

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO
DEPARTAMENTO DE COMUNICAÇÕES

**UMA ABORDAGEM UNIFICADA PARA ALGORITMOS
DE EQUALIZAÇÃO AUTODIDATA**

Autor

Aline de Oliveira Neves

Orientador

Prof. Dr. João Marcos Travassos Romano

Co-Orientador

Profa. Dra. Maria D. Miranda

Banca Examinadora:

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

Prof. Dr. João Cesar Moura Mota (DETI/UFC)

Prof. Dr. Max Gerken (EP/USP)

Prof. Dr. Dalton Soares Arantes (FEEC/UNICAMP)

Prof. Dr. Max Henrique Machado Costa (FEEC/UNICAMP)

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

Campinas, Dezembro de 2001

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

N414a Neves, Aline de Oliveira
Uma abordagem unificada para algoritmos de equalização autodidata / Aline de Oliveira Neves.-- Campinas, SP: [s.n.], 2001.

Orientadores: João Marcos Travassos Romano; Maria D. Miranda.

Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Processamento de sinal adaptativo. 2. Filtros adaptativos. I. Romano, João Marcos Travassos. II. Miranda, Maria D. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

Resumo

Neste trabalho propõe-se uma abordagem unificada para os critérios e algoritmos mais estudados e utilizados em equalização autodidata. Parte-se do algoritmo do módulo constante (CMA), uma referência clássica em equalização autodidata, e do algoritmo super-exponencial, por seu comprovado bom desempenho. Visam-se, em termos de critérios, mostrar as equivalências e diferenças existentes e, em termos de algoritmos, avaliar diversos parâmetros de desempenho, tais como velocidade de convergência e desajuste. Inclui-se, nesta análise, o algoritmo do módulo constante generalizado (GCMA), recentemente proposto e ainda não estudado dentro deste contexto. Faz-se uma reinterpretação deste algoritmo à luz da forma como ele é deduzido: utilizando-se um método de aproximação estocástica não usual. Apresentam-se considerações detalhadas sobre os parâmetros deste novo algoritmo e comparações de desempenho através de simulações.

Abstract

This work proposes a unified approach for the most studied and used criteria and algorithms in blind equalization. We start with the constant modulus algorithm (CMA), a classical reference in blind equalization, and the super-exponential algorithm, for its proven good performance. The aim, in terms of criteria, is to show equivalences and differences and, in terms of algorithms, it is to analyze parameters of performance such as convergence speed and misadjustment. We include, in this analysis, the generalized constant modulus algorithm (GCMA), recently proposed and still not studied in this context. We reinterpret this algorithm in light of the way it is obtained: using an unusual stochastic approximation method. Detailed considerations are provided on its parameters and comparisons of performance are presented through simulations.

“De tudo ficaram três coisas: a certeza de que estamos começando, a certeza de que é preciso continuar e a certeza de que podemos ser interrompidos antes de terminar. Fazer da interrupção um caminho novo, fazer da queda um passo de dança, do medo uma escada, do sonho uma ponte, da procura um encontro. E assim terá valido a pena existir.”

Fernando Sabino

*Aos meus pais,
Aloisio e Lourdiney*

Agradecimentos

Agradeço

À FAPESP, pelo apoio financeiro.

Aos meus pais, Aloisio e Lourdiney, cujo apoio e amor sempre me incentivaram. Certamente não há palavras suficientes para agradecer-lhes e nem para dizer-lhes o quão importante são para a minha vida.

Às minhas irmãs, Larissa e Leticia, que tornam minha vida mais alegre e às quais eu amo muito, pelo enorme apoio e carinho.

Ao Cristiano, pelo carinho, companheirismo, incentivo, apoio e ajuda, tanto em discussões técnicas que resultaram em sugestões muito interessantes para este trabalho como em questões de ordem prática.

Ao meu orientador, Prof. João Marcos T. Romano, pelo apoio desde a época da graduação, pela orientação segura e tranqüila, pelas discussões enriquecedoras. Agradeço também pela preocupação constante não só com o desenvolvimento do trabalho mas também com a minha formação e carreira.

À minha co-orientadora, Profa. Maria D. Miranda, pelas proveitosas discussões e sugestões, pelo trabalho de vir de São Paulo para que pudéssemos conversar, pela preocupação com o andamento do trabalho e pelo rigor na definição de conceitos e notações que me permitiram um grande amadurecimento técnico.

Ao Prof. Dalton Arantes, pelas sugestões e discussões enriquecedoras durante todo o

desenvolvimento deste trabalho.

Ao Prof. João César M. Mota, pelo grande interesse demonstrado e pelas sugestões muito proveitosas.

Ao Prof. Max Costa, pelas inúmeras sugestões e pela paciência e cuidado na fase final de correção desta tese.

Ao Prof. Max Gerken, pelas valiosas sugestões que não só contribuíram na fase final deste trabalho mas também abriram caminhos para a continuação deste estudo.

Ao Charles Cavalcante, pelo apoio e ajuda, principalmente com relação à edição desta tese em Latex.

À Cynthia Junqueira, pela paciência em ler esta tese antes da defesa, pelas sugestões e discussões muito interessantes.

Ao Danilo Zanatta, pelo apoio técnico.

Ao Maurício Sol pela ajuda no esclarecimento de várias dúvidas no decorrer deste trabalho.

Ao Romis Attux, pelas sugestões e discussões muito proveitosas.

Aos demais amigos do DSPCom: Cristiano Cruz, João Destro, Moisés Ribeiro, Murilo Loiola, Rafael Ferrari e Ricardo Suyama, pelo incentivo e amizade.

Ao Prof. Rodrigo Cavalcanti, pelo grande apoio, incentivo e interesse.

Ao André Fortunato, pela presença amiga na defesa desta tese.

Ao Marco Aurélio Catalão e Rodrigo Coppola, pela presença sempre alegre.

Aos amigos do DECOM.

À Eloisa, pela inestimável ajuda nas questões de ordem prática.

À Lúcia, Celi, Lena, Noêmia, Mazé, Washington, Warley e todos os funcionários da FEEC, cuja ajuda foi imprescindível.

A toda a minha família.

Aos amigos da turma EE95 que, mesmo distantes, sempre me apoiaram e incentivaram.

A Deus, sem o qual nada teria sido possível.

Lista de Abreviaturas

- ASE: Algoritmo Super-Exponencial
- AWGN: Ruído Aditivo, Branco e Gaussiano (*Additive White Gaussian Noise*)
- BPSK: *Binary Phase Shift Keying*
- CAG: Controlador Automático de ganho
- CM: Módulo Constante (*Constant Modulus*)
- CMA: Algoritmo do Módulo Constante (*Constant Modulus Algorithm*)
- EOS: Estatísticas de Ordem Superior
- EQM: Erro Quadrático Médio
- fdp: função densidade de probabilidade
- FIR: Resposta ao Impulso Finita (*Finite Impulse Response*)
- GCM: Módulo Constante Generalizado (*Generalized Constant Modulus*)
- GCMA: Algoritmo do Módulo Constante Generalizado (*Generalized Constant Modulus Algorithm*)
- i.i.d: independente e identicamente distribuído
- IIR: Resposta ao Impulso Infinita (*Infinite Impulse Response*)
- IIS: Interferência Intersimbólica
- LMS: *Least Mean Square*
- MA : Média Móvel *Moving Average*
- PAM: Modulação por Amplitude de Pulso (*Pulse Amplitude Modulation*)
- PSK: *Phase Shift Keying*
- QAM: Modulação por Amplitude em Quadratura (*Quadrature Amplitude Modulation*)
- QPSK: *Quadrature Amplitude Modulation*

- RLS: *Recursive Least Squares*
- SIR: Relação Sinal-Interferência (*Signal to Interference Ratio*)
- SNR: Relação Sinal-Ruído (*Signal to Noise Ratio*)
- SVR: Relação Sinal-Varição de Envelope (*Signal to Envelope Variation Ratio*)
- SW: Shalvi-Weinstein
- TDMA: Múltiplo Acesso por Divisão no Tempo (*Time Division Multiple Access*)
- ZF: *Zero-Forcing*

Índice

1	Introdução	1
2	Métodos para Equalização Autodidata	7
2.1	Introdução	7
2.2	Equalização Supervisionada	9
2.2.1	Filtragem de Wiener	11
2.2.2	Algoritmo do Gradiente Determinístico	12
2.2.3	O Algoritmo LMS	13
2.3	Equalizadores <i>Zero-Forcing</i>	14
2.4	Algoritmos de Bussgang	17
2.4.1	Definição do Problema	17
2.4.2	Estimador Não-Linear de Memória Nula	19
2.4.3	Técnicas de Bussgang para Equalização Autodidata	21
2.4.4	Considerações sobre Convergência	23
2.4.5	Extensão para Canais em Banda-Base Complexos	26
2.4.6	Algoritmo de Decisão Direta	26
2.4.7	Algoritmo de Sato	28
2.4.8	Algoritmos de Godard	29
2.5	Algoritmos de Shalvi-Weinstein	32
2.5.1	Fundamentos Matemáticos	32
2.5.2	Definição do Problema	33
2.5.3	Critérios com Restrição	34
2.5.4	Critérios Normalizados	36
2.5.5	Critérios sem Restrição	38
2.5.6	O Algoritmo Super-Exponencial	39
2.5.7	Revisitando a Literatura	49
2.6	Critério do Módulo Constante Generalizado	52
2.6.1	O critério	53
2.6.2	O algoritmo: Questão da aproximação estocástica	53
2.7	Conclusão	55

3	Comparação Entre os Critérios	57
3.1	Revisitando Resultados da Literatura	57
3.1.1	Primeira Abordagem	57
3.1.2	Resultados Obtidos por P. Regalia	59
3.1.3	Regiões de Atração de Pontos Estacionários no Critério de Shalvi-Weinstein	63
3.1.4	Relação com Wiener e ZF	65
3.2	GCM neste contexto	68
3.3	Equivalências do Ponto de Vista do Vetor Gradiente	70
3.4	Conclusão	72
4	Comparação Entre os Algoritmos	75
4.1	Formas de Análise de Desempenho	75
4.2	Análise de Convergência dos Algoritmos	77
4.3	Desempenho dos Algoritmos	81
4.3.1	Modulação BPSK	84
4.3.2	Modulação QPSK	89
4.4	Estudo do GCMA	94
4.4.1	Considerações sobre o efeito do fator de esquecimento λ	100
4.4.2	Alterando a inicialização	103
4.4.3	Alterando o passo de adaptação	105
4.5	Conclusão	107
5	Conclusões e Perspectivas	109
	Apêndice	113

Capítulo 1

Introdução

A capacidade de se comunicar sempre foi muito prezada pelo homem. A comunicação à distância, entretanto, sempre trouxe várias dificuldades. Desde a invenção do telefone no século XIX até os dias de hoje, as formas de comunicação evoluíram de maneira espantosa. Mais recentemente, o desenvolvimento das comunicações digitais, representaram um novo marco nesta evolução, por sua simplicidade e baixa complexidade quando comparado aos sistemas analógicos de transmissão . Esta nova forma de transmissão, traz como vantagens o oferecimento de novos serviços aos usuários, um aumento na capacidade de transmissão e uma melhora na qualidade do sinal, que agora pode ser mantida por distâncias maiores. Hoje em dia, com o uso de tecnologias de *banda larga*, busca-se a transmissão simultânea de voz, dados e imagens.

