado pelo equalizador, de acordo com a configuração de seus coeficientes \mathbf{w} (em relação ao canal em questão). Sob essa perspectiva, a configuração ótima dos pesos \mathbf{w} é aquela que atinge o valor mínimo do custo CM, isto é, aquela associada a um dos mínimos globais dessa função, cujos coeficientes assumem os valores aqui denotados por \mathbf{w}_{CMg} (este representa, portanto, a configuração dos pesos referente a um dos mínimos globais CM). Alternativamente, esse mesmo valor de custo pode ser obtido através da formulação polinomial com restrições (2.5), pois existe uma correspondência entre os coeficientes:

$$\begin{aligned} \mathbf{w}_{CMg}^H &= \left[w_{CMg,0}^* \quad w_{CMg,1}^* \quad \cdots \quad w_{CMg,N}^* \right] \\ &\quad \Updownarrow \\ \boldsymbol{\theta}_{CMg}^H &= \left[|w_{CMg,0}|^2 \quad w_{CMg,0} w_{CMg,1}^* \quad \cdots \quad |w_{CMg,N}|^2 \right], \end{aligned} \quad (2.22)$$

de maneira que, para $\boldsymbol{\theta}_r = \boldsymbol{\theta}_{CMg}$, as condições são equivalentes para a representação do mínimo custo CM. Dessa forma, serão estas as configurações de \mathbf{w} , na formulação clássica, ou $\boldsymbol{\theta}_r$, na formulação polinomial, que garantem o melhor desempenho possível de ser atingido em termos do custo CM.

Dependendo das peculiaridades da resposta ao impulso do canal de transmissão, ele pode fazer com que o sinal recebido apresente níveis de IIS mais severas e, portanto, mais difíceis de serem compensadas por uma estrutura com resposta finita (como um filtro FIR). Para canais desse tipo, o valor mínimo do custo CM tende a ser maior do que nos casos onde

a IIS é combatida de maneira mais trivial. Isto mostra que a configuração dos parâmetros do filtro que melhor atende ao custo CM – ou seja, \mathbf{w}_{CMg} ou $\boldsymbol{\theta}_{CMg}$ – ainda pode deixar um nível considerável de erro residual, sendo este mais elevado para os canais cuja inversão é mais difícil. Dessa forma, o valor do custo CM associado aos mínimos globais pode indicar o tipo de canal que se está lidando (fácil ou difícil de desconvoluir), ou ainda, pode ser considerado um *índice de equalizabilidade*, isto é, uma medida do quanto um equalizador FIR de determinada ordem em sua configuração ótima – no sentido do custo CM – se presta à inversão do canal dado. Em outras palavras, equalizabilidade diz respeito ao desempenho atingível na desconvolução de um canal por determinada estrutura de filtragem.

Como sugerido em [Sousa Júnior, 2011], uma das possíveis aplicações para o limitante inferior $J_{LI}(\boldsymbol{\theta}_o)$ se dá através de sua interpretação como um índice de equalizabilidade não supervisionado. No entanto, para que esse valor seja representativo do comportamento do custo CM em seus mínimos globais, é preciso assumir certa proximidade entre os valores de parâmetros $\boldsymbol{\theta}_o$ e $\boldsymbol{\theta}_{CMg}$.

Além disso, como tende a existir certa proximidade entre os mínimos CM e as melhores soluções de Wiener (1.13), conforme constatado nos trabalhos [Zeng et al., 1998, 1999; Schniter e Johnson Jr., 2000; Suyama, 2003], convém questionar se também existe uma relação entre o limitante inferior $J_{LI}(\boldsymbol{\theta}_o)$ e o desempenho atingível em termos de EQM no caso supervisionado, considerando sempre o atraso k mais adequado – o que reforça a ideia de índice de equalizabilidade não supervisionado.

Podemos sumarizar as hipóteses lançadas sobre as relações envolvendo as funções custo supracitadas através da Fig. 2.2. Esta busca exprimir, de maneira um pouco abstrata, o grau de proximidade desejado entre os coeficientes em cada abordagem. Nela, o lado esquerdo representa o espaço de soluções ótimas do equalizador no domínio original para os critérios de Wiener e CM (contido nas esferas) e a proximidade desejada entre eles. Do lado direito, como uma imagem espelhada, temos as soluções ótimas no domínio do filtro polinomial para os critérios LI e CM polinomial. Este último, além de possuir conexões ao critério CM, também apresenta semelhanças ao critério de Wiener, estabelecendo vínculos a essas duas abordagens. De fato, essa é a relação que permite que $J_{LI}(\boldsymbol{\theta}_o)$ seja satisfatoriamente interpretado como um índice de equalizabilidade.

Esta discussão, por ser especulativa em alguns pontos, tornar-se-á mais clara através da análise dos resultados de simulação apresentados a seguir. Consideraremos inicialmente o caso mais simples, em que os sinais e sistemas estão definidos no domínio dos reais \mathcal{R} , no qual verificaremos a validade da proposição de $J_{LI}(\boldsymbol{\theta}_o)$ como limitante inferior do critério CM e também como índice de equalizabilidade não supervisionado. Após isso, analisaremos o caso no qual as variáveis assumem valores complexos \mathcal{C} .

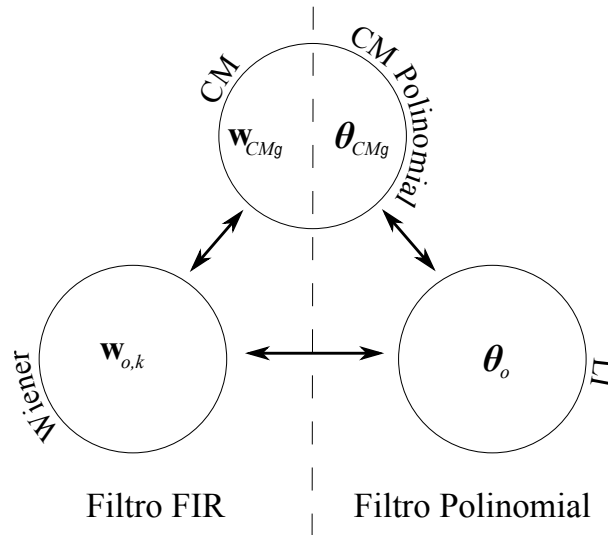


Figura 2.2: Relação entre os coeficientes ótimos associados aos critérios CM, Wiener e LI.

2.2.2 Resultados - Caso no Domínio dos Reais

Com o intuito de verificar a solidez de $J_{LI}(\boldsymbol{\theta}_o)$ como limitante inferior do critério CM e também como índice de equalizabilidade não supervisionado, iremos revisitar a análise conduzida em [Sousa Júnior, 2011] e tecer comparações com os valores da função custo correspondentes aos mínimos globais do critério CM e das soluções de Wiener (considerando o melhor atraso k de equalização).

Na análise feita em [Sousa Júnior, 2011], foram considerados como cenários de simulação os casos em que o canal possui comprimento 2, 3 e, por geração aleatória de coeficientes (do canal), comprimento entre 2 e 5. Da mesma maneira, utilizaremos como cenários de teste canais com as mesmas ordens, porém com a diferença que, nos casos com canais com 2 e 3 coeficientes, o equalizador não possui a mesma ordem do canal, e, nos casos com canais de 4 e 5 coeficientes, a análise abordará individualmente cada caso (em [Sousa Júnior, 2011], as simulações para estes canais se deu de forma concorrente). Além disso, serão considerados também os cenários em que os sinais transmitidos possuem modulação multinível ou, alternativamente, de módulo não constante, para canais de primeira e segunda ordens.

Para efeitos de simulação, estimamos o valor mínimo da função custo CM através do emprego do CMA – com passo de adaptação $\mu = 0.0008$ – ajustado de forma que a inicialização dos seus coeficientes $\mathbf{w}(0)$ correspondessem à melhor solução de Wiener. Esta estratégia também foi adotada em [Sousa Júnior, 2011], que, à luz da proximidade entre os mínimos de Wiener e CM [Schniter e Johnson Jr., 2000; Zeng et al., 1998, 1999; Suyama, 2003], pode ser considerada eficiente para se obter a convergência global. Mesmo assim, ressaltamos que, para canais mais “patológicos”, com zeros da função de transferência próximos ao círculo de raio unitário, esta estratégia pode levar o CMA à convergência local.

Todavia, devido à maior escassez desses tipos de canais, consideramos que esta estratégia fornecerá uma boa estimativa dos pontos de mínimo do custo CM.

Antes de partimos para o estudo dos cenários cujos canais possuem apenas zeros em sua função de transferência, faz-se interessante analisarmos alguns casos em que o canal possui somente polos, pois, como assumimos o filtro equalizador dotado apenas de zeros, existirá, nesse caso, a possibilidade de se alcançar a solução ZF, isto é, a inversão perfeita do canal. Nos demais casos que consideraremos, para canais e equalizadores dotados somente de zeros, essa condição não poderá ser atingida por questões estruturais do equalizador, podendo originar diferentes comportamentos nos valores mínimos dos custos em questão.

Dessa forma, no primeiro cenário, estudaremos os casos nos quais o canal pode ser expresso pela seguinte função de transferência:

$$H(z) = \frac{1}{1 + \alpha z^{-1}}, \quad (2.23)$$

onde α é percorrido de $-0,9$ a $0,9$ em passos de $0,041$. Estes canais, conforme comentado, possuem apenas polos. O sinal original é assumido *i.i.d.* de modulação BPSK, sem ruído, sendo transmitidas 50.000 amostras. Emprega-se, como equalizador, um filtro FIR de dois coeficientes (ou $N = 1$), o que implica, no domínio de Volterra, num filtro polinomial de ordem $M = 2$. Assim, com base nas amostras à disposição, para cada valor de α , estimou-se os valores para $J_{LI}(\theta_o)$ e os valores mínimos do custo CM e Wiener, sendo exibidos na Fig. 2.3 os resultados obtidos. Notamos que, como existe a possibilidade de inversão perfeita do canal, o menor custo Wiener e CM são nulos para qualquer valor de α (as curvas estão sobrepostas). Como $J_{LI}(\theta_o)$ é um limitante inferior do custo CM, também assume apenas valores nulos.

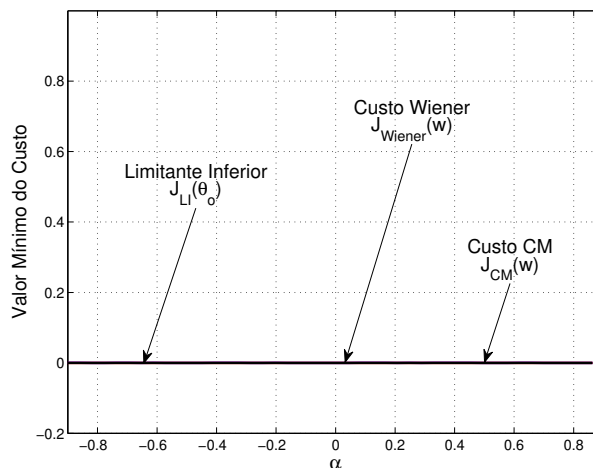


Figura 2.3: Comparação entre os valores mínimos dos custos CM e Wiener e o limitante inferior $J_{LI}(\theta_o)$ do custo CM para canais somente com polos.

Podemos concluir, portanto, que, no caso onde é possível alcançar a solução ZF, os valores mínimos dos custos Wiener e CM são nulos e coincidem com o limitante inferior. No entanto, em situações em que a estrutura de equalização não permite a inversão perfeita do canal, poderemos ter uma situação bastante diferente, como veremos a seguir.

Consideramos, agora, cenários nos quais os canais possuem apenas zeros na função de transferência. Abordaremos, primeiramente, os canais de primeira ordem do tipo:

$$H(z) = 1 + \alpha z^{-1}, \quad (2.24)$$

onde α varia de 0 a 3 (e, a cada valor de α , $H(z)$ é normalizado por sua norma euclidiana). O equalizador adotado é um filtro FIR de três coeficientes e o sinal transmitido é composto de 50.000 amostras *i.i.d.* com modulação BPSK sem adição de ruído. Vale lembrar que o emprego do equalizador de ordem $N = 2$ implica em um filtro polinomial de ordem $M = 5$ no domínio de Volterra. Foram utilizadas todas as amostras disponíveis para a estimação – via média amostral – de \mathbf{R}_ξ e \mathbf{p}_ξ , definidos em (2.16) e (2.15), respectivamente.

Os valores obtidos para $J_{LI}(\boldsymbol{\theta}_o)$ e os valores mínimos do custo CM e Wiener encontram-se ilustrados na Fig. 2.4(a). É nítida a atuação de $J_{LI}(\boldsymbol{\theta}_o)$ como um limitante inferior do custo CM. Ademais, os valores mínimos dos custos CM e Wiener tendem a acompanhar o nível de IIS ainda presente – dada a configuração canal+equalizador. Um exemplo interessante é o caso $\alpha = 1$, no qual os valores observados são máximos, coincidindo justamente com a situação na qual o canal é mais severo, i.e., onde existe um zero sobre o círculo de raio unitário na função de transferência $H(z)$. Na mesma figura, é possível observar ainda que este é também o ponto onde existe maior distância entre o limitante e o mínimo custo CM, ou seja, o caso onde $\boldsymbol{\theta}_o$ mais diverge de $\boldsymbol{\theta}_{CMg}$. Mesmo assim, nota-se ainda uma significativa proximidade entre o perfil do limitante e o do menor erro de Wiener, o que indica que o primeiro atuou adequadamente como uma medida de equalizabilidade.

O próximo cenário contempla canais de segunda ordem, cuja função de transferência definiremos como:

$$H(z) = 1 + \alpha z^{-1} + \beta z^{-2}, \quad (2.25)$$

em que α e β variam de 0 a 2 (sendo $H(z)$ posteriormente normalizado). Desta vez, o equalizador possui ordem $N = 1$, menor que a do canal. Seguiu-se o mesmo procedimento adotado no primeiro cenário para a obtenção dos custos mínimos, isto é, com 50.000 amostras *i.i.d.* e modulação BPSK. Os resultados para os valores mínimos das funções custo CM e Wiener, bem como os valores do limitante inferior $J_{LI}(\boldsymbol{\theta}_o)$, estão expressos na Fig. 2.4(b). Observa-se um comportamento similar ao do caso anterior, sendo os custos mais elevados para canais mais difíceis de serem equalizados. Note que $J_{LI}(\boldsymbol{\theta}_o)$ segue exatamente essa tendência, o que reafirma a validade do limitante inferior como índice de equalizabilidade.

Dando continuidade aos testes, foram gerados aleatoriamente 500 canais de terceira

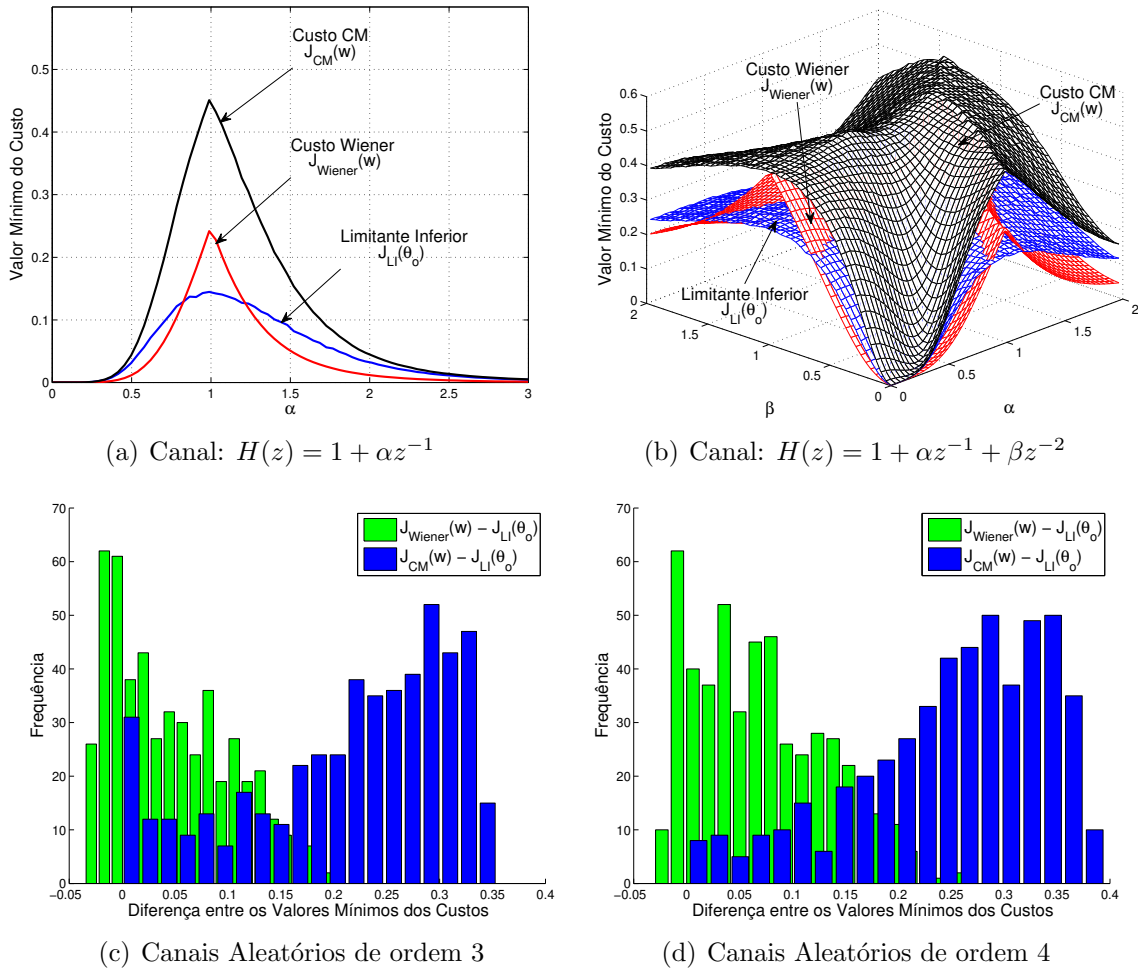


Figura 2.4: Comparação entre os valores mínimos dos custos CM e Wiener e o limitante inferior $J_{LI}(\theta_o)$ do custo CM para sinais com modulação BPSK.

ordem – cujos coeficientes foram extraídos de forma independente de acordo com uma distribuição uniforme entre -1 e $+1$ – e, para equalizadores de mesma ordem, foram calculados, para cada caso, os valores do limitante e dos mínimos custos CM e de Wiener a partir de 50.000 amostras *i.i.d.* com modulação BPSK. Assumiu-se, em seguida, o mesmo procedimento para canais de quarta ordem. Os resultados estão expressos nos histogramas das Figs. 2.4(c) e 2.4(d) para ambos os casos. Eles apresentam a frequência dos valores das diferenças entre os custos dos dois critérios clássicos e o custo associado ao limitante. A existência de valores unicamente positivos na diferença $J_{CM}(w) - J_{LI}(\theta_o)$ não deixa dúvidas da atuação de $J_{LI}(\theta_o)$ como limitante inferior da função custo CM. Em especial, destacamos também a maior proximidade de $J_{LI}(\theta_o)$ do custo mínimo de Wiener (observe que existe maior adensamento das diferenças $J_{Wiener}(w) - J_{LI}(\theta_o)$ em torno de zero), o que é auspicioso do ponto de vista de quantificação da equalizabilidade.

Uma análise complementar ainda pode ser traçada se considerarmos as modulações com módulo não constante (multinível), isto é, com múltiplos pontos na constelação. Assumindo a modulação 4-PAM para 50.000 amostras *i.i.d.* e o canal definido em (2.24), para α variando de 0 a 3 (sendo o canal subseqüentemente normalizado), poderemos tecer uma análise similar ao caso anterior com modulação BPSK. Admitindo um equalizador de mesma ordem do canal, i.e., $N = 1$, temos como resultado a Fig. 2.5(a). Nela, $J_{LI}(\boldsymbol{\theta}_o)$ ainda reflete fielmente sua propriedade de limitante inferior do custo CM, mas retém um nível de erro residual (ou valor do custo) mais elevado para todas as configurações de canal em relação ao caso BPSK, notável principalmente pelo seu distanciamento em relação ao custo Wiener. Esse fenômeno também afeta o mínimo custo CM, já que este *offset* de erro residual resulta dos frequentes desvios sofridos em relação à constante R_2 quando se adota uma constelação de múltiplos pontos. Para entendermos melhor esse efeito, podemos tomar como referência a Fig. 2.6, onde temos, no caso de um sinal modulado em 4-PAM, a constelação observada do módulo ao quadrado do sinal equalizado, $|y(n)|^2$, considerando um caso de equalização perfeita do canal. É possível notar que, em relação a R_2 , existirá sempre uma distância não nula, contribuindo para o acúmulo de um resíduo médio no custo $J_{CM}(\mathbf{w})$, dando origem, justamente, ao *offset* observado. Trata-se, portanto, de um comportamento inerente à própria definição do custo. Ainda assim, é possível distinguir claramente níveis mais elevados para canais associados à equalização considerada difícil.

Da mesma maneira, observamos que as mesmas conclusões podem ser tiradas se considerarmos o canal dado em (2.25) – e assumindo o equalizador de ordem $N = 2$ –, cujos resultados são mostrados na Fig. 2.5(b).

Igualmente interessante é a consideração do comportamento dos valores mínimos dos custos em questão para um caso com constelação ainda mais numerosa: a modulação 8-PAM. Para os mesmos canais dados em (2.24) e (2.25), obtivemos os resultados expressos nas Figs. 2.5(c) e 2.5(d), respectivamente. É notório o incremento dos valores mínimos do custo CM e também do limitante inferior, um resultado direto do maior acúmulo de erro na função custo CM (isso corresponderia a um maior número de pontos na constelação na Fig. 2.6 e conseqüente maior erro residual no custo CM). Como observado no caso da modulação 4-PAM, isto causa uma maior distância entre $J_{LI}(\boldsymbol{\theta}_o)$ e o mínimo custo Wiener, mesmo assim, a noção de um índice de equalizabilidade ainda perdura. No entanto, notamos que, à medida que se aumenta o número de pontos na constelação, o *offset* residual tende a crescer cada vez mais, tornando os valores de $J_{LI}(\boldsymbol{\theta}_o)$ menos distinguíveis entre canais associados à uma menor ou maior dificuldade de equalização.

Em suma, percebe-se que a medida $J_{LI}(\boldsymbol{\theta}_o)$ mostra-se um limitante inferior sólido do custo CM, concordando com os resultados iniciais apresentados no trabalho de [Sousa Júnior, 2011]. Todavia, quando interpretada como um índice de equalizabilidade não supervisionado, principalmente em relação aos novos resultados considerando modulações multinível,

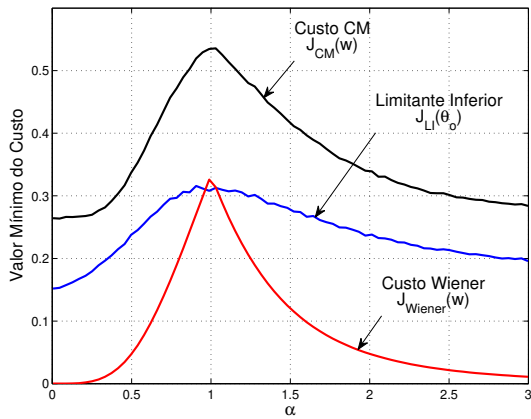
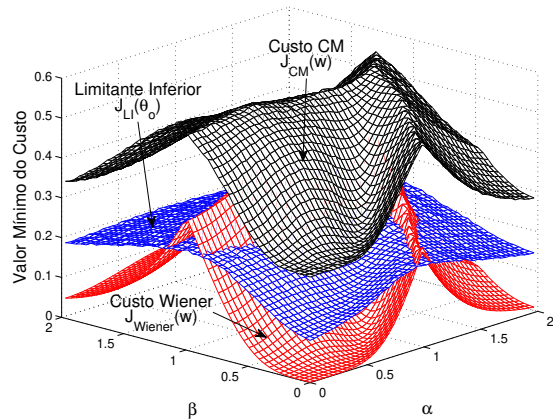
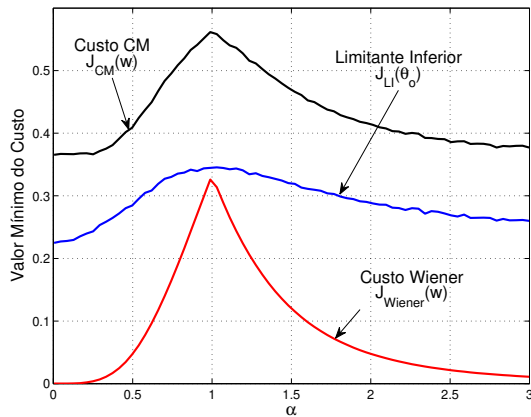
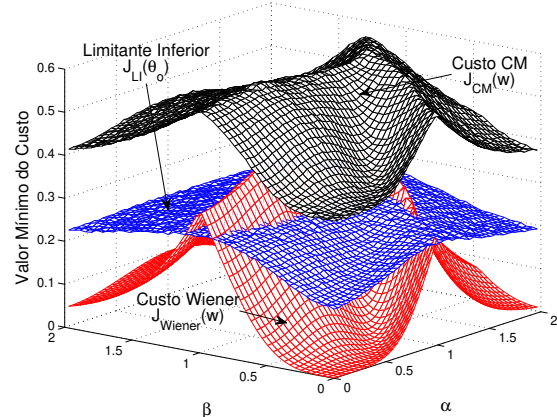
(a) Modulação 4-PAM e $H(z) = 1 + \alpha z^{-1}$ (b) Modulação 4-PAM e $H(z) = 1 + \alpha z^{-1} + \beta z^{-2}$ (c) Modulação 8-PAM e $H(z) = 1 + \alpha z^{-1}$ (d) Modulação 8-PAM e $H(z) = 1 + \alpha z^{-1} + \beta z^{-2}$

Figura 2.5: Comparação entre os valores mínimos dos custos CM e Wiener e o limitante inferior do custo CM $J_{LI}(\theta_o)$ para modulações 4- e 8-PAM.

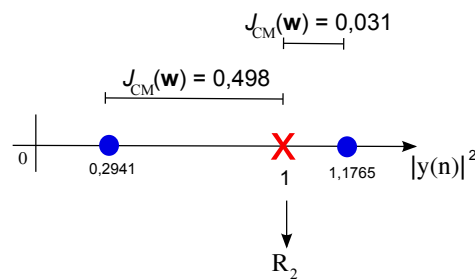


Figura 2.6: Constelação do sinal $|y(n)|^2$ supondo modulação 4-PAM.

observamos uma predileção para modulações com menor número de pontos na constelação, visto que nesses casos, existe maior proximidade entre o limitante e os valores mínimos dos custos CM e Wiener.

Estenderemos agora nossa análise para sinais e sistemas inseridos num contexto ainda não coberto pelo trabalho de [Sousa Júnior, 2011]: o domínio dos complexos.

2.2.3 Resultados - Caso no Domínio dos Complexos

A dedução conduzida nas seções anteriores contida no domínio de sinais e sistemas complexos atestam que $J_{LI}(\boldsymbol{\theta}_o)$ representa um limitante inferior do custo CM. Contudo, observamos que o caso complexo possui algumas peculiaridades que podem causar certas diferenças quando se busca interpretar $J_{LI}(\boldsymbol{\theta}_o)$ como um índice de equalizabilidade.

Conforme discutido na Seção 2.2.1, é essencial na aplicação do conceito de índice de equalizabilidade a $J_{LI}(\boldsymbol{\theta}_o)$ que haja certa proximidade entre $\boldsymbol{\theta}_o$ e $\boldsymbol{\theta}_{CMg}$. No entanto, em relação ao caso de valores reais, a abordagem no domínio complexo apresenta um maior número de coeficientes na formulação polinomial e também relações mais limitadoras do ponto de vista das restrições dos valores de \mathbf{w} sobre $\boldsymbol{\theta}_r$, resultantes da própria característica dos valores complexos. Isto abre espaço para que os coeficientes livres de $\boldsymbol{\theta}_o$ possam distanciar-se mais de $\boldsymbol{\theta}_{CMg}$. A seguir, investigaremos em maiores detalhes, através de simulações, qual é o impacto dessas constatações.

Em um primeiro cenário de análise, consideramos um canal complexo com a seguinte função de transferência:

$$H(z) = 1 + j\alpha z^{-1}, \quad (2.26)$$

onde α , como no cenário anteriormente visto, varia de 0 a 3, com subsequente normalização de $H(z)$, e j denota a unidade imaginária. O sinal é composto de 50.000 amostras *i.i.d.* pertencentes à constelação formada pela modulação complexa 4-QAM e o equalizador de coeficientes \mathbf{w} é um filtro FIR de ordem igual a do canal, i.e., $N = 1$. O filtro polinomial associado a esse equalizador, diferentemente do caso real, possui ordem $M = 3$. Os valores obtidos para $J_{LI}(\boldsymbol{\theta}_o)$ e os valores mínimos das funções custo CM e Wiener estão representados na Fig. 2.7(a). É interessante notar que, nesse caso, a curva atribuída ao custo CM aproxima-se mais daquela do custo Wiener. De fato, o caso complexo permite que o custo CM seja menos restritivo, no sentido de que as correções em direção a um módulo constante admitem uma rotação na constelação do sinal equalizado, o que dá maior flexibilidade para a atuação do filtro equalizador em relação ao custo Wiener. Voltando-nos novamente para a Fig. 2.7(a), temos que, para os canais associados à uma maior dificuldade de equalização ($0,8 < \alpha < 1,2$), o valor mínimo do custo CM forma uma espécie de platô, cuja origem ainda será investigada em trabalhos futuros por nossa equipe de pesquisa. A respeito de $J_{LI}(\boldsymbol{\theta}_o)$, podemos dizer que este exprime uma maior paridade ao custo Wiener

(e também ao custo CM) quando o canal é menos severo. Nesse sentido, podemos afirmar que, de maneira similar ao caso com valores reais, θ_o e θ_{CMg} estarão mais próximos nesses casos e mais distantes para canais mais severos. Mesmo assim, ainda é possível extrair da figura que o conceito de limitante inferior é válido, bem como perceber que a ideia de índice de equalizabilidade também se preserva, visto a tendência de $J_{LI}(\theta_o)$ em seguir o comportamento verificado nas outras curvas.