Estas novas técnicas, entretanto, também trazem novos problemas. O meio pelo qual se propaga o sinal transmitido com a informação, chamado de canal, introduz distorções lineares ao sinal. Estas distorções podem ser provenientes de multi-percurso e de desvanecimento, existentes por exemplo em sistemas de comunicações móveis, e da limitação em largura de banda do canal.

Estas distorções lineares são acentuadas pela interferência intersimbólica (IIS) que, resumidamente, consiste na extensão de um símbolo transmitido sobre um intervalo de tempo maior do que sua duração original, fazendo com que interfira no próximo símbolo transmitido. Assim, o sinal é corrompido ao ponto de sua recuperação correta se tornar

inviável.

A equalização do canal tem se mostrado um método muito eficaz na remoção da IIS, compensando as distorções introduzidas pelo canal. O dispositivo que realiza esta operação é chamado de equalizador, usualmente implementado como um filtro digital de resposta ao impulso finita.

Entretanto, na grande maioria das aplicações, o canal não é conhecido ou varia com o tempo. Isto requer um equalizador que possa ser ajustado, o que pode ser implementado por meio de um filtro adaptativo.

Tradicionalmente, para se ajustar os coeficientes do equalizador, utiliza-se uma sequência de treinamento, ou seja, uma sequência transmitida que é conhecida e utilizada no receptor como referência. Esta técnica é chamada equalização supervisionada. O objetivo então é se adaptar os coeficientes do equalizador de forma a minimizar o erro existente entre a sua saída e o símbolo que foi transmitido, otimizando assim uma função custo tal como o erro quadrático médio. Os coeficientes do filtro são ajustados através de um algoritmo obtido para minimizar, recursivamente, a função custo em questão. Os algoritmos mais conhecidos e utilizados são o LMS (*Least Mean Square*) e o RLS (*Recursive Least Squares*).

Durante este período de treinamento não existe transmissão efetiva de informação, já que tudo o que é transmitido é conhecido pelo receptor. Após esta fase inicial, o equalizador continua seu processo de adaptação utilizando, como sinal de referência, não mais a sequência de treinamento, mas sim o próprio símbolo obtido na saída do equalizador e decidido. Normalmente, este processo é necessário a fim de garantir a convergência para o melhor filtro possível ou para rastrear pequenas variações do canal.

Entretanto, nem sempre é possível ter acesso a uma sequência de treinamento. Dependendo da aplicação, esta pode representar um custo adicional significativo ou pode até impossibilitar o funcionamento do sistema. Por exemplo, em sistemas *broadcast* multi-usuário, é altamente indesejável que o transmissor tenha que parar a transmissão para todos os seus usuários, somente para treinar um novo usuário que se conecte à rede. Em sistemas de comunicações móveis, um desvanecimento severo pode inutilizar uma sequência de treinamento transmitida, fazendo com que o equalizador não

seja treinado convenientemente, resultando em uma queda na qualidade do sinal que chega ao usuário. Existe portanto uma necessidade prática para um tipo especial de equalizador, chamado equalizador autodidata, que não precisa de sequência de treinamento. Seu objetivo é recuperar o sinal de entrada (desconhecido) de um canal (também desconhecido), baseado somente no sinal recebido (saída do canal) e em características estatísticas e probabilísticas do sinal inicialmente transmitido. Eliminando a sequência de treinamento e maximizando a capacidade do canal para transmissão somente de informação, a equalização autodidata representa uma solução eficiente, em termos de largura de banda, para compensar distorções.

Devido à inexistência da sequência de treinamento, a função custo a ser otimizada utiliza como sinal de referência uma estimativa não-linear do sinal transmitido, o que introduz problemas que não estão presentes no caso da equalização supervisionada, como a existência de mínimos locais. Os últimos anos têm visto uma quantidade significativa de trabalhos nesta área, estudando algoritmos já existentes e propondo novos. Entretanto, estes algoritmos praticamente não são utilizados na prática. Algumas das razões que levam a isto são a baixa velocidade de convergência dos algoritmos, o que pode causar problemas principalmente em sistemas móveis que possuem canais variantes, e a diferença de desempenho, complexidade e robustez quando comparados a algoritmos supervisionados [Ding e Li 01].

Os primeiros algoritmos autodidatas propostos na literatura pertencem à chamada família de *algoritmos de Bussgang*. Dentro destes estão o algoritmo de decisão direta, normalmente utilizado após uma fase inicial de treinamento, que foi um dos primeiros algoritmos em equalização autodidata; o algoritmo de Sato, que foi o primeiro capaz de operar em condições de “olho fechado” e os algoritmos de Godard que possuem, como caso particular, o algoritmo do módulo constante (CMA). Este último se tornou uma referência clássica em equalização autodidata, sendo o algoritmo mais estudado e utilizado na prática. Ele possui como principal característica, otimizar uma função custo que penaliza desvios do sinal na saída do equalizador com relação a um módulo constante, definido pelas estatísticas do sinal inicialmente transmitido.

Entretanto, percebeu-se que talvez não houvesse a necessidade de especificar o valor do módulo do sinal na saída do equalizador. No caso de modulações PSK (*Phase Shift Keying*), por exemplo, onde o sinal possui módulo constante, qualquer sinal na saída do

equalizador que mantenha esta propriedade representa uma equalização bem sucedida. Deste fato então foi proposto, primeiramente em [Cavalcanti et al. 98], o algoritmo do módulo constante generalizado (GCMA), que flexibiliza o raio da constelação do sinal de saída.

Esse estudo, inicialmente, se dedica a re-situar este algoritmo, tendo em vista algoritmos já bastante estudados na literatura como o CMA e o algoritmo super-exponencial (ASE) [Shalvi e Weinstein 90]. Percebeu-se, que o critério GCM e o critério de Shalvi-Weinstein são mais semelhantes do que se imaginava num primeiro momento. Buscou-se então dar uma abordagem unificada para estes três critérios, CM, GCM e Shalvi-Weinstein, à luz do trabalho realizado por P. Regalia em [Regalia99]. Já em termos de algoritmos, foi observado que o GCMA é deduzido utilizando-se um método de aproximação estocástica pouco usual, o que traz a necessidade de um estudo mais detalhado de suas características e seus parâmetros. Foram realizadas também comparações de desempenho entre o CMA, GCMA e ASE em termos de velocidade de convergência e desajustes.

Assim, o trabalho foi organizado da seguinte forma:

Capítulo 2- Métodos para Equalização Autodidata: É feita uma revisão dos algoritmos de Bussgang, nos quais se incluem o algoritmo de decisão direta, o algoritmo de Sato e o de Godard (do qual um caso particular é o CMA). Em seguida, tratamos dos critérios e algoritmos de Shalvi-Weinstein, mostrando detalhadamente a dedução do algoritmo super-exponencial. Neste ponto é feita uma breve revisão bibliográfica dos principais estudos realizados sobre este último. Finalmente, introduzimos o critério do módulo constante generalizado (GCM) e seu algoritmo, esclarecendo a questão da aproximação estocástica utilizada na sua dedução.

Capítulo 3- Comparação Entre os Critérios: Analisam-se os diversos critérios em questão, buscando mostrar suas principais equivalências e diferenças. Para isto, iniciamos com uma revisão dos resultados já obtidos na literatura, inclusive comparando os critérios autodidatas com critérios supervisionados, como o de Wiener. Incluímos nesta análise o critério GCM e propomos novas formas de se enxergar estas equivalências.

Capítulo 4- Comparação Entre os Algoritmos: É feita uma análise dos algoritmos,

tanto em termos de posições de mínimos locais, como em termos de desempenho, velocidade de convergência e desajuste, quando utilizados para equalizar canais de fase não mínima, com adição de ruído. Aproveitamos para estudar também mais detalhadamente o algoritmo GCMA e o porquê de seu bom desempenho.

Capítulo 5- Conclusão e Perspectivas: Apresentam-se a conclusão deste trabalho e perspectivas para estudos futuros.

Capítulo 2

Métodos para Equalização Autodidata

Este capítulo aborda brevemente o problema da equalização em comunicações digitais, partindo da técnica supervisionada e seguindo com uma discussão mais detalhada sobre os critérios e algoritmos autodidatas, incluindo os algoritmos de Bussgang, o critério de Shalvi-Weinstein e o algoritmo super-exponencial. Por fim, introduzimos o critério e o algoritmo do módulo constante generalizado (GCMA), que será objeto de um estudo detalhado nos próximos capítulos.

2.1 Introdução

Uma das formas de equalização mais utilizadas na prática é a chamada equalização supervisionada, na qual existe uma sequência de dados que já é conhecida *a priori* pelo receptor. Este então a utiliza como referência para o ajuste de seus coeficientes, procurando minimizar o erro entre a sua saída e estes dados já conhecidos. Esta sequência é chamada de sequência de treinamento. Note que, neste caso, não existe transmissão efetiva de informação, pois o receptor já conhece tudo o que está sendo transmitido. Após esta fase inicial, se a saída do equalizador for próxima ao sinal transmitido originalmente, o sistema pode ser chaveado para o chamado *modo de decisão direta*, quando

se inicia realmente a transmissão. Nesta fase, os coeficientes do equalizador podem continuar sendo adaptados para rastrear alguma pequena variação do canal ou para procurar atingir uma melhor configuração, mas o receptor já não possui aquela sequência de referência, buscando minimizar o erro entre a saída do equalizador e a saída de um dispositivo de decisão.