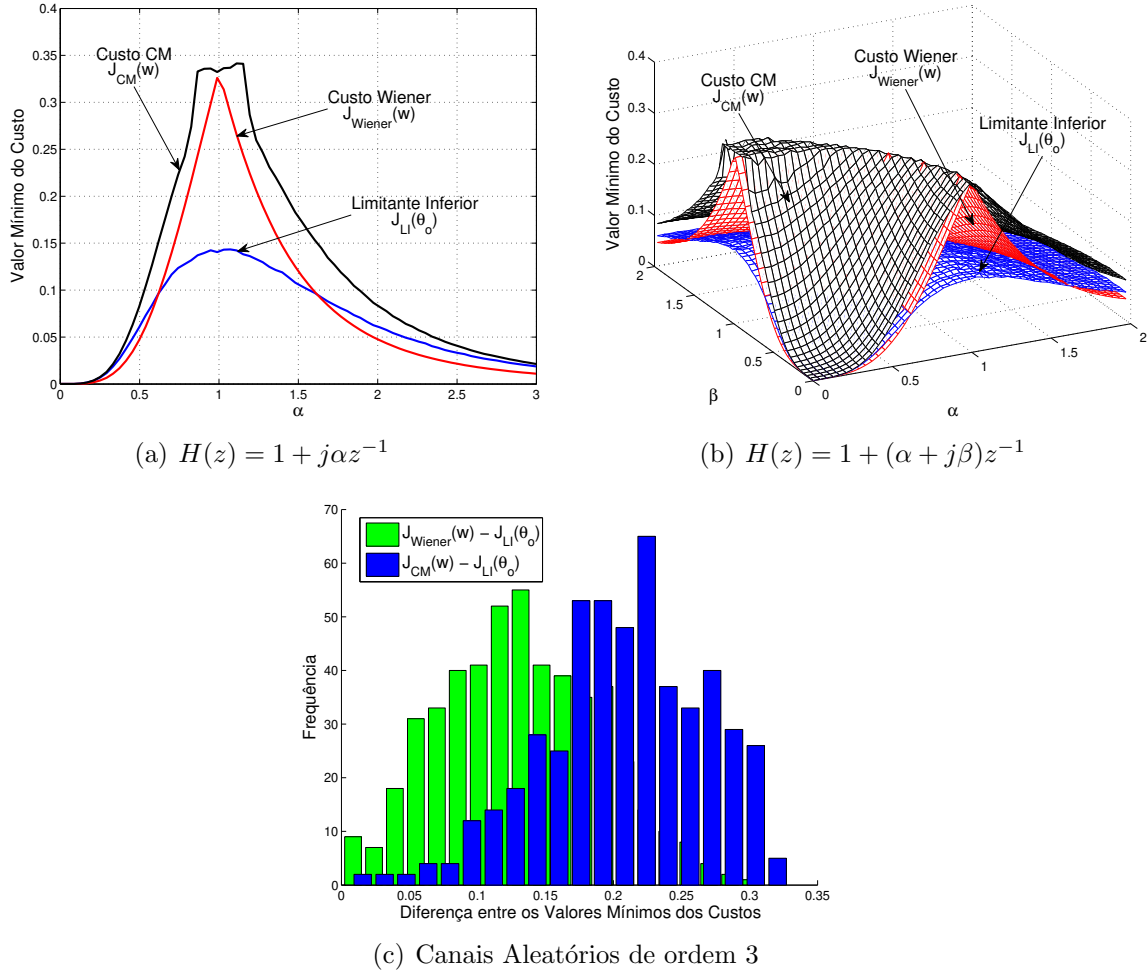


Figura 2.7: Comparação entre os valores mínimos dos custos CM e Wiener e o limitante inferior do custo CM $J_{LI}(\theta_o)$ para para sinais com modulação 4-QAM.

Em um segundo cenário complexo, avaliamos ainda canais de primeira ordem mas que admitem maior variação na parte real, de forma que a função de transferência resultante seja

$$H(z) = 1 + (\alpha + j\beta)z^{-1}, \quad (2.27)$$

com α e β excursionando de 0 a 2. Em relação ao cenário anterior, todas os demais características são mantidas. Os resultados podem ser observados na Fig. 2.7(b). O

comportamento é bastante similar ao caso anterior, mas o platô no mínimo custo CM existe apenas quando a parte imaginária é mais preponderante. Para os valores de α e β em que $\sqrt{\alpha^2 + \beta^2} = 1$, as curvas associadas aos valores mínimos dos custos CM e Wiener indicam um canal mais severo. Esse comportamento tende a ser seguido por $J_{LI}(\boldsymbol{\theta}_o)$, confirmando a ideia de índice de equalizabilidade.

Em se tratando de canais mais longos, recorreremos novamente à representação por histogramas. Foram gerados 500 canais de terceira ordem com coeficientes complexos, sendo, tanto a parte real quanto a imaginária, extraídos a partir de uma distribuição uniforme entre -1 e $+1$. Mantivemos o sinal *i.i.d.* com modulação 4-QAM (50.000 amostras) para simulação e obtenção dos valores mínimos para os custos CM, Wiener e $J_{LI}(\boldsymbol{\theta}_o)$. O resultado envolvendo as diferenças dos valores mínimos das funções custo em questão e o limitante inferior é mostrado na Fig. 2.7(c). Comparando-se ao caso real, Fig. 2.4(c), notamos uma notável disparidade, sobretudo na distribuição da diferença $J_{Wiener}(\mathbf{w}) - J_{LI}(\boldsymbol{\theta}_o)$, cujo pico de adensamento não mais recai sobre as proximidades do valor nulo. É interessante notar também que esta mesma diferença apresenta apenas valores positivos, o que indica que $J_{LI}(\boldsymbol{\theta}_o)$ também é inferior ao custo Wiener neste caso. Isto indica que, mesmo para canais mais severos, os valores de $J_{LI}(\boldsymbol{\theta}_o)$ não se sobressaem tanto, tornando mais difícil de se avaliar o significado do índice de equalizabilidade.

Por fim, analisamos também o cenário cuja transmissão do sinal se dá por uma modulação 16-QAM. Todos os casos de simulação para a modulação 4-QAM foram repetidos alterando-se apenas o sinal transmitido. Os resultados foram ilustrados nas Figs. 2.8(a), 2.8(b) e 2.8(c). Note que está presente o mesmo fenômeno observado quando da implementação de um sinal 4- e 8-PAM no caso real, isto é, o limitante inferior e os valores mínimos da função custo CM passam a admitir uma espécie de *offset* que desloca a curva verticalmente.

De uma maneira geral, podemos concluir que a abordagem no domínio dos valores complexos preserva completamente a relação de $J_{LI}(\boldsymbol{\theta}_o)$ como limitante inferior do custo CM, no entanto, alguns prejuízos são sofridos quando este é interpretado como índice de equalizabilidade, já que, neste cenário, o limitante apresenta valores mais rasos, não existindo grandes variações quando o canal passa de fácil para difícil de equalizar. Além disso, o emprego de sinais com modulação multinível (ou, alternativamente, com constelações de múltiplos pontos) provoca ainda um deslocamento vertical sobre $J_{LI}(\boldsymbol{\theta}_o)$, como também foi observado no caso real. Mesmo assim, a tendência geral de menor custo é sempre preservada, sendo possível discernir o nível de equalizabilidade atingível.

A relação verificada entre o limitante inferior e os valores mínimos dos custos CM e Wiener tem como pano de fundo, como discutido na seção 2.2.1, uma conexão intrínseca

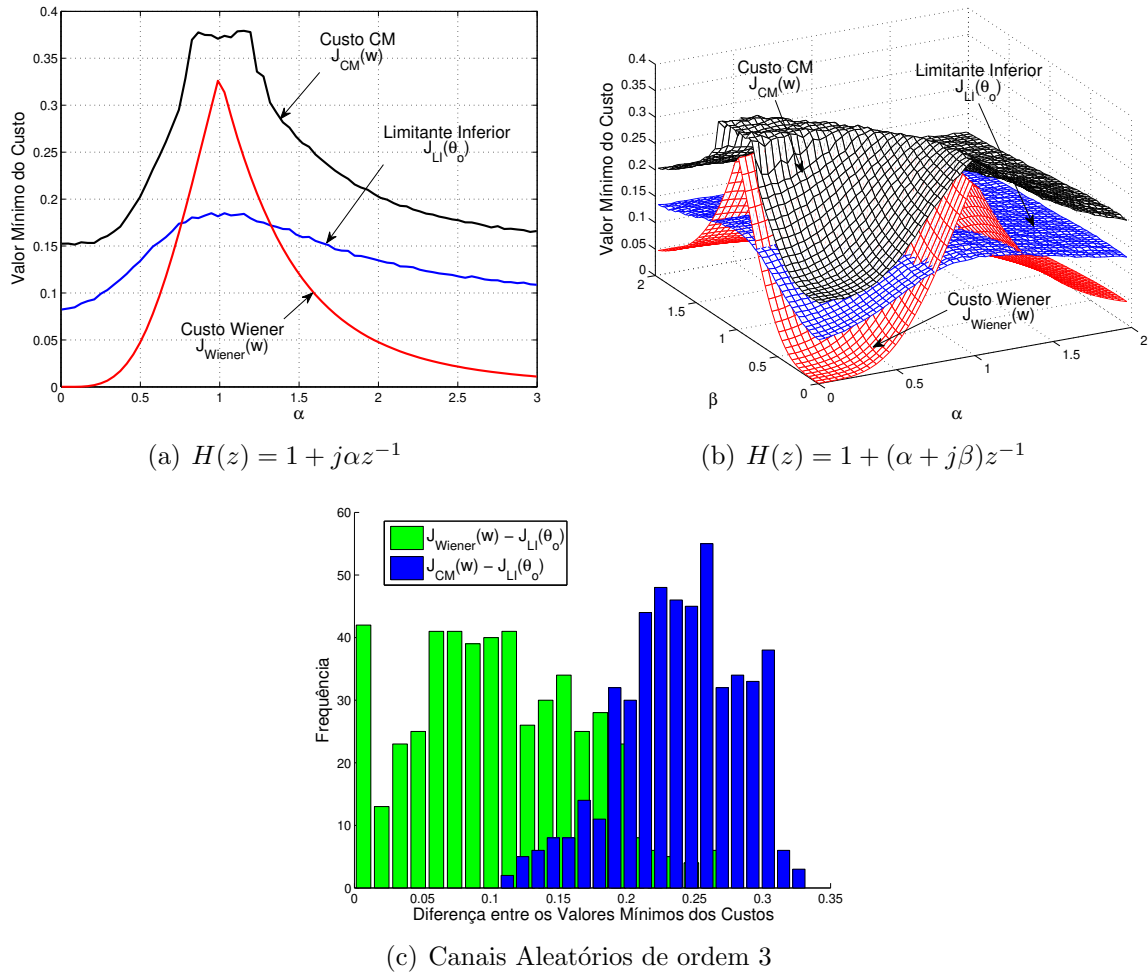


Figura 2.8: Comparação entre os valores mínimos dos custos CM e Wiener e o limitante inferior do custo CM $J_{LI}(\theta_o)$ para para sinais com modulação 16-QAM.

entre os coeficientes que compõem o filtro em cada abordagem. Valendo-se dessa informação, uma outra aplicação envolvendo a formulação polinomial do critério CM pode ser vislumbrada se considerarmos um problema que é inerente a critérios cegos: a questão da inicialização dos pesos \mathbf{w} do filtro.

2.3 Inicialização do Algoritmo do Módulo Constante

Algoritmos adaptativos não supervisionados dependem de uma adequada escolha de condições iniciais para atingirem um desempenho próximo ao ideal. Um dos fatores mais determinantes nesse problema é a inicialização dos coeficientes do equalizador. No âmbito da abordagem de Wiener, esse problema não é tão crucial, mas, para o caso não supervisionado, inicializações inadequadas podem levar à convergência para mínimos locais, i.e., soluções

subótimas [Ding et al., 1991]. Nesse sentido, é fundamental a definição de uma metodologia de inicialização para as condições de operação destes algoritmos.

Proposta por [Foschini, 1985] no contexto de inicializações para o CMA, a metodologia *center-spike* é considerada padrão atualmente e tem como principal característica um único elemento de valor unitário no “centro de massa” do vetor de coeficientes do filtro, os demais elementos são todos nulos. Nessa abordagem, soluções com atrasos k intermediários são privilegiadas, o que tende a estabelecer “condições equivalentes” tanto para canais de fase mínima como de fase máxima; contudo, muitas vezes, tais condições não são suficientes para garantir a convergência do CMA para mínimos globais.

Por meio da derivação do limitante inferior apresentado anteriormente, chega-se a uma solução ótima irrestrita em forma fechada, dada pela Eq. (2.19). A otimalidade dessa solução indica uma região que, em vista das proximidades apontadas na Fig. 2.2, se mapeada de maneira simples no conjunto de soluções atingíveis pelo equalizador em \mathbf{w} , deve gerar inicializações potencialmente muito interessantes para os algoritmos cegos, como o CMA.

Nesse espírito, uma heurística para inicialização do CMA foi proposta em [Sousa Júnior, 2011], a qual levou a uma taxa de convergência global superior (para um conjunto representativo de canais) aos métodos *center-spike* e inicialização aleatória. Esses resultados foram bastante animadores e nos motivou a propor uma nova heurística, como veremos a seguir.

2.3.1 Heurísticas de Inicialização

À luz das relações entre os coeficientes da formulação polinomial restrita e irrestrita, bem como da relativa proximidade entre estes e os mínimos globais do custo CM (vide Fig. 2.2), surge uma interessante abordagem que busca traduzir a solução fechada obtida no domínio de Volterra para o domínio clássico dos coeficientes do equalizador. Nesse sentido, as heurísticas aqui apresentadas buscam, de diferentes maneiras, realizar um mapeamento $\boldsymbol{\theta}_o \rightarrow \mathbf{w}$.

Nesta seção, limitaremos o escopo de análise apenas ao caso real, visto que, no domínio complexo, o mapeamento $\boldsymbol{\theta}_o \rightarrow \mathbf{w}$ conduz a um sistema de equações não lineares que traz ambiguidades de sinal tanto na parte real quanto na parte imaginária, comprometendo totalmente a qualidade dos valores \mathbf{w} obtidos.

Antes de prosseguirmos na exposição das heurísticas, convém fazermos algumas distinções sobre os elementos que compõem o vetor de coeficientes do filtro polinomial.

Como veremos adiante, a ideia que permeia a implementação das heurísticas supõe uma proximidade entre as soluções $\boldsymbol{\theta}_o$ e $\boldsymbol{\theta}_{CMg}$. Se esta relação for verdadeira, existe um mapeamento entre os coeficientes no domínio de Volterra e os coeficientes do filtro FIR – como na relação exposta em (2.22) – que se estabelece justamente pelas restrições de \mathbf{w} sobre $\boldsymbol{\theta}_r$, Eq. (2.11). Assim, se compararmos os coeficientes do filtro polinomial restrito (com

seus vínculos com \mathbf{w}) e irrestrito, $\boldsymbol{\theta}_r$ e $\boldsymbol{\theta}$, respectivamente, é possível fazer uma associação entre eles, capaz de explicitar as possíveis conexões entre $\boldsymbol{\theta}_o$ e \mathbf{w} :

$$\begin{aligned} \boldsymbol{\theta}_r^T &= [\quad w_0^2 \quad 2w_0w_1 \quad w_1^2 \quad 2w_0w_2 \quad \dots \quad w_N^2 \quad] \\ &\quad \quad \quad \Downarrow \quad \Downarrow \quad \Downarrow \quad \Downarrow \quad \quad \quad \Downarrow \\ \boldsymbol{\theta}^T &= [\quad \theta_0 \quad \theta_1 \quad \theta_2 \quad \theta_3 \quad \dots \quad \theta_{N(N+3)/2} \quad]. \end{aligned} \quad (2.28)$$

Observamos que há alguns termos de $\boldsymbol{\theta}$ associados a combinações puramente quadráticas do tipo w_i^2 , para $i = 1, \dots, N$, como θ_0 a w_0^2 , θ_2 a w_1^2 e $\theta_{N(N+3)/2}$ a w_N^2 . Da mesma forma, existem também coeficientes ligados a produtos cruzados de elementos distintos de \mathbf{w} , como θ_1 a $2w_0w_1$, θ_3 a $2w_0w_2$ e etc. Torna-se então possível dividir o vetor de coeficientes $\boldsymbol{\theta}$ pela distinção entre esses dois casos, originando os dois novos vetores:

$$\begin{aligned} \boldsymbol{\theta}[w_i^2] &= [\quad \theta_0 \quad \theta_2 \quad \dots \quad \theta_{N(N+3)/2} \quad] \\ &= [\quad \theta[w_0^2] \quad \theta[w_1^2] \quad \dots \quad \theta[w_N^2] \quad]; \end{aligned} \quad (2.29)$$

$$\begin{aligned} \boldsymbol{\theta}[2w_iw_j] &= [\quad \theta_1 \quad \theta_3 \quad \dots \quad \theta_{(N(N+3)/2)-1} \quad] \\ &= [\quad \theta[2w_0w_1] \quad \theta[2w_0w_2] \quad \dots \quad \theta[2w_{N-1}w_N] \quad], \end{aligned} \quad (2.30)$$

em que $\boldsymbol{\theta}[w_i^2]$ é o conjunto dos elementos de $\boldsymbol{\theta}$ que associam-se aos termos w_i^2 presentes em $\boldsymbol{\theta}_r$, para $i = 0, \dots, N$, e $\boldsymbol{\theta}[2w_iw_j]$ os elementos de $\boldsymbol{\theta}$ que relacionam-se a $2w_iw_j$ em $\boldsymbol{\theta}_r$, com $i \neq j$, assumindo os valores $i, j = 0, \dots, N$.

As heurísticas que serão apresentadas a seguir tomam como base justamente estes termos e buscam explorar, cada uma à sua maneira, as peculiaridades associadas a eles.

Heurística dos Termos Quadráticos

Para efetuar o mapeamento $\boldsymbol{\theta}_o \rightarrow \mathbf{w}$, a heurística proposta em [Sousa Júnior, 2011] prioriza as relações presentes em $\boldsymbol{\theta}[w_i^2]$: por isso, chamamo-na aqui de *heurística dos termos quadráticos* (HTQ).

Basicamente, a heurística funciona da seguinte forma: obtém-se a magnitude de todos os coeficientes \mathbf{w} aplicando-se a raiz quadrada sobre os elementos de $\boldsymbol{\theta}[w_i^2]$; após isso, as correções de sinal são feitas a partir de $\boldsymbol{\theta}[2w_iw_j]$. Este procedimento pode ser melhor entendido conforme expresso no seguinte algoritmo:

Heurística dos Termos Cruzados

Em uma análise mais detalhada dos coeficientes $\boldsymbol{\theta}_o$ e $\boldsymbol{\theta}_{CMg}$, constatamos que a aproximação entre os elementos correspondentes de cada formulação se dá de forma desigual, de modo que alguns termos distanciam-se mais do que outros. Notando uma maior concordância nos elementos representados por $\boldsymbol{\theta}[2w_iw_j]$, isto é, os elementos de $\boldsymbol{\theta}_o$ associados

Algoritmo 2.1 HTQ - Heurística dos Termos Quadráticos

Seleciona-se o elemento de maior magnitude em $\theta[w_i^2]$ para a determinação de w_i :

$$w_i \leftarrow \sqrt{\max_i |\theta[w_i^2]|}$$

Para os demais elementos de \mathbf{w} :

for j from 0 to N , $j \neq i$ **do**
 $w_j \leftarrow \sqrt{|\theta[w_j^2]|} \text{sgn}(\theta[2w_i w_j])$
end for

as relações cruzadas presentes em θ_{CMg} , propusemos uma nova heurística, chamada de *heurística dos termos cruzados* (HTC):

Algoritmo 2.2 HTC - Heurística dos Termos Cruzados

Seleciona-se o elemento de maior magnitude em $\theta[w_i^2]$ para a determinação de w_i :

$$w_i \leftarrow \sqrt{\max_i |\theta[w_i^2]|}$$

Para os demais elementos de \mathbf{w} :

for j from 0 to N , $j \neq i$ **do**
 $w_j \leftarrow \theta[w_i w_j] / (2w_i)$
end for

Em suma, é feita uma busca pela maior magnitude de $\theta[w_i^2]$ para determinar o coeficiente w_i (da mesma forma que o HTQ); após isso, os coeficientes restantes de \mathbf{w} são obtidos a partir dos termos associados ao produto cruzado $\theta[2w_i w_j]$. É importante salientar que o sinal do coeficiente de maior magnitude w_i é considerado sempre positivo (tanto no HTQ quanto no HTC), o que não acarreta em nenhuma perda de generalidade em vista da ambiguidade de sinal inerente aos algoritmos de Bussgang [Romano et al., 2010].

Os coeficientes \mathbf{w} resultantes são então empregados como inicialização $\mathbf{w}(0)$ para o CMA, cuja convergência para um mínimo global dependerá essencialmente de quão próximo $\mathbf{w}(0)$ está de \mathbf{w}_{CMg} .

2.3.2 Resultados

A fim de verificar o desempenho das heurísticas de inicialização, consideraremos quatro tipos de canais: um somente com polos na função de transferência e três somente com zeros (de primeira, segunda e quinta ordem). O equalizador adotado possui sempre, por simplicidade, a mesma ordem do canal. Em todos os casos, as heurísticas HTQ e HTC são

comparadas a dois outros métodos: a inicialização padrão center-spike [Foschini, 1985] e uma inicialização aleatória baseada em coeficientes com distribuição uniforme entre -1 e $+1$. O desempenho de todos os métodos é analisado por meio do resultado da convergência do CMA, com $\mu = 0,0008$, para cada inicialização, depois de 30.000 iterações (a fim de garantir que o CMA convergiu para alguma solução). Assume-se a convergência global quando a distância euclidiana entre os coeficientes obtidos pelo CMA e os coeficientes ótimos \mathbf{w}_{CMg} – assumidos serem o resultado da convergência do CMA inicializado na melhor solução de Wiener – for menor que 0,1 (descontando-se a ambiguidade de sinal).

No primeiro caso, para um sinal *i.i.d.* de modulação BPSK, o canal dotado somente de polos possui a seguinte função de transferência:

$$H(z) = \frac{1}{1 + \alpha z^{-1}}, \quad (2.31)$$

onde α é percorrido de $-0,9$ a $0,9$ em passos de $0,041$. Este caso, em particular, admite uma solução para os coeficientes do equalizador que é capaz de inverter perfeitamente o canal, ou seja, é possível de se alcançar a solução ZF. A partir das inicializações geradas para o CMA, foi possível constatar a taxa de convergência global e seu associado valor médio do custo CM, conforme apresentado na Tab. 2.1. O primeiro ponto que podemos observar dos

Tabela 2.1: Taxa de convergência global e valor médio do custo CM para canais somente com polos.

Inicialização	Taxa de Convergência Global	Valor Médio do Custo CM
HTQ	100%	0
HTC	100%	0
Center-Spike	100%	0
Aleatória	47,73%	0,1853

resultados é o fato de que ambas as heurísticas atingiram 100% de convergência global com valor de custo CM nulo. Isto demonstra que o mapeamento $\boldsymbol{\theta}_o \rightarrow \mathbf{w}$ realizado por cada uma das heurísticas extrai informações coerentes dos parâmetros $\boldsymbol{\theta}_o$. Em uma análise mais cautelosa, observamos que o mapeamento das heurísticas leva diretamente à solução do mínimo global \mathbf{w}_{CMg} , isto é, não é necessário inicializar o CMA para atingir a convergência global, uma vez que a saída das heurísticas já oferece a solução desejada. Vale ressaltar que este é um caso específico de inversão perfeita do canal, com custo CM mínimo nulo (conforme vimos na Fig. 2.3), o que faz com que haja uma correspondência ótima entre os coeficientes $\boldsymbol{\theta}_o$ e \mathbf{w} , como apontado pela Eq. (2.28) – motivo pelo qual ambas as heurísticas provêm soluções equivalentes.

Outro ponto a destacar desses resultados é o bom desempenho da inicialização center-

spike. Como não há um coeficiente central no vetor de pesos do equalizador (possui apenas dois coeficientes), a inicialização escolhida foi $\mathbf{w}(0) = [1 \ 0]^T$, o que favoreceu a convergência global em relação ao tipo de canal que se lidava. Para a inicialização aleatória, a sorte não foi a mesma, atingindo uma taxa de convergência global em torno de 50% com erro residual não nulo.

Partiremos, agora, para cenários em que a estrutura do equalizador não permite uma inversão perfeita do canal. Para um sinal BPSK, o canal possui a seguinte função de transferência

$$H(z) = 1 + \alpha z^{-1}, \quad (2.32)$$

onde α é percorrido de 0 a 3 em passos de 0,041 (sendo $H(z)$ normalizado em seguida). A Tab. 2.2 mostra as taxas de convergência global para todos os quatro métodos e também o correspondente valor médio referente ao custo CM (2.1). Dos resultados obtidos, a HTC

Tabela 2.2: Taxa de convergência global e valor médio do custo CM para canais de primeira ordem.

Inicialização	Taxa de Convergência Global	Valor Médio do Custo CM
HTQ	98,65%	0,1958
HTC	100%	0,1957
Center-Spike	32,43%	0,3193
Aleatória	39,19%	0,2888

alcançou o melhor desempenho possível garantindo 100% de convergência global, seguido logo pela HTQ, que mostrou um desempenho bastante próximo, ao passo que os outros dois métodos mantiveram uma taxa inferior a 40%. Uma análise do valor médio do custo CM também mostra expressiva vantagem no uso das duas heurísticas.

Em outro cenário, o canal possui ordem dois, com a seguinte função de transferência:

$$H(z) = 1 + \alpha z^{-1} + \beta z^{-2}, \quad (2.33)$$

em que α e β vão de 0 a 2 a passos de tamanho 0,041 (com subsequentemente normalização de $H(z)$). Os resultados estão apresentados na Tab. 2.3, cuja análise também contribui para a validade das heurísticas, principalmente a HTC, cuja taxa de convergência é duas vezes maior que a taxa obtida pelos métodos center-spike e inicialização aleatória. O bom desempenho da HTC também espelha-se no seu reduzido valor médio do custo CM, sendo o menor dentre os casos observados. De maneira curiosa, a inicialização aleatória, apesar de ter uma taxa de convergência pouco superior ao método center-spike, possui valor médio do custo CM superior a este. Isto pode ser explicado pelo fato de que as condições iniciais aleatórias, para casos em que não há convergência para o mínimo global, podem conduzir

o CMA para mínimos mais inadequados em termos de custo CM, aumentando, assim, seu valor médio.

Tabela 2.3: Taxa de convergência global e valor médio do custo CM para canais de segunda ordem.

Inicialização	Taxa de Convergência Global	Valor Médio do Custo CM
HTQ	56,64%	0,3498
HTC	78,22%	0,3243
Center-Spike	35,44%	0,3673
Aleatória	37,61%	0,3809

Para o canal de quinta ordem, devido à dimensão do problema em mãos, geramos 100 canais cujos coeficientes foram obtidos aleatoriamente de acordo com uma distribuição uniforme entre -1 e $+1$. As taxas de convergência global e seus valores médios do custo CM, apresentados na Tab. 2.4, indicam mais uma vez a relevância prática das heurísticas, sobretudo a HTC. As inicializações center-spike e aleatória atingem uma convergência global menor que 20%. De uma maneira geral, observamos, mais uma vez, a queda generalizada nas taxas de convergência global em relação ao cenário anterior. Isso se justifica pelo aumento da dimensão do problema: um maior número de coeficientes para o filtro implica em maiores chances de se desviar da solução desejada.

Tabela 2.4: Taxa de convergência global e valor médio do custo CM para canais de quinta ordem.

Inicialização	Taxa de Convergência Global	Valor Médio do Custo CM
HTQ	31%	0,4090
HTC	40%	0,3954
Center-Spike	17%	0,4317
Aleatória	13%	0,4453

Para analisar também o desempenho da heurística frente à sinais com modulação multinível, repetimos a simulação para o caso com canal de segunda ordem porém com sinais 4-PAM e 8-PAM. Os resultados estão mostrados nas Tabs. 2.5 e 2.6. De uma maneira geral, em relação ao caso BPSK (Tab. 2.3), notamos uma sensível redução na taxa de convergência global para ambas as heurísticas, um reflexo direto do maior distanciamento entre θ_o e θ_{CMg} (como observado no limitante inferior). Mesmo assim, o aprimoramento de desempenho observado frente os demais métodos de inicialização ainda é notável. Vale observar também que o aumento do número de pontos na constelação do sinal modulado faz com que

o valor médio do custo CM se eleva cada vez mais para todos os métodos de inicialização, como esperado. Uma atenção especial pode ser ainda dedicada à HTC, que conseguiu se sobressair em todos os casos avaliados, tanto em termos de taxa de convergência global, como em valor médio do custo CM.

Tabela 2.5: Taxa de convergência global e valor médio do custo CM para canais de segunda ordem e modulação 4-PAM.

Inicialização	Taxa de Convergência Global	Valor Médio do Custo CM
HTQ	47,02%	0,4563
HTC	69,01%	0,4392
Center-Spike	35,03%	0,4604
Aleatória	32,69%	0,4703

Tabela 2.6: Taxa de convergência global e valor médio do custo CM para canais de segunda ordem e modulação 8-PAM.

Inicialização	Taxa de Convergência Global	Valor Médio do Custo CM
HTQ	43,82%	0,5009
HTC	63,56%	0,4896
Center-Spike	34,40%	0,5027
Aleatória	30,07%	0,5108

2.4 Conclusão

Foram apresentados neste capítulo algumas extensões da abordagem inicialmente proposta em [Sousa Júnior, 2011]. Partimos da derivação da formulação polinomial do critério CM e, por meio da relaxação das restrições sobre o vetor de parâmetros do filtro polinomial, chegou-se a um novo critério, capaz de estabelecer vínculos fortes com a teoria de Wiener.

Em especial, a partir desta segunda formulação, intitulada formulação sem restrições, foi possível obter um limitante inferior para a função custo CM, mesmo no domínio dos complexos, situação ainda não explorada no trabalho original [Sousa Júnior, 2011]. Este resultado foi alcançado graças à existência de uma solução fechada também nessa nova abordagem.

À luz das conexões com os custos de Wiener e CM, avaliamos a interpretação do limitante inferior como um índice de equalizabilidade. Além de analisarmos casos similares àqueles apresentados na proposição inicial, estendemos o repertório de aplicações para modulações multinível e também para cenários inseridos no domínio complexo.

Os resultados revelaram que o índice de equalizabilidade não supervisionado é válido para uma ampla gama de cenários, e é capaz prover informações relevantes sobre o tipo de canal que se lida e também qual o maior desempenho passível de ser alcançado por uma estrutura FIR. Para modulações multinível, tanto no caso real quanto no complexo, notamos o surgimento de um *offset* somado ao índice, mas que não altera as informações relativas sobre a equalizabilidade, o que confirma a solidez da métrica.

Por fim, propomos uma nova heurística de inicialização para o CMA, baseando-se naquela apresentada em [Sousa Júnior, 2011]. Em uma análise comparativa que também inclui os métodos de inicialização center-spike e aleatória, a heurística apresentada mostrou um aumento na taxa de convergência global, porém, da mesma forma que as demais metodologias de inicialização, o desempenho tende a reduzir à medida que se aumenta a ordem do equalizador. Ainda assim, foi verificado que o desempenho da nova heurística é cerca de duas vezes melhor que o método center-spike, em termos de taxa de convergência global.

Num âmbito geral, observamos que a formulação sem restrições pôde dar luz à contribuições importantes, como o índice de equalizabilidade e a heurística de inicialização do CMA. Mesmo assim, um passo adicional pode ser dado se tentarmos reincorporar nessa formulação as restrições anteriormente abandonadas, o que é plausível de ser alcançado se adotarmos o método dos multiplicadores de Lagrange. Este será o assunto do próximo capítulo.