Quando a transmissão de uma sequência de treinamento se torna um procedimento muito custoso, difícil ou até mesmo impossível, outras formas de adaptação dos coeficientes do equalizador devem ser encontradas. Define-se então a chamada equalização autodidata ou cega, quando o receptor não conhece, em momento algum, o que está sendo transmitido *a priori*. Deve-se portanto buscar outros critérios a serem otimizados pois o erro entre a saída do equalizador e a sequência de treinamento não está mais disponível.

Os algoritmos utilizados em equalização autodidata podem ser divididos basicamente em duas famílias: algoritmos baseados em estatísticas de ordem superior (EOS) e algoritmos baseados em estatísticas ciclo-estacionárias. No primeiro caso, ainda pode-se subdividi-los em algoritmos que utilizam EOS explicitamente ou implicitamente. Nestes últimos incluem-se os algoritmos de Bussgang, cuja característica principal é a minimização de uma função custo não-convexa, o que leva a problemas de mínimos locais, com a vantagem de apresentar uma baixa complexidade computacional. Os algoritmos que utilizam EOS explicitamente, utilizam cumulantes de ordem superior ou suas transformadas discretas de Fourier, conhecidas como poli-espectros. A propriedade destes últimos de preservar informações de fase, faz com que sejam apropriados para a desconvolução autodidata. As principais características deste tipo de algoritmo são a alta complexidade computacional e a estimação do canal sem o uso de uma função custo, o que evita o problema de mínimos locais. Neste trabalho, trataremos somente dos algoritmos autodidatas baseados em EOS implícita ou explicitamente.

Uma limitação comum da identificação e equalização autodidata de canais através de algoritmos baseados em EOS é a baixa taxa de convergência. Esta limitação pode impedir seu uso em situações que requerem aquisições rápidas do sinal transmitido. Nestes casos, entretanto, pode-se utilizar as características ciclo-estacionárias presentes nos sinais recebidos.

Os algoritmos autodidatas a serem utilizados dependem de uma característica importante do equalizador. Para que as estatísticas ciclo-estacionárias dos sinais possam ser exploradas, estes devem ser amostrados numa taxa maior do que a de um intervalo de símbolo, resultando nos chamados equalizadores fracionários. Isto já não é necessário no caso do uso de algoritmos baseados em estatísticas de ordem superior. Neste caso, os sinais podem ser amostrados a cada intervalo de símbolo resultando em equalizadores chamados *baud-rate*. Somente estes últimos serão estudados neste trabalho.

Para que um canal seja corretamente equalizado, existe a necessidade de se identificar tanto a magnitude como a fase da função de transferência do sistema desconhecido. A magnitude pode ser identificada usando momentos de segunda ordem do sinal de saída do canal. Entretanto, a identificação da fase é mais complicada, necessitando do uso de momentos de ordem superior. Justamente devido a isto, a distribuição de probabilidade do sinal de entrada do sistema não pode ser gaussiana, já que esta é completamente caracterizada por sua média e variância (momento de segunda ordem), possuindo cumulantes de ordem superior nulos [Benveniste et al. 80].

Neste capítulo, será feita uma breve revisão sobre equalização supervisionada, seguida da descrição detalhada dos critérios e algoritmos autodidatas objetos de estudo deste trabalho: os algoritmos de Bussgang, o de Shalvi-Weinstein e o algoritmo do módulo constante generalizado (GCMA).

2.2 Equalização Supervisionada

A Figura 2.1 em seguida ilustra, de forma bastante simples, o sistema em estudo, em banda base.

O sinal $a(n)$ é transmitido por um canal com resposta ao impulso h finita (FIR), que introduz distorções lineares, o que resulta na chamada interferência intersimbólica (IIS). Este canal será modelado como um filtro linear transversal, invariante no tempo e, por enquanto, sem a adição de ruído. A fim de procurar recuperar o sinal que foi transmitido, utiliza-se então um equalizador adaptativo que, para este estudo, também será um filtro linear transversal como mostrado na figura 2.2.

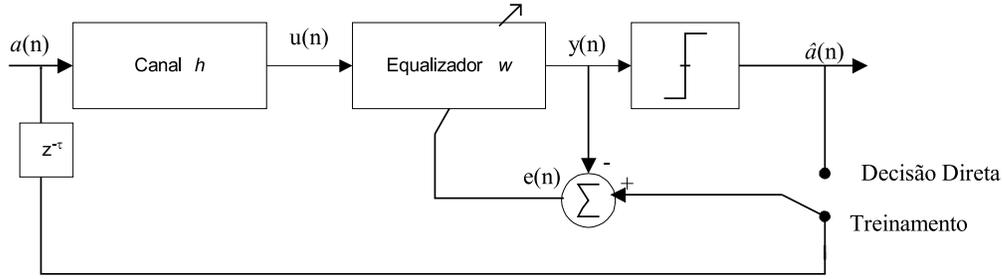


Figura 2.1: Sistema em estudo, caso supervisionado

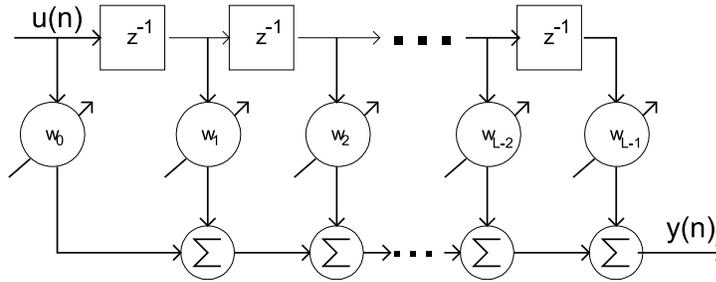


Figura 2.2: Equalizador como um filtro transversal

Portanto:

$$\begin{aligned} u(n) &= \mathbf{a}^T(n)\mathbf{h} \\ y(n) &= \mathbf{w}^T(n)\mathbf{u}(n) \end{aligned} \quad (2.1)$$

onde $\mathbf{a}(n) = [a(n) \ a(n-1) \ \dots \ a(n-M+1)]^T$ e $\mathbf{h} = [h_0 \ h_1 \ \dots \ h_{M-1}]^T$, com $(M-1)$ sendo a ordem do canal, $\mathbf{u}(n) = [u(n) \ u(n-1) \ \dots \ u(n-L+1)]^T$ com $(L-1)$ sendo a ordem do equalizador, $\mathbf{w}(n) = [w_0(n)w_1(n) \ \dots \ w_{L-1}(n)]^T$ e $y(n)$ é a saída do equalizador. O super-escrito T denota transposição.

O dispositivo de decisão tem por objetivo remover alguma IIS residual que ainda exista no sinal de saída do equalizador. Se o canal tiver sido equalizado corretamente e o diagrama de olho neste ponto estiver aberto, o decisor conseguirá recuperar a sequência transmitida sem erros. Basicamente, este dispositivo procura qual o símbolo da constelação do sinal transmitido mais próximo da atual saída $y(n)$. Por exemplo, se $a(n)$ pertencer à constelação BPSK (*Binary Phase Shift Keying*) $\{-1, +1\}$, valores de $y(n)$ positivos serão decididos por $+1$ e valores negativos serão -1 .

2.2.1 Filtragem de Wiener

Durante o período de transmissão da sequência de treinamento, conhecida no receptor, o ajuste dos coeficientes do equalizador é feito de forma a minimizar o erro quadrático médio (EQM) entre a saída do equalizador $y(n)$ e a resposta desejada $d(n)$ (sinal transmitido a ser recuperado):

$$J_{EQM} = E [|e(n)|^2] = E [|d(n) - y(n)|^2] = E [|d(n) - \mathbf{w}^T(n)\mathbf{u}(n)|^2] \quad (2.2)$$

onde $E[\cdot]$ é o operador esperança e $d(n)$ é dado por $a(n-\tau)$ sendo τ um atraso constante. J_{EQM} é chamada função custo.

Para obter os valores dos coeficientes do equalizador, w_i , $i = 0, \dots, L-1$, que minimizam (2.2), precisamos então derivar J_{EQM} em relação a $\mathbf{w}(n)$. Como este é complexo, podemos escrever:

$$w_i = a_i + jb_i \quad (2.3)$$

e portanto o gradiente de J_{EQM} será dado por

$$\nabla_i J_{EQM} = \frac{\partial J}{\partial a_i} + j \frac{\partial J}{\partial b_i} \quad (2.4)$$

Usando então (2.2) teremos:

$$\begin{aligned} \frac{\partial E [|e(n)|^2]}{\partial \mathbf{w}} &= E \left[\frac{\partial e(n)e^*(n)}{\partial \mathbf{w}} \right] = -2E[e(n)\mathbf{u}^*(n)] = \\ &= -2E[\mathbf{u}^*(n)(d(n) - \mathbf{w}^T(n)\mathbf{u}(n))] = 0 \end{aligned} \quad (2.5)$$

Definindo $\mathbf{R}_u = E[\mathbf{u}^*(n)\mathbf{u}^T(n)]$ e $p_{ud} = E[d(n)\mathbf{u}^*(n)]$, onde o super-escrito * denota conjugação, temos de (2.5) que:

$$E[\mathbf{u}^*(n)d(n)] - E[\mathbf{u}^*(n)\mathbf{u}^T(n)\mathbf{w}] = 0 \Rightarrow \mathbf{w} = \mathbf{R}_u^{-1}\mathbf{p}_{ud} \quad (2.6)$$

O desenvolvimento realizado acima é chamado de filtragem de Wiener [Haykin96] e através dele é possível encontrar o valor ótimo para \mathbf{w} . A equação (2.6) é chamada de equação de Wiener-Hopf. É interessante observar que existe somente uma solução

que anula o gradiente da função custo J_{EQM} , dada por (2.6). Portanto, esta função é *unimodal*.