Técnica para Ajuste dos Parâmetros das Restrições Quadráticas

A versão irrestrita da formulação polinomial do critério CM apresentada no capítulo anterior mostrou haver certa relação entre os mínimos por ela originados e os ótimos globais da função custo CM. Nesta análise, que baseou-se no trabalho de [Sousa Júnior, 2011], figuraram também como pano de fundo os vínculos entre os mínimos CM e as melhores soluções do critério de Wiener, inicialmente estudadas em [Zeng et al., 1998, 1999; Schniter e Johnson Jr., 2000].

Em especial, esta última relação serviu como motivação para que novos estudos, sob uma perspectiva mais teórica, fossem conduzidos. Destacamos, como importante passo dado nessa direção, a análise conduzida em [Suyama, 2003], em que, a fim de se alcançar uma relação mais precisa entre as posições dos mínimos do critério CM e das soluções de Wiener, buscou-se uma expressão capaz de retratar de forma analítica as soluções para o critério CM. De forma bastante interessante, esta proposição resgata algumas definições da formulação polinomial sem restrições por meio do seu tratamento como um problema de minimização sujeito a uma restrição.

Nesse sentido, a abordagem seguida em [Suyama, 2003] permite restaurar, com os parâmetros no domínio de Volterra [Mathews e Sicuranza, 2000], o vínculo direto com o critério CM através da obtenção dos coeficientes do filtro em cada um dos pontos onde o custo CM apresenta derivada nula. Entretanto, a partir da análise feita, não é possível estabelecer, de maneira simples e imediata, uma distinção clara entre quais soluções associam-se aos mínimos globais, mínimos locais ou pontos de sela do custo CM. Em particular, seria bastante interessante, tendo em vista o melhor desempenho atingível pelo critério, identificar as soluções vinculadas aos mínimos globais, o que certamente contribuiria, sob um prisma mais prático, para a obtenção de maiores taxas de convergência global para o CMA.

Neste capítulo, de forma a dar continuidade aos esforços relatados em [Suyama, 2003],

buscaremos utilizar o método dos multiplicadores de Lagrange para identificar as soluções ótimas do critério CM, bem como propor, de um ponto de vista mais prático, uma técnica iterativa para obtenção dessa solução. Destacamos, todavia, que o estudo ainda não foi plenamente desenvolvido, havendo espaço para posterior aprofundamento das análises teóricas e práticas. Mesmo assim, os resultados alcançados até o momento, como será mostrado, são relevantes e, ademais, férteis em perspectivas de trabalhos futuros.

3.1 Formulação Polinomial do Critério CM Usando Restrições Quadráticas

Conforme apresentado no Cap. 2, vimos que a formulação polinomial irrestrita (2.17) permite o cômputo da solução ótima em forma fechada (2.19), nos moldes da teoria de Wiener. Todavia, exigiu-se em contrapartida o abandono das restrições sobre o filtro polinomial que asseguravam a representação fidedigna do critério CM. Isso fez com que houvesse, como verificado na análise do índice de equalizabilidade e também na heurística de inicialização, uma maior disparidade – de forma mais acentuada em alguns casos – entre as soluções desejadas de cada formulação.

Como forma de minorar o distanciamento observado, encontramos no trabalho de [Suyama, 2003] uma abordagem capaz de retratar o problema como uma otimização sujeita a restrições, passível de ser adotada através do método dos multiplicadores de Lagrange. Torna-se então viável conciliar a conveniência da existência de uma solução fechada à representação fiel do critério CM. Veremos, a seguir, os detalhes dessa formulação.

Iremos nos ater, primeiramente, a um caso estruturalmente simples em que o equalizador possui apenas dois coeficientes, para, posteriormente, estendermos a formulação para o caso genérico do equalizador de ordem N .

Em vista da dificuldade no mapeamento complexo dos coeficientes no domínio de Volterra para o domínio original do equalizador, restringiremos nossa análise apenas ao caso em que os sinais e sistemas incluem-se no conjunto dos reais \mathcal{R} .

3.1.1 Equalizador de Primeira Ordem

Assumindo que sinais e sistemas estão contidos no domínio dos reais \mathcal{R} , consideramos inicialmente um equalizador de ordem $N = 1$, i.e., com dois coeficientes: w_0 e w_1 . À luz das formulações polinomiais restrita e irrestrita, Eqs. (2.5) e (2.17), respectivamente, tem-se que esse caso leva à seguinte relação entre os coeficientes dos filtros polinomiais em cada

abordagem:

$$\begin{aligned} \boldsymbol{\theta}_r^T &= [w_0^2 \quad 2w_0w_1 \quad w_1^2] \\ &\quad \Downarrow \quad \Downarrow \quad \Downarrow \\ \boldsymbol{\theta}^T &= [\theta_0 \quad \theta_1 \quad \theta_2]. \end{aligned} \quad (3.1)$$

Os três elementos que compõem $\boldsymbol{\theta}_r$ representam as restrições dos parâmetros do filtro polinomial com respeito aos coeficientes \mathbf{w} do equalizador. Por sua vez, os valores de \mathbf{w} em $\boldsymbol{\theta}_r$ podem ser associados segundo uma relação quadrática:

$$(2w_0w_1)^2 = 4(w_0^2)(w_1^2), \quad (3.2)$$

o que corresponderia, tendo em vista o vínculo entre os elementos de $\boldsymbol{\theta}_r$ e $\boldsymbol{\theta}$ mostrados na Eq. (3.1), a

$$\theta_1^2 = 4\theta_0\theta_2. \quad (3.3)$$

Podemos observar que a relação entre os coeficientes de $\boldsymbol{\theta}$ mostrada pela Eq. (3.3) somente será válida se $\boldsymbol{\theta} = \boldsymbol{\theta}_r$. Dessa forma, a expressão contribui diretamente para a definição de uma restrição sobre $\boldsymbol{\theta}$ capaz de refazer os laços com o critério CM.

Podemos reescrever a Eq. (3.3) como:

$$\begin{aligned} G(\boldsymbol{\theta}) &= \theta_1^2 - 4\theta_0\theta_2 = 0 \\ &= \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 \end{bmatrix} \begin{bmatrix} 0 & 0 & -2 \\ 0 & 1 & 0 \\ -2 & 0 & 0 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = 0 \\ &= \boldsymbol{\theta}^T \mathbf{C} \boldsymbol{\theta} = 0, \end{aligned} \quad (3.4)$$

em que $G(\boldsymbol{\theta})$ é chamada de função restrição e \mathbf{C} de matriz de restrição (com dimensões $(M+1) \times (M+1)$). Isto nos permite definir o problema de minimização com restrições da seguinte forma:

$$\begin{aligned} \min_{\boldsymbol{\theta}} J_{LI}(\boldsymbol{\theta}) &= \min_{\boldsymbol{\theta}} E [(v(n) - 1)^2] \\ \text{sujeito a } G(\boldsymbol{\theta}) &= \boldsymbol{\theta}^T \mathbf{C} \boldsymbol{\theta} = 0, \end{aligned} \quad (3.5)$$

em que uma solução pode ser buscada através do método de multiplicadores de Lagrange [Dorny, 1975].

Método dos Multiplicadores de Lagrange

Tendo em vista o problema apresentado em (3.5), a otimização com restrições pode ser realizada relacionando-se a função custo $J_{LI}(\boldsymbol{\theta})$ e a função restrição $G(\boldsymbol{\theta})$ através da

expressão:

$$\frac{\partial J_{LI}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \lambda \frac{\partial G(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \quad (3.6)$$

em que λ é o chamado *multiplicador de Lagrange*.

Calculando os gradientes, temos:

$$\begin{aligned} \frac{\partial J_{LI}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \lambda \frac{\partial G(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \mathbf{0} \\ (2\mathbf{R}_\xi \boldsymbol{\theta} - 2\mathbf{p}_\xi) - (2\lambda \mathbf{C} \boldsymbol{\theta}) &= \mathbf{0} \\ (\mathbf{R}_\xi - \lambda \mathbf{C}) \boldsymbol{\theta} - \mathbf{p}_\xi &= \mathbf{0}, \end{aligned} \quad (3.7)$$

o que nos leva à seguinte solução

$$\boldsymbol{\theta}_\lambda = (\mathbf{R}_\xi - \lambda \mathbf{C})^{-1} \mathbf{p}_\xi. \quad (3.8)$$

Obtemos assim uma solução em forma fechada para o problema restrito, que nos remete à solução de Wiener, e, ao mesmo tempo, é capaz de portar informações estatísticas de até quarta ordem em respeito ao sinal recebido através da matriz de correlação \mathbf{R}_ξ . Todavia, existe uma dependência explícita sobre os valores passíveis de serem assumidos pelo multiplicador de Lagrange λ , que funciona como um termo de correção sobre a matriz de correlação \mathbf{R}_ξ no domínio de Volterra.

De posse dessa relação, podemos exprimir o valor do custo em função de $\boldsymbol{\theta}_\lambda$ (e, por conseguinte, de λ). Porém, convém antes reescrevermos a Eq. (3.7) como:

$$\begin{aligned} (\mathbf{R}_\xi - \lambda \mathbf{C}) \boldsymbol{\theta}_\lambda &= \mathbf{p}_\xi \\ \mathbf{R}_\xi \boldsymbol{\theta}_\lambda - \lambda \mathbf{C} \boldsymbol{\theta}_\lambda &= \mathbf{p}_\xi, \end{aligned} \quad (3.9)$$

multiplicando ambos os lados por \mathbf{R}_ξ^{-1} , temos

$$\begin{aligned} \mathbf{R}_\xi^{-1} \mathbf{R}_\xi \boldsymbol{\theta}_\lambda - \lambda \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda &= \mathbf{R}_\xi^{-1} \mathbf{p}_\xi \\ \boldsymbol{\theta}_\lambda - \lambda \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda &= \boldsymbol{\theta}_o \\ \boldsymbol{\theta}_\lambda &= \boldsymbol{\theta}_o + \lambda \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda, \end{aligned} \quad (3.10)$$

onde usou-se a relação (2.19) no segundo passo.

Finalmente, com a Eq. (3.10) em mãos, podemos obter o valor da função custo $J_{LI}(\boldsymbol{\theta}_\lambda)$ através de (2.18). Considerando apenas valores reais para $\boldsymbol{\theta}_\lambda$, tem-se

$$\begin{aligned} J_{LI}(\boldsymbol{\theta}_\lambda) &= \boldsymbol{\theta}_\lambda^T \mathbf{R}_\xi \boldsymbol{\theta}_\lambda - 2\boldsymbol{\theta}_\lambda^T \mathbf{p}_\xi + 1 \\ &= (\boldsymbol{\theta}_o + \lambda \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T \mathbf{R}_\xi (\boldsymbol{\theta}_o + \lambda \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda) - 2(\boldsymbol{\theta}_o + \lambda \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T \mathbf{p}_\xi + 1 \\ &= J_{LI}(\boldsymbol{\theta}_o) + \lambda (\boldsymbol{\theta}_\lambda^T - \mathbf{p}_\xi^T \mathbf{R}_\xi^{-1}) \mathbf{C} \boldsymbol{\theta}_\lambda. \end{aligned} \quad (3.11)$$

A dedução completa pode ser conferida no Apêndice B.

Observamos pela Eq. (3.11) que $J_{LI}(\boldsymbol{\theta}_\lambda)$ é o próprio valor do limitante inferior $J_{LI}(\boldsymbol{\theta}_o)$ acrescido de um termo dependente de λ . Sob outro prisma, podemos dizer que o segundo termo da Eq. (3.11) atua como um fator de correção sobre $J_{LI}(\boldsymbol{\theta}_o)$ para que $J_{LI}(\boldsymbol{\theta}_\lambda)$ seja uma representação fidedigna dos valores do custo CM. No entanto, isto será verdade apenas se λ for ajustado corretamente.

Antes de focarmos nossa análise sobre o multiplicador de Lagrange, apresentaremos a formulação do problema para o equalizador de ordem N .

3.1.2 Equalizador de Ordem N

A extensão para o caso genérico com equalizador de ordem N não se dá de maneira trivial, pois engendra na formulação múltiplas restrições quadráticas, como aquela definida em (3.4). Veremos a seguir como proceder neste caso.

De maneira similar ao equalizador de primeira ordem, podemos fazer a associação entre os coeficientes dos filtros polinomiais restrito e irrestrito da seguinte forma:

$$\begin{aligned} \boldsymbol{\theta}_r^T &= [\quad w_0^2 \quad 2w_0w_1 \quad 2w_0w_2 \quad \dots \quad w_N^2 \quad] \\ &\quad \Downarrow \quad \Downarrow \quad \Downarrow \quad \quad \quad \Downarrow \\ \boldsymbol{\theta}^T &= [\quad \theta_0 \quad \theta_1 \quad \theta_2 \quad \dots \quad \theta_{N(N+3)/2} \quad]. \end{aligned} \quad (3.12)$$

Devido ao maior número de coeficientes, se fizermos associações como aquela dada pela Eq. (3.2), teremos $N(N+1)/2$ relações quadráticas:

$$\begin{aligned} (2w_0w_1)^2 &= 4(w_0^2)(w_1^2) \\ (2w_0w_2)^2 &= 4(w_0^2)(w_2^2) \\ &\quad \vdots \\ (2w_{N-1}w_N)^2 &= 4(w_{N-1}^2)(w_N^2), \end{aligned} \quad (3.13)$$

implicando em $N(N+1)/2$ restrições do tipo:

$$\begin{aligned} \theta_1^2 &= 4\theta_0\theta_2 \\ \theta_3^2 &= 4\theta_0\theta_{N+1} \\ &\quad \vdots \\ \theta_{(N(N+3)/2)-1}^2 &= 4\theta_{(N(N+3)/2)-2}\theta_{N(N+3)/2}. \end{aligned} \quad (3.14)$$

Empregando a notação matricial usada em (3.4), podemos expressar as restrições como:

$$G_i(\boldsymbol{\theta}) = \boldsymbol{\theta}^T \mathbf{C}_i \boldsymbol{\theta} = 0, \quad \text{para } i = 1, 2, \dots, N(N+1)/2 \quad (3.15)$$

em que \mathbf{C}_i são as matrizes de restrição, com dimensões $(M+1) \times (M+1)$ – lembrando que $M = N(N+3)/2$ no caso real.

Dessa forma, o problema pode ser posto como:

$$\begin{aligned} \min_{\boldsymbol{\theta}} J_{LI}(\boldsymbol{\theta}) &= \min_{\boldsymbol{\theta}} E[(v(n) - 1)^2] \\ \text{sujeito a } G_i(\boldsymbol{\theta}) &= \boldsymbol{\theta}^T \mathbf{C}_i \boldsymbol{\theta} = 0, \end{aligned} \quad (3.16)$$

para $i = 1, 2, \dots, N(N+1)/2$. Novamente, a solução pode ser obtida pelo método dos multiplicadores de Lagrange; contudo, deve-se levar em conta todas as restrições.

Método dos Multiplicadores de Lagrange com Múltiplas Restrições

A resolução do problema apresentado na Eq. (3.16) utiliza o mesmo princípio de otimização para o caso de uma única restrição, mas adota para cada restrição um multiplicador de Lagrange. Assim, a expressão a ser resolvida agora é:

$$\frac{\partial J_{LI}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \sum_{i=1}^{N(N+1)/2} \lambda_i \frac{\partial G_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}. \quad (3.17)$$

Através do cálculo do gradiente, chegamos à seguinte relação:

$$\boldsymbol{\theta}_\lambda = \left(\mathbf{R}_\xi - \sum_{i=1}^{N(N+1)/2} \lambda_i \mathbf{C}_i \right)^{-1} \mathbf{p}_\xi, \quad (3.18)$$

que, analogamente à Eq. (3.10), também pode ser expressa como

$$\boldsymbol{\theta}_\lambda = \boldsymbol{\theta}_o + \mathbf{R}_\xi^{-1} \left(\sum_{i=1}^{N(N+1)/2} \lambda_i \mathbf{C}_i \right) \boldsymbol{\theta}_\lambda, \quad (3.19)$$

cujo valor de custo associado, como na Eq. (3.11), é

$$J_{LI}(\boldsymbol{\theta}_\lambda) = J_{LI}(\boldsymbol{\theta}_o) + (\boldsymbol{\theta}_\lambda^T - \mathbf{p}_\xi^T \mathbf{R}_\xi^{-1}) \left(\sum_{i=1}^{N(N+1)/2} \lambda_i \mathbf{C}_i \right) \boldsymbol{\theta}_\lambda. \quad (3.20)$$

É importante notar que a validade das Eqs. (3.18) a (3.20) dependem do correto ajuste de todos os valores de λ_i , o que pode ser demasiadamente complicado em vista da dimensão do problema. Veremos na próxima seção uma maneira de identificar os valores desejados de λ_i e, a partir das características gerais observadas nos resultados, proporemos uma técnica que efetua seu ajuste ótimo de forma mais simples.

3.2 Técnica de Obtenção dos Coeficientes de Lagrange

Conforme mostrado na seção anterior, foi possível alcançar uma expressão em forma fechada que determina a solução para a formulação polinomial com restrições – Eq. (3.18). Sua importância vem do fato de sintetizar numa relação razoavelmente simples os pontos de equilíbrio do critério CM. Contudo, em contrapartida, existem agora os coeficientes de Lagrange λ_i que precisam ser ajustados de forma a garantir a validade das relações.

De uma maneira intuitiva, podemos dizer que os valores λ_i desejados são aqueles que satisfazem as restrições do problema e, concomitantemente, minimizam o custo sem restrição $J_{LI}(\boldsymbol{\theta})$, conforme estabelecido na Eq. (3.17). Entretanto, como mostrado em [Suyama, 2003], estes valores não são únicos, e cada um aponta para diferentes mínimos e pontos de sela da função custo CM. Mais especificamente, devido à ambiguidade de sinal inerente às restrições quadráticas na formulação, cada solução em λ_i estará associada a duas possíveis soluções para o critério CM no domínio do filtro original \mathbf{w} .

Nesta seção, como uma extensão dos resultados observados em [Suyama, 2003], buscaremos explorar dois tópicos pertinentes: identificar a solução em λ_i que corresponde aos mínimos globais do critério CM e, como consequência das particularidades concernentes à esta solução, propor uma técnica para alcançar estes parâmetros λ_i , mesmo para casos de equalizadores de ordem maior que um. Enfatizamos que tais abordagens ainda não foram analisadas de maneira completa, principalmente em vista das dificuldades inerentes à complexidade do problema, como veremos em breve. Não obstante, a relevância dos resultados alcançados são bastante favoráveis à proposta, sendo mesmo possível vislumbrar possibilidades concretas para o surgimento de novos algoritmos que assegurem a convergência global do CMA.

Partiremos do estudo do caso de um equalizador de primeira ordem. Isso nos permitirá tecer importantes constatações para o desenvolvimento da técnica proposta. A sequência natural é a extensão para casos de maior ordem, em que consideraremos, como exemplo, um equalizador de segunda ordem.

Por simplicidade de notação, assumiremos que $\boldsymbol{\lambda}$ é o vetor contendo os multiplicadores de Lagrange $[\lambda_1, \lambda_2, \dots, \lambda_{N(N+1)/2}]$.

3.2.1 Identificação do Coeficiente Ótimo de Lagrange

Na formulação polinomial sem restrições, vimos que os coeficientes do filtro polinomial $\boldsymbol{\theta}$ podem assumir, a princípio, quaisquer valores, pois não carregam nenhum tipo de limitação (explícita ou implícita). Por outro lado, no caso restrito, as limitações são claras, de forma que os valores possíveis para $\boldsymbol{\theta}_r$ constituem um subconjunto dos possíveis valores de $\boldsymbol{\theta}$. Nesse sentido, a ideia por trás do emprego do método dos multiplicadores de Lagrange vai

justamente na direção de reduzir o conjunto dos possíveis valores de $\boldsymbol{\theta}$ para aqueles de $\boldsymbol{\theta}_r$ através do ajuste pelos coeficientes $\boldsymbol{\lambda}$. Entretanto, como o processo de otimização também é englobado pela abordagem, $\boldsymbol{\lambda}$ deve satisfazer, simultaneamente, a minimização do custo $J_{LI}(\boldsymbol{\theta})$ e a própria restrição dada pela Eq. (3.16). Isto torna o escopo de valores desejados para $\boldsymbol{\lambda}$ bastante reduzido.

Um melhor entendimento do comportamento de $\boldsymbol{\lambda}$ pode ser alcançado se conseguirmos conciliar em uma só expressão a minimização de $J_{LI}(\boldsymbol{\theta})$ e a restrição quadrática. Dado que existe solução fechada para a formulação com restrições quadráticas, a expressão almejada ganha corpo se fizermos a substituição de (3.18) em (3.15), que leva a:

$$\begin{aligned} G_i(\boldsymbol{\theta}_\lambda) &= \boldsymbol{\theta}_\lambda^T \mathbf{C}_i \boldsymbol{\theta}_\lambda = 0 \\ &= \mathbf{p}_\xi^T \left(\mathbf{R}_\xi - \sum_{j=1}^{N(N+1)/2} \lambda_j \mathbf{C}_j \right)^{-T} \mathbf{C}_i \left(\mathbf{R}_\xi - \sum_{j=1}^{N(N+1)/2} \lambda_j \mathbf{C}_j \right)^{-1} \mathbf{p}_\xi = 0, \end{aligned} \quad (3.21)$$

para $i = 1, 2, \dots, N(N+1)/2$.

Por abranger todas as condições necessárias estabelecidas na definição do problema, a Eq. (3.21) é capaz de exprimir quais são os valores desejados para $\boldsymbol{\lambda}$. Porém, sob um prisma mais cauteloso, é possível verificar que (3.21) trata-se de um sistema de equações em função dos multiplicadores $\boldsymbol{\lambda}$ cuja solução não é obtida facilmente [Suyama, 2003]. No entanto, a situação torna-se mais simples para o caso de um equalizador de ordem $N = 1$.

Equalizador de Primeira Ordem

O equalizador de dois coeficientes possui a vantagem de apresentar apenas uma restrição, simplificando o problema em mãos. Nesse caso, a Eq. (3.21) torna-se:

$$\begin{aligned} G(\boldsymbol{\theta}_\lambda) &= \boldsymbol{\theta}_\lambda^T \mathbf{C} \boldsymbol{\theta}_\lambda = 0 \\ &= \mathbf{p}_\xi^T (\mathbf{R}_\xi - \lambda \mathbf{C})^{-T} \mathbf{C} (\mathbf{R}_\xi - \lambda \mathbf{C})^{-1} \mathbf{p}_\xi = 0, \end{aligned} \quad (3.22)$$

que, se expandida, revela um polinômio de quarta ordem em função de λ como numerador e um polinômio de sexta ordem em λ como denominador [Suyama, 2003]. Para anularmos esse quociente, ou o denominador tende ao infinito, o que é obtido somente quando $\lambda \rightarrow \pm\infty$, ou o numerador é equivalente a zero. Como o primeiro caso não apresenta uma solução de interesse (está associado ao máximo da função custo CM), o maior foco será destinado ao polinômio do numerador, que possui quatro raízes [Suyama, 2003].

Para fins ilustrativos, tomaremos como exemplo o seguinte canal linear de fase mínima:

$$H(z) = 1 + 0,6z^{-1}. \quad (3.23)$$

Nesse cenário, estimamos a matriz de correlação \mathbf{R}_ξ e o vetor de correlação cruzada \mathbf{p}_ξ a partir de 30.000 amostras *i.i.d.* de um sinal BPSK, sem ruído. Usando a matriz de restrição \mathbf{C} definida em (3.4), percorremos λ de -6 a 6 em intervalos de um centésimo para a obtenção de $G(\boldsymbol{\theta}_\lambda)$. No entanto os valores obtidos apresentaram grande faixa, sendo a visualização prejudicada.

Para contornar este problema, fizemos uso de uma função de suavização capaz de atenuar os valores demasiadamente elevados:

$$G_s(\boldsymbol{\theta}_\lambda) = \frac{1}{1 + G^2(\boldsymbol{\theta}_\lambda)}. \quad (3.24)$$

No entanto, destacamos que, para $G(\boldsymbol{\theta}_\lambda) = 0$ (situação desejada), temos $G_s(\boldsymbol{\theta}_\lambda) = 1$, que é também o valor máximo de $G_s(\boldsymbol{\theta}_\lambda)$. Essa nova perspectiva deve ser considerada na análise.

A Fig. 3.1(a) mostra o comportamento de $G_s(\boldsymbol{\theta}_\lambda)$ variando-se λ para o cenário em questão. Nela, podemos observar um reflexo do perfil do polinômio de quarto grau, cujas raízes serão, como dito anteriormente, os casos em que $G_s(\boldsymbol{\theta}_\lambda) = 1$. Dessa forma, a região de interesse em λ é aquela compreendida entre 0 e 2. Faz-se importante notar também que, para valores de λ com magnitude mais elevada, a função $G_s(\boldsymbol{\theta}_\lambda)$ tende à unidade: será correspondente ao máximo da função custo CM (vetor \mathbf{w} nulo), associada à solução do polinômio de sexta ordem do denominador da Eq. (3.22), que, como dissemos, não fornece uma solução de interesse.

Voltando-nos novamente para a região $0 < \lambda < 2$, temos, na Fig. 3.1(b), quatro possíveis soluções para o multiplicador de Lagrange, denotadas por círculos coloridos, isto é, $\lambda_{(1)}, \dots, \lambda_{(4)}$. Esses são os pontos que, concomitantemente, atendem à restrição e anulam o gradiente de $J_{LI}(\boldsymbol{\theta})$. Tais soluções são passíveis de serem obtidas analiticamente [Suyama, 2003], mas, por simplicidade, optamos por atingir tais valores por meio de um detector de picos de $G_s(\boldsymbol{\theta}_\lambda)$, isto é, uma busca exaustiva unidimensional pelos máximos de $G_s(\boldsymbol{\theta}_\lambda)$.

O detector percorre um vetor que possui uma determinada faixa de valores escolhidos para λ , comparando, em seguida, os respectivos valores de $G_s(\boldsymbol{\theta}_\lambda)$. Dado um valor atual de λ correspondente à um dos elementos do vetor, compara-se seu associado valor de $G_s(\boldsymbol{\theta}_\lambda)$ com o daqueles de elementos adjacentes. Se, para o valor atual de λ , $G_s(\boldsymbol{\theta}_\lambda)$ for maior que o de seus adjacentes, o detector identifica um pico. A precisão deste método dependerá da escolha da faixa de valores de λ . Para o canal em questão, passos de um centésimo para λ entre 0 e 2 são suficientes. Após a detecção dos picos, temos as seguintes soluções para λ :

$$\begin{aligned} \lambda_{(1)} &= 0,29; \\ \lambda_{(2)} &= 0,61; \\ \lambda_{(3)} &= 1,30; \\ \lambda_{(4)} &= 1,50. \end{aligned} \quad (3.25)$$

Como passo seguinte, obtemos, para cada solução em λ , através da Eq. (3.8), o filtro polinomial θ_λ correspondente.

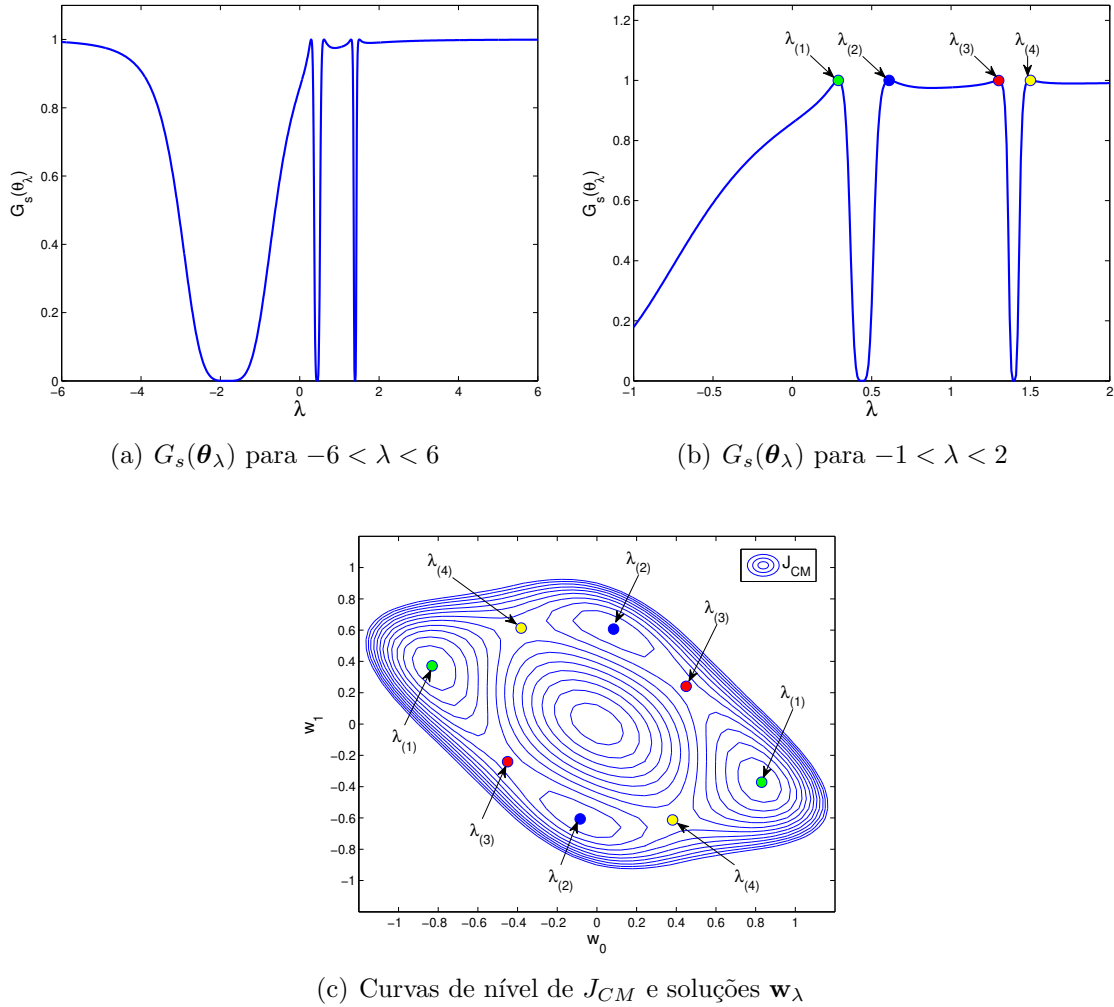


Figura 3.1: Análise de $G_s(\theta_\lambda)$ e soluções para λ em um canal $H(z) = 1 + 0,6z^{-1}$.

Resta-nos agora fazer o mapeamento $\theta_\lambda \rightarrow \mathbf{w}_\lambda$ que leva para o domínio original do equalizador. Para isso, pode-se empregar tanto a heurística dos termos quadráticos (HTQ, definido no Alg. 2.1) quanto a dos termos cruzados (HTC, Alg. 2.2), dado que, respeitadas as restrições, o mapeamento é equivalente em ambos os casos. A Fig. 3.1(c) mostra a posição dos coeficientes \mathbf{w}_λ (círculos coloridos) associados às diferentes soluções para λ , e as curvas de nível da função custo CM. Neste caso, consideramos a ambiguidade de sinal inerente ao critério CM para fazer o mapeamento, por isso a duplicidade de soluções para \mathbf{w}_λ . Observamos que $\lambda_{(1)}$ associa-se aos mínimos globais do critério CM, enquanto $\lambda_{(2)}$ se vincula aos mínimos locais. Aos pontos de sela se associam as raízes $\lambda_{(3)}$ e $\lambda_{(4)}$.