Entretanto não é factível utilizar (2.6) para equalização em tempo real, pois existe a necessidade de estimar \mathbf{R}_u e \mathbf{p}_{ud} , desconhecidos. Além disso, solucionar as equações de Wiener-Hopf pode ser muito custoso computacionalmente, principalmente no caso de equalizadores com muitos coeficientes e operando a altas taxas. Procura-se então um algoritmo que realize uma otimização iterativa, de forma que a função custo $J_{EQM}(\mathbf{w}(n+1))$ seja menor do que $J_{EQM}(\mathbf{w}(n))$ e portanto, com o passar do tempo, aproxime-se do ponto onde $J_{EQM}(\mathbf{w})$ é mínima. Uma alternativa é o chamado algoritmo do gradiente determinístico ou *steepest descent*.

2.2.2 Algoritmo do Gradiente Determinístico

Partindo-se do fato de que o valor da função J_{EQM} no ponto $\mathbf{w}(n+1)$ pode ser obtido a partir do seu comportamento no ponto $\mathbf{w}(n)$ através de uma expansão em série de Taylor, o chamado algoritmo do gradiente determinístico (também chamado *steepest descent*), é obtido truncando-se esta expansão na derivada de primeira ordem, ou seja, no cálculo do gradiente da função. Intuitivamente, é razoável que sucessivas correções no sentido contrário ao gradiente da função, ou seja, na direção de sua maior declividade, acabe levando ao ponto de mínimo desta. Assim, dada uma função custo J , a atualização dos coeficientes do equalizador \mathbf{w} poderia ser expressa como:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \nabla J(n) \quad (2.7)$$

onde μ é um passo de adaptação, constante, real e positivo e $\nabla J(n)$ é o gradiente da função $J(n)$.

Calculando então o gradiente de (2.2), função custo a ser otimizada, obtém-se a expressão do algoritmo do gradiente estocástico (para maiores detalhes desta dedução, olhar [Haykin96]):

$$\mathbf{w}(n+1) = \mathbf{w}(n) + 2\mu(\mathbf{p}_{ud} - \mathbf{R}_u \mathbf{w}(n)) \quad (2.8)$$

O fator multiplicativo 2 pode ser incluído no próprio passo de adaptação do algo-

ritmo, μ . O algoritmo acima tem como único ponto de convergência (mínimo global) a solução de Wiener. Note que, apesar do algoritmo ser iterativo, pressupõe-se que os dados já foram adquiridos e \mathbf{R}_u e \mathbf{p}_{ud} já tenham sido estimados. Entretanto, isto pode requerer um custo computacional muito alto em certos casos, além de dificultar o processamento em tempo real. Surge então a idéia de algoritmos obtidos através de aproximação estocástica, que será vista na próxima seção.

2.2.3 O Algoritmo LMS

A idéia da aproximação estocástica foi proposta inicialmente em [Robbins e Monro 51], e constitui a base de algoritmos em filtragem adaptativa, assim como em outras áreas como identificação e controle adaptativo, reconhecimento de padrões por redes neurais, além de outros. A idéia é bastante simples: substituir o operador esperança pelos valores instantâneos dos sinais envolvidos. Tomando por base o algoritmo do gradiente determinístico (2.8), faremos as seguintes alterações:

$$\begin{aligned}\mathbf{R}_u &\rightarrow \mathbf{R}_{u_{ap}} = \mathbf{u}^*(n)\mathbf{u}^T(n) \\ \mathbf{p}_{ud} &\rightarrow \mathbf{p}_{ud_{ap}} = d(n)\mathbf{u}^*(n)\end{aligned}\quad (2.9)$$

Substituindo então (2.9) em (2.8) temos:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu (d(n)\mathbf{u}^*(n) - \mathbf{u}^*(n)\mathbf{u}^T(n)\mathbf{w}(n)) \quad (2.10)$$

O que resulta em:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \mathbf{u}^*(n) (d(n) - \mathbf{u}^T \mathbf{w}(n)) = \mathbf{w}(n) + \mu e(n)\mathbf{u}^*(n) \quad (2.11)$$

Este é o chamado algoritmo LMS (*Least Mean Square*), cujo nome deve-se ao fato de que o algoritmo minimiza o erro quadrático médio, por meio de uma busca estocástica.

Através da equação final do algoritmo (2.11), vemos que este é bastante simples, permitindo o processamento em tempo real e, portanto, sendo atraente do ponto de

vista prático. Devido à aproximação utilizada, o algoritmo não converge mais para um ponto fixo (solução de Wiener), como no caso do algoritmo do gradiente determinístico, mas ficará oscilando em torno do ponto estacionário.

A primeira vista, pode parecer que o algoritmo LMS não possui bom desempenho já que as estimativas instantâneas de \mathbf{R}_u e \mathbf{p}_{ud} possuem variâncias relativamente altas. Entretanto, é preciso lembrar que o LMS é recursivo e sendo assim, o próprio algoritmo, de certa forma, realiza uma média destas estimativas a cada iteração.

2.3 Equalizadores *Zero-Forcing*

Iniciando então o estudo dos equalizadores autodidatas, vale ressaltar que nos restringiremos ao espaço *baud-rate*, ou seja, a saída do canal, $u(t)$ é amostrada na taxa de símbolos, assumindo sincronização perfeita, podendo ser escrita como:

$$u(n) = \sum_{i=-\infty}^{\infty} h(i)a(n-i) + r(n) \quad (2.12)$$

onde $a(n)$ é o símbolo transmitido e $r(n)$ é o ruído. A sequência $u(n)$ é então discreta e estacionária.

Considere o modelo em banda base de um sistema de comunicações digitais como mostrado na Figura 2.3.

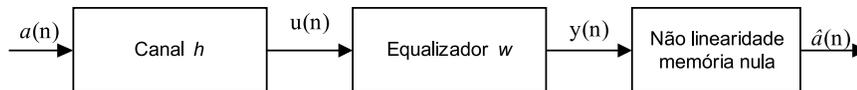


Figura 2.3: Sistema em estudo, caso autodidata

Este modelo consiste em um canal com resposta ao impulso finita (FIR), dada por h , sendo esta desconhecida. Quando o canal não é ideal, h possui mais do que uma componente não nula. Isto introduz uma distorção indesejável ao sinal, fazendo com que $u(n)$ dependa de símbolos transmitidos em instantes anteriores. Este fenômeno é conhecido como interferência intersimbólica (IIS) e pode corromper severamente o sinal

transmitido. A IIS é usualmente causada por canais limitados em largura de banda, multi-percurso ou canais com desvanecimento. Um simples dispositivo de decisão sem memória não é capaz de recuperar a sequência original em casos de IIS severa. Equalizadores, por outro lado, têm se mostrado bastante eficientes na sua remoção. Estes, por sua vez, nada mais são do que filtros lineares com resposta ao impulso finita (FIR) w , tendo como objetivo gerar um sinal $y(n)$ que seja uma estimativa confiável do sinal inicialmente transmitido:

$$y(n) = \hat{a}(n) = a(n - \tau) \quad (2.13)$$

onde τ é um atraso constante.

Nesta primeira análise, o efeito do ruído será ignorado, supondo-se que a relação sinal-ruído (SNR-*Signal to Noise Ratio*) seja bastante elevada. O problema então consiste em recuperar $a(n)$, conhecendo-se somente a sequência $u(n)$ ou, de maneira equivalente, encontrar um equalizador autodidata que possua uma resposta impulsiva inversa à do canal desconhecido, sem se ter acesso à informação originalmente transmitida, ou seja, sem sequência de treinamento. Para isto, bastaria identificar a função de transferência do canal $h(z)$ dada por:

$$h(z) = \sum_{i=0}^{\infty} h_i z^{-i} \quad (2.14)$$

Uma vez identificando o canal $h(z)$, o equalizador pode ser construído conforme o critério *zero-forcing* (ZF), o que resultará no seguinte equalizador ótimo [Ding e Li 2001]:

$$w(z) = \frac{z^{-\tau}}{h(z)} \quad (2.15)$$

Este equalizador leva a resposta combinada do sistema, dada pela convolução da resposta impulsiva do canal com a do equalizador, a ser puramente um atraso τ com IIS nula, como mostrado abaixo:

$$\sum_i w_i h_{l-i} = \delta_l \quad (2.16)$$

onde δ_l é o chamado *delta de Kronecker*: $\delta_l = \begin{cases} 1, & l = \tau \\ 0, & l \neq \tau \end{cases}$. Chamando o vetor resposta combinada do sistema de \mathbf{s} , (2.16) equivale a termos $\mathbf{s} = [0 \dots 0 \ 1 \ 0 \dots 0]$. É fácil mostrar que, neste caso, a sequência transmitida $a(n)$ será recuperada perfeitamente. Podemos escrever:

$$\sum_i w_i u(n-i) = \sum_i \sum_k w_i h_k a(n-i-k) \quad (2.17)$$

Fazendo então $k = l-i$ e alterando a ordem dos somatórios temos:

$$\sum_i w_i u(n-i) = \sum_l a(n-l) \sum_i w_i h_{l-i} \quad (2.18)$$

Usando então a equação (2.16) em (2.18) e aplicando a definição do delta de Kronecker, temos:

$$\sum_i w_i u(n-i) = \sum_l \delta_l a(n-l) = a(n) \quad (2.19)$$

que é o resultado desejado.

Entretanto, como h é desconhecido, não é possível utilizar a expressão (2.15) ou (2.16) para determinar o filtro inverso. É importante frisar que, na equação (2.19) acima, o índice i pertence ao intervalo $(-\infty, \infty)$, ou equivalentemente, o equalizador dado por (2.15) possui resposta ao impulso infinita (IIR). Isto implica em um equalizador não causal e com um número infinito de coeficientes, podendo ser chamado de equalizador duplamente infinito. Como o canal em questão é FIR, esta é a única forma de se conseguir invertê-lo perfeitamente. Entretanto, na prática, o número de coeficientes do equalizador deve ser finito e o filtro, logicamente, deve ser causal. Assim, aproximações à resposta IIR são necessárias. Estas últimas são feitas baseadas em otimizações de funções custo. Em seguida veremos detalhadamente vários critérios possíveis, gerando equalizadores implementáveis.

Vale ressaltar também que equalizadores ZF tendem a ter um desempenho ruim quando o ruído é significativo e quando a resposta impulsiva do canal $\mathbf{h}(z)$ possui zeros próximos do círculo unitário no plano z . Isto ocorre devido ao fenômeno de amplificação do ruído (*noise enhancement*) causado pelo equalizador ZF, cujo princípio é inverter o canal sem levar em conta o ruído.