De maneira bastante interessante, associada aos mínimos globais do critério CM temos

a solução em λ com menor valor absoluto, isto é, $\lambda_{(1)}$. Em vista da Eq. (3.8), isso significa, intuitivamente, que os mínimos globais são aqueles que necessitam de uma menor correção sobre a matriz de correlação \mathbf{R}_ξ . De outro ponto de vista, em referência à Eq. (3.10), esta é “a solução em $\boldsymbol{\theta}_\lambda$ que menos se desvia de $\boldsymbol{\theta}_o$ ”. Isto nos remete à proximidade entre $\boldsymbol{\theta}_o$ e $\boldsymbol{\theta}_{CMg}$ discutida no capítulo anterior. Com base nesta constatação, podemos conjecturar, empiricamente, que serão sempre as soluções de menor valor $\lambda_{(1)}$ que estarão associadas aos mínimos ótimos do critério CM.

Teste de Identificação - Equalizador de Primeira Ordem

Para comprovarmos a validade da conjectura proposta, buscamos expandir os cenários de teste empregando canais do tipo:

$$H_1(z) = 1 + \alpha z^{-1}; \quad (3.26)$$

$$H_2(z) = 1 + \alpha z^{-1} + \beta z^{-2}, \quad (3.27)$$

em que α e β variam de 0 a 3, a passos de 0,041, sendo $H_1(z)$ e $H_2(z)$ subsequentemente normalizados. Usamos nas simulações 30.000 amostras de um sinal *i.i.d.* com modulação BPSK para estimação de \mathbf{R}_ξ e \mathbf{p}_ξ , e percorremos λ de -6 a 6 (em intervalos de um centésimo – com ressalvas para alguns casos, como comentado abaixo) para identificar os picos em $G_s(\boldsymbol{\theta}_\lambda)$.

Durante a execução da simulação, principalmente para canais cujos zeros encontravam-se bastante próximos à origem, foi necessário reduzir o tamanho de passo de varredura de λ para até 1×10^{-6} , pois havia grande proximidade entre os picos de $G_s(\boldsymbol{\theta}_\lambda)$, exigindo, portanto, maior precisão. Isso trará consequências para a técnica que iremos propor, conforme veremos adiante.

Para cada canal, as raízes $\lambda_{(1)}$, $\lambda_{(2)}$, $\lambda_{(3)}$ e $\lambda_{(4)}$ – i.e., as quatro soluções de interesse para λ – foram usadas para se obter os respectivos coeficientes $\mathbf{w}_{\lambda(1)}$, $\mathbf{w}_{\lambda(2)}$, $\mathbf{w}_{\lambda(3)}$ e $\mathbf{w}_{\lambda(4)}$ no domínio original do filtro (utilizando-se o algoritmo HTC). Como estes vinculam-se a pontos de mínimo e de sela da superfície do custo CM, para identificar aquele correspondente ao mínimo global \mathbf{w}_{CMg} , bastou verificar qual possui o menor custo CM associado. Assim, comparamos os valores de custo CM correspondentes, obtidos analiticamente através da expressão [Suyama, 2003]:

$$J_{CM}(\mathbf{c}_\lambda) = 1 - 2 \sum_{i=0}^{L-1} c_\lambda^2(i) + 3 \sum_{i=0}^{L-1} \sum_{j=0, j \neq i}^{L-1} c_\lambda^2(i) c_\lambda^2(j) + \sum_{i=0}^{L-1} c_\lambda^4(i), \quad (3.28)$$

em que $c_\lambda(n)$ é

$$c_\lambda(n) = \sum_{l=0}^{\infty} w_{l,\lambda} h(n-l), \quad (3.29)$$

ou seja, a resposta combinada canal+equalizador \mathbf{w}_λ , sendo L o comprimento máximo de $c_\lambda(n)$ (assumiu-se, como na formulação polinomial, $R_2 = 1$).

Dessa forma, verificamos se o menor valor de $J_{CM}(\mathbf{c}_\lambda)$ computado era de fato aquele associado à raiz $\lambda_{(1)}$. Em caso positivo, ter-se-ia uma evidência em favor da conjectura levantada.

Como resultado, para todos os tipos de canais de primeira e segunda ordem, as raízes de $G(\boldsymbol{\theta}_\lambda)$ apresentaram apenas valores positivos, e, com sucesso, as simulações apontaram uma relação direta entre a raiz $\lambda_{(1)}$ e o ótimo global \mathbf{w}_{CMg} em 100% dos casos.

Apesar de contarmos apenas com resultados empíricos, em que contemplamos somente canais lineares de primeira e segunda ordem, temos um forte indício de uma correspondência entre $\lambda_{(1)}$ e a solução ótima do critério CM. Por este motivo, chamaremos $\lambda_{(1)}$ de multiplicador ótimo de Lagrange, ou λ_o . Contudo, uma investigação mais profunda se faz necessária neste âmbito, tanto em termos teóricos quanto práticos, abrangendo, neste último caso, cenários variados e canais de diferentes ordens. Este estudo fica como perspectiva para trabalho futuro.

Tentaremos agora, seguindo a linha de raciocínio mostrada, tecer alguns comentários para o caso de um equalizador de segunda ordem.

Equalizador de Segunda Ordem

Considerando um equalizador de ordem $N = 2$, temos agora em mãos um maior número de restrições a serem consideradas, mais especificamente, três. Isto torna o problema mais complexo, prejudicando até mesmo uma possível obtenção de soluções analíticas para o caso em questão [Suyama, 2003]. Mesmo assim, alguns pontos ainda podem ser melhor explorados, como veremos a seguir.

Matematicamente, as relações capazes de gerar as soluções para $\boldsymbol{\lambda} = [\lambda_1 \ \lambda_2 \ \lambda_3]$ são definidas conforme a Eq. (3.21), que, neste caso, torna-se:

$$\begin{aligned} G_i(\boldsymbol{\theta}_\lambda) &= \boldsymbol{\theta}_\lambda^T \mathbf{C}_i \boldsymbol{\theta}_\lambda = 0 \\ &= \mathbf{p}_\xi^T \left(\mathbf{R}_\xi - \sum_{j=1}^3 \lambda_j \mathbf{C}_j \right)^{-T} \mathbf{C}_i \left(\mathbf{R}_\xi - \sum_{j=1}^3 \lambda_j \mathbf{C}_j \right)^{-1} \mathbf{p}_\xi = 0, \end{aligned} \quad (3.30)$$

para $i = 1, 2, 3$. Trata-se de um sistema não linear, sendo os valores de $\boldsymbol{\lambda}$ dependentes entre

si. As matrizes de restrição \mathbf{C}_i , como anteriormente, resultam das restrições $G_i(\boldsymbol{\theta})$:

$$\begin{aligned}
 G_1(\boldsymbol{\theta}) &= \theta_1^2 - 4\theta_0\theta_3 \\
 &= \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & -2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{bmatrix} \\
 &= \boldsymbol{\theta}^T \mathbf{C}_1 \boldsymbol{\theta} = 0;
 \end{aligned} \tag{3.31}$$

$$\begin{aligned}
 G_2(\boldsymbol{\theta}) &= \theta_2^2 - 4\theta_0\theta_5 \\
 &= \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{bmatrix} \\
 &= \boldsymbol{\theta}^T \mathbf{C}_2 \boldsymbol{\theta} = 0;
 \end{aligned} \tag{3.32}$$

$$\begin{aligned}
 G_3(\boldsymbol{\theta}) &= \theta_4^2 - 4\theta_3\theta_5 \\
 &= \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -2 & 0 & 0 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{bmatrix} \\
 &= \boldsymbol{\theta}^T \mathbf{C}_3 \boldsymbol{\theta} = 0.
 \end{aligned} \tag{3.33}$$

Caso seguissemos um procedimento similar ao visto para o equalizador de primeira ordem, deveríamos fazer uma varredura nos possíveis valores de λ_1 , λ_2 e λ_3 a fim de encontrar os nulos de (3.30). Isto exigiria um enorme esforço, pois significaria realizarmos uma busca exaustiva em três dimensões.

Para tentar tornar o problema mais tratável, buscaremos fazer algumas considerações e hipóteses acerca das restrições. Como indicado pela Fig. 3.2, as restrições quadráticas $G_i(\boldsymbol{\theta})$ asseguram as relações entre pares de coeficientes em \mathbf{w} – vide Eq. (3.13). Se forem satisfeitas duas dessas restrições, por exemplo, $G_1(\boldsymbol{\theta})$ e $G_3(\boldsymbol{\theta})$, isto significa que os pares $[w_0 \ w_1]$ e $[w_1 \ w_2]$ atendem às condições impostas na formulação. Todavia, esses pares

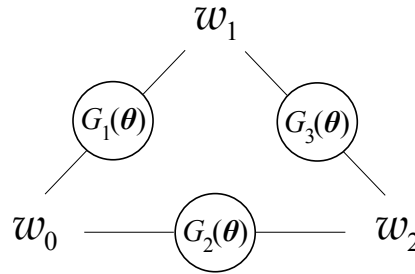


Figura 3.2: Relação entre os coeficientes no domínio original do equalizador e as correspondentes restrições.

possuem como elemento comum o coeficiente w_1 , que, por sua vez, deve respeitar ambas as restrições $G_1(\theta)$ e $G_3(\theta)$ simultaneamente. Para que isso seja verdade, uma das possíveis soluções é que o par $[w_0 \ w_2]$ atenda, implicitamente, à restrição $G_2(\theta)$. Nesse sentido, caso tal suposição seja atendida, podemos dizer que a restrição $G_2(\theta)$ é redundante.

Se supusermos que esta condição é sempre atendida, torna-se desnecessário considerar uma das restrições, o que simplifica consideravelmente o problema em mãos. Contudo, nem sempre isso é verdade.

Para entendermos melhor a ideia, podemos fixar qualquer valor para w_1 e ajustar w_0 e w_2 de acordo com as restrições $G_1(\theta)$ e $G_3(\theta)$, respectivamente. Nesse caso, a restrição $G_2(\theta)$ não necessariamente é atendida. É preciso que w_1 assuma um valor adequado de forma a propiciar também o atendimento a $G_2(\theta)$. Dessa maneira, a hipótese de que esta última restrição é redundante não é válida. Mesmo assim, para fins de simplificação do problema, assumiremos que ela seja verdadeira, pois, caso o valor de w_1 fixo seja próximo àquele que atende a $G_2(\theta)$, teremos uma solução também próxima à desejada.

De outro ponto de vista, com maior ênfase no ajuste dos parâmetros λ , podemos aprimorar o entendimento dessa suposição a partir de uma outra interpretação da Fig. 3.2: para cada restrição $G_i(\theta)$ existe um multiplicador de Lagrange λ_i associado, cujo valor deverá satisfazer a relação entre a tríplice $[w_0 \ w_1 \ w_2]$. Se desconsiderarmos uma das restrições, por exemplo, $G_2(\theta)$, isto equivaleria a fazer $\lambda_2 = 0$. É possível afirmar que ainda existe uma solução para λ_1 e λ_3 capaz de manter a relação entre o conjunto $[w_0 \ w_1 \ w_2]$ que atenda às três restrições. Entretanto, haverá abertura, como comentado anteriormente, para surgimento de outras soluções.

De um prisma mais geral, se considerarmos nossa hipótese válida e excluirmos uma restrição a cada tríplice, a ampliação desse conceito para um equalizador genérico de ordem N resultaria numa significativa redução de restrições, de $N(N+1)/2$ para N .

Voltando para o caso em que $N = 2$, a utilização de apenas 2 restrições facilita consideravelmente a análise e propicia, ao menos, a visualização de alguns resultados. Assim, assumindo nossa hipótese como verdadeira, buscaremos agora verificar, como feito para o equalizador de primeira ordem, se também é possível associar a uma das raízes das restrições

$G_i(\boldsymbol{\theta})$ os mínimos globais do critério CM.

Teste de Identificação - Equalizador de Segunda Ordem

Consideramos inicialmente o canal de fase mínima dado por (3.23) e um sinal *i.i.d.* de modulação BPSK com 30.000 amostras. Após estimar \mathbf{R}_ξ e \mathbf{p}_ξ , fizemos $\lambda_3 = 0$, isto é, desconsideramos a terceira restrição (3.33), e percorremos, numa grade bidimensional, λ_1 e λ_2 de -4 a 4 .

Devido ao comportamento algumas vezes explosivo de $G_i(\boldsymbol{\theta})$, usamos novamente a função de suavização, que pode ser expressa como:

$$G_s(\boldsymbol{\theta}_\lambda) = \frac{1}{1 + \sum_{i=1}^2 G_i^2(\boldsymbol{\theta}_\lambda)}. \quad (3.34)$$

O resultado pode ser verificado na Fig. 3.3(a). Observamos que existem vários casos em que $G_s(\boldsymbol{\theta}_\lambda) = 1$, o que torna a busca por tais valores mais difícil de ser realizada. Contudo, em particular, observamos nas proximidades da origem a formação de um pico bastante distinto.

Com base no caso do equalizador de primeira ordem, esperamos que o ótimo $\boldsymbol{\lambda}_o$ seja aquele que possua valores mínimos para cada λ_i . Isto implica que, caso seja válida a extensão desse resultado para os equalizadores de maior ordem, $\boldsymbol{\lambda}_o$ será justamente aquele associado ao pico próximo à origem na Fig. 3.3(a).

Anteriormente, para $N = 1$, buscamos encontrar λ_o a partir da comparação dos valores do custo CM associado às soluções $\lambda_{(1)}, \dots, \lambda_{(4)}$. No entanto, a dimensão do problema em mãos e a hipótese de se abandonar uma das restrições inviabilizam a repetição deste procedimento. Em vez disso, buscaremos então identificar $\boldsymbol{\lambda}_o$ a partir de uma comparação com o valor desejado em $\boldsymbol{\theta}$. Para isto, usaremos uma função de suavização que mede a distância entre os coeficientes correspondentes de $\boldsymbol{\theta}_\lambda$ e da solução ótima do critério CM no domínio de Volterra, isto é, $\boldsymbol{\theta}_{CMg}$:

$$D_s(\boldsymbol{\theta}_\lambda) = \frac{1}{1 + \|\boldsymbol{\theta}_\lambda - \boldsymbol{\theta}_{CMg}\|^2}. \quad (3.35)$$

Assim, quando a distância $\|\boldsymbol{\theta}_\lambda - \boldsymbol{\theta}_{CMg}\|^2$ for nula, $D_s(\boldsymbol{\theta}_\lambda)$ será igual à unidade. O valor de $\boldsymbol{\theta}_{CMg}$ foi obtido a partir do resultado da convergência do CMA (com tamanho de passo 0,0008) após 30.000 iterações, com inicialização na melhor solução de Wiener. Após isso, foi feito o mapeamento inverso $\mathbf{w}_{CMg} \rightarrow \boldsymbol{\theta}_{CMg}$.

Varrendo novamente a grade de valores para λ_1 e λ_2 , plotamos a função $D_s(\boldsymbol{\theta}_\lambda)$ na Fig. 3.3(b). Notamos que existe um único pico distinto, que representa o caso em que há equivalência entre $\boldsymbol{\theta}_\lambda$ e $\boldsymbol{\theta}_{CMg}$. Basta agora verificar se o pico em $D_s(\boldsymbol{\theta}_\lambda)$ coincide com aquele

de $G_s(\boldsymbol{\theta}_\lambda)$ próximo à origem. Nas Figs. 3.3(c) e 3.3(d) temos, respectivamente, as curvas de nível de $G_s(\boldsymbol{\theta}_\lambda)$ e $D_s(\boldsymbol{\theta}_\lambda)$ próximas à origem. Em ambas, temos um círculo verde que representa o ponto de máximo em $D_s(\boldsymbol{\theta}_\lambda)$ e, de fato, pela Fig. 3.3(c), notamos uma grande proximidade entre os picos.

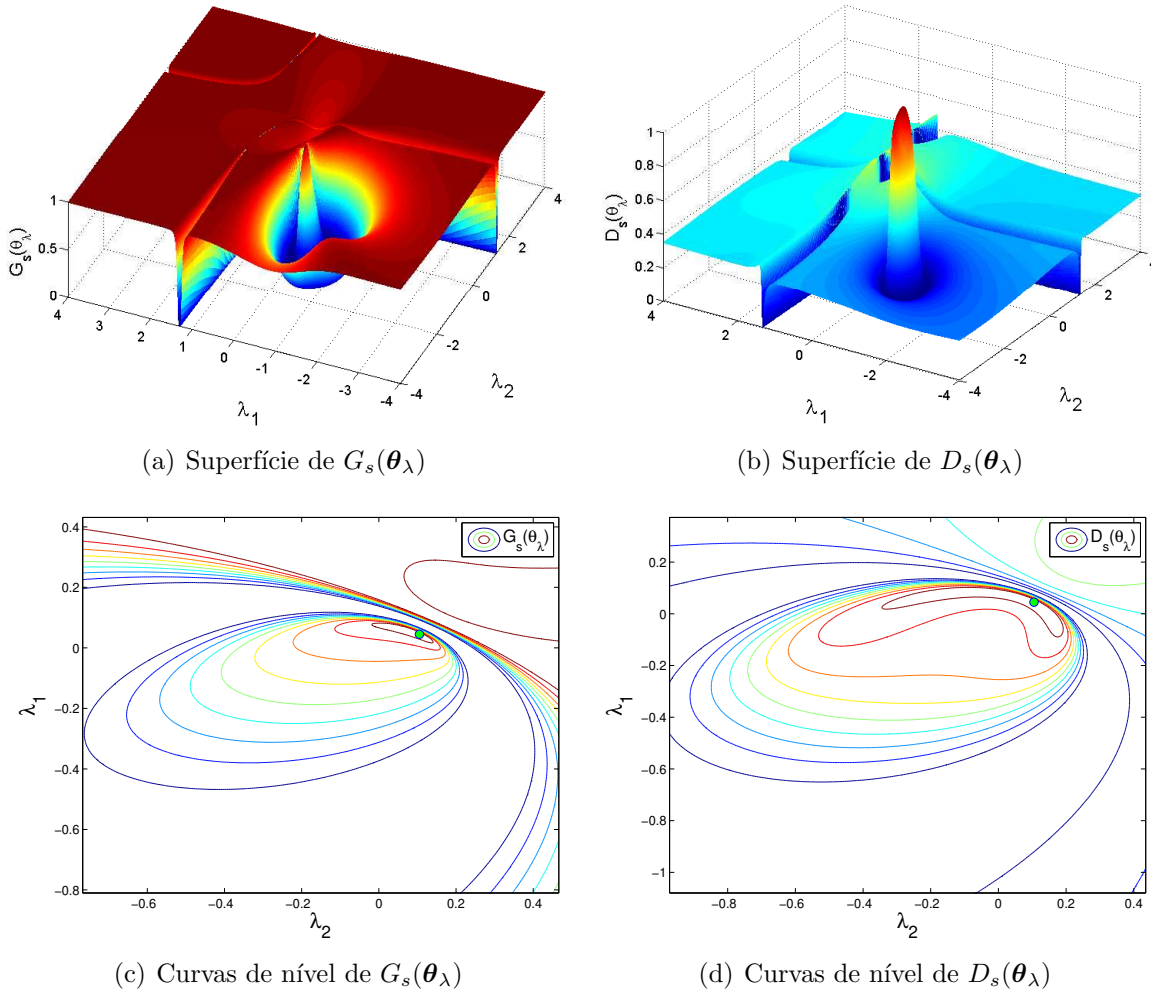


Figura 3.3: Superfície e curvas de nível de $G_s(\boldsymbol{\theta}_\lambda)$ e $D_s(\boldsymbol{\theta}_\lambda)$ fazendo-se $\boldsymbol{\lambda}$ variável.

Esta breve análise é um indício que se mostra bastante favorável à extensão dos resultados observados no caso do equalizador de primeira ordem, pois corrobora a conjectura de que a solução em $\boldsymbol{\lambda}$ associada aos mínimos globais do critério CM é aquela que possui menores valores nos elementos λ_i .

A rigor, faz-se necessário um estudo mais aprofundado sobre a hipótese de descarte de uma das restrições antes de lançarmos algumas conclusões, de forma que seja possível responder a indagações como “qual a qualidade das soluções obtidas?”, “para que tipo de canais essa conjectura é válida?”, “existe uma restrição mais adequada para descarte?”, e assim por diante. Da mesma forma, também se exigem análises mais extensivas para a

comprovação de qual é a solução ótima $\boldsymbol{\lambda}_o$ quando o equalizador original possui ordem $N > 1$. Todavia, consideramos que as informações coletadas até o momento são suficientes para a proposição de uma técnica de obtenção de $\boldsymbol{\lambda}_o$.

Mesmo restando ainda muito a ser feito, a abordagem a ser seguida, adicionalmente, nos ajudará a fortalecer os argumentos que, nesta seção, apoiam-se apenas em indicativos.

3.2.2 Técnica Baseada no Método do Gradiente

A possibilidade de se criar um vínculo entre as raízes de menor valor de $G_i(\boldsymbol{\theta}_\lambda)$ e a solução ótima do critério CM, conforme foi indicado na seção anterior, é bastante motivadora, pois, na busca por $\boldsymbol{\lambda}_o$, podemos, por exemplo, reduzir consideravelmente o espaço de busca para valores positivos próximos à origem. Além disso, com base nessas informações, outras alternativas ainda menos custosas podem ser seguidas.

Tendo isso em mente, tomaremos como base uma expressão que centraliza todos os conceitos discutidos no capítulo: a Eq. (3.21). No entanto, esta pode se mostrar mais proveitosa se utilizarmos seus termos em forma quadrática – conforme veremos à frente. Dessa forma, a nova função com todas as condições e restrições do problema pode ser escrita como

$$J_G(\boldsymbol{\theta}_\lambda) = \sum_{i=1}^{N(N+1)/2} G_i^2(\boldsymbol{\theta}_\lambda), \quad (3.36)$$

em que consideramos o caso genérico de um equalizador de ordem N e o vetor com os multiplicadores de Lagrange $\boldsymbol{\lambda} = [\lambda_1 \ \lambda_2 \ \dots \ \lambda_{N(N+1)/2}]$.

Apesar de $G_i(\boldsymbol{\theta}_\lambda)$ não ser bem comportada – o que exigiu, anteriormente, o emprego de $G_s(\boldsymbol{\theta}_\lambda)$ para sua análise – temos que, próximo à menor raiz em $\boldsymbol{\lambda}$, $J_G(\boldsymbol{\theta}_\lambda)$ é uma função contínua e suave, como pode ser observado pelas Figs. 3.4(a) e 3.4(b), que retratam os casos do equalizador de primeira e segunda ordem no cenário da seção anterior, com canal dado por (3.23).

Como mostrado na Fig. 3.4(a), outro ponto bastante relevante é o fato de $J_G(\boldsymbol{\theta}_\lambda)$ possuir um mínimo relativamente distinto, associado à solução ótima $\boldsymbol{\lambda}_o$, com vizinhança contínua e suave. Ainda vale ressaltar que, em ambos os casos, $\boldsymbol{\lambda}_o$ possui relativa proximidade com respeito à origem.

Tais fatores nos motivam a propor uma técnica baseada no método do gradiente para a obtenção dos multiplicadores ótimos de Lagrange. Temos, para isso, um excelente ponto de referência para inicialização de $\boldsymbol{\lambda}$, que é a origem. Nesse sentido, uma escolha adequada do tamanho de passo μ , nesta abordagem, pode garantir a convergência para o ponto de mínimo mais próximo, isto é, idealmente, $\boldsymbol{\lambda}_o$.

Baseado no algoritmo do gradiente determinístico, podemos fazer a adaptação de $\boldsymbol{\lambda}$

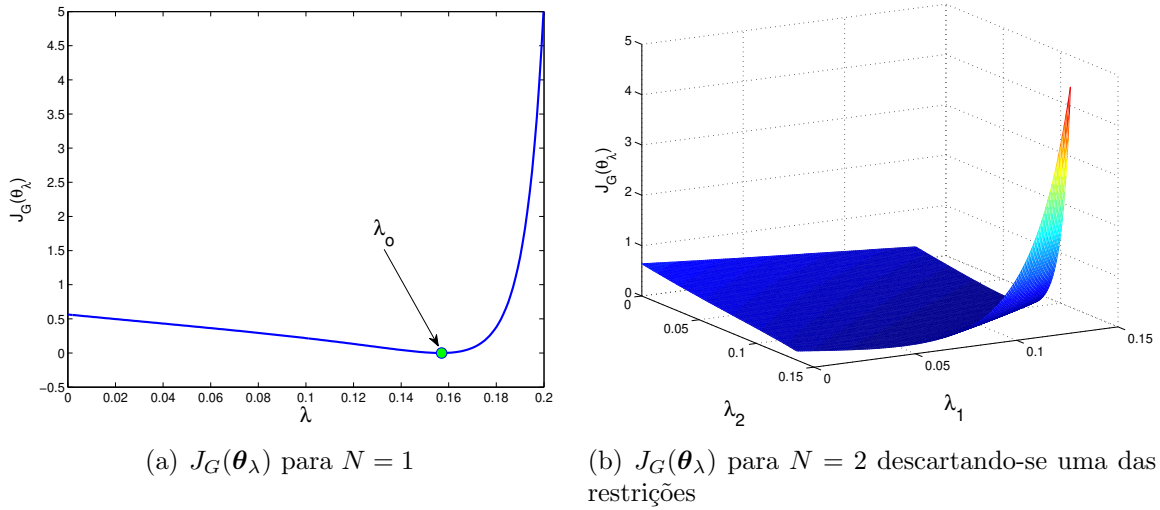


Figura 3.4: $J_G(\boldsymbol{\theta}_\lambda)$ para o canal $H(z) = 1 + 0,6z^{-1}$ e equalizadores de ordem $N = 1$ e $N = 2$.

como na Eq. (1.15), resultando em:

$$\boldsymbol{\lambda}(n+1) = \boldsymbol{\lambda}(n) - \mu \frac{\partial J_G(\boldsymbol{\theta}_\lambda)}{\partial \boldsymbol{\lambda}(n)}, \quad (3.37)$$

em que a derivada parcial de $J_G(\boldsymbol{\theta}_\lambda)$ em relação a $\boldsymbol{\lambda}(n)$ é expressa como

$$\frac{\partial J_G(\boldsymbol{\theta}_\lambda)}{\partial \boldsymbol{\lambda}(n)} = 2 \sum_{i=1}^{N(N+1)/2} G_i(\boldsymbol{\theta}_\lambda) \frac{\partial G_i(\boldsymbol{\theta}_\lambda)}{\partial \boldsymbol{\lambda}(n)}, \quad (3.38)$$

onde foram consideradas todas as $N(N+1)/2$ restrições.

Por sua vez, devemos agora calcular a derivada de $G_i(\boldsymbol{\theta}_\lambda)$ em relação a $\boldsymbol{\lambda}$. Para isso, faremos uso da Eq. (3.19) combinada à (3.21), o que nos dá

$$\begin{aligned} G_i(\boldsymbol{\theta}_\lambda) &= (\boldsymbol{\theta}_o + \mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda)^T \mathbf{C}_i (\boldsymbol{\theta}_o + \mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda) \\ &= \boldsymbol{\theta}_o^T \mathbf{C}_i \boldsymbol{\theta}_o + \boldsymbol{\theta}_o^T \mathbf{C}_i \mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda + (\mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda)^T \mathbf{C}_i \boldsymbol{\theta}_o \\ &\quad + (\mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda)^T \mathbf{C}_i \mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda, \end{aligned} \quad (3.39)$$

em que

$$\boldsymbol{\Gamma} = \sum_{j=1}^{N(N+1)/2} \lambda_j \mathbf{C}_j. \quad (3.40)$$

A relação em (3.39) pode ser entendida como um polinômio em função de $\boldsymbol{\lambda}$, facilitando o cálculo de sua derivada. Assim, como mostrado no Apêndice C, $\partial J_G(\boldsymbol{\theta}_\lambda)/\partial \lambda_i$ pode ser

definido como:

$$\frac{\partial J_G(\boldsymbol{\theta}_\lambda)}{\partial \lambda_i} = 4\boldsymbol{\theta}_\lambda^T \left(\sum_{j=1}^{N(N+1)/2} \mathbf{C}_j \boldsymbol{\theta}_\lambda \boldsymbol{\theta}_\lambda^T \mathbf{C}_j \right) \mathbf{R}_\xi^{-1} \mathbf{C}_i \boldsymbol{\theta}_\lambda. \quad (3.41)$$

À luz do que foi exposto, estes resultados permitem a proposição do seguinte algoritmo:

Algoritmo 3.1 Algoritmo para obtenção dos multiplicadores ótimos de Lagrange

Inicialização do Algoritmo:

Estima-se \mathbf{R}_ξ e \mathbf{p}_ξ .

$\boldsymbol{\lambda}(0) \leftarrow \mathbf{0}$

Adaptação de $\boldsymbol{\lambda}$:

for j from 0 to $T_\lambda - 1$ **do**

$$\boldsymbol{\theta}_\lambda(j) \leftarrow \left(\mathbf{R}_\xi - \sum_{k=1}^{N(N+1)/2} \lambda_k(j) \mathbf{C}_k \right)^{-1} \mathbf{p}_\xi$$

$$K \leftarrow \sum_{k=1}^{N(N+1)/2} \mathbf{C}_k \boldsymbol{\theta}_\lambda(j) \boldsymbol{\theta}_\lambda^T(j) \mathbf{C}_k$$

for i from 1 to $N(N+1)/2$ **do**

$$\lambda_i(j+1) \leftarrow \lambda_i(j) - \mu \boldsymbol{\theta}_\lambda^T(j) K \mathbf{R}_\xi^{-1} \mathbf{C}_i \boldsymbol{\theta}_\lambda(j)$$

end for

end for

em que T_λ é o número máximo de iterações e μ o tamanho de passo para adaptação (note que o fator de escala 4 na Eq. (3.41) foi incorporado à μ).

Uma característica fundamental desse algoritmo é sua grande sensibilidade à variações no tamanho de passo. Seu ajuste exige grande cautela, pois deve ser pequeno o suficiente para evitar que o algoritmo convirja para outros mínimos. Somado a isso, também é desejável pequenos valores para se reduzir os efeitos de *misadjustment*.

De um ponto de vista mais amplo, observamos que há neste algoritmo promissoras alternativas de implementação. Uma proposta interessante é a de realizar a estimativa de \mathbf{R}_ξ e \mathbf{p}_ξ de maneira adaptativa, de forma concorrente à técnica de ajuste de $\boldsymbol{\lambda}$, o que pode dar origem a novos algoritmos adaptativos nos moldes, por exemplo, do RLS (Seção 1.3.2). Contudo, essa ideia fica como uma perspectiva para trabalhos futuros.