2.4 Algoritmos de Bussgang

Os algoritmos em equalização autodata são normalmente desenvolvidos de forma a minimizar funções custo diferentes da EQM, utilizada na dedução do LMS. Estas funções não envolvem diretamente os símbolos $a(n)$ mas refletem a quantidade de IIS existente na saída do equalizador em um instante n . A média destas funções devem ser especificadas de forma que seu ponto de mínimo corresponda a vetores \mathbf{w} que levem à mínima IIS ou mínimo EQM. Em seguida veremos várias possibilidades de funções custo, que resultam nos chamados algoritmos de Bussgang.

2.4.1 Definição do Problema

Baseado no sistema em banda base mostrado na figura 2.3, consideramos um canal com resposta ao impulso finita (FIR), ou seja, linear, estável, causal e, além disso, invariante no tempo e sem nullos espectrais, ou seja, sem zeros sobre o círculo unitário, podendo ser de fase não mínima. Nesta primeira análise, o efeito do ruído será ignorado, supondo-se que a relação sinal-ruído é bastante elevada. Assume-se também que existe um controle automático de ganho (CAG) perfeito de forma a manter a variância da saída do canal $u(n)$ constante e igual à variância do sinal da entrada $a(n)$. Supomos, por enquanto, que os sinais sejam reais.

O problema consiste em recuperar $a(n)$, dada a sequência $u(n)$ e algumas características estatísticas de $a(n)$. É necessário portanto definir o modelo probabilístico deste sinal $a(n)$: a sequência de símbolos é independente e identicamente distribuída (i.i.d.), com média nula e variância unitária, ou seja:

$$E[a(n)] = 0 \quad e \quad E[a(n)a(k)] = \begin{cases} 1, & k = n \\ 0, & k \neq n \end{cases}$$

Além disso, a sequência é considerada estacionária no sentido estrito e a função densidade de probabilidade do símbolo $a(n)$ é simétrica e uniforme. Devido a esta simetria, o equalizador autodata não consegue distinguir entre $\pm a(n - \tau)$. Assim, a função custo deve ser par. As duas soluções $y(n) = a(n - \tau)$ e $y(n) = -a(n - \tau)$

representam mínimos globais aceitáveis.

Considerando o equalizador como um filtro linear transversal, já mostrado na figura 2.2, FIR, vimos anteriormente que é impossível conseguir equalizar perfeitamente um canal com resposta impulsiva também finita. Procura-se portanto um processo iterativo que forneça uma aproximação para o filtro inverso do canal procurado. Depois de n iterações, teremos uma sequência aproximadamente desconvoluída:

$$y(n) = \sum_{i=0}^L \hat{w}_i(n)u(n-i) = \mathbf{u}^T(n)\hat{\mathbf{w}}(n) \quad (2.20)$$

sendo $(L-1)$ a ordem do equalizador, $\mathbf{u}(n) = [u(n) \ u(n-1) \dots u(n-L+1)]^T$ e $\hat{\mathbf{w}}(n) = [\hat{w}_0(n) \ \hat{w}_1(n) \dots \hat{w}_{L-1}(n)]^T$, o que representa uma aproximação para os valores dos coeficientes do filtro inverso. Podemos escrever, de outra forma, que

$$y(n) = a(n) + \nu(n) \quad (2.21)$$

onde $\nu(n)$ é o chamado ruído convolucional, representando a interferência intersimbólica residual que resulta do uso de um filtro inverso aproximado.

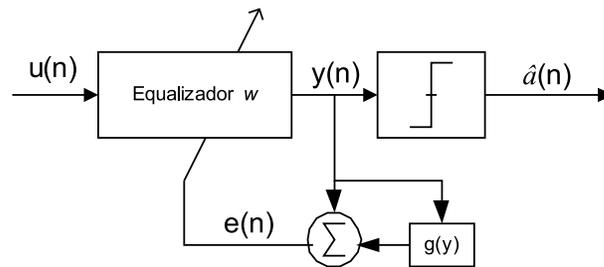


Figura 2.4: Sistema em estudo, caso autodidata

Finalmente, para produzir uma estimativa $\hat{a}(n)$ do símbolo transmitido, aplica-se um estimador não linear de memória nula à saída $y(n)$ do equalizador. Este dispositivo pode ser visto na Figura 2.4.

$$\hat{a}(n) = g(y(n)) \quad (2.22)$$

Utiliza-se então um sistema adaptativo de forma que, a partir da estimativa no instante n , $\hat{a}(n)$, consiga-se uma melhor no instante $n+1$. Isto é feito buscando-se minimizar o

erro quadrático médio dado por:

$$E [|e(n)|^2] = E [|\hat{a}(n) - y(n)|^2] \quad (2.23)$$

Note que, no caso do LMS, a função custo dada por (2.2), é uma função quadrática com relação aos coeficientes do equalizador, sendo portanto convexa e possuindo somente um ponto de mínimo bem definido. Neste caso entretanto, a função custo $J = E [|e(n)|^2]$ dada por (2.23), não será convexa. Isto significa que a superfície de erro resultante do processo de iteração poderá possuir mínimos locais além do mínimo global. Além disso, mais de um mínimo global pode existir, correspondendo a seqüências $y(n)$ que são equivalentes do ponto de vista do critério autodidata escolhido.

2.4.2 Estimador Não-Linear de Memória Nula

Os algoritmos de Bussgang são derivados utilizando-se o critério de maximização a posteriori para estimar $a(n - \tau)$ a partir de $y(n)$. Para estimar os símbolos transmitidos em meio a IIS residual na saída do equalizador, primeiramente é preciso determinar a distribuição de probabilidade desta IIS residual. Assim, em seguida, trataremos das aproximações necessárias.

Partindo da equação (2.21) temos que:

$$y(n) = \sum_i w_i u(n - i) + \sum_i [\hat{w}_i(n) - w_i(n)] u(n - i) \quad (2.24)$$

Portanto

$$\nu(n) = \sum_i [\hat{w}_i(n) - w_i(n)] u(n - i) = \sum_i \sum_k h_k [\hat{w}_i(n) - w_i(n)] a(n - i - k) \quad (2.25)$$

Fazendo $n - i - k = l$,

$$\nu(n) = \sum_l a(l) \nabla(n - l), \quad \text{onde } \nabla(n) = \sum_k h_k [\hat{w}_{n-k}(n) - w_{n-k}(n)] \quad (2.26)$$

A seqüência $\nabla(n)$ representa a resposta impulsiva residual do sistema canal-equalizador

devido à equalização imperfeita. Podemos então calcular as características estatísticas do ruído convolucional $\nu(n)$ o que resulta em [Haykin96]:

$$\begin{aligned} E[\nu(n)] &= 0 \\ E[\nu(n)\nu(n-k)] &= \begin{cases} \sum_l \nabla^2(n-l), & k=0 \\ 0, & k \neq 0 \end{cases} \\ E[a(n)\nu(n-k)] &= \nabla(-k) \end{aligned} \quad (2.27)$$

Das equações em (2.27) acima pode-se concluir que $\nu(n)$ é um processo branco e de média nula. De acordo com a equação (2.26), o ruído convolucional também pode ser visto como a soma de variáveis i.i.d. resultantes da transmissão de símbolos diferentes. Assim, se a resposta impulsiva residual $\nabla(n)$ for suficientemente longa, o teorema central do limite faz com que seja plausível considerar $\nu(n)$ gaussiano.

Como $a(n)$ é branco por hipótese e o ruído convolucional pode ser considerado aproximadamente branco por dedução, e como estas duas sequências podem ser consideradas descorrelacionadas já que a variância de $\nu(n)$ é grande se comparada à magnitude da correlação cruzada $E[a(n)\nu(n-k)]$, temos que a sua soma, $y(n)$, resulta em um processo também aproximadamente branco. Isto sugere que $a(n)$ e $\nu(n)$ podem ser considerados independentes. Portanto, $\nu(n)$ pode ser modelado como um processo aditivo, branco, gaussiano e de média nula, estatisticamente independente de $a(n)$.

Estas aproximações para o ruído convolucional e seu uso em um processo iterativo, levam a um estimador sub-ótimo para a sequência de dados. Considerando que o processo adaptativo converge, pode-se supor que, ao final do processo, a IIS é suficientemente pequena para tornar o modelo aplicável. Entretanto, no início do processo, a IIS ainda é bastante grande e as sequências $a(n)$ e $\nu(n)$ são correlacionadas, fazendo com que o ruído convolucional seja mais uniforme do que gaussiano [Godfrey e Rocca 81].

A fim de simplificar a notação, daqui em diante, os coeficientes do equalizador serão denotados sempre por w_i em vez de \hat{w}_i , pois consideraremos o equalizador sempre com um número finito de coeficientes a não ser que explicitamente se diga o contrário.

Visto isso, podemos então partir para a obtenção do estimador não linear. Simplicidade, robustez e a discussão acima sugerem o uso de um estimador de memória nula. Procura-se então um estimador de *Bayes* que é ótimo no sentido do erro quadrático médio mas, devido às aproximações realizadas acima, será sub-ótimo. Além disso, a estimação é naturalmente condicional pois depende do critério de otimização. Um critério particularmente interessante é o de minimizar o valor médio quadrático do erro entre o símbolo transmitido $a(n)$ e o símbolo estimado $\hat{a}(n)$. A escolha deste critério de otimização leva ao estimador de média condicional que consiste basicamente em, dado $y(n)$, a estimação $\hat{a}(n)$ da variável aleatória $a(n)$ é escrita como $E[a|y]$ [Haykin96].

Este estimador depende basicamente da função densidade de probabilidade da sequência transmitida originalmente e do nível do ruído convolucional no instante n . Note que, no início do processo, a sequência estimada $\hat{a}(n)$ não é confiável. Mesmo assim, como não existe uma sugestão melhor, podemos utilizá-la para melhorar a aproximação do filtro inverso. Isto corresponde a solucionar o seguinte problema:

$$\min E[(y(n) - \hat{a}(n))^2] = \min E[(\mathbf{u}^T(n)\mathbf{w}(n) - g(y(n)))^2] \quad (2.28)$$

onde a minimização é feita com relação a $\mathbf{w}(n)$.

2.4.3 Técnicas de Bussgang para Equalização Autodidata

Resolver a equação (2.28) de forma iterativa através de um algoritmo do tipo gradiente estocástico (seção 2.2), possibilita a obtenção de uma estrutura transversal adaptativa para o equalizador. Como no caso do algoritmo LMS, mas tendo $\hat{a}(n) = g(y(n))$ como resposta desejada, teremos a seguinte equação de atualização de coeficientes ($i = 0, \dots, L - 1$):

$$w_i(n+1) = w_i(n) - \mu(y(n) - g(y(n)))u(n-i) \quad (2.29)$$

Note que este algoritmo difere do LMS (eq. 2.11) somente pela função não linear $g(y(n))$. De fato, os equalizadores supervisionados utilizam a sequência de treinamento no lugar do estimador de média condicional $g(y(n))$.

Tanto o estimador como o passo de adaptação μ deveriam mudar suavemente, conforme o nível do ruído convolucional. Entretanto, estimar este nível não é uma tarefa simples. Portanto normalmente se usa somente um estimador ajustado para o maior nível de ruído, que por sua vez acontece no início do processo. Neste caso, deve ser observado que os estimadores de média condicionais possuem ganhos menores do que um para compensar o ruído alto. Assim uma pequena correção de ganho deve ser feita para se atingir a equalização completa com um estimador não-linear fixo.

De fato espera-se, no ponto de equilíbrio do algoritmo, que $y(n) = a(n - \tau)e^{j\theta}$, onde τ é um atraso constante e θ é um deslocamento de fase também constante. Observe entretanto que, como $a(n)$ é estacionário e o canal é invariante no tempo, a saída do canal $u(n)$ também é estacionária e portanto sua função densidade de probabilidade (f.d.p.) é invariante sobre um atraso τ . Além disso, um deslocamento de fase constante também não possui nenhuma consequência imediata quando a f.d.p. da sequência de dados de entrada do sistema se mantém simétrica sob rotação, que é o caso de $a(n)$ considerada aqui. Assim, a fim de realizar as contas a seguir, fez-se a seguinte simplificação: no ponto de equilíbrio do algoritmo, $y(n) = a(n)$ [Haykin96]. Temos então, de (2.29), que:

$$E[u(n - i)(a(n) - g(a(n)))] = \sum_k h_k E[a(n - i - k)(a(n) - g(a(n)))] = h_i(1 - E[a(n)g(a(n))]) = 0, \forall i \quad (2.30)$$

E portanto,

$$E[a(n)g(a(n))] = 1 \quad (2.31)$$

É interessante notar que, quando o ruído convolucional é pequeno, o algoritmo adaptativo autodidata se aproxima do resultante da minimização do erro quadrático médio (LMS). Neste caso, o estimador não-linear se aproxima de um dispositivo de decisão como o usado no algoritmo de decisão direta, que será discutido mais detalhadamente na seção 2.4.6.

2.4.4 Considerações sobre Convergência

Para que o algoritmo iterativo dado em (2.29) convirja, espera-se que o valor de $w(n)$ se aproxime de algum valor constante conforme n se torne grande. Assim, a condição de convergência pode ser escrita como:

$$E[u(n-i)y(n)] = E[u(n-i)g(y(n))] \quad (2.32)$$

Multiplicando por w_{i-k} , somando sobre i e considerando que a ordem do equalizador $(L-1)$ seja grande o suficiente, aproximadamente temos que (2.32) se resume a:

$$E[y(n)y(n-k)] = E[y(n-k)g(y(n))] \quad (2.33)$$

Processos estocásticos que satisfazem a equação (2.33) acima são chamados *processos de Bussgang*. Portanto a sequência equalizada $y(n)$ é *Bussgang*, mesmo em uma situação mais realista onde L é finito. Quanto maior L , melhor a aproximação. Note que um processo de Bussgang continua satisfazendo a equação (2.33) se for multiplicado por uma constante. No caso mostrado aqui, esta constante possui valor unitário devido a (2.31).

Isto explica porque o algoritmo autodidata descrito aqui é chamado de *algoritmo de Bussgang*. Não existe garantia quanto à sua convergência, no caso em que o filtro é causal e L é finito. Isto porque a função custo (2.28) neste caso não é convexa e portanto podem existir mínimos locais que terminam não reduzindo a IIS como desejado.

Entretanto, para o caso do equalizador duplamente infinito, foi provado, por Benveniste et al. [Benveniste et al. 80] que o algoritmo converge. Considere primeiramente uma família de distribuições sub e super-gaussianas, definidas da seguinte forma:

$$f_x(x) = \exp(-\omega(x))$$

onde $\omega(x)$ é uma função par e estritamente crescente. A função densidade de probabilidade $f_x(x)$ é sub-gaussiana (super-gaussiana) se $\omega'(x)/x$ for estritamente crescente (decrecente) para $x \geq 0$. A distribuição uniforme é incluída, por definição, na família

de distribuições sub-gaussiana. Como exemplo temos a seguinte f.d.p.: $f_x(x) = Ke^{|x/\beta|^v}$ que é sub-gaussiana quando $v > 2$. Para $v = \infty$, temos uma distribuição uniforme.

Esta definição também pode ser escrita da seguinte forma: o kurtosis associado a uma variável aleatória x é dado por

$$K(x) = E[|x|^4] - 2E^2[|x|^2] - |E[x^2]|^2 \quad (2.34)$$

É fácil verificar então que se x for gaussiana, $K(x) = 0$. Tem-se então que distribuições sub-gaussianas possuem $K(x) < 0$ e distribuições super-gaussianas possuem $K(x) > 0$.

Visto isto, foi provado o seguinte teorema, chamado teorema de *Benveniste-Goursat-Rouget* (a função $\psi(y)$ definida a seguir é o negativo da definida em [Benveniste et al. 80]):

Teorema *Tendo $\psi(y) = g(y) - y$, a convergência do algoritmo de Bussgang é garantida, a menos de um atraso constante τ e de uma possível inversão de sinal, se a distribuição da sequência de dados $a(n)$ for sub-gaussiana e a segunda derivada de $\psi(y)$ for negativa no intervalo $[0, \infty)$.*

As condições citadas no teorema são satisfeitas pela descrição do algoritmo realizada até aqui: $a(n)$ possui distribuição sub-gaussiana (uniforme) e $\psi(y)$ possui derivada segunda negativa.

Entretanto, é importante notar que o teorema só vale para equalizadores duplamente infinitos. Até hoje, ainda não encontrou-se uma função não linear de memória nula $g(\cdot)$ para a qual a convergência para o mínimo global fosse garantida. Godfrey e Rocca [Godfrey e Rocca 81] demonstraram a convergência dos equalizadores de Bussgang para mínimos locais, baseado no fato de $y(n)$ ser *Bussgang*. Note, no entanto, que nem esta última consideração ainda não foi mostrada ser válida para o caso do equalizador finito. Sabe-se que algoritmos de Bussgang produzem saídas também Bussgang somente se o equalizador for duplamente infinito.

Benveniste et al. concluem ainda que, se for possível ter uma idéia da posição do

coeficiente de maior valor do equalizador finito, então, na prática, inicializar o procedimento com 1 nesta posição e todos os outros coeficientes nulos resolve o problema, diminuindo a ocorrência de convergência para mínimos locais. Entretanto isto nem sempre é possível pois normalmente o canal é completamente desconhecido. Este procedimento é citado também por Godard [Godard80], Foschini [Foschini85] e Li e Ding [Li e Ding 95]. O primeiro utiliza uma inicialização bastante parecida na análise de propriedades de convergência do seu algoritmo (seção 2.4.8), concluindo que esta é uma condição suficiente mas não necessária para a convergência. O segundo começa realizando uma análise do algoritmo de Godard para o caso do equalizador duplamente infinito e, ao realizar uma extensão de suas conclusões para o caso do equalizador finito, sugere que se busque periodicamente deixar o coeficiente de maior valor no centro de massa do vetor de coeficientes, pois isto poderia aproximar uma situação de comprimento infinito do equalizador. O terceiro mostra que esta estratégia geralmente deve funcionar, a não ser para casos bastante específicos, analisando sempre o caso do equalizador finito. Esta forma de inicialização pode também resolver o problema da ambiguidade de fase que surge devido à simetria da densidade de probabilidade de $a(n)$. Basta então que o coeficiente não nulo do equalizador também possua o sinal desejado. Este método de inicialização é conhecido como *center-spike*.

Resumidamente, ainda baseando-se no trabalho de Benveniste et al. [Benveniste et al. 80], é interessante citar os seguintes fatos:

-Estatísticas de segunda ordem de $u(n)$ só fornecem informações sobre magnitude da resposta impulsiva do canal, não sendo suficiente para a equalização autodidata de canais de fase não-mínima.

-Um canal de fase não-mínima não pode ser identificado a partir de suas saídas se o sinal de entrada for i.i.d. e gaussiano, pois neste caso a distribuição de probabilidade é inteiramente definida pelas estatísticas de segunda ordem.

-Equalizadores ZF não podem ser implementados para canais cuja resposta impulsiva possua zeros no círculo unitário no plano z

2.4.5 Extensão para Canais em Banda-Base Complexos

Neste caso, a sequência de dados transmitida $a(n)$, a resposta impulsiva do canal h e o sinal recebido $u(n)$ possuem valores complexos. Assim os algoritmos de Bussgang serão dados da seguinte forma:

$$\begin{aligned}
 y(n) &= y_I(n) + jy_Q(n) = \sum_{i=0}^L w_i u(n-i) \\
 \hat{a}(n) &= \hat{a}_I(n) + j\hat{a}_Q(n) = g(y_I(n)) + jg(y_Q(n)) \\
 e(n) &= y(n) - \hat{a}(n) \\
 w_i(n+1) &= w_i(n) - \mu u^*(n-i)e(n), \quad i = 0, \dots, L
 \end{aligned} \tag{2.35}$$

onde o super-escrito $*$ indica conjugação e os sub-escritos I e Q indicam a componente em fase e em quadratura, respectivamente.