A seguir, analisaremos alguns resultados da implementação deste algoritmo para os casos em que o equalizador original tem ordem $N = 1$ e $N = 2$. Buscaremos também averiguar em maiores detalhes alguns aspectos ainda não suficientemente explorados durante o capítulo.

3.2.3 Resultados

A eficácia do algoritmo para obtenção dos multiplicadores ótimos de Lagrange será avaliada para cenários com canais lineares e sinal transmitido *i.i.d.* com modulação BPSK. Serão utilizadas 30.000 amostras para estimação de \mathbf{R}_ξ e \mathbf{p}_ξ em todas as simulações.

Exploraremos primeiro o caso no qual o equalizador original é de primeira ordem e, posteriormente, de segunda ordem.

Resultados - Equalizador de Primeira Ordem

A situação mais simples de análise do algoritmo proposto é, sem dúvida, o caso em que existe somente uma restrição a ser considerada, isto é, quando $N = 1$.

Como retratado na Seção 3.2.1, verificamos uma associação do valor de $\lambda_{(1)}$ aos mínimos globais do critério CM. Seguindo esta linha, o desempenho do algoritmo pode ser avaliado em termos da convergência do coeficiente $\lambda(n)$ a $\lambda_{(1)}$, que é o ótimo λ_o .

Para o canal de fase mínima dado em (3.23), a convergência pode ser obtida em poucas iterações, como mostrado na Fig. 3.5. Neste caso, foi escolhido um passo de adaptação $\mu = 1/\gamma_{max} = 0,16$, sendo γ_{max} o maior autovalor de \mathbf{R}_ξ .

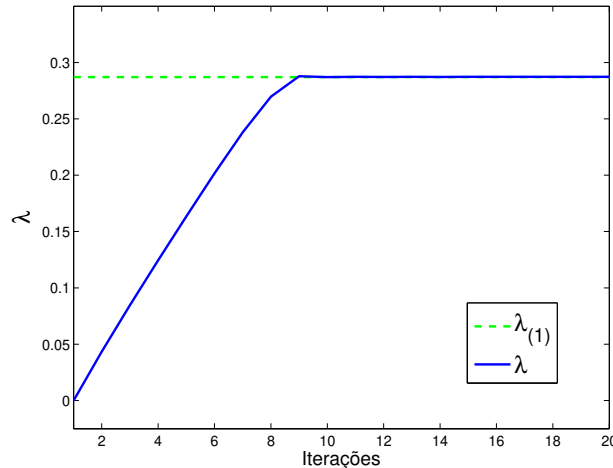


Figura 3.5: Convergência de $\lambda(n)$ à $\lambda_{(1)}$ para o canal $H(z) = 1 + 0,6z^{-1}$.

Como ilustrado na Fig. 3.5, uma escolha adequada do tamanho de passo μ deve buscar satisfazer o compromisso entre a velocidade e a garantia de convergência. Em alguns casos onde não se tem informação *a priori* sobre o canal, sugere-se a adoção de medidas mais cautelosas, com tamanho de passo que priorize a garantia de convergência. Essa será a postura tomada nas simulações a seguir.

Para um escopo mais amplo de cenários, avaliaremos a convergência de $\lambda(n)$ para diversos tipos de canais. Conduziremos as simulações varrendo os canais definidos nas Eqs. (3.26) e (3.27), para α e β indo de 0 a 3, a passos de 0,041. Contudo, como analisar a

velocidade de convergência para todos esses casos é inviável, focaremos apenas na questão da convergência para o mínimo desejado, isto é, λ_o . Optamos, assim, por um tamanho de passo considerado pequeno, escolhido empiricamente como $\mu = 1/(10\gamma_{2,max})^2$, em que $\gamma_{2,max}$ é o maior autovalor de \mathbf{R}_ξ^{-1} . Escolhemos também como número máximo de iterações $T_\lambda = 10.000$. Consideraremos que houve convergência para o mínimo desejado se $0,9\lambda_o < \lambda(T_\lambda) < 1,1\lambda_o$. Os resultados estão expressos na Tab. 3.1.

Tabela 3.1: Taxa de convergência para λ_o com equalizador original de ordem $N = 1$.

Canal	Taxa de Convergência para λ_o
$H_1(z) = 1 + \alpha z^{-1}$	100%
$H_2(z) = 1 + \alpha z^{-1} + \beta z^{-2}$	96,53%

Os resultados alcançados pelo algoritmo proposto são bastante favoráveis. Para os canais de primeira ordem $H_1(z)$ houve convergência para a solução desejada em todos os casos. Já para o canal $H_2(z)$, apesar de μ portar certas informações sobre o canal, houve casos em que seu tamanho foi excessivo, levando $\lambda(n)$ a convergir para outros mínimos. Nesse sentido, o ajuste empírico do valor de μ ainda requer uma análise mais cuidadosa, que pretendemos empreender na sequência deste trabalho. Não obstante, a taxa de convergência desejada de 96,53% é um excelente indicador do potencial da técnica proposta.

Resultados - Equalizador de Segunda Ordem

No caso do equalizador de maior ordem – no caso, $N = 2$ –, buscaremos averiguar não apenas o comportamento do algoritmo proposto, mas também, com base nas suposições feitas na Seção 3.2.1, se as conjecturas que permeiam o número reduzido de restrições são de fato válidas nas simulações.

Iniciaremos a análise do algoritmo tomando como base um cenário específico, onde pretendemos avaliar sucintamente o comportamento dos coeficientes de Lagrange em termos de sua convergência. Contudo, para verificar a otimalidade da solução, não poderemos proceder como no caso do equalizador de primeira ordem – em que visualizamos a convergência de $\lambda(n)$ para o caso ótimo λ_o – pois, quando temos $N = 2$, o espaço de busca para a obtenção de λ_o passa a ser multidimensional (se consideradas todas as restrições, o espaço é tridimensional) e, portanto, extremamente custosa. Como alternativa, recorreremos a uma comparação entre soluções obtidas no contexto do equalizador original \mathbf{w} em relação aos coeficientes ótimos do critério CM, ou seja, \mathbf{w}_{CMg} – que podem ser estimados como o resultado da convergência do CMA com inicialização na melhor solução de Wiener. Por fim, faremos uma análise mais genérica, em diversos cenários, avaliando se houve a convergência desejada por parte do algoritmo.

Para o canal dado em (3.23), aplicamos o algoritmo nas situações de interesse: (i) utilizando-se todas as restrições e (ii) utilizando apenas duas restrições (foi descartada a terceira restrição $G_3(\boldsymbol{\theta})$). Em ambos os casos, foi escolhido um passo de adaptação $\mu = 0,0005$ (pois admite um bom compromisso entre velocidade e garantia de convergência) e $T_\lambda = 200$.

As Figs. 3.6(a) e 3.6(b) apresentam a convergência dos parâmetros $\boldsymbol{\lambda}(n)$ ao longo das iterações para casos com três e duas restrições, respectivamente. Notamos que, para duas restrições (Fig. 3.6(b)), o comportamento de λ_1 e λ_2 é bastante similar ao de seus correspondentes quando se tem 3 restrições. Contudo, os valores observados são ligeiramente diferentes – o que não é surpresa, visto a necessidade em se compensar a ausência da terceira restrição. É possível notar também que a convergência no caso com três restrições é levemente mais rápida (≈ 100 iterações) que aquela com apenas duas (≈ 120 iterações).

Como próximo passo, verificamos se a solução encontrada em \mathbf{w}_λ é ótima por sua comparação à \mathbf{w}_{CMg} . O resultado pode ser conferido nas Figs. 3.6(c) e 3.6(d) considerando três e duas restrições, respectivamente, onde os coeficientes \mathbf{w}_λ foram obtidos através do mapeamento $\boldsymbol{\theta}_\lambda \rightarrow \mathbf{w}_\lambda$ usando o HTC (Alg. 2.2). Em ambas as figuras, as linhas contínuas coloridas representam os coeficientes \mathbf{w}_λ obtidos pelo mapeamento e os tracejados formam os valores de referência \mathbf{w}_{CMg} . É possível notar que houve grande proximidade entre \mathbf{w}_λ e \mathbf{w}_{CMg} nos dois casos, mas não existiu equivalência perfeita entre eles.

Se computarmos o custo CM associado às soluções \mathbf{w}_λ com três e duas restrições (após convergência) e também à \mathbf{w}_{CMg} usando (3.28), chegamos aos valores apresentados na Tab. 3.2. A partir dela, temos o indício de que usar três restrições pode conduzir à soluções com maior qualidade. Mesmo assim, é indubitável o fato de ambas apresentarem soluções bastante próximas ao ótimo global do critério CM. Vale notar também que o custo CM para \mathbf{w}_{CMg} é levemente inferior, o que mostra como \mathbf{w}_λ pode estar sujeito a maiores imprecisões, seja pela estimação de \mathbf{R}_ξ e \mathbf{p}_ξ , seja pelo *misadjustment* causado pelo tamanho de passo escolhido.

Tabela 3.2: Valor do custo CM para \mathbf{w}_λ (com três e duas restrições) e para \mathbf{w}_{CMg} .

Coeficientes	Valor do custo CM
\mathbf{w}_{CMg} (estimado)	0,1061
\mathbf{w}_λ (3 Restrições)	0,1067
\mathbf{w}_λ (2 Restrições)	0,1068

Avaliaremos agora em um âmbito mais geral se estes resultados podem ser estendidos para os canais de primeira e segunda ordem das Eqs. (3.26) e (3.27), respectivamente. Não consideraremos neste caso a velocidade de convergência. Avaliaremos apenas se a

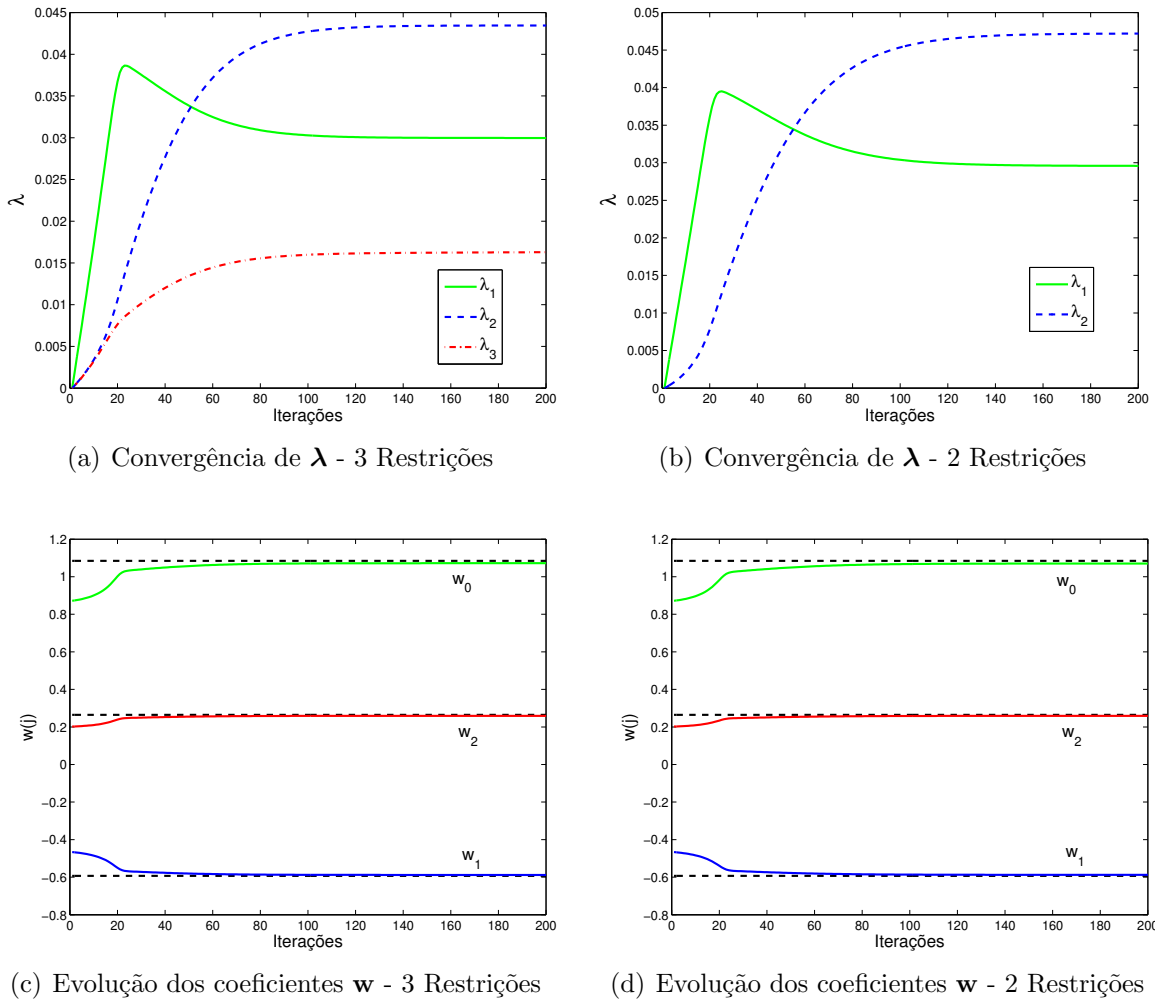


Figura 3.6: Convergência dos parâmetros λ e dos coeficientes \mathbf{w}_λ associados, considerando 3 e 2 restrições.

convergência de $\lambda(n)$ está de fato associada ao ótimo λ_o . Para isso, verificamos se a distância euclidiana entre os coeficientes correspondentes de \mathbf{w}_λ e \mathbf{w}_{CMg} é menor que 0,1. Em caso positivo, assumimos que $\lambda(n)$ convergiu para λ_o . Foram escolhidos $\mu = 1/(10\gamma_{2,max})^2$, em que $\gamma_{2,max}$ é o maior autovalor de \mathbf{R}_ξ^{-1} e $T_\lambda = 10.000$.

A Tab. 3.3 mostra o resultado obtido para o canal de primeira ordem $H_1(z)$. Notamos que, quando consideradas todas as três restrições, a taxa de convergência para λ_o é próxima de 90%, enquanto que com apenas duas restrições a taxa cai para 67,57%.

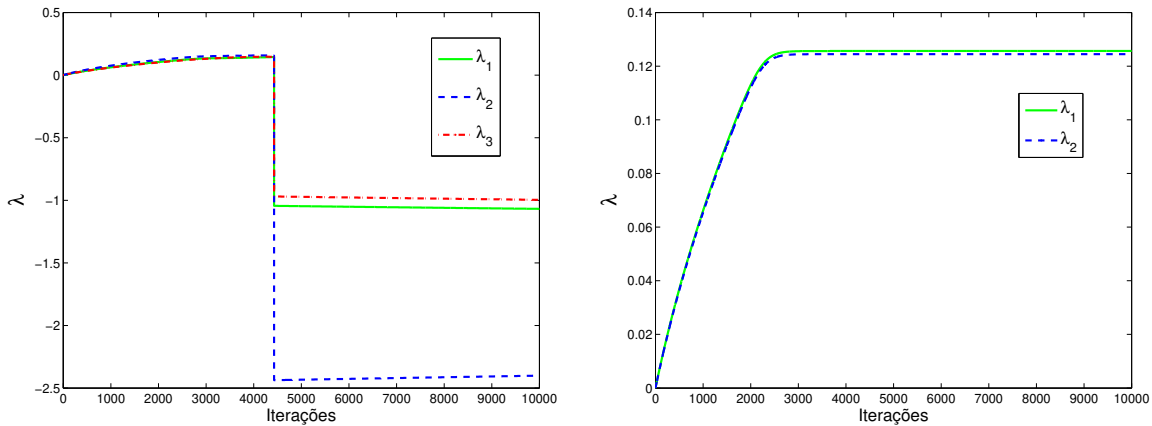
Para os canais em que não houve convergência para λ_o , temos em comum, tanto para a abordagem com três restrições quanto para aquela com apenas duas, o fato de que os zeros da função de transferência $H_1(z)$ estão próximos do círculo de raio unitário. Contudo, o motivo que leva à convergência indesejada é bastante distinto.

Tomemos como exemplo o canal $H_1(z) = 1 + 0,989z^{-1}$, um dos casos de simulação que

Tabela 3.3: Taxa de convergência para λ_o e valor médio do custo CM para equalizador de ordem $N = 2$ e canal de primeira ordem $H_1(z)$.

Coeficientes	Taxa de Convergência para λ_o	Valor Médio do custo CM
\mathbf{w}_{CMg}	-	0,1118
\mathbf{w}_λ (3 Restrições)	89,19%	0,1287
\mathbf{w}_λ (2 Restrições)	67,57%	0,1412

ilustra bem a situação. Na abordagem com três restrições, Fig. 3.7(a), notamos que o tamanho de passo μ não está ajustado corretamente, fazendo com que o algoritmo convirja para um mínimo indesejado. Por outro lado, quando se consideram apenas duas restrições, Fig. 3.7(b), observamos que λ converge para o mínimo mais próximo, mas tal solução não se associa a λ_o .



(a) Convergência de λ - 3 Restrições - Canal $H_1(z) = 1 + 0,989z^{-1}$

(b) Convergência de λ - 2 Restrições - Canal $H_1(z) = 1 + 0,989z^{-1}$

Figura 3.7: Convergência dos parâmetros λ para o canal $H_1(z) = 1 + 0,989z^{-1}$.

Essa análise põe em cheque a validade geral da noção de descartar uma das restrições, revelando uma situação “problemática”. Como mencionamos na Seção 3.2.1, ao considerarmos apenas duas das restrições, permite-se que haja múltiplas soluções em λ . Conforme mostrado, existem soluções próximas à origem que não se vinculam a λ_o , especialmente quando o canal em questão tem seus zeros próximos ao círculo de raio unitário.

De maneira mais robusta, considerando-se todas as restrições, falha-se apenas pelo ajuste inadequado do tamanho de passo, o que parece indicar que a solução ótima λ_o é a mais próxima à origem.

Por fim, apresentamos os resultados para o canal de segunda ordem $H_2(z)$ na Tab. 3.4. A taxa de convergência para a abordagem considerando todas as restrições é reduzida para

aproximadamente 65%. Novamente, observamos que os casos com convergência indesejada têm a si associados um tamanho de passo inadequado, com evoluções de $\boldsymbol{\lambda}(n)$ semelhantes à mostrada na Fig. 3.7(a). Na abordagem que carrega apenas duas restrições, a queda na taxa de convergência para $\boldsymbol{\lambda}_o$ é ainda mais acentuada, pois lidamos com um número maior de casos em que a garantia de duas restrições não implica no atendimento de uma terceira.

Tabela 3.4: Taxa de convergência para λ_o e valor médio do custo CM para equalizador de ordem $N = 2$ e canal de segunda ordem $H_2(z)$.

Coeficientes	Taxa de Convergência para $\boldsymbol{\lambda}_o$	Valor Médio do custo CM
\mathbf{w}_{CMg}	-	0,2477
\mathbf{w}_λ (3 Restrições)	64,66%	0,3257
\mathbf{w}_λ (2 Restrições)	22,53%	0,3181

Como fato curioso na Tab. 3.4, temos que o valor médio do custo CM para a abordagem com três restrições é o mais elevado, mesmo com melhor taxa de convergência. Ao averiguarmos alguns de seus casos, notamos que $\boldsymbol{\lambda}(n)$ pode convergir para soluções bastante diferentes de $\boldsymbol{\lambda}_o$, que são as convergências indesejadas. Estas corresponderão, após o mapeamento, a valores de \mathbf{w}_λ associados a mínimos locais e, não menos provável, a pontos de sela do critério CM, cujo custo associado é demasiadamente elevado, o que explica os valores observados na tabela. Em contrapartida, a abordagem com apenas duas restrições, após o mapeamento, apresenta soluções para \mathbf{w}_λ mais próximas à \mathbf{w}_{CMg} , mas que, por sua vez, sequer atendem às premissas da definição do problema em (3.16), não representando, portanto, os mínimos do critério CM.

Em suma, estes resultados trouxeram alguns esclarecimentos quanto às suposições feitas sobre o número de restrições a ser empregado e também contribuiu para fortalecer a ideia-chave que figura por trás da técnica utilizada, propiciando indícios sobre a proximidade de $\boldsymbol{\lambda}_o$ em relação à origem, mesmo para equalizadores de ordem $N = 2$.

3.3 Conclusão

Apresentamos neste capítulo a formulação polinomial do critério CM usando restrições quadráticas, cuja solução pôde ser obtida através do Método dos multiplicadores de Lagrange [Suyama, 2003]. Foi mostrado o caso em que o equalizador original é de primeira ordem e, em seguida, como uma extensão deste, o caso genérico de um equalizador de ordem N .

Um importante resultado advindo dessa formulação foi o de encontrar uma expressão em forma fechada no domínio de Volterra capaz de prover as soluções para o critério CM.

No entanto, para que tais soluções fossem válidas, exigiu-se em contrapartida o correto ajuste dos multiplicadores de Lagrange, tarefa esta que resguarda algumas dificuldades.

Com o intuito de simplificar a obtenção dos multiplicadores de Lagrange desejados, estudamos quais eram as soluções que apresentavam um vínculo direto com os mínimos globais do critério CM, onde foi possível identificar, nitidamente, para o caso do equalizador de primeira ordem, uma correspondência unívoca com as soluções mais próximas à origem. Já para o equalizador de ordem N , foi necessário adotar algumas suposições para se utilizar um número mais reduzido de restrições, a fim de que pudéssemos ter ao menos um indício de que também existia uma solução equivalente ao caso de primeira ordem.

Baseando-se nas propriedades observadas durante a identificação da solução ótima, propomos uma técnica baseada no método do gradiente para o ajuste dos multiplicadores de Lagrange, capaz de fazer a adaptação mesmo para casos em que o equalizador possui ordem elevada. Esta técnica permite, caso seja alcançada a convergência desejada, a obtenção direta da solução global do critério CM.

Os resultados considerando um equalizador de primeira ordem foram bastante convincentes e favoráveis à técnica proposta, mas percebe-se que ainda existe espaço para melhoras quanto ao ajuste do tamanho do passo de adaptação.

Para o equalizador de segunda ordem, os resultados nos permitiram constatar que a hipótese de se empregar um menor número de restrições falha em grande parte dos casos analisados. Por outro lado, há bons indicativos do uso da técnica quando todas as restrições são consideradas. Todavia, tornou-se mais forte a exigência de uma escolha mais adequada para o passo de adaptação. Este foi escolhido empiricamente e ainda se faz necessária uma análise mais profunda, como um estudo de estabilidade da técnica ou até mesmo do emprego de metodologias que independam do tamanho de passo, como o método de Newton. Mas deixaremos tais abordagens como perspectivas futuras.

As simulações que conduzimos cobriram alguns casos para canais lineares de primeira e segunda ordem, e não foram mais extensivas em razão de limitações de tempo. Pelo mesmo motivo, algumas análises e suposições mantiveram-se preliminares. Mesmo assim, julgamos que os resultados alcançados até o momento fornecem indícios da validade e da relevância da técnica proposta.

Destacamos ainda, como trabalhos futuros, a possibilidade de surgimento de novos algoritmos que façam em paralelo o ajuste dos multiplicadores de Lagrange e também da estimação da matriz de correlação e do vetor de correlação cruzada, aos moldes, por exemplo, do RLS.

Critérios Cegos Baseados na Teoria da Informação

Nos capítulos anteriores, tomamos como elemento comum às proposições apresentadas a formulação polinomial do critério CM. Conforme vimos, sua importância reside na capacidade de explorar explicitamente as estatísticas de quarta ordem a cerca do sinal recebido presentes na formulação original do critério CM.

Tendo por base tais estatísticas, verificamos um desempenho satisfatório nos cenários observados. Entretanto, se admitirmos situações e contextos distintos, como aqueles marcados pela presença de ruído com distribuições esparsas, uniforme ou impulsivas, ou até mesmo sinais de modulação multinível (ou multimódulo) tais como M -PAM e M -QAM, nota-se que os algoritmos clássicos, principalmente o CMA, exibem um elevado erro residual e também reduzida velocidade de convergência [Santamaría et al., 2002a; Lázaro et al., 2005; Mendes Filho et al., 2012].

Surgiu assim o interesse por uma representação estatística que retratasse o sinal com maior completude, sendo capaz de propiciar maior robustez e velocidade aos algoritmos para equalização do canal. Essa linha de raciocínio levou ao advento dos critérios baseados na teoria da informação, conforme foi exposto na Seção 1.5.

Mantendo ainda o estudo no âmbito não supervisionado, buscaremos avaliar neste capítulo alguns desses importantes critérios pertencentes à esta recente linha de pesquisa, também chamada pela sigla ITL (do inglês *Information Theoretic Learning*). Embora conte com alguns resultados bastante marcantes alcançados pela equipe coordenada pelo Prof. Principe [Principe, 2010], acreditamos haver ainda algumas lacunas, de um ponto de vista mais analítico, no que diz respeito às relações entre os próprios critérios ITL e também destes com o clássico critério CM, conforme pretendemos explorar adiante.

A análise que virá a seguir, mais precisamente, abordará uma comparação sob diversas óticas entre um grupo de critérios de interesse, assim como o estudo de suas funções custo,

de suas superfícies de erro e também de seu desempenho, como veremos a seguir.

4.1 Critérios Selecionados

Nos critérios associados ao ramo de ITL, o uso dos estimadores não paramétricos de densidade baseados em janelamento de Parzen [Principe, 2010] permite que se obtenha uma estimativa razoavelmente confiável da PDF dos sinais em questão e, por conseguinte, propicia uma exploração mais abrangente das estatísticas envolvidas. Esta abordagem pode ser conduzida tanto num viés supervisionado quanto cego. Neste trabalho, em consonância com a temática central, estudaremos apenas o caso não supervisionado.

Três critérios serão considerados: a entropia da dispersão de ordem p [Santamaría et al., 2002a], o casamento de PDF's por distância quadrática [Santamaría et al., 2002b] e o casamento de PDF's por correlação [Lázaro et al., 2003]. A definição de cada um deles será apresentada a seguir; após isso, partiremos para as comparações.

4.1.1 Entropia da Dispersão de Ordem p

A família de critérios sugerida por Godard [Godard, 1980] serviu de inspiração para a extensão das novas metodologias baseadas em ITL para o ramo não supervisionado. Fundamentalmente, a ideia apresentada em [Santamaría et al., 2002a] foi a de conciliar a medida de entropia de Rényi (1.52) com a função custo de dispersão de ordem p (1.38), dando origem a uma nova família de funções custo:

$$J_{H_{\alpha,p}}(\mathbf{w}) = H_{\alpha}(|y(n)|^p - R_p) = H_{\alpha}(|y(n)|^p), \quad (4.1)$$

onde p é um inteiro positivo e R_p é uma constante definida por (1.39), que, devido à invariância da entropia ao valor médio, pode ser desprezada. Por conveniência, denominamos este custo entropia da dispersão de ordem p ou, simplesmente, RD (do inglês *Rényi's entropy of the p -order Dispersion*).

Em particular, o caso em que esta função custo recebeu maior atenção foi, indubitavelmente, aquele caracterizado pelos valores $p = 2$ e $\alpha = 2$. É interessante notar que tais considerações estabelecem conexões com o custo CM (1.41) e, concomitantemente, permitem simplificações nas análises, bem como maior facilidade na obtenção de um algoritmo baseado no método do gradiente estocástico.

Todavia, o cálculo da entropia conforme a Eq. (1.52) ainda é demasiadamente complicado. Para contornar esta dificuldade, como exposto na Seção 1.5, é possível recorrer a uma estimação da entropia através do uso de funções de kernel. Em particular, como sinalizado em [Santamaría et al., 2002a], o kernel Gaussiano $G_{\sigma}(\cdot)$ definido em (1.51) se adequa a um

contexto genérico de equalização. Se traduzirmos todas essas considerações na função custo da Eq. (4.1), chegamos a

$$\begin{aligned}\hat{J}_{H_{2,2}}(\mathbf{w}) &= \hat{H}_2(|y(n)|^2) \\ &= -\log \left(\frac{1}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} G_{\sigma\sqrt{2}}(|y(n-j)|^2 - |y(n-i)|^2) \right),\end{aligned}\quad (4.2)$$

cujas deduções seguem os mesmos passos da Eq. (1.54) – trata-se da mesma função custo usada como exemplo anteriormente.

Ainda, conforme observado em [Santamaría et al., 2002a], é possível obter algumas simplificações com respeito ao $\hat{J}_{H_{2,2}}(\mathbf{w})$, resultando no estimador do potencial de informação quadrático $V_2(\cdot)$, como na Eq. 1.56. A nova função custo pode ser expressa como:

$$\begin{aligned}\hat{J}_{RD}(\mathbf{w}) &= V_2(|y(n)|^2) \\ &= \frac{1}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} G_{\sigma\sqrt{2}}(|y(n-j)|^2 - |y(n-i)|^2),\end{aligned}\quad (4.3)$$

que, fundamentalmente, suscita uma equivalência entre seus pontos de máximo e os pontos de mínimo de (4.2), significando que o problema em mão agora é de maximização – e não minimização, como anteriormente.

Todos esses esforços contribuem, na verdade, para a utilização final do critério num algoritmo baseado no método do gradiente estocástico, que fará uso direto das derivadas de $\hat{J}_{RD}(\mathbf{w})$:

$$\begin{aligned}\nabla_{\mathbf{w}} \hat{J}_{RD}(\mathbf{w}) &= \frac{1}{L^2 \sigma^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} G_{\sigma\sqrt{2}}(v_y(i, j)) v_y(i, j) \\ &\quad \times (\mathbf{x}(n-i)y^*(n-i) - \mathbf{x}(n-j)y^*(n-j)),\end{aligned}\quad (4.4)$$

em que $v_y(i, j) = (|y(n-j)|^2 - |y(n-i)|^2)$. Vale ressaltar que, nos casos em que $i = j$, a derivada é nula, de forma que não é necessário o seu cômputo. Isto significa que o tamanho L da janela de amostras a ser considerada deve ser maior que 1.

Tais definições nos permitem escrever o algoritmo da entropia de dispersão, SRD (do inglês *Stochastic gradient algorithm for RD*), como:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \nabla_{\mathbf{w}} \hat{J}_{RD}(\mathbf{w}).\quad (4.5)$$

A importância do SRD resulta de sua capacidade de conciliar a simplicidade do critério CM com o uso mais amplo das estatísticas do sinal. Salientamos, contudo, que este algoritmo carrega algumas deficiências. Possui, por exemplo, maior vulnerabilidade de con-

vergência para a solução trivial, isto é, com sinal de saída nulo ($y(n) = 0$), sendo necessário tomar alguns cuidados adicionais, conforme veremos adiante.

A próxima técnica que iremos apresentar não faz uso direto dos conceitos e métricas clássicos em Teoria da Informação, mas incorpora explicitamente a ideia de levar em consideração as diversas estatísticas do sinal de interesse.

4.1.2 Casamento de PDF's por Distância Quadrática

Outra interessante abordagem não supervisionada é aquela que busca o casamento entre a PDF relativa ao sinal de saída do equalizador $y(n)$ e aquela do sinal transmitido $s(n)$. O critério resultante, que tem como essência o teorema BGR [Benveniste et al., 1980], já foi introduzido previamente na Seção 1.5; entretanto, resgataremos seus principais conceitos para facilitar nossa análise subsequente.