2.4.6 Algoritmo de Decisão Direta

O algoritmo de decisão direta foi um dos primeiros algoritmos autodidatas, proposto primeiramente por Lucky [Lucky68], sendo um dos algoritmos mais utilizados, seja junto com a equalização supervisionada, seja junto com outros algoritmos autodidatas mais robustos. Ele consiste de uma estrutura bastante simples, utilizando somente um dispositivo de decisão como estimador não-linear de memória nula, como já mostrado na Figura 2.3. Este, no entanto, pode ser considerado como uma aproximação grosseira do estimador de média condicional visto na seção 2.4.2 e portanto, o algoritmo de decisão direta pode ser considerado como um caso particular dos algoritmos de Bussgang.

O critério consiste em minimizar a seguinte função custo, caso o sinal transmitido seja BPSK:

$$J = E [|sgn(y(n)) - y(n)|^2] \tag{2.36}$$

onde $sgn(y_n) = \begin{cases} 1, & y_n > 0 \\ -1, & y_n < 0 \end{cases}$. Caso $y_n = 0$, o símbolo recuperado é escolhido aleatoriamente entre $\{+1, -1\}$.

Calculando o gradiente, chega-se a um algoritmo semelhante ao LMS, mas ao invés do sinal desejado no cálculo do erro, utiliza-se a saída do dispositivo de decisão, dado pela função sinal. Por representar uma aproximação grosseira do estimador de média condicional, o algoritmo possui um desempenho bastante limitado em termos de robustez, possuindo muitos mínimos locais indesejáveis.

No caso da equalização supervisionada onde existe uma sequência de treinamento, o chaveamento para o modo de decisão direta normalmente leva os coeficientes do equalizador a continuarem convergindo até o mínimo global. Por que não existe divergência? Odile Macchi [Macchi e Eweda84] prova que o algoritmo de decisão direta converge para a solução desejada, ou seja, mínimo global, quando o diagrama de olho na saída do equalizador estiver aberto, após uma fase inicial de treinamento. Como a função da sequência de treinamento é justamente ajustar os coeficientes do equalizador de forma a diminuir a IIS, abrindo o olho, então no momento em que existe o chaveamento para o modo de decisão direta, o algoritmo continua convergindo.

O mesmo procedimento pode ser utilizado no caso dos algoritmos autodidatas. Pode-se inicializar o processo com algum algoritmo mais robusto (alguns dos quais serão vistos adiante) e, após a redução da IIS na saída do equalizador, chaveia-se para o modo de decisão direta a fim de se concluir a convergência ou então para se rastrear pequenas variações do canal.

Mazo [Mazo80] estuda o caso do equalizador operando sempre no modo de decisão direta, ou seja, o caso do equalizador autodidata. Ele mostra, para um canal bastante simples que poderia ser facilmente equalizado (solução ZF), como o algoritmo pode ficar preso em mínimos locais indesejáveis, situação na qual a IIS não é eliminada, dependendo da inicialização dos coeficientes do equalizador. Mesmo se este fosse duplamente infinito, continuaria havendo a possibilidade de convergência para mínimos locais. A imperfeição está na função custo utilizada. Seguindo a classificação feita em [Li et al.96], o algoritmo de decisão direta possui mínimos locais dependentes da função custo e não simplesmente provenientes do fato do equalizador ser implementado como um filtro com um número finito de coeficientes.

2.4.7 Algoritmo de Sato

O algoritmo proposto por Sato foi um dos primeiros sistemas recursivos de identificação autodidata a ser amplamente utilizado e estudado, tendo sido o primeiro capaz de operar em condições de olho fechado. Apesar de não ter sido justificado teoricamente em [Sato75], resultados convincentes mostravam a convergência do equalizador proposto.

O algoritmo de Sato, proposto para ser utilizado em sistemas M-PAM (*Pulse Amplitude Modulation* M-ário), ou seja, sistemas cuja transmissão se dava por somente uma portadora (reais) e com o objetivo de ser mais robusto do que o algoritmo de decisão direta, consiste em minimizar a seguinte função custo não-convexa:

$$J = E [(\hat{a}(n) - y(n))^2] \quad (2.37)$$

onde $\hat{a}(n)$ é estimado através de uma não linearidade de memória nula dada por:

$$\hat{a}(n) = \gamma \operatorname{sgn}[y(n)] \quad (2.38)$$

onde $\operatorname{sgn}[\cdot]$ é a função sinal definida como $\operatorname{sgn}[x] = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}$ e γ é escolhida de acordo com (2.31) e pode ser definida como:

$$\gamma = \frac{E[a^2(n)]}{E[|a(n)|]} \quad (2.39)$$

O algoritmo de Sato pode ser considerado um caso particular dos algoritmos de Bussgang, tendo como função não linear $g(y) = \gamma \operatorname{sgn}(y)$. A função sinal é uma aproximação grosseira do estimador de média condicional. Note também que, para o caso PAM binário, este algoritmo coincide com o algoritmo de decisão direta.

Inicialmente portanto o algoritmo trata o sinal digital como sendo ‘binário’, estimando o bit mais significativo. Os bits restantes do sinal são tratados como ruído aditivo. O algoritmo então usa o resultado obtido neste passo inicial para modificar o sinal de erro obtido de um algoritmo de decisão direta convencional.

O teorema de *Benveniste-Goursat-Rouget* vale também para o algoritmo de Sato, mesmo com a função $\psi(\cdot)$ sendo não diferenciável. Em [Benveniste et al. 80] é provado que o algoritmo converge para distribuições sub-gaussianas. Como já discutido para o caso dos algoritmos de Bussgang, esta análise de convergência só vale para o caso do equalizador duplamente infinito.

Como no caso do algoritmo de decisão direta, o algoritmo de Sato converge para mínimos locais indesejáveis mesmo no caso do equalizador duplamente infinito, quando a entrada é PAM (*Pulse Amplitude Modulation*) [Ding et al. 93]. Assim, seus mínimos locais também são dependentes da função custo.

2.4.8 Algoritmos de Godard

Godard [Godard80] foi o primeiro a propor uma família de algoritmos autodidatas de módulo constante para uso em sistemas de comunicações digitais bidimensionais (QAM - *Quadrature Amplitude Modulation* por exemplo). Ele procurou heurísticamente por funções custo independentes da fase do sinal na saída do equalizador, buscando chegar à convergência deste independentemente da aquisição da portadora.

O algoritmo de Godard procura minimizar a seguinte função custo não convexa:

$$J_G = E [(|y(n)|^p - R_p)^2] \quad (2.40)$$

onde p é um inteiro positivo e R_p é uma constante que controla a amplificação do equalizador, escolhida de forma a dar um incremento nulo aos seus coeficientes quando a equalização perfeita é alcançada. Ela é definida como:

$$R_p = \frac{E [|a(n)|^{2p}]}{E [|a(n)|^p]} \quad (2.41)$$

O algoritmo de Godard procura penalizar desvios da saída do equalizador, $y(n)$, com relação a um módulo constante e, portanto, sua função custo é muitas vezes chamada de critério do módulo constante (CM-*Constant Modulus*). A adaptação dos coeficientes

é feita de acordo com o algoritmo do gradiente estocástico mostrado abaixo:

$$\begin{aligned}\mathbf{w}(n+1) &= \mathbf{w}(n) - \mu \mathbf{u}^*(n)e(n) \\ e(n) &= y(n)|y(n)|^{p-2} (|y(n)|^p - R_p)\end{aligned}\tag{2.42}$$

onde μ é o passo de adaptação do algoritmo.

O algoritmo possui como característica a capacidade de desacoplar os problemas de IIS e de recuperação da fase da portadora. Entretanto, justamente por não requerer a recuperação da fase da portadora, ele tende a convergir lentamente.

Godard analisou principalmente os casos em que $p = 1$ e $p = 2$. No caso em que $p = 1$ temos $R_1 = E[|a(n)|^2] / E[|a(n)|]$ e o algoritmo resultante pode ser visto como uma modificação do algoritmo de Sato para a utilização com sinais complexos. Para $p = 2$, que é mais fácil de implementar, o algoritmo resultante é também chamado de algoritmo do módulo constante (CMA- *Constant Modulus Algorithm*) e tem sido amplamente utilizado e estudado. A propriedade de recuperação do módulo, existente neste caso, foi observada por Godard mas o nome CMA foi primeiramente empregado por Treichler e Agee em [Treichler e Agee 83], que propuseram o algoritmo de forma independente, em um contexto de sinais FM (modulação em frequência) e PAM. O CMA foi um algoritmo que obteve rapidamente uma grande aceitação por sua simplicidade, eficácia e robustez quanto a erros de fase da portadora, permitindo aos equalizadores adaptarem separadamente do mecanismo de recuperação da portadora.

Para este caso ($p = 2$), Godard mostrou que se $E[|a|^4] < 2E^2[|a|^2]$, o mínimo global da função custo (2.40) corresponde à zerar a IIS. Isto não significa, entretanto, que não existam outros mínimos locais estáveis. Resultados de simulações mostraram que a inicialização do equalizador com somente um coeficiente diferente de zero (*center spike*) é uma condição suficiente para a convergência do algoritmo para a solução desejada.

Uma análise mais detalhada para o caso do equalizador duplamente infinito foi realizada por Foschini [Foschini85], concluindo que o algoritmo de Godard não possui mínimos locais indesejáveis. Mais tarde, Ding et al. [Ding et al. 91, Ding e Kennedy 92, Ding et al. 92, Ding e Johnson 93] mostraram que, para o caso do equalizador finito, existem sim mínimos locais e o desempenho do algoritmo passa a depender da inicia-

lização. Esta convergência para soluções indesejadas pode acontecer mesmo em casos, sem ruído, onde a IIS introduzida pelo canal poderia ser completamente eliminada por um equalizador FIR, como concluído por Ding, em [Ding et al. 91]. Em [Li e Ding 95], é mostrado que a inicialização por *center-spike* reduz a probabilidade de convergência para mínimos locais, como o próprio Godard, de certa forma, já havia observado. Como não existem mínimos locais quando o equalizador é duplamente infinito, temos que os mínimos locais existentes são dependentes do comprimento do equalizador, ou seja, aparecem devido ao número de coeficientes deste ser finito [Li et al.96].