Proposto inicialmente em [Santamaría et al., 2002b] e posteriormente estendido para lidar com modulações multinível em [Lázaro et al., 2005], o critério QD (ou da distância quadrática) busca minimizar a seguinte função custo:

$$J_{QD}(\mathbf{w}) = \int_{-\infty}^{\infty} (p_{Y^q}(z) - p_{S^q}(z))^2 dz, \quad (4.6)$$

sendo $p_{Y^q}(z)$ e $p_{S^q}(z)$ as PDF's das VA's Y^q e S^q que, por sua vez, associam-se aos sinais $|y(n)|^q$ e $|s(n)|^q$, respectivamente.

A distância euclidiana mostrada em (4.6) pode ter seus termos expandidos conforme abaixo:

$$J_{QD}(\mathbf{w}) = \int_{-\infty}^{\infty} p_{Y^q}^2(z) dz - 2 \int_{-\infty}^{\infty} p_{Y^q}(z)p_{S^q}(z) dz + \int_{-\infty}^{\infty} p_{S^q}^2(z) dz, \quad (4.7)$$

o que facilitará a manipulação algébrica para estimação das PDF's, pois, se utilizarmos o estimador por janelamento de Parzen [Principe, 2010] com kernels Gaussianos, teremos que a multiplicação entre duas PDF's resulta na mesma condição favorável e simplificadora da relação dada pela Eq. (1.55).

Visando novamente uma possível aplicação do método do gradiente estocástico, outra simplificação importante a se fazer é assumir $q = 2$. Assim, podemos estimar a função custo

como:

$$\begin{aligned}
\hat{J}_{QD}(\mathbf{w}) &= \frac{1}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} G_{\sigma\sqrt{2}}(|y(n-j)|^2 - |y(n-i)|^2) \\
&\quad - \frac{2}{LN_s} \sum_{i=1}^{N_s} \sum_{j=0}^{L-1} G_{\sigma\sqrt{2}}(|y(n-j)|^2 - |s_i|^2) \\
&\quad + \frac{1}{N_s^2} \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} G_{\sigma\sqrt{2}}(|s_j|^2 - |s_i|^2),
\end{aligned} \tag{4.8}$$

sendo N_s a quantidade de símbolos complexos definidos na constelação da modulação em questão e s_i os valores complexos associados ao i -ésimo símbolo.

Notamos que o primeiro termo em $\hat{J}_{QD}(\mathbf{w})$ busca corrigir os desvios entre as amostras de $y(n)$ para diferentes instantes de tempo, correspondendo exatamente à função custo $\hat{J}_{RD}(\mathbf{w})$ (contudo, este último se vincula a um problema de maximização). Enquanto isso, o segundo termo é o responsável por fazer correções na PDF associada ao sinal $y(n)$ na direção da PDF desejada. Por fim, o último termo independe de \mathbf{w} e pode ser desconsiderado durante seu ajuste pelo método do gradiente.

De maneira similar à Eq. (4.4), o gradiente da Eq. (4.8) é expresso como:

$$\begin{aligned}
\nabla_{\mathbf{w}} \hat{J}_{QD}(\mathbf{w}) &= \frac{1}{L^2 \sigma^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} G_{\sigma\sqrt{2}}(v_y(i, j)) v_y(i, j) \\
&\quad \times (\mathbf{x}(n-i)y^*(n-i) - \mathbf{x}(n-j)y^*(n-j)) \\
&\quad + \frac{2}{LN_s \sigma^2} \sum_{i=1}^{N_s} \sum_{j=0}^{L-1} G_{\sigma\sqrt{2}}(v_s(i, j)) v_s(i, j) \\
&\quad \times (\mathbf{x}(n-j)y^*(n-j)).
\end{aligned} \tag{4.9}$$

sendo $v_y(i, j) = (|y(n-j)|^2 - |y(n-i)|^2)$ e $v_s(i, j) = (|y(n-j)|^2 - |s_i|^2)$.

De posse das derivadas da função custo, o algoritmo para adaptação de \mathbf{w} denominado SQD (do inglês *Stochastic gradient algorithm for the Quadratic Distance*) pode ser expresso como

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \nabla_{\mathbf{w}} \hat{J}_{QD}(\mathbf{w}). \tag{4.10}$$

Convém lembrar que o sinal negativo indica que se trata de um problema de minimização.

Por fim, como último critério a ser analisado, apresentaremos a seguir uma metodologia que é considerada uma simplificação do QD.

4.1.3 Casamento de PDF's por Correlação

Uma análise mais cautelosa da função custo $J_{QD}(\mathbf{w})$ expressa em (4.7) permite-nos atribuir diferentes níveis de importância a cada um dos seus termos. Particularmente, o termo com PDF's cruzadas é o responsável por medir a distância efetiva entre elas, pois, como comentado em [Lázaro et al., 2005], os demais termos funcionam apenas como um fator de normalização para as diferentes PDF's.

No trabalho conduzido em [Lázaro et al., 2003], o mesmo termo cruzado é utilizado como uma medida de correlação para a definição de uma nova função custo, em que se busca maximizar:

$$J_{MQD}(\mathbf{w}) = \int_{-\infty}^{\infty} p_{Y^q}(z)p_{S^q}(z) dz. \quad (4.11)$$

Por ser uma modificação do custo $J_{QD}(\mathbf{w})$, chamaremos esta abordagem de MQD (do inglês *Modified Quadratic Distance*).

Se supormos que os custos $J_{QD}(\mathbf{w})$ e $J_{MQD}(\mathbf{w})$ possuem soluções equivalentes para \mathbf{w} (visto que ambos buscam um casamento de PDF's), isto implica numa importante redução do custo computacional envolvido na estimação da distância entre as PDF's.

Novamente, usando o método de janelamento de Parzen com kernels Gaussianos e $q = 2$, temos

$$\hat{J}_{MQD}(\mathbf{w}) = \frac{1}{LN_s} \sum_{i=1}^{N_s} \sum_{j=0}^{L-1} G_{\sigma\sqrt{2}}(|y(n-j)|^2 - |s_i|^2). \quad (4.12)$$

O gradiente dessa função é dado por

$$\begin{aligned} \nabla_{\mathbf{w}} \hat{J}_{MQD}(\mathbf{w}) = & -\frac{1}{LN_s\sigma^2} \sum_{i=1}^{N_s} \sum_{j=0}^{L-1} G_{\sigma\sqrt{2}}(v_s(i, j)) v_s(i, j) \\ & \times (\mathbf{x}(n-j)y^*(n-j)), \end{aligned} \quad (4.13)$$

sendo $v_s(i, j)$ definido como na Eq. (4.9).

Finalmente, a adaptação de \mathbf{w} é feita considerando o método do gradiente estocástico, dando origem ao algoritmo MSQD (do inglês *Modified SQD*), conforme abaixo:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \nabla_{\mathbf{w}} \hat{J}_{MQD}(\mathbf{w}). \quad (4.14)$$

Este algoritmo, se comparado com o SRD e com o SQD, suscita interessantes relações que surgem da comparação entre eles e o CMA. Considerando também os critérios associados a cada um deles, buscaremos nas próximas seções averiguar suas conexões tendo em vista três perspectivas: suas funções custo, suas superfícies de erro e, também, o desempenho em diferentes cenários.

4.2 Pontos de Contato Entre as Funções Custo

A abordagem seguida nos critérios ITL levanta propostas inovadoras sobre como extrair a informação. Mesmo assim, principalmente no caso não supervisionado, há fortes ligações com critérios que pertencem à teoria clássica, em especial, com o critério CM.

4.2.1 Conexões ao Critério CM

Veremos nesta seção que muitos dos conceitos envolvendo o critério CM encontram-se presentes nos critérios apresentados na seção anterior. Buscaremos, a seguir, explorar estas conexões em maior profundidade.

De maneira mais direta, o critério RD carrega em sua própria definição – Eq. (4.1) com $p = 2$ – o termo do módulo constante presente no custo CM. Já para os outros dois critérios ITL expostos, esta relação não é tão trivial.

A princípio, o critério QD, em sua definição rigosa – Eq. (4.7) –, parece não ter ligação direta com o CM. Contudo, na estimativa dada por (4.8), temos no segundo termo algumas semelhanças. Se supusermos, por exemplo, que o sinal transmitido possui módulo constante – $\hat{R}_2 = |s(n)|^2$ – e que apenas uma amostra seja considerada na estimação da PDF ($L = 1$), o termo analisado torna-se

$$-2G_{\sigma\sqrt{2}}(|y(n)|^2 - \hat{R}_2). \quad (4.15)$$

Fazendo uma expansão em série de Taylor em torno de zero, temos, até segunda ordem:

$$-2G_{\sigma\sqrt{2}}(|y(n)|^2 - \hat{R}_2) \approx \frac{1}{\sigma\sqrt{\pi}} \left(\frac{(|y(n)|^2 - \hat{R}_2)^2}{4\sigma^2} - 1 \right). \quad (4.16)$$

Sabendo que nem um fator de escala nem a subtração de uma constante alteram um ponto de mínimo, caso $\hat{R}_2 = R_2$, teremos que a aproximação do segundo termo de $\hat{J}_{QD}(\mathbf{w})$ mostrada em (4.16) conduz exatamente às mesmas soluções do critério CM.

Ainda, voltando-nos para o primeiro termo de $\hat{J}_{QD}(\mathbf{w})$, Eq. (4.8), verificamos que este é exatamente o custo RD estimado $\hat{J}_{RD}(\mathbf{w})$, Eq. (4.3), cuja origem remonta ao CM. Assim, estabelecem-se vínculos com o CM no cerne de ambas as formulações – vale lembrar que o terceiro termo de $\hat{J}_{QD}(\mathbf{w})$ pode ser descartado, pois não contribui para o ajuste dos coeficientes do filtro.

Por fim, o critério MQD, Eq. (4.11), pode ser visto como um caso específico do QD, levando-se em conta apenas o segundo termo de (4.8), que, como mostramos, também evoca o critério CM. Assim, de uma maneira geral, podemos afirmar que os custos RD, QD e MQD são, de certa forma, extensões do custo CM para o arcabouço de ITL.

4.2.2 Relações entre os Critérios ITL

De outro ponto de vista, além de apresentarem conexões com o critério CM, as três técnicas enfocadas também estão interrelacionadas. Se observarmos os custos estimados de RD, QD e MQD, Eqs. (4.3), (4.8) e (4.12), respectivamente, podemos relacioná-los da seguinte forma:

$$\hat{J}_{QD}(\mathbf{w}) = \hat{J}_{RD}(\mathbf{w}) - 2\hat{J}_{MQD}(\mathbf{w}), \quad (4.17)$$

em que, novamente, desconsideramos o terceiro termo de (4.8). Uma relação semelhante a essa foi apresentada em [Santamaría et al., 2002b], sem, contudo, considerar o custo estimado $\hat{J}_{MQD}(\mathbf{w})$.

É importante ressaltar que, nos critérios originais, $\hat{J}_{RD}(\mathbf{w})$ e $\hat{J}_{MQD}(\mathbf{w})$ devem ser maximizados enquanto $\hat{J}_{QD}(\mathbf{w})$ deve ser minimizado. Contudo, na Eq. (4.17), se tomarmos como referência a minimização de $\hat{J}_{QD}(\mathbf{w})$, teremos que $\hat{J}_{RD}(\mathbf{w})$ também está sendo minimizado. Iremos discutir isso a seguir.

Mais atentamente, em vista da correspondência de $\hat{J}_{RD}(\mathbf{w})$ com o primeiro termo de $\hat{J}_{MQD}(\mathbf{w})$, podemos traçar um caminho inverso das estimativas para notar que o critério original para $J_{H_{2,2}}(\mathbf{w})$ – Eq. (4.1) – corresponderia a maximizar o primeiro termo de $J_{MQD}(\mathbf{w})$, Eq. (4.7). Esta analogia nos permite inferir que tal critério tende a maximizar a PDF associada a $|y(n)|^2$. Ora, se a PDF em questão for, por exemplo, uniforme, ao fazermos a maximização, tenderemos a estreitar sua base para conseguir um maior adensamento dessa distribuição. No caso limite, esta maximização terá seu ápice quando a PDF resultante for um impulso. Assim, se considerarmos sinais com módulo constante, este será justamente o perfil desejado para a equalização. Contudo, vale lembrar que a solução trivial deve ser evitada (PDF associada a $|y(n)|^2$ sendo um impulso na origem).

Por outro lado, no critério QD, o termo do custo associado a $\hat{J}_{RD}(\mathbf{w})$ está sujeito a uma minimização, provocando efeito inverso ao comentado anteriormente, isto é, no exemplo da distribuição uniforme, tenderá a tornar a base da PDF associada a $|y(n)|^2$ mais larga, o que parece não contribuir aos princípios de equalização. Devemos lembrar, contudo, que este termo age em conjunto com outro termo, associado a $\hat{J}_{MQD}(\mathbf{w})$, e, além disso, tem menor influência sobre o custo $\hat{J}_{QD}(\mathbf{w})$ – note em (4.17) que o custo $\hat{J}_{MQD}(\mathbf{w})$ tem um fator multiplicativo de dois. Dessa forma, a solução para o critério QD deverá aproximar, com maior prioridade, os símbolos $|y(n)|$ e $|s_i|$ – devido a $\hat{J}_{MQD}(\mathbf{w})$ –, e, em segundo plano, distanciar $|y(n)|$ de suas versões atrasadas – devido a $\hat{J}_{RD}(\mathbf{w})$. Isto ficará mais claro quando analisarmos as superfícies da função custo.

A Eq. (4.17) ainda pode nos oferecer algumas informações sobre as relações entre QD e MQD. Espera-se que as soluções de ambos os critérios sejam equivalentes, visto que os dois sugerem o mesmo casamento entre PDF's. Entretanto, tendo por base (4.17), isso só será verdade se $\hat{J}_{RD}(\mathbf{w}) \approx 0$ em cada solução, o que, de certa maneira, pode ser plausível

quando existem mínimos globais para o custo QD (ou máximos globais para o MQD) capazes de equalizar o canal de maneira perfeita (i.e., quando existem soluções ZF). Nos demais casos (principalmente para mínimos locais do custo QD), $\hat{J}_{RD}(\mathbf{w})$ exercerá certa influência, comprometendo a igualdade entre os dois referidos critérios.

As análises envolvendo as funções custo dos critérios podem ser complementadas e melhor compreendidas se investigarmos também as superfícies de erro, que será o assunto da próxima seção.

4.3 Análise das Superfícies de Erro

Uma maneira mais pragmática de avaliar a relação entre os custos se dá através da visualização de suas superfícies, percorrendo-se os parâmetros em \mathbf{w} e computando-se o valor de custo correspondente para cada caso. A análise pode ser bastante útil, mas, para que esteja restrita a três dimensões, deve-se permitir a variação de apenas dois coeficientes do filtro. Devido a isso, restringiremos a análise para os casos em que $N = 1$, ou seja, quando o equalizador é de primeira ordem.

Escolhemos como cenário de simulação, para todas as superfícies, 100 amostras de um sinal BPSK *i.i.d.* sem ruído e o canal de fase mínima

$$H(z) = 1 + 0,6z^{-1}. \quad (4.18)$$

Para os custos RD, QD e MQD, adotamos $L = 100$ para a janela de símbolos (i.e., usamos todas as amostras disponíveis) e $\sigma = 1$ como tamanho do kernel.

Como forma de facilitar a visualização de todos os pontos de máximo ou de mínimo, mostraremos as curvas de nível das superfícies obtidas em vez da superfície em si. Além disso, optamos também pela inclusão das curvas referentes ao critério CM a fim de facilitar a comparação.

Para o primeiro critério analisado, o RD, plotamos as curvas de nível do custo estimado $\hat{J}_{RD}(\mathbf{w})$, Eq. (4.3), na Fig. 4.1(a), em linhas escuras. Encontram-se também ilustradas as curvas de nível do custo CM, mas cobrimos apenas as regiões próximas aos mínimos, com traços mais claros, em azul. Observamos que $\hat{J}_{RD}(\mathbf{w})$ possui um único ponto de máximo na origem: a solução trivial. Mesmo assim, sua forma tende a privilegiar valores de \mathbf{w} que encontram-se nas proximidades do eixo definido pelos mínimos globais e mínimos locais do custo CM. Portanto, torna-se plausível que, se for imposta uma restrição sobre \mathbf{w} que impeça a sua convergência para origem, é possível atingir soluções que estabeleçam algum vínculo com o custo CM.

Conforme sugerido pelos autores em [Santamaría et al., 2002a] e [Santamaría et al., 2002b], usar restrições como a fixação do maior coeficiente na unidade ou um vetor \mathbf{w}

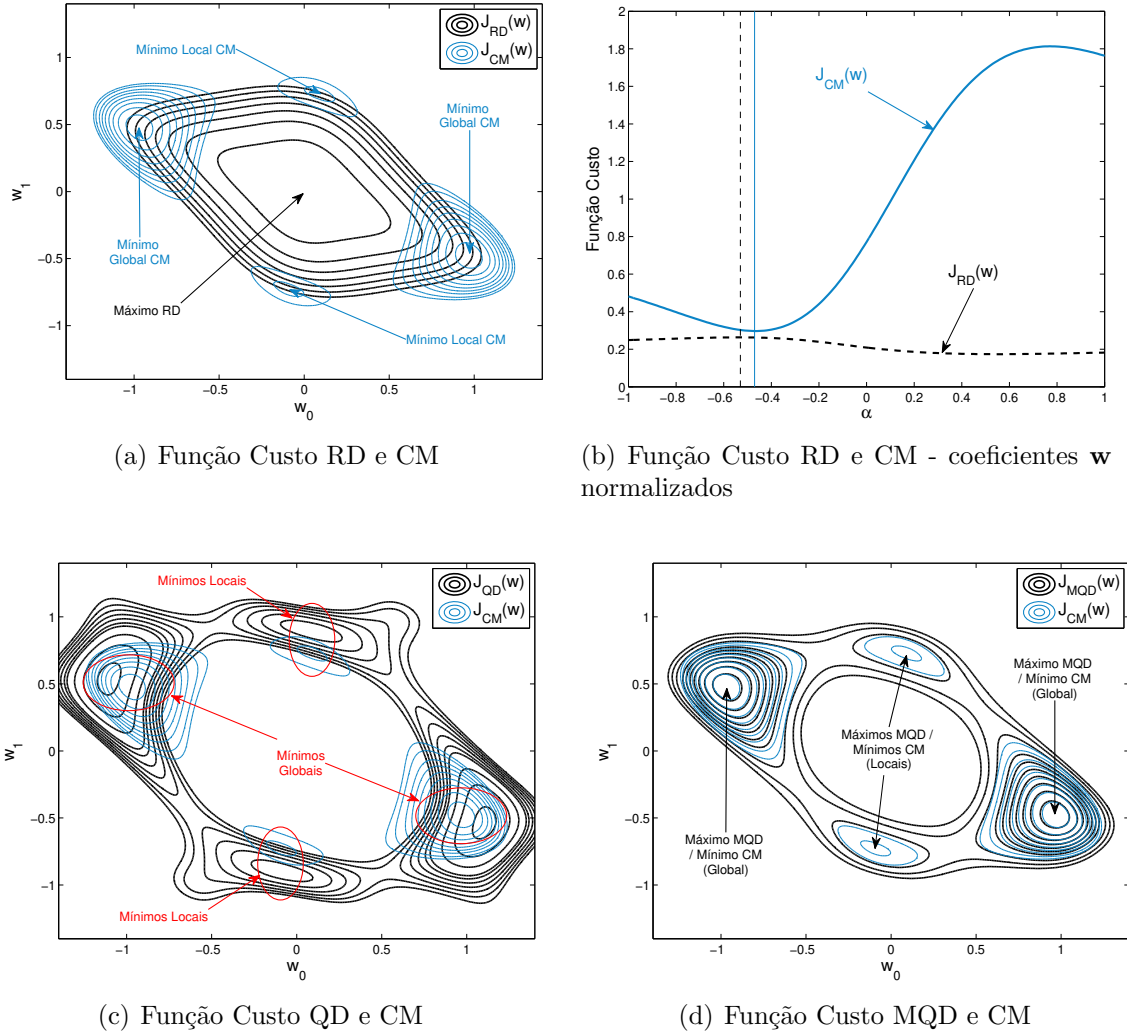


Figura 4.1: Curvas de nível das funções custo - canal $H(z) = 1 + 0,6z^{-1}$.

de norma unitária é suficiente para que o critério opere adequadamente. Para averiguar melhor, escolhemos um equalizador de primeira ordem do tipo

$$\mathbf{w} = [1 \ \alpha]^T, \quad (4.19)$$

em que α varia de -1 a 1 , com posterior normalização pela norma euclidiana. Os valores obtidos para os custos $\hat{J}_{RD}(\mathbf{w})$ e $J_{CM}(\mathbf{w})$ estão apresentados na Fig. 4.1(b). Neste caso, o custo RD estimado apresenta valores mais rasos que o custo CM, mas, ainda assim, é possível notar sua tendência a refletir o comportamento inverso de $J_{CM}(\mathbf{w})$ (resultado de $\hat{J}_{RD}(\mathbf{w})$ ser um problema de maximização enquanto $J_{CM}(\mathbf{w})$ é de minimização). Tais observações contribuem favoravelmente para a nossa interpretação, na Seção 4.2, de que o critério RD estabelece vínculos com o critério CM; contudo, fica nítido que suas soluções

ótimas não coincidem (traços verticais na Fig. 4.1(b)), o que se justifica se levarmos em conta que $\hat{J}_{RD}(\mathbf{w})$ utiliza um conjunto mais extensivo de estatísticas a respeito do sinal em relação ao critério CM.

Como próxima função custo a ser avaliada, o casamento de PDF's por distância quadrática, dado pela Eq. (4.8), também se assemelha ao custo CM. Suas curvas de nível estão mostradas na Fig. 4.1(c). Nela, observamos que $\hat{J}_{QD}(\mathbf{w})$ possui uma forma dotada de certas irregularidades, mas conta com mínimos globais e locais bastante distintos, sem sofrer do problema de convergência para a solução trivial que assola o critério baseado na entropia. Ainda, a relativa proximidade entre as respectivas soluções do QD e aquelas do CM evidenciam a relação entre os custos.

No terceiro caso avaliado, comparamos as curvas de nível das funções custo MQD e CM. Como ilustrado na Fig. 4.1(d), é possível observar grande consonância entre elas: por existir equivalência entre as constantes \hat{R}_2 e R_2 , quase não há distinção entre os pontos de mínimo do CM e os máximos de MQD (lembrando que o MQD é um problema de maximização), principalmente em relação às soluções globais. Em vista dessa maior proximidade, podemos dizer que o MQD é aquele que melhor preserva as relações com o custo CM. De fato, como visto na Seção 4.2, temos agora uma confirmação mais palpável da validade da aproximação usada para representar o MQD através de sua expansão em série de Taylor – Eq. (4.16) –, capaz de refletir de forma bastante precisa as conexões ao critério CM.

De uma forma geral, com base no que foi exposto na seção anterior e presente, podemos concluir que os três critérios, RD, QD e MQD, expressam algum tipo de similaridade com o custo CM. Um fato particularmente interessante é o modo como as correspondentes soluções dos critérios tendem a alinhar-se no mesmo “eixo de soluções”. Por exemplo, se traçarmos uma reta que liga os dois mínimos globais do critério CM, veremos que todas as soluções globais dos critérios ITL pertencem (ou encontram-se muito próximos) a esta mesma reta. Como esta linha passa pela origem, as soluções que estiverem exatamente alinhadas irão divergir apenas por fatores de escala ou, equivalentemente, por diferenças de ganho. Muito provavelmente, isso ocorre para que haja melhor casamento das informações das EOS adicionais presentes nos critérios ITL; entretanto, como veremos mais à frente, um ajuste de ganho não implicará, necessariamente, em melhora de desempenho. Visto que os critérios ITL abrangem um conteúdo estatístico mais amplo a respeito dos sinais, este resultado indica que a solução apontada pelo critério CM também carrega informações bastante significativas para a equalização. De fato, conforme mostramos nos capítulos anteriores, o CM porta estatísticas de segunda e, implicitamente, quarta ordem. Do prisma do teorema SW [Shalvi e Weinstein, 1990], tais informações são suficientes para que haja o casamento entre as PDF's do sinal transmitido e equalizado num caso estruturalmente ideal, o que, de certa forma, atrai grande concordância entre todos os critérios ITL aqui apresentados.

De outro ponto de vista, além das conexões com o critério CM, apontamos também na seção anterior as relações entre os próprios critérios RD, QD e MQD, que podem ser sumarizados pela Eq. (4.17). Particularmente, por se basearem no casamento de PDF's, os critérios QD e MQD, deveriam, a princípio, possuir soluções coincidentes. Para estudar este caso, devemos comparar as duas superfícies: na Fig. 4.1(d), podemos tomar como referência as posições dos mínimos do custo CM que encontram-se praticamente sobrepostas aos máximos do MQD, usando-as como comparação com as posições dos mínimos do QD. Assim, voltando-nos para a Fig. 4.1(c), notamos que a posição das soluções não coincidem. Isto significa que o termo associado a $\hat{J}_{RD}(\mathbf{w})$ presente na definição do custo $\hat{J}_{QD}(\mathbf{w})$ possui influência bastante significativa. Concluimos, portanto, em harmonia com nossos comentários da Seção 4.2, que o MQD não reflete o mesmo tipo de casamento que o QD para casos em que não há equalização perfeita, mas é capaz de chegar a uma solução similar. Se nos recordamos do resultado de que as soluções diferem apenas por um ajuste de ganho, teremos que as PDF's resultantes da equalização vão diferir apenas por um ajuste de escala.

Em suma, vimos que as soluções dos critérios ITL apresentam certa proximidade com respeito às soluções do CM, mas, nos falta ainda a compreensão de qual é o desempenho observado para cada abordagem. Face ao custo computacional mais elevado que abordagens de ITL representam, espera-se, desde a concepção desta nova linha de pesquisa, uma melhora significativa na qualidade do processo de equalização, principalmente nos casos onde o CM tende a mostrar dificuldades, i.e., em cenários que contem com ruídos de distribuição não Gaussiana e sinais de modulação multinível [Santamaría et al., 2002a; Lázaro et al., 2005]. Para avaliar melhor esta questão, na próxima seção, focaremos nossos esforços na análise do desempenho desses algoritmos para alguns cenários práticos que cubram tais situações.

4.4 Análise Comparativa de Desempenho

Existem várias maneiras de se avaliar o desempenho de um equalizador. Uma das mais consolidadas na literatura é aquela que busca medir o nível de interferência intersimbólica (ou IIS) residual produzida. Esta medida mostra-se bastante útil, pois é capaz de exprimir tanto a qualidade da solução obtida como a velocidade da evolução dos parâmetros do filtro; contudo, só é passível de ser avaliada mediante o conhecimento pleno do canal. Sua expressão é:

$$IIS_{dB} = 10 \log_{10} \frac{\left(\sum_{i=0}^{L_c} |c_i|^2 \right) - \max_i |c_i|^2}{\max_i |c_i|^2}, \quad (4.20)$$

em que \mathbf{c} é a resposta ao impulso combinada canal+equalizador (sob a hipótese de causalidade) e L_c é o comprimento máximo de \mathbf{c} .

É interessante notar que a medida de desempenho IIS independe de fatores de escala em relação ao equalizador \mathbf{w} – devido à razão presente em (4.20) e pelo fato de a convolução subjacente à obtenção de \mathbf{c} ser linear. Isto implica que as diferenças de ganho observadas na seção anterior não terão efeito ao compararmos o desempenho em termos da IIS residual.

Nesta seção, analisaremos o desempenho dos algoritmos SRD, SQD, MSQD e CMA, considerando sete diferentes cenários. No primeiro deles, de forma a complementar a análise prestada na seção anterior, utilizaremos o mesmo canal de fase mínima dado em (4.18); nos demais, serão enfatizados os cenários práticos encontrados na literatura, com modulações multinível e ruídos diferentes do Gaussiano.

Antes de nos atermos às simulações, resta ainda definir alguns parâmetros.

4.4.1 Ajuste dos Parâmetros

Em relação aos algoritmos oriundos do ramo de ITL, conforme apontado em [Santamaría et al., 2002a,b; Lázaro et al., 2003], a escolha de valores menores para o tamanho da janela de amostras L (como $L = 1$ ou $L = 2$) é bastante adequado, pois, além de resultar num menor custo computacional, ainda é capaz de preservar um bom desempenho. Assim, adotaremos em nossas simulações ambos os casos, $L = 1$ e $L = 2$. Entretanto, para $L = 1$, o SRD viola sua condição de existência: portanto, não será considerado. Ainda, para este mesmo valor de L , o primeiro termo em (4.9) deixa de existir, fazendo com que o algoritmo SQD reduza-se ao MSQD (a diferença é um mero fator constante): por isso, consideraremos apenas um deles.

Outro parâmetro de interesse a ser ajustado é o tamanho do kernel σ , cuja influência é determinante para a suavidade das funções custo que dele dependem. O ajuste de seu valor estará vinculado ao compromisso entre velocidade de convergência e precisão da solução obtida nos coeficientes do filtro \mathbf{w} .

Um kernel Gaussiano com σ grande possui base mais ampla (vide Fig. 4.2 para $\sigma = 5$), o que permite, considerando que sua entrada pode ser $(|y(n-j)|^2 - |y(n-i)|^2)$ ou $(|y(n-j)|^2 - |s_i|^2)$, que mesmo símbolos mais distantes tenham participação efetiva na composição da função custo: isto tende a torná-la mais suave, privilegiando uma convergência mais rápida. Por outro lado, um tamanho menor para o kernel Gaussiano (Fig. 4.2 para $\sigma = 1$) prioriza a relação entre símbolos que estão próximos, aumentando a precisão da solução obtida [Lázaro et al., 2005].

Apesar de tamanhos variáveis de kernel serem recomendados [Principe, 2010], sua implementação eleva a complexidade computacional. Dessa forma, uma solução de menor custo é aquela em que se busca utilizar um valor fixo de σ que satisfaça razoavelmente as demandas para cada caso. No problema de equalização, costuma-se utilizar $\sigma = 1$ ou $\sigma = 15$ [Santamaría et al., 2002a; Lázaro et al., 2005]. Assim, em nossas simulações, adotaremos

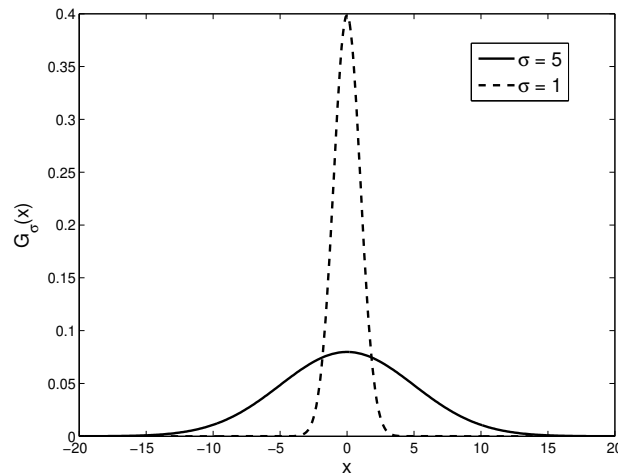


Figura 4.2: Kernel Gaussiano para $\sigma = 5$ e $\sigma = 1$.

sempre $\sigma = 1$.

Por fim, resta-nos ajustar o tamanho de passo de adaptação μ dos algoritmos. Este ajuste pode ser feito segundo diferentes critérios de análise. Nesta pesquisa, além de estarmos interessados na velocidade de convergência, desejamos, sobretudo, averiguar a qualidade da solução obtida. Assim, para que a análise seja confiável, usaremos como critério para o ajuste de μ a estabilidade da solução do algoritmo. Nesse sentido, uma das maneiras de se garantir essa condição é admitir que haja apenas pequenas oscilações sobre os coeficientes \mathbf{w} . Tal ideia vincula-se intimamente com o conceito de *misadjustment*.

Não obstante, assumir apenas um *misadjustment* estável não implica em um equilíbrio tão justo quanto possível perante os diferentes critérios. Para nivelá-los, é necessário adotar também um mesmo patamar de *misadjustment*, o que pode ser ajustado e aferido, por exemplo, através da medição da variância de cada um dos coeficientes do filtro \mathbf{w} após a convergência do algoritmo.

Para entendermos melhor o procedimento de ajuste, usaremos como exemplo o caso do canal de fase mínima (4.18) e o equalizador FIR de dois coeficientes. Inicialmente, ajustamos o tamanho de passo μ fazendo uso apenas do algoritmo CMA até observarmos um nível razoavelmente pequeno de *misadjustment*, medida através da variância dos coeficientes \mathbf{w} após a convergência (usamos, neste caso, 1.000 amostras para estimação). Satisfeito este requisito, os tamanhos de passo μ dos demais algoritmos são ajustados até que se observe variância equivalente (ou similar). A Tab. 4.1 mostra os valores obtidos para o caso em questão.

É possível observar na Tab. 4.1 que o algoritmo SQD é o que mais se distancia dos demais. Isto ocorre porque este algoritmo usa um fator para a normalização unitária dos coeficientes, tornando o ajuste por μ menos preciso. De uma maneira geral, para os demais casos de simulação, foram admitidas nas simulações variâncias da ordem de 1×10^{-4} ou

Tabela 4.1: Variância dos coeficientes \mathbf{w} após a convergência dos algoritmos.

	CMA	SRD (L=2)	SQD (L=2)	MSQD (L=2)	MSQD (L=1)
$var(w_0)$	$0,4429e-3$	$0,1334e-3$	$0,4452e-3$	$0,4218e-3$	$0,4525e-3$
$var(w_1)$	$0,6093e-3$	$0,5037e-3$	$0,7212e-3$	$0,5723e-3$	$0,6038e-3$

1×10^{-5} .

Esclarecidos os métodos adotados para o ajuste de parâmetros, voltamo-nos para as simulações para análise de desempenho.

4.4.2 Desempenho da Equalização

No primeiro cenário, analisamos o mesmo canal utilizado na Seção 4.3 – definido em (4.18) – a fim de complementarmos a análise iniciada nas superfícies de erro. Assumimos também o mesmo sinal BPSK *i.i.d.* e sem ruído. Os coeficientes do filtro foram inicializados em $\mathbf{w}(n) = [1 \ 0]$ e os tamanhos de passo μ foram ajustados a fim de alcançar um nível de *misadjustment* similar. Seus valores estão mostrados na Tab. 4.2 (para o sinal BPSK). Para os algoritmos SQD e MSQD, assumiu-se $N_s = 1$, com $|s_1|^2 = 1$. Para o algoritmo SRD, adotou-se a normalização dos coeficientes \mathbf{w} como restrição para evitar a convergência para a solução trivial. Vale lembrar que os tamanhos de kernel serão sempre $\sigma = 1$. Os resultados foram gerados a partir da média de 50 simulações independentes de Monte Carlo.

Tabela 4.2: Tamanhos de passo μ para o canal $H(z) = 1 + 0,6z^{-1}$.

Tamanho de Passo μ					
Sinal	CMA	SRD (L=2)	SQD (L=2)	MSQD (L=2)	MSQD (L=1)
BPSK	0,022	0,08	0,05	0,08	0,08
4-QAM	0,003	0,4	0,0075	0,01	0,01
16-QAM	0,0008	0,03	0,003	0,004	0,004

A Fig. 4.3(a) mostra o desempenho obtido pelos algoritmos em termos de IIS. Como podemos observar, todos eles convergem para um mesmo patamar associado às soluções ótimas em cada critério. Como a diferença entre as soluções em \mathbf{w} difere apenas por um fator de escala, conforme comentado anteriormente, isso prova que, sob a perspectiva da medida de IIS, as soluções alcançadas para CMA, SQD e MSQD são equivalentes. Notamos que apenas o SRD desvia-se um pouco desse patamar (note que, na centésima iteração, o SRD é o que possui IIS mais elevada), visto que este é o único algoritmo que apresenta uma ligeira divergência em relação à solução do CMA, como vimos na Fig. 4.1(b). Sobre

a velocidade de convergência, temos uma grande semelhança entre os algoritmos CMA e MSQD (nos casos $L = 1$ e $L = 2$), cujos desempenhos são praticamente idênticos, de forma a concordar com a semelhança observada entre suas funções custo (vide Fig. 4.1(d)). Dentre os algoritmos, o CMA parece convergir um pouco mais rápido para a solução no início, mas a alcança somente após aproximadamente 85 iterações. O SRD, apesar de inicialmente mostrar-se mais lento, consegue chegar à solução com apenas 60 iterações, sendo o mais rápido. Por último, o SQD mostra-se mais lento, exigindo mais que 100 iterações para sua convergência, o que pode ser decorrente das irregularidades presentes na sua função custo (Fig. 4.1(c)). Vale lembrar que tais velocidades de convergência são observadas considerando o tamanho de passo μ que implica num nível similar de *misadjustment*.

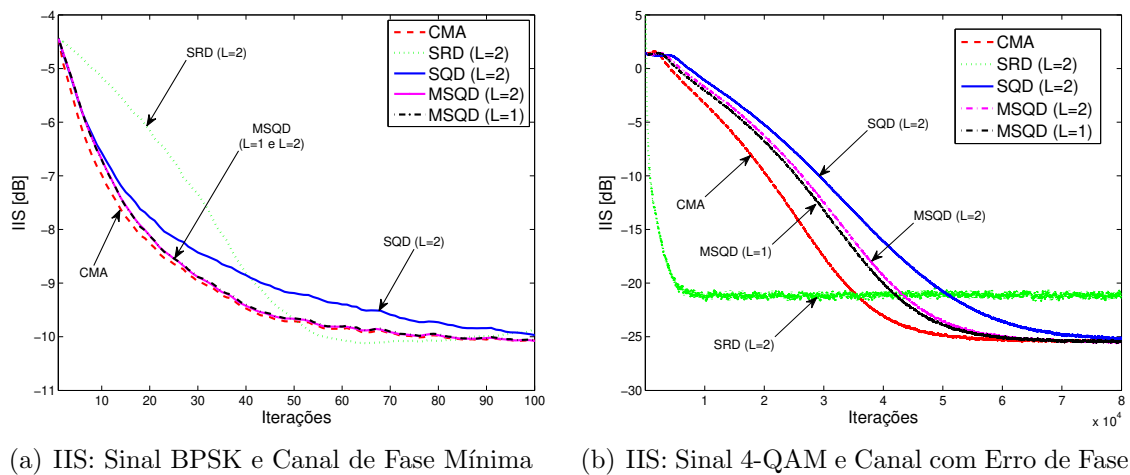


Figura 4.3: Desempenho dos algoritmos em termos de IIS.

Partindo para um cenário mais prático, utilizamos um sinal complexo modulado em 4-QAM e o seguinte canal, com erro de fase de $\pi/4$ [Santamaría et al., 2002b]:

$$H(z) = e^{j\pi/4} (0,2258 + 0,5161z^{-1} + 0,6452z^{-2} + 0,5161z^{-3}), \quad (4.21)$$

sendo adicionado ruído Gaussiano branco com razão sinal ruído (SNR, do inglês *Signal-to-Noise Ratio*) de 30 dB. O sinal transmitido $s(n)$ teve sua amplitude ajustada de forma que $R_2 = E[|s(n)|^4] / E[|s(n)|^2] = 1$. Adotamos um filtro equalizador com 21 coeficientes e inicialização center-spike [Foschini, 1985]. Para os algoritmos SQD e MSQD, assumiu-se $N_s = 1$, com $|s_1|^2 = 1$. Os tamanhos de passo obtidos para o mesmo *misadjustment* estão apresentados na Tab. 4.2. É interessante notar que, neste caso, o SRD admite um maior tamanho de passo, o que é uma característica benéfica para uma maior velocidade de convergência. É interessante comentar que o fato de lidar mais facilmente com maiores valores para μ levou os autores do SRD a chamá-lo de algoritmo rápido [Santamaría et al.,

2002a].

Os resultados em termos de IIS durante a adaptação dos coeficientes \mathbf{w} pelos algoritmos estão mostrados na Fig. 4.3(b). Considerando como critério de avaliação a qualidade da solução obtida, observamos que os algoritmos CMA, MSQD e SQD convergem todos para o mesmo nível de IIS (de aproximadamente -25 dB). Isto mostra que a convergência se deu para soluções equivalentes em cada critério. Enquanto isso, para o SRD, a pequena diferença de desempenho observada no primeiro cenário tornou-se mais acentuada, fazendo com que este se limitasse à uma IIS residual mais elevada, em torno de -21 dB. Por outro lado, sob o prisma de análise da velocidade de convergência, o SRD mostra seu ritmo acelerado, atingindo sua solução com aproximadamente 10.000 iterações. O próximo algoritmo a convergir é o CMA, necessitando em torno de 55.000 iterações. Novamente, as versões do MSQD para $L = 1$ e $L = 2$ são bastante similares, com convergência por volta de 60.000 iterações. O SQD mais uma vez é o mais lento, atingindo a marca de 80.000 iterações para convergência.

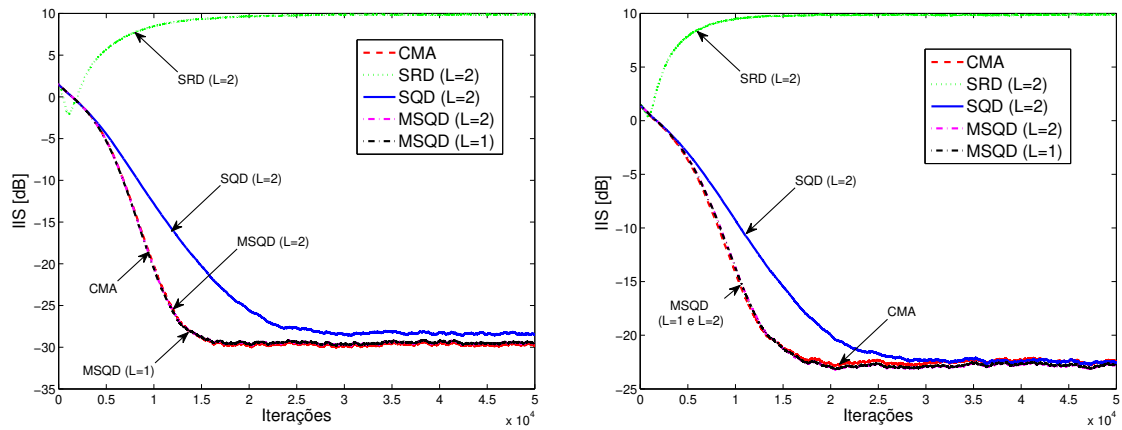
O próximo passo nos cenários de simulação é abranger o caso de sinais com modulação multinível, situação que, como sugerida por [Santamaría et al., 2002a; Lázaro et al., 2005], causa certas dificuldades para o algoritmo CMA. Dessa maneira, adotamos um sinal complexo com modulação 16-QAM. Utilizamos o mesmo canal de terceira ordem do cenário anterior, mas sem o desvio de fase [Lázaro et al., 2005]:

$$H(z) = (0,2258 + 0,5161z^{-1} + 0,6452z^{-2} + 0,5161z^{-3}). \quad (4.22)$$

Novamente existe a presença de ruído Gaussiano branco, com $\text{SNR} = 30$ dB. A constelação do sinal teve sua amplitude ajustada para que $R_2 = 1$. Mantivemos o mesmo equalizador de 21 coeficientes e a inicialização pelo método center-spike.

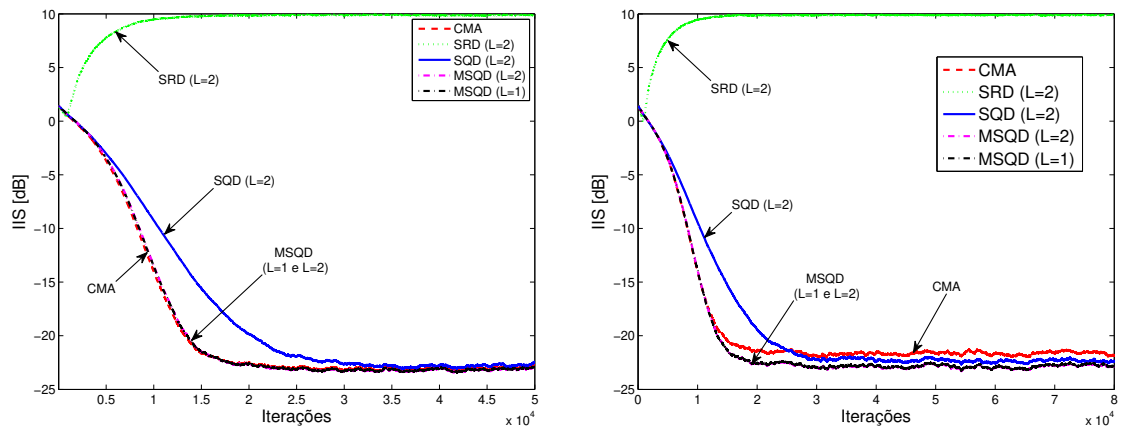
Devido ao maior número de pontos na constelação, foi necessário considerar $N_s = 4$ para os algoritmos SQD e MSQD, sendo os valores de $|s_i|^2$ obtidos a partir da constelação ajustada. Os valores assumidos do tamanho de passo dos algoritmos estão mostrados na Tab. 4.2. Mais uma vez, o SRD é o maior deles. Os resultados do desempenho podem ser conferidos na Fig. 4.4(a).

Como podemos notar, a qualidade das soluções obtidas na Fig. 4.4(a) é um pouco mais diversificada que aquelas dos cenários anteriores. Por exemplo, o SRD converge agora para um máximo local, com nível de IIS residual de 10 dB, incapaz de equalizar o canal. Isto indica que este algoritmo pode não ser tão adequado para lidar com sinais com modulação multinível, pois sempre irá existir um custo (erro) residual devido à constante distância entre os símbolos $|y(n)|^2$ e suas versões atrasadas. Possivelmente pelo mesmo motivo, o SQD (que possui um termo idêntico ao SRD), mesmo convergindo para uma solução mais adequada, tem uma IIS residual mais elevada que os algoritmos CMA e MSQD. É interessante notar



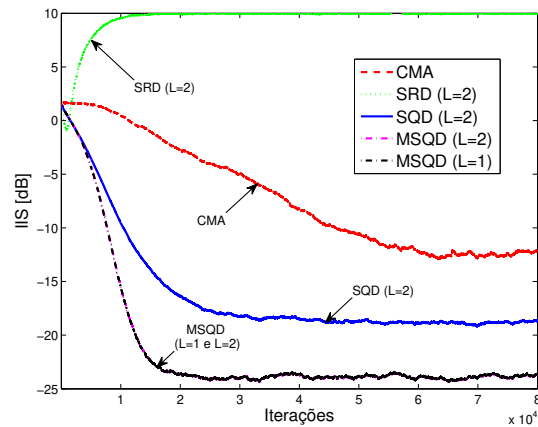
(a) IIS: Sinal 16QAM com Ruído Gaussiano e SNR = 30dB

(b) IIS: Sinal 16QAM com Ruído Gaussiano e SNR = 10dB



(c) IIS: Sinal 16QAM com Ruído Uniforme e SNR = 10dB

(d) IIS: Sinal 16QAM com Ruído Laplaciano e SNR = 10dB



(e) IIS: Sinal 16QAM com Ruído Impulsivo e SNR = 10dB

Figura 4.4: Desempenho dos algoritmos em termos de IIS para diferentes tipos de ruído.

que o CMA, apesar de considerar apenas um valor médio constante para representar o módulo da constelação – i.e., R_2 – atinge o mesmo desempenho que o MSQD. Em relação à velocidade de convergência, temos em primeiro lugar o SRD, que alcança convergência com apenas 15.000 iterações; contudo, a solução obtida é indesejada. Se considerarmos apenas os casos em que há de fato equalização, o CMA e o MSQD são os mais rápidos. Note ainda que a utilização de $L = 1$ ou $L = 2$ mostra ser indiferente.

Outro cenário interessante de se avaliar é aquele em que o ruído tem presença mais intensa. Repetindo o mesmo cenário envolvendo o sinal 16-QAM, aumentamos a intensidade do ruído Gaussiano até conseguir a SNR de 10 dB (todos os demais parâmetros foram mantidos). O resultado pode ser observado na Fig. 4.4(b). Notamos que não houve grande variação na velocidade de convergência em relação ao caso anterior (vide Fig. 4.4(a)), mas o desempenho dos algoritmos, de um modo geral, caiu. Além disso, não é mais possível distinguir claramente a diferença entre o nível de IIS alcançado pelos algoritmos CMA, SQD e MSQD: todos parecem estar no mesmo patamar de $-22,5$ dB. Contudo, após 20.000 iterações, o CMA revela uma discreta piora em relação ao MSQD, que parece ser um pouco mais robusto ao ruído.

Até o momento, constatamos que o CMA mantém uma robustez que, de certa forma, é similar à de algoritmos do ramo de ITL. Porém, como dito em [Santamaría et al., 2002a; Lázaro et al., 2005] e afirmado em [Principe, 2010], frente a sinais que divergem do caso clássico, os algoritmos de ITL ganham maior destaque. Para averiguarmos tal afirmação, consideraremos três tipos de ruído: uniforme, Laplaciano e impulsivo.

No primeiro caso, um ruído uniformemente distribuído entre $-0,5$ e $0,5$ foi gerado, substituindo o ruído Gaussiano usado no cenário do sinal modulado em 16-QAM. Todos os outros parâmetros do sistema foram mantidos para auxiliar a comparação. A intensidade do ruído foi ajustada para que a SNR resultante fosse de 10 dB. O desempenho da IIS medida nas simulações está ilustrado na Fig. 4.4(c). Notamos que quase não há divergência entre o caso correspondente com ruído Gaussiano (Fig. 4.4(b)), apenas que o CMA, na verdade, parece ser um pouco mais robusto ao ruído uniforme. Mas, como veremos, esta situação não se mantém para outros tipos de ruído.

Para o segundo tipo de ruído, o Laplaciano, foram necessárias algumas manipulações de VA's para sua obtenção: geramos inicialmente um ruído uniforme entre $-0,5$ e $0,5$ e, em seguida, para cada valor extraído desta distribuição, que chamaremos de U , usamos a relação:

$$\nu_{Laplace} = -b \operatorname{sgn}(U) \ln(1 - 2|U|), \quad (4.23)$$

em que b foi ajustado de forma a garantir uma SNR de 10 dB, no mesmo cenário do sinal 16-QAM.

Com este novo ruído, o desempenho do CMA passou a sofrer alguns prejuízos na qua-

lidade da solução, como podemos observar pela Fig. 4.4(d). Em relação ao SQD e ao MSQD, o CMA passou a ter uma IIS residual cerca de 1,5 dB maior. Os algoritmos ITL, por sua vez, têm o mesmo comportamento dos casos de ruído Gaussiano e uniforme com SNR equivalente. Em relação à velocidade de convergência, de uma maneira geral, não houve alteração.

Por fim, analisamos o cenário considerando o ruído impulsivo. Este foi gerado através do modelo de misturas de Gaussianas apresentado em [Santamaría et al., 2006]. Basicamente, a ideia consiste em combinar duas Gaussianas com diferentes variâncias:

$$\nu_{Impulsivo} = \begin{cases} G_{\sigma_1}(\nu), & \text{se } U < \epsilon \\ G_{\sigma_2}(\nu), & \text{se } U \geq \epsilon \end{cases}, \quad (4.24)$$

em que U é uma ocorrência extraída aleatoriamente de uma distribuição uniforme entre 0 e 1, $G_{\sigma_1}(\nu)$ e $G_{\sigma_2}(\nu)$ são valores aleatórios extraídos de distribuições Gaussianas de variância σ_1^2 e σ_2^2 , respectivamente, e ϵ é um valor entre 0 e 1 (de preferência, $\epsilon \ll 1$).

Com base neste modelo, escolhemos $\epsilon = 0.05$, $\sigma_1^2 = 1000\sigma_2^2$, sendo σ_2^2 ajustado para que se obtivesse uma SNR equivalente a 10 dB para o sinal com modulação 16-QAM. O resultado pode ser verificado na Fig. 4.4(e). Notamos que o CMA tem seu desempenho degradado de maneira mais intensa que nos casos anteriores. Em relação ao caso com ruído Gaussiano (Fig. 4.4(b)), a IIS residual passa de $-22,5$ dB para $-12,5$ dB, implicando na perda substancial de 10 dB. Além disso, sua velocidade de convergência também se mostrou bastante afetada: antes atingia convergência em aproximadamente 20.000 iterações, agora são necessárias cerca de 65.000 iterações, ou seja, o CMA tornou-se cerca de 3,25 vezes mais lento. Por sua vez, o algoritmo SRD manteve o mesmo desempenho, convergindo, como nos casos anteriores, para uma solução indesejada; contudo, observamos que a influência do ruído acelerou ligeiramente sua convergência.

Outro algoritmo que sofreu na presença do ruído impulsivo foi o SQD. Apesar de sua velocidade de convergência não ser alterada em relação ao caso com ruído Gaussiano, a IIS residual foi para aproximadamente -18 dB. Muito provavelmente, em vista das relações que este algoritmo tem com o SRD e MSQD – como foi mostrado na Eq. (4.17) –, este resultado é uma consequência das contribuições associadas ao termo correspondente ao SRD, pois, como podemos notar na Fig. 4.4(e), o MSQD, além de não ser prejudicado pelo ruído, parece ter convergido para solução com menor nível de IIS que o caso Gaussiano. Dessa forma, a contribuição do termo do MSQD no algoritmo SQD tenderia apenas a valorizar a qualidade de sua solução, restando ao termo relativo ao SRD a justificativa para a redução de desempenho observada.

Em termos mais gerais, os resultados envolvendo o desempenho dos algoritmos – do ponto de vista da IIS residual – nos permitem tecer algumas conclusões importantes. Em

primeiro lugar, nos cenários clássicos, o CMA tende a exibir maior grau de similaridade com os algoritmos SQD e MSQD, possuindo velocidade de convergência e nível de IIS residual similares. Por outro lado, o algoritmo SRD consegue atingir, na maioria das vezes, maior velocidade de convergência, a custo, porém, de uma redução na qualidade da solução (que por vezes conduz à soluções indesejadas).

Outro ponto de destaque é o parâmetro L no algoritmo MSQD que, seja $L = 1$ ou $L = 2$, pouca ou quase nenhuma diferença é percebida no seu desempenho. Entretanto, em relação ao SQD (o caso MSQD para $L = 1$ coincide com o algoritmo SQD para o mesmo L), o incremento em L é notável. De fato, quando $L = 2$, todos os termos no custo QD – Eq. (4.8) – são considerados, implicando em alterações de maior impacto no algoritmo. Mesmo assim, conforme observamos, os reflexos na mudança de $L = 1$ para $L = 2$ recaem principalmente na velocidade de convergência.

Os sinais com modulações multinível, no cenário avaliado, pareceram não afetar o desempenho do CMA, que manteve grande robustez e velocidade de convergência, mesmo em relação ao SQD e MSQD, que consideram em sua formulação os pontos pertencentes à constelação de maneira individual.

Podemos dizer ainda que, em relação ao CMA, a maior vantagem do uso dos algoritmos pertencentes ao ramo de ITL avaliados se deu nos cenários em que o ruído aditivo é do tipo Laplaciano ou impulsivo, casos para os quais o CMA apresentou perda no desempenho em termos de IIS residual.

Por fim, embora tenhamos analisado um número limitado de cenários, os resultados alcançados foram bastante esclarecedores e representativos da real aplicabilidade desta tripla de critérios ITL. Vimos que a proposição dos critérios cegos em ITL é de fato capaz de levar a índices superiores de qualidade de equalização em relação ao CMA, mas, isto é válido somente para casos mais restritos de cenários com ruídos não clássicos (como os ruídos Laplaciano e impulsivo). Nos demais casos, observamos grande semelhança com o comportamento do CMA.

4.5 Conclusão

No presente capítulo, buscamos tecer uma análise comparativa de um conjunto de critérios cegos ligados ao paradigma de aprendizado baseado na teoria da informação (ou ITL): o critério da entropia da dispersão e os critérios do casamento de PDF's por distância quadrática e por correlação.

Sob três diferentes perspectivas de estudo, isto é, da análise das funções custo, das superfícies de erro e do desempenho, buscamos verificar quais eram as conexões com o critério CM e também as relações que os critérios ITL estabeleciam entre si.

Em termos das funções custo, pudemos verificar que elementos do custo CM funcionavam

como peças chave para a formação dos critérios ITL. Além disso, verificamos que as três funções custo baseadas na teoria da informação podiam ser relacionadas entre si numa única equação, o que nos permitiu tirar algumas conclusões sobre a posição relativa de suas soluções.

Na análise da superfície de erro, tornou-se evidente a semelhança entre as funções custo em ITL e o custo CM, sendo que suas soluções, com ressalvas para o custo da entropia de dispersão, indicavam ser diferentes apenas por um mero ajuste de escala no cenário escolhido.

Finalmente, analisamos o desempenho da qualidade da equalização a partir da medida de interferência intersimbólica residual. Exploramos justamente as situações em que, como destacado na literatura, tende-se a ter um maior contraste entre as características dos critérios analisados. Os resultados mostraram que os algoritmos de ITL, com exceção daquele baseado na entropia da dispersão, possuem desempenho muito próximo ao do CMA, sendo que consideramos o tamanho de passo como aquele que assegura níveis semelhantes de oscilação dos coeficientes do filtro. Contudo, os critérios pertencentes ao ramo de ITL tiveram melhor desempenho na presença de ruído Laplaciano e, principalmente, impulsivo. Em relação ao algoritmo derivado da entropia da dispersão, observamos uma maior velocidade de convergência, porém, por outro lado, maior susceptibilidade à convergência para mínimos locais, principalmente nos casos em que o sinal possui modulação multinível.

Em suma, temos, para cenários não clássicos, maior robustez à ruídos com distribuições não gaussianas por parte dos algoritmos de ITL. Esta vantagem pode ser atribuída ao emprego das funções kernel, que funciona como uma espécie de filtro passa-baixas capaz de remover grandes perturbações no sistema. Por outro lado, em abordagens clássicas, não se nota grande diferença da qualidade da solução obtida em relação ao algoritmo CMA, o que confirma as semelhanças verificadas entre os critérios analisados. De fato, devido às limitações presentes na estrutura de equalização – no caso, um filtro FIR – os critérios do ramo de ITL pouco tem a adicionar, visto que, usando apenas estatísticas de segunda e quarta ordem, como no CMA, já temos informações suficientes para a equalização. Nesse sentido, seria interessante avaliar também o uso dos critérios ITL em estruturas menos limitadas, como filtros IIR e redes neurais, o que se insere no leque de perspectivas futuras.

Conclusões e Perspectivas

Neste trabalho, abordamos o problema de equalização não supervisionada em dois diferentes âmbitos: o primeiro deles concentrou-se no estudo da formulação polinomial do critério CM, enquanto o segundo voltou-se para a análise de critérios pertencentes à classe de ITL, isto é, cujo aprendizado baseia-se na teoria da informação.

No Capítulo 1, foram introduzidos os principais conceitos do problema de equalização, onde se situam, no caso supervisionado, o critério de Wiener e os algoritmos LMS e RLS [Haykin, 1996]; já no caso não supervisionado ou cego, insere-se, desempenhando um papel central neste trabalho, o critério CM [Godard, 1980], cuja análise mostrou ser, ao longo de três décadas, bastante desafiadora. Em particular, é percorrido um perfil histórico desse método, em que são descobertas a presença de mínimos locais e, também, de maneira bastante interessante, relações com os critérios de Shalvi-Weinstein e até mesmo de Wiener.

Em especial, essa última relação serviu como motivação para uma nova formulação, capaz de estreitar ainda mais os vínculos entre os critérios de Wiener e CM: a proposição da formulação polinomial do critério CM [Sousa Júnior, 2011]. O Capítulo 2 tratou de apresentar essa abordagem, onde foi considerada, como complementação ao trabalho original, a extensão para os cenários em que sinais e sistemas estavam inseridos no domínio complexo. Assim, mesmo nesse domínio, foi possível demonstrar a existência de um limitante inferior para o custo CM, como aquele existente no domínio dos reais. O passo seguinte foi analisar a validade do limitante – também interpretado como um índice de equalizabilidade cego – para novos cenários de simulação, incluindo, principalmente, os casos em que a modulação é complexa ou, até mesmo, multinível. Os resultados mostraram-se bastante satisfatórios, revelando que o limitante inferior é capaz de preservar a tendência geral do comportamento dos custos mínimos de Wiener e CM, fortalecendo a ideia de sua interpretação como índice de equalizabilidade.

Na Seção 2.3, os esforços são direcionados a um problema mais específico: a inicialização do critério CM. Conforme explorado em [Sousa Júnior, 2011], a formulação polinomial é

capaz de agregar as informações estatísticas de quarta ordem em relação ao sinal recebido de maneira mais explícita. Isto permite que o conjunto de soluções obtidas para o filtro definido na nova formulação – pertencente ao chamado domínio de Volterra – porte informações bastante relevantes capazes de auxiliar na escolha dos valores iniciais dos parâmetros do equalizador no CMA. Foi necessário, portanto, que fosse feito um mapeamento do filtro no domínio de Volterra para aquele do equalizador original. Baseando-se na heurística de inicialização introduzida em [Sousa Júnior, 2011], uma nova proposta é obtida, a qual é capaz de elevar a taxa de convergência global do CMA, em comparação, inclusive, a metodologias bastante tradicionais, como a center-spike.

Tendo como motivação esses promissores resultados, buscou-se um maior aprofundamento teórico sobre a formulação polinomial do custo CM, o que engloba a incorporação de restrições quadráticas, conforme proposto inicialmente em [Suyama, 2003]. Nesse sentido, foi apresentada no Capítulo 3 a definição do problema utilizando o método dos multiplicadores de Lagrange. Uma pequena contribuição foi feita nessa abordagem ao estendermos a definição para o caso de um filtro FIR de comprimento genérico. Na análise que se seguiu, foram exploradas as relações entre as soluções no espaço de multiplicadores de Lagrange e as soluções do critério CM. Nessa análise, houve fortes indícios de uma correspondência entre a solução de menor amplitude para os multiplicadores e as soluções globais do critério CM. Devido à complexidade do problema em mãos, foi proposta uma técnica baseada no método do gradiente capaz de tornar a identificação da solução no espaço de multiplicadores de Lagrange mais simples que uma busca exaustiva. As simulações nos cenários selecionados indicaram que o novo método pode ser bastante eficiente, contudo restam ainda alguns refinamentos, principalmente no ajuste do tamanho de passo utilizado pela técnica.

É importante ressaltar que a formulação polinomial do custo CM, apresentada nos Capítulos 2 e 3, possui uma estrutura de filtragem contida no domínio de Volterra, que, em relação ao filtro original, tem ordem maior, aumentando consideravelmente à medida que se aumenta o número de coeficientes no caso original. Dessa forma, a complexidade do problema em mãos pode ser um grande agravante, no entanto, as vantagens obtidas na abordagem muitas vezes pode justificar o incremento computacional.

Por fim, o Capítulo 4 apresentou uma análise comparativa entre três algoritmos pertencentes ao ramo de ITL e também o CMA. Partindo do estudo das funções custo, encontraram-se diversos aspectos em comum entre cada abordagem, resultando em uma única expressão que relaciona as três funções custo da classe de critérios baseados na teoria da informação. Além disso, foi possível identificar nestas funções custo a presença de elementos do custo CM clássico. As análises seguintes – em termos de superfície de erro e desempenho de equalização – deram suporte às relações constatadas anteriormente. Sob o prisma de desempenho na redução da interferência intersimbólica, o comportamento dos algoritmos foi, de uma forma geral, similar para cenários clássicos. Contudo, quando se

lida, por exemplo, com ruídos não Gaussianos, observa-se maior vantagem no emprego de algoritmos em ITL, principalmente no caso de ruído impulsivo.

Podemos concluir, portanto, que o estudo conduzido revelou algumas contribuições importantes para o problema de equalização não supervisionada. Em especial, foram providos elementos que auxiliaram para uma melhor compreensão e esclarecimento sobre o critério CM e de abordagens derivadas deste, bem como sobre alguns critérios cegos baseados na teoria da informação.

Perspectivas para Trabalhos Futuros

Das abordagens seguidas neste trabalho surgiram algumas interessantes possibilidades a serem exploradas em trabalhos futuros:

- Análise do limitante inferior para um conjunto mais diversificado de cenários, incluindo aqueles com canais não lineares. Insere-se também neste estudo o comportamento dos custos mínimos CM e de Wiener, que visará esclarecer, por exemplo, a formação de platôs nas superfícies constituídas a partir da variação do canal.
- Estudo de uma heurística de inicialização capaz de efetuar o mapeamento no domínio de Volterra para o domínio original considerando um filtro complexo. Também seria interessante, em vista da complexidade do problema para casos de maior ordem, gerar estimativas de inicialização mais complexas a partir de casos mais simples, isto é, fazer uma extrapolação de um resultado mais simples obtido no domínio de Volterra para um filtro no domínio original com maior número de coeficientes.
- Aprofundar os estudos da técnica para obtenção dos multiplicadores de Lagrange na formulação polinomial sujeita a restrições quadráticas para um conjunto mais amplo de cenários e aprimorar o ajuste do tamanho de passo utilizado no algoritmo. Se tais ajustes forem bem sucedidos, será possível também gerar uma estimativa baseada no algoritmo RLS para a matriz de correlação e para o vetor de correlação cruzada no domínio de Volterra, contribuindo para um ajuste adaptativo mais dinâmico para a técnica utilizada.
- Explorar combinações interessantes para a técnica proposta baseada no método dos coeficientes de Lagrange, como, por exemplo, pelo uso subsequente de um filtro baseado no algoritmo DD. Isto permitirá algumas simplificações na técnica proposta, visto que caberá a ela somente a tarefa de levar o sinal para a condição de “olho aberto”, o que pode ser feito sem a necessidade de operar em ordens muito elevadas para o filtro polinomial.

- Avaliar o comportamento dos critérios da classe de ITL para estruturas mais complexas, como filtros IIR e redes neurais.

Publicações

Trabalhos publicados durante o mestrado:

Fantinato, D. G. ; Attux, R. R. F. ; Sousa Júnior, C. ; Suyama, R. ; Neves, A. ; Romano, J. M. T. . Two Contributions Derived from a Polynomial Formulation of the Constant Modulus Criterion. Em *XXXI Simpósio Brasileiro de Telecomunicações (SBrT 2013)*, Fortaleza, 2013.

Fantinato, D. G. ; Attux, R. R. F. ; Neves, A. ; Suyama, R. ; Romano, J. M. T. . Blind Deconvolution of Correlated Sources Based on Second-Order Statistics. Em *XXXI Simpósio Brasileiro de Telecomunicações (SBrT 2013)*, Fortaleza, 2013.

Nose Filho, K. ; Fantinato, D. G. ; Attux, R. R. F. ; Neves, A. ; Romano, J. M. T. . A Novel Entropy-Based Equalization Performance Measure and Relations to Lp-Norm Deconvolution. Em *XXXI Simpósio Brasileiro de Telecomunicações (SBrT 2013)*, Fortaleza, 2013.

Fantinato, D. G. ; Silva, D. G. ; Nadalin, E. Z. ; Neves, A. ; Montalvão Filho, J. R. ; Attux, R. R. F. ; Romano, J. M. T. . Blind Separation of Convolutional Mixtures over Galois Fields. Em *IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2013)*, Southampton, 2013.

Bocato, L. ; Silva, D. G. ; Fantinato, D. G. ; Nose Filho, K. ; Ferrari, R. ; Attux, R. R. F. ; Neves, A. ; Montalvão Filho, J. R. ; Romano, J. M. T. . Error Entropy Criterion in Echo State Network Training. *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, páginas 35–40, 2013.

Bibliografia

- Aronszajn, N. The Theory of Reproducing Kernels and Their Applications. *Cambridge Philos. Soc. Proc*, 39:133–153, 1943.
- Bellini, S. e Rocca, F. Blind Deconvolution: Polyspectra or Bussgang Techniques? *Elsevier Science Publishers*, páginas 251–263, 1986.
- Benveniste, A., Goursat, M., e Ruget, G. Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications. *IEEE Trans. on Automatic Control*, AC-25(3):385–399, 1980.
- Chapeau-Blondeau, F. e Rousseau, D. The Minimum Description Length Principle for Probability Density Estimation by Regular Histograms. *Physica A: Statistical Mechanics and its Applications*, 388:3969–3984, 2009.
- Cover, T. M. e Thomas, J. A. *Elements of Information Theory*. John Wiley, 1991.
- Ding, Z., Kennedy, R., Anderson, B., e Johnson, C. Ill-Convergence of Godard Blind Equalizers in Data Communication Systems. *IEEE Trans. on Signal Processing*, 39:1313–1327, 1991.
- Ding, Z., Kennedy, R. A., e Anderson, B. D. O. Local Convergence of Sato Blind Equalizer and Generalizations Under Practical Constraints. *IEEE Trans. on Information Theory*, 39(1):129–144, 1993.
- Dorny, C. N. *A Vector Space Approach to Models and Optimization*. Wiley-Interscience, New York, 1975.
- Erdogmus, D. e Principe, J. C. From Linear Adaptive Filtering to Nonlinear Information Processing. *IEEE Signal Processing Magazine*, 23(2):14–33, 2006.
- Fisher, J. W. *Nonlinear Extensions to the Minimum Average Correlation Energy Filter*. Tese de doutorado, University of Florida, 1997.

- Foschini, G. Equalization Without Altering or Detecting Data. *AT&T Technical Journal*, 64(8):1885–1911, 1985.
- Godard, D. Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems. *IEEE Trans. on Communications*, COM-28(11):1867–1875, 1980.
- Godfrey, R. e Rocca, F. Zero Memory Non-Linear Deconvolution. *Geophysical Prospecting*, 29:189–228, 1981.
- Golub, G. H. e Van Loan, C. F. *Matrix Computations*. Johns Hopkins Series in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 1996.
- Haykin, S. *Adaptive Filter Theory*. Prentice Hall, 3a. edição, 1996.
- Haykin, S. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 2a. edição, 1998.
- Haykin, S. *Unsupervised Adaptive Filtering Volume 2: Blind Deconvolution*. Wiley, 2000.
- Hérault, J., Jutten, C., e Ans, B. Détection de Grandeus Primitives Dans Un Message Composite par Une Architecture de Calcul Neuromimétique en Apprentissage Non Supervise. *Actes du Xème Colloque GRETSI*, páginas 1017–1022, 1985.
- Hyvärinen, A., Karhunen, J., e Oja, E. *Independent Component Analysis*. Wiley, 2001.
- Johnson Jr., C. R., Schniter, P., Endres, T. J., Behm, J. D., Brown, D. R., e Casas, R. A. Blind Equalization Using the Constant Modulus Criterion: A Review. *Proceedings of the IEEE*, 86(10):1927–1950, 1998.
- Li, Y. e Ding, Z. Convergence Analysis of Finite Length Blind Adaptive Equalizers. *IEEE Trans. on Signal Processing*, 43(9):2120–2129, 1995.
- Lucky, R. Automatic Equalization for Digital Communication. *Bell Systems Technical Journal*, 44:547–588, 1965.
- Lázaro, M., Santamaría, I., Pantaleón, C., Erdogmus, D., e Principe, J. Matched PDF-Based Blind Equalization. *IEEE Intern. Conf. on Acoustics, Speech, and Signal Proc.*, 4:IV – 297–300, 2003.
- Lázaro, M., Santamaría, I., Erdogmus, D., Hild, K., Pantaleon, C., e Principe, J. Stochastic Blind Equalization Based on PDF Fitting Using Parzen Estimator. *IEEE Trans. on Signal Processing*, 53(2):696–704, 2005.
- Macchi, O. *Adaptive Processing*. Jhon Wiley & Sons, 1996.

- Macchi, O. e Eweda, E. Convergence Analysis of Self-Adaptive Equalizers. *IEEE Trans. on Information Theory*, IT-30(2):161–176, 1984. Special Issue on Linear Adaptive Filtering.
- Mathews, V. J. e Sicuranza, G. L. *Polynomial Signal Processing*. Wiley, 2000.
- Mendes Filho, J., Miranda, M. D., e Silva, M. T. M. Accelerating the Convergence of a Decision-Based Algorithm for Blind Equalization of QAM Signals. *3rd International Workshop on Cognitive Information Processing (CIP)*, páginas 1–6, 2012.
- Neves, A., Attux, R. R. F., Suyama, R., Miranda, M. D., e Romano, J. M. T. Sobre Critérios para Equalização Não-Supervisionada. *Revista Controle & Automação*, 17(3): 278–299, 2006.
- Parzen, E. On the Estimation of a Probability Density Function and the Mode. *Ann. Math. Statist.*, 33:1065–1067, 1962.
- Principe, J. C. *Information Theoretic Learning Renyi's Entropy and Kernel Perspectives*. Springer, 2010.
- Regalia, P. A. On the Equivalence Between the Godard and Shalvi-Weinstein Schemes of Blind Equalization. *Signal Processing*, 73(1-2):185–190, 1999.
- Rényi, A. Some Fundamental Questions of Information Theory. *Selected Papers of Alfred Renyi*, 2:526–552, 1976.
- Robbins, H. e Monro, S. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22:400–407, 1951.
- Romano, J. M. T., Attux, R. R. F., Cavalcante, C. C., e Suyama, R. *Unsupervised Signal Processing: Channel Equalization and Source Separation*. CRC Press, 2010.
- Santamaría, I., Pantaleon, C., Vielva, L., e Principe, J. C. A Fast Algorithm for Adaptive Blind Equalization Using Order- α Renyi's Entropy. *Proceedings of the International Conference on Acoustics*, 3:III–2657–III–2660, 2002a.
- Santamaría, I., Pantaleon, C., Vielva, L., e Principe, J. C. Adaptive Blind Equalization Through Quadratic PDF Matching. *Proc. Eur. Signal Process. Conf.*, II:289–292, 2002b.
- Santamaría, I., Pokharel, P., e Principe, J. Generalized Correlation Function: Definition, Properties and Application to Blind Equalization. *IEEE Trans. on Signal Processing*, 54: 2187–2197, 2006.
- Sato, Y. A Method of Self-Recovering Equalization for Multilevel Amplitude-Modulation Systems. *IEEE Trans. on Communications*, 23(6):679–682, 1975.

- Schniter, P. e Johnson Jr., C. R. Bounds for the MSE Performance of Constant Modulus Estimators. *IEEE Trans. on Information Theory*, 46(7):2544–2560, 2000.
- Shalvi, O. e Weinstein, E. New Criteria for Blind Deconvolution of Nonminimum Phase Systems (Channels). *IEEE Trans. on Information Theory*, 36(2):312–321, 1990.
- Shalvi, O. e Weinstein, E. Super-Exponential Methods for Blind Deconvolution. *IEEE Trans. on Information Theory*, 39:504–519, 1993.
- Shannon, C. E. A Mathematical Theory of Communications. *Bell Systems Technical Journal*, 27:623–656, 1948.
- Sousa Júnior, C. *Análise de Estabilidade de Lyapunov de Algoritmos Adaptativos com Contribuições ao Estudo do Critério de Módulo Constante*. Tese de doutorado, Universidade Estadual de Campinas, 2011.
- Suyama, R. *Sobre o Critério do Módulo Constante para Equalização Não-Supervisionada e Suas Relações com a Teoria de Wiener*. Dissertação de mestrado, Universidade Estadual de Campinas, 2003.
- Suyama, R., Duarte, L. T., Neves, A., Ferrari, R., Attux, R. R. F., Cavalvante, C. C., Junqueira, C., e Romano, J. Unsupervised Signal Processing: Concepts, Applications and Trends. Em *Telecommunications: Advances and Trends in Transmission, Networking and Applications*. UNIFOR, 2006.
- Treichler, J. e Agee, B. New Approach to Multipath Correction of Constant Modulus Signals. *IEEE Trans. Acoust., Speech and Signal Processing*, ASSP-31(2):459–472, 1983.
- Vapnik, V. *The Nature of Statistical Learning Theory*. Springer Verlag, New York, 1995.
- Widrow, B., McCool, J., e Ball, M. The Complex LMS Algorithm. *Proc. IEEE*, 63:719–720, 1975.
- Zeng, H. H., Tong, L., e Johnson Jr., C. R. Relationships Between the Constant Modulus and Wiener Receivers. *IEEE Trans. on Information Theory*, 44(4):1523–1538, 1998.
- Zeng, H. H., Tong, L., e Johnson Jr., C. R. An Analysis of Constant Modulus Receivers. *IEEE Trans. on Signal Processing*, 47(11):2990–2999, 1999.

Equação Normal Obtida por Diferenciação Complexa

Seja $f(\mathbf{w})$ uma função escalar de um vetor \mathbf{w} complexo com dimensões $(N+1) \times 1$. Seja o i -ésimo elemento w_i do vetor \mathbf{w} escrito como:

$$w_i = a_i + jb_i, \quad (\text{A.1})$$

onde a_i é a parte real de w_i e b_i sua parte imaginária. Assumindo $f(\mathbf{w})$ uma função diferenciável em relação à todas as $2(N+1)$ variáveis a_i e b_i , para $i = 0, 2, \dots, N$, podemos definir a derivada de $f(\mathbf{w})$ em relação ao vetor \mathbf{w} como sendo

$$\frac{df}{d\mathbf{w}} = \begin{bmatrix} \frac{\partial f}{\partial a_0} + j \frac{\partial f}{\partial b_0} \\ \frac{\partial f}{\partial a_1} + j \frac{\partial f}{\partial b_1} \\ \vdots \\ \frac{\partial f}{\partial a_N} + j \frac{\partial f}{\partial b_N} \end{bmatrix}. \quad (\text{A.2})$$

Assim, tomando como referência o custo Wiener dado por:

$$J_{Wiener}(\mathbf{w}) = \sigma_d^2 - \mathbf{p}_x^H \mathbf{w} - \mathbf{w}^H \mathbf{p}_x + \mathbf{w}^H \mathbf{R}_x \mathbf{w}, \quad (\text{A.3})$$

iremos obter sua derivada complexa, computando-a elemento a elemento [Haykin, 1996].

O primeiro termo do lado direito da Eq. (A.3), σ_d^2 , independe de \mathbf{w} , portanto, possui derivada nula.

O segundo termo pode ser escrito como:

$$\begin{aligned}\mathbf{p}_x^H \mathbf{w} &= \sum_{i=0}^N p_{x,i}^* w_i \\ &= \sum_{i=0}^N p_{x,i}^* (a_i + j b_i).\end{aligned}\tag{A.4}$$

Dessa forma,

$$\frac{\partial \mathbf{p}_x^H \mathbf{w}}{\partial a_i} = p_{x,i}^* \quad \frac{\partial \mathbf{p}_x^H \mathbf{w}}{\partial b_i} = j p_{x,i}^*,\tag{A.5}$$

para $i = 0, 1, \dots, N$. Com base na Eq. (A.2), a relação expressa em (A.5), leva a

$$\frac{d}{d\mathbf{w}} \mathbf{p}_x^H \mathbf{w} = \mathbf{0},\tag{A.6}$$

em que $\mathbf{0}$ é o vetor nulo de dimensão $(N+1) \times 1$.

De maneira similar, o terceiro termo de (A.3) pode ser escrito como:

$$\begin{aligned}\mathbf{w}^H \mathbf{p}_x &= \sum_{i=0}^N p_{x,i} w_i^* \\ &= \sum_{i=0}^N p_{x,i} (a_i - j b_i),\end{aligned}\tag{A.7}$$

cujas derivadas em relação aos termos a_i e b_i ficam

$$\frac{\partial \mathbf{w}^H \mathbf{p}_x}{\partial a_i} = p_{x,i} \quad \frac{\partial \mathbf{w}^H \mathbf{p}_x}{\partial b_i} = -j p_{x,i},\tag{A.8}$$

resultando em:

$$\frac{d}{d\mathbf{w}} \mathbf{w}^H \mathbf{p}_x = 2\mathbf{p}_{x,i}.\tag{A.9}$$

Por fim, o último termo é quadrático em relação à \mathbf{w} . Sua derivada pode ser obtida usando a regra do produto:

$$\frac{d}{d\mathbf{w}} \mathbf{w}^H \mathbf{R}_x \mathbf{w} = \frac{d}{d\mathbf{w}} \mathbf{w}^H \mathbf{c}_1 + \frac{d}{d\mathbf{w}} \mathbf{c}_2^H \mathbf{w},\tag{A.10}$$

em que $\mathbf{c}_1 = \mathbf{R}_x \mathbf{w}$ e $\mathbf{c}_2^H = \mathbf{w}^H \mathbf{R}$. Com base nos resultados obtidos em (A.9) e (A.6), o resultado pode ser expresso como:

$$\begin{aligned}\frac{d}{d\mathbf{w}} \mathbf{w}^H \mathbf{R}_x \mathbf{w} &= 2\mathbf{c}_1 + \mathbf{0} \\ &= 2\mathbf{R}_x \mathbf{w}.\end{aligned}\tag{A.11}$$

Finalmente, o vetor gradiente da função custo de Wiener pode ser escrito como:

$$\nabla J_{Wiener}(\mathbf{w}) = \frac{dJ_{Wiener}(\mathbf{w})}{d\mathbf{w}} = -2\mathbf{p}_x + 2\mathbf{R}_x\mathbf{w}. \quad (\text{A.12})$$

Seja \mathbf{w}_o o vetor de coeficientes ótimos que anulam o vetor gradiente $\nabla J_{Wiener}(\mathbf{w})$, obtemos a equação normal (que dá origem à solução de Wiener):

$$\mathbf{R}_x\mathbf{w}_o = \mathbf{p}_x. \quad (\text{A.13})$$

Função Custo Polinomial Usando Restrições Quadráticas

Desejamos obter uma nova expressão para a função custo do limitante inferior, definido como

$$J_{LI}(\boldsymbol{\theta}) = \boldsymbol{\theta}^T \mathbf{R}_\xi \boldsymbol{\theta} - 2\boldsymbol{\theta}^T \mathbf{p}_\xi + 1, \quad (\text{B.1})$$

que exprima o valor para o caso em que $\boldsymbol{\theta} = \boldsymbol{\theta}_\lambda$, i.e., a solução para o filtro polinomial no domínio de Volterra usando restrições quadráticas:

$$\boldsymbol{\theta}_\lambda = \boldsymbol{\theta}_o + \lambda \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda. \quad (\text{B.2})$$

Substituindo (B.2) em (B.1), temos

$$\begin{aligned} J_{LI}(\boldsymbol{\theta}_\lambda) &= \boldsymbol{\theta}_\lambda^T \mathbf{R}_\xi \boldsymbol{\theta}_\lambda - 2\boldsymbol{\theta}_\lambda^T \mathbf{p}_\xi + 1 \\ &= (\boldsymbol{\theta}_o + \lambda \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T \mathbf{R}_\xi (\boldsymbol{\theta}_o + \lambda \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda) - 2(\boldsymbol{\theta}_o + \lambda \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T \mathbf{p}_\xi + 1 \\ &= \boldsymbol{\theta}_o^T \mathbf{R}_\xi (\boldsymbol{\theta}_o + \lambda \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda) + \lambda (\mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T \mathbf{R}_\xi (\boldsymbol{\theta}_o + \lambda \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda) - 2\boldsymbol{\theta}_o^T \mathbf{p}_\xi \\ &\quad - 2\lambda (\mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T \mathbf{p}_\xi + 1 \\ &= \boldsymbol{\theta}_o^T \mathbf{R}_\xi \boldsymbol{\theta}_o + \lambda \boldsymbol{\theta}_o^T \mathbf{R}_\xi \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda + \lambda (\mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T \mathbf{R}_\xi \boldsymbol{\theta}_o + \lambda^2 (\mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T \mathbf{R}_\xi \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda \\ &\quad - 2\boldsymbol{\theta}_o^T \mathbf{p}_\xi - 2\lambda (\mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T \mathbf{p}_\xi + 1. \end{aligned} \quad (\text{B.3})$$

Rearranjando os termos:

$$\begin{aligned} J_{LI}(\boldsymbol{\theta}_\lambda) &= (\boldsymbol{\theta}_o^T \mathbf{R}_\xi \boldsymbol{\theta}_o - 2\boldsymbol{\theta}_o^T \mathbf{p}_\xi + 1) + \lambda \left(\boldsymbol{\theta}_o^T + \lambda (\mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T \right) \mathbf{R}_\xi \mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda \\ &\quad + \lambda (\mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T (\mathbf{R}_\xi \boldsymbol{\theta}_o - 2\mathbf{p}_\xi) \\ &= J_{LI}(\boldsymbol{\theta}_o) + \lambda \boldsymbol{\theta}_o^T \mathbf{I} \mathbf{C} \boldsymbol{\theta}_\lambda + \lambda (\mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T (\mathbf{R}_\xi \boldsymbol{\theta}_o - 2\mathbf{p}_\xi), \end{aligned} \quad (\text{B.4})$$

em que \mathbf{I} é a matriz identidade. Lembrando que $\boldsymbol{\theta}_o = \mathbf{R}_\xi^{-1} \mathbf{p}_\xi$, podemos ainda escrever:

$$\begin{aligned}
J_{LI}(\boldsymbol{\theta}_\lambda) &= J_{LI}(\boldsymbol{\theta}_o) + \lambda \boldsymbol{\theta}_\lambda^T \mathbf{C} \boldsymbol{\theta}_\lambda + \lambda (\mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T (\mathbf{R}_\xi \mathbf{R}_\xi^{-1} \mathbf{p}_\xi - 2 \mathbf{p}_\xi) \\
&= J_{LI}(\boldsymbol{\theta}_o) + \lambda \boldsymbol{\theta}_\lambda^T \mathbf{C} \boldsymbol{\theta}_\lambda + \lambda (\mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T (-\mathbf{p}_\xi) \\
&= J_{LI}(\boldsymbol{\theta}_o) + \lambda \left(\boldsymbol{\theta}_\lambda^T \mathbf{C} \boldsymbol{\theta}_\lambda - (\mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda)^T \mathbf{p}_\xi \right) \\
&= J_{LI}(\boldsymbol{\theta}_o) + \lambda \left(\boldsymbol{\theta}_\lambda^T \mathbf{C} \boldsymbol{\theta}_\lambda - \mathbf{p}_\xi^T (\mathbf{R}_\xi^{-1} \mathbf{C} \boldsymbol{\theta}_\lambda) \right) \\
&= J_{LI}(\boldsymbol{\theta}_o) + \lambda \left(\boldsymbol{\theta}_\lambda^T - \mathbf{p}_\xi^T \mathbf{R}_\xi^{-1} \right) \mathbf{C} \boldsymbol{\theta}_\lambda.
\end{aligned} \tag{B.5}$$

Gradiente da Função Custo de Restrições Quadráticas

Desejamos obter a seguinte derivada parcial:

$$\frac{\partial J_G(\boldsymbol{\theta}_\lambda)}{\partial \lambda_j} = 2 \sum_{i=1}^{N(N+1)/2} G_i(\boldsymbol{\theta}_\lambda) \frac{\partial G_i(\boldsymbol{\theta}_\lambda)}{\partial \lambda_j}. \quad (\text{C.1})$$

Sabendo que $G_i(\boldsymbol{\theta}_\lambda)$ pode ser expresso como

$$\begin{aligned} G_i(\boldsymbol{\theta}_\lambda) &= \boldsymbol{\theta}_\lambda^T \mathbf{C}_i \boldsymbol{\theta}_\lambda \\ &= (\boldsymbol{\theta}_o + \mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda)^T \mathbf{C}_i (\boldsymbol{\theta}_o + \mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda) \\ &= \boldsymbol{\theta}_o^T \mathbf{C}_i \boldsymbol{\theta}_o + \boldsymbol{\theta}_o^T \mathbf{C}_i \mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda + (\mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda)^T \mathbf{C}_i \boldsymbol{\theta}_o \\ &\quad + (\mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda)^T \mathbf{C}_i \mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda, \end{aligned} \quad (\text{C.2})$$

em que

$$\boldsymbol{\Gamma} = \sum_{k=1}^{N(N+1)/2} \lambda_k \mathbf{C}_k, \quad (\text{C.3})$$

a próxima etapa é calcular a derivada parcial de $G_i(\boldsymbol{\theta}_\lambda)$ em relação a λ_j :

$$\begin{aligned} \frac{\partial G_i(\boldsymbol{\theta}_\lambda)}{\partial \lambda_j} &= \boldsymbol{\theta}_o^T \mathbf{C}_i \mathbf{R}_\xi^{-1} \mathbf{C}_j \boldsymbol{\theta}_\lambda + (\mathbf{R}_\xi^{-1} \mathbf{C}_j \boldsymbol{\theta}_\lambda)^T \mathbf{C}_i \boldsymbol{\theta}_o \\ &\quad + (\mathbf{R}_\xi^{-1} \mathbf{C}_j \boldsymbol{\theta}_\lambda)^T \mathbf{C}_i \mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda + (\mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda)^T \mathbf{C}_i \mathbf{R}_\xi^{-1} \mathbf{C}_j \boldsymbol{\theta}_\lambda \\ &= (\boldsymbol{\theta}_o + \mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda)^T \mathbf{C}_i \mathbf{R}_\xi^{-1} \mathbf{C}_j \boldsymbol{\theta}_\lambda \\ &\quad + (\mathbf{R}_\xi^{-1} \mathbf{C}_j \boldsymbol{\theta}_\lambda)^T \mathbf{C}_i (\boldsymbol{\theta}_o + \mathbf{R}_\xi^{-1} \boldsymbol{\Gamma} \boldsymbol{\theta}_\lambda) \\ &= \boldsymbol{\theta}_\lambda^T \mathbf{C}_i \mathbf{R}_\xi^{-1} \mathbf{C}_j \boldsymbol{\theta}_\lambda + (\mathbf{R}_\xi^{-1} \mathbf{C}_j \boldsymbol{\theta}_\lambda)^T \mathbf{C}_i \boldsymbol{\theta}_\lambda. \end{aligned} \quad (\text{C.4})$$

Assumindo $\mathbf{R}_\xi^{-1}\mathbf{C}_j\boldsymbol{\theta}_\lambda$ inversível, pode-se mostrar que, tanto a matriz original quanto sua inversa, são simétricas. Dessa forma, podemos escrever

$$\frac{\partial G_i(\boldsymbol{\theta}_\lambda)}{\partial \lambda_j} = 2\boldsymbol{\theta}_\lambda^T \mathbf{C}_i \mathbf{R}_\xi^{-1} \mathbf{C}_j \boldsymbol{\theta}_\lambda. \quad (\text{C.5})$$

Finalmente, as relações obtidas (C.2) e (C.5) podem ser usadas para obter a derivada parcial definida em (C.1):

$$\begin{aligned} \frac{\partial J_G(\boldsymbol{\theta}_\lambda)}{\partial \lambda_j} &= 2 \left(G_1(\boldsymbol{\theta}_\lambda) \frac{\partial G_1(\boldsymbol{\theta}_\lambda)}{\partial \lambda_j} + \dots + G_{N(N+1)/2}(\boldsymbol{\theta}_\lambda) \frac{\partial G_{N(N+1)/2}(\boldsymbol{\theta}_\lambda)}{\partial \lambda_j} \right) \\ &= 2 \left(\boldsymbol{\theta}_\lambda^T \mathbf{C}_1 \boldsymbol{\theta}_\lambda (2\boldsymbol{\theta}_\lambda^T \mathbf{C}_1 \mathbf{R}_\xi^{-1} \mathbf{C}_j \boldsymbol{\theta}_\lambda) + \dots + \boldsymbol{\theta}_\lambda^T \mathbf{C}_{N(N+1)/2} \boldsymbol{\theta}_\lambda (2\boldsymbol{\theta}_\lambda^T \mathbf{C}_{N(N+1)/2} \mathbf{R}_\xi^{-1} \mathbf{C}_j \boldsymbol{\theta}_\lambda) \right) \\ &= 4\boldsymbol{\theta}_\lambda^T \left(\mathbf{C}_1 \boldsymbol{\theta}_\lambda \boldsymbol{\theta}_\lambda^T \mathbf{C}_1 + \dots + \mathbf{C}_{N(N+1)/2} \boldsymbol{\theta}_\lambda \boldsymbol{\theta}_\lambda^T \mathbf{C}_{N(N+1)/2} \right) \mathbf{R}_\xi^{-1} \mathbf{C}_j \boldsymbol{\theta}_\lambda \\ &= 4\boldsymbol{\theta}_\lambda^T \left(\sum_{k=1}^{N(N+1)/2} \mathbf{C}_k \boldsymbol{\theta}_\lambda \boldsymbol{\theta}_\lambda^T \mathbf{C}_k \right) \mathbf{R}_\xi^{-1} \mathbf{C}_j \boldsymbol{\theta}_\lambda. \end{aligned} \quad (\text{C.6})$$