É interessante notar que, se as componentes em fase e em quadratura forem independentes entre si, a condição $E[|a|^4] < 2E^2[|a|^2]$ colocada por Godard pode ser escrita como:

$$E[(a_1^2 + a_2^2)^2] < 2(E[a_1^2 + a_2^2])^2 \Rightarrow 2E[a_1^4] + 2E^2[a_1^2] < 8E^2[a_1^2] \Rightarrow E[a_1^4] < 3E^2[a_1^2] \quad (2.43)$$

Isto significa que o kurtosis, como já definido em (2.34), será negativo. Assim sendo, o algoritmo de Godard só funciona para o caso em que a f.d.p. de $a(n)$ for sub-Gaussiana.

Através de simulações, Godard também chegou à conclusão que o valor $p = 2$ resultava em uma convergência mais rápida do que $p = 1$, mas o desempenho do algoritmo para $p = 3$ foi decepcionante.

Além disso, o algoritmo de Godard é o mais robusto dentre os algoritmos de Bussgang com respeito ao *offset* de fase da portadora. Isto acontece devido à função custo e à derivação do algoritmo serem baseadas somente na amplitude do sinal recebido. Após a convergência, em condições de regime, o algoritmo de Godard também é o que atinge o menor valor de erro quadrático médio dentre os algoritmos de Bussgang.

2.5 Algoritmos de Shalvi-Weinstein

Shalvi e Weinstein, em [Shalvi e Weinstein 90], propõem uma nova classe de algoritmos, na qual não existe restrição com relação à distribuição de probabilidade dos dados de entrada do sistema $a(n)$, diferentemente do CMA por exemplo, contanto que esta não seja gaussiana.

Os critérios propostos em [Shalvi e Weinstein 90,94] podem ser divididos, basicamente, em três famílias: os critérios com restrição, os critérios normalizados e os critérios sem restrição. Os critérios com restrição garantem que o sinal na saída do equalizador terá a mesma potência do sinal inicialmente transmitido. Os critérios normalizados podem ou não conter esta mesma restrição de potência, como será visto adiante. Já os critérios sem restrição na verdade incorporam um termo de penalidade à própria função custo a ser otimizada, através da soma de um novo termo.

Como estes critérios utilizam estatísticas de ordem superior dos sinais envolvidos, iniciaremos com uma breve revisão destes conceitos.

2.5.1 Fundamentos Matemáticos

Seja uma sequência de variáveis aleatórias x_1, x_2, \dots, x_M , com a seguinte função característica conjunta,

$$\phi(\omega) = E \left[e^{j\omega^T \mathbf{x}} \right] = E \left[e^{j \sum_{i=1}^M \omega_i x_i} \right], \quad (2.44)$$

o cumulante conjunto de ordem $p = k_1 + \dots + k_M$ é definido em [Nikias93] como:

$$cum(x_1^{k_1}, x_2^{k_2}, \dots, x_M^{k_M}) = (-j)^p \frac{\partial^p \ln \phi(\omega)}{\partial \omega_1^{k_1} \partial \omega_2^{k_2} \dots \partial \omega_M^{k_M}} \Bigg|_{\omega_1 = \omega_2 = \dots = \omega_M = 0} \quad (2.45)$$

onde \ln é o logaritmo natural.

Portanto, se as variáveis aleatórias possuírem média nula, temos:

$$cum(x_1, x_2) = E[x_1 x_2]$$

$$\begin{aligned} cum(x_1, x_2, x_3) &= E[x_1 x_2 x_3] \\ cum(x_1, x_2, x_3, x_4) &= E[x_1 x_2 x_3 x_4] - E[x_1 x_2] E[x_3 x_4] - E[x_1 x_3] E[x_2 x_4] - E[x_1 x_4] E[x_2 x_3] \end{aligned} \quad (2.46)$$

Note que, até ordem 3, os cumulantes são iguais aos momentos. Daí por diante os dois se tornam bastante diferentes. As seguintes propriedades podem ser verificadas [Nikias e Petropulu 93]:

(p.1) Linearidade: $cum(\sum_i b_i x_i; \dots) = \sum_i b_i cum(x_i; \dots)$.

(p.2) Se x_1, x_2, \dots, x_M puderem ser divididos em dois ou mais grupos estatisticamente independentes, então seu cumulante conjunto é nulo.

(p.3) Se x_1, x_2, \dots, x_M são conjuntamente gaussianas então seu cumulante conjunto é nulo sempre que $M > 2$.

A fim de facilitar a notação, denotaremos

$$cum(\underbrace{x, x, \dots, x}_p, \dots) = cum(x : p, \dots) .$$

O cumulante de ordem p de x será denotado por $C_p^x = cum(x : p)$ e o cumulante de ordem (p, q) de x será denotado por $C_{p,q}^x = cum(x : p; x^* : q)$. Assim, se x for real, $C_{p,q}^x = C_{p+q}^x$.

2.5.2 Definição do Problema

As hipóteses sobre o sistema em estudo continuam sendo as mesmas já definidas para os algoritmos de Bussgang (seção 2.4.1), mas existem algumas alterações a serem consideradas quanto ao modelo probabilístico do sinal de entrada do sistema, $a(n)$. Este continua sendo i.i.d., mas além disso supõe-se simplesmente que seu cumulante de ordem (1,1), ou seja, a variância do sinal $a(n)$, exista e que o sinal também possua algum cumulante de ordem superior não nulo. Isto implica em $a(n)$ não poder ser gaussiano (propriedade (p.3)). O sinal pode ser uma seqüência de variáveis aleatórias real/complexa, contínua/discreta, apesar de, em sistemas de comunicações digitais,

normalmente esta ser discreta.

As hipóteses quanto ao canal e ao equalizador continuam sendo as mesmas. Continuamos também querendo recuperar o sinal de entrada $a(n)$ a menos de um atraso constante e um deslocamento de fase, conhecendo apenas algumas características estatísticas deste e o sinal na saída do canal $u(n)$. Assim, se

$$\mathbf{s}(n) = \mathbf{h}(n) \otimes \mathbf{w}(n) = \sum_i w_i h(n - i) \quad (2.47)$$

é a resposta combinada do sistema desconhecido (convolução canal-equalizador), então queremos encontrar $\mathbf{w}(n)$ de forma que

$$s(n) = e^{j\theta} \delta(n - \tau) \quad (2.48)$$

onde τ é um atraso constante e θ é um deslocamento de fase também constante. Note que o atraso constante não pode ser identificado devido à estacionariedade do processo observado $u(n)$. O deslocamento em fase também é inevitável se a distribuição de probabilidade de $a(n)$ for invariante à rotação.

2.5.3 Critérios com Restrição

A relação entrada-saída do sistema pode ser expressa da seguinte forma:

$$y(n) = \sum_i s_i a(n - i) \quad (2.49)$$

Fazendo uso da propriedade (p.1) pode-se escrever [Shalvi e Weinstein 94]:

$$\begin{aligned} C_{p,q}^y &= \text{cum}(\underbrace{y(n); \dots; y(n)}_p; \underbrace{y^*(n); \dots; y^*(n)}_q) \\ &= \text{cum}(\sum_{i_1} s_{i_1} a(n - i_1); \dots; \sum_{i_p} s_{i_p} a(n - i_p); \sum_{k_1} s_{k_1}^* a^*(n - k_1); \dots; \sum_{k_q} s_{k_q}^* a^*(n - k_q)) \\ &= \sum_{i_1} \dots \sum_{i_p} \sum_{k_1} \dots \sum_{k_q} s_{i_1} \dots s_{i_p} s_{k_1}^* \dots s_{k_q}^* \text{cum}(a(n) : p; a^*(n) : q) \end{aligned} \quad (2.50)$$

Como $a(n)$ é uma sequência i.i.d, usando a propriedade (p.2) tem-se:

$$\text{cum}(a(n-i_1); \dots; a(n-i_p); a^*(n-k_1); \dots; a^*(n-k_q)) = \begin{cases} C_{p,q}^a, & i_1 = \dots = i_p = k_1 = \dots = k_q \\ 0, & \text{caso contrario} \end{cases} \quad (2.51)$$

Portanto, substituindo (2.51) em (2.50)

$$C_{p,q}^y = C_{p,q}^a \sum_l s_l^p (s_l^*)^q \quad (2.52)$$

A equação (2.52) constitui a base para o seguinte teorema [Shalvi e Weinstein 90 e 94]:

Teorema *Suponha que $C_{1,1}^y = E[|a|^2] > 0$ existe e que $C_{p,q}^a$ também existe e é diferente de zero para quaisquer inteiros p e q tais que $p + q > 2$. Se*

$$C_{1,1}^y = C_{1,1}^a \quad (2.53)$$

então

$$|C_{p,q}^y| \leq |C_{p,q}^a| \quad (2.54)$$

com igualdade se e somente se a equalização for perfeita (solução ZF), ou seja, se $\mathbf{s}(n)$ for como mostrado na equação (2.48).

A demonstração deste teorema é bastante simples: comparando (2.53) com (2.52) temos que $\sum_l |s_l|^2 = 1$. Assim, $|s_l| \leq 1 \forall l$ e portanto $|s_l|^{p+q} \leq |s_l|^2 \forall l$. Em (2.52) novamente tem-se:

$$|C_{p,q}^y| = |C_{p,q}^a| \left| \sum_l s_l^p (s_l^*)^q \right| \leq |C_{p,q}^a| \sum_l |s_l|^{p+q} \leq |C_{p,q}^a| \sum_l |s_l|^2 = |C_{p,q}^a| \quad (2.55)$$

onde a igualdade vale se e somente se \mathbf{s} possuir somente um elemento não nulo e com magnitude igual a 1 (de forma que $\sum_l |s_l|^2 = 1$).

O teorema descrito acima é uma simplificação do teorema de *Benveniste-Goursat-Rouget* pois este último requer a equalização da distribuição de probabilidade (todos os cumulantes) enquanto que o teorema acima afirma que a condição suficiente para equalização é que as variâncias (cumulantes de ordem (1,1)) e algum outro cumulante

de ordem superior não-nulo de y sejam iguais aos de a .

Este teorema dá origem ao seguinte critério, chamado critério de Shalvi-Weinstein (SW) com restrição [Shalvi e Weinstein 94]:

$$\max |C_{p,q}^y| \quad (2.56)$$

$$\text{Sujeito a: } C_{1,1}^y = C_{1,1}^a$$

Sabendo que o cumulante de ordem (2,2) é igual ao kurtosis, definido em (2.34), o critério definido acima é mais geral do que o definido em [Shalvi e Weinstein 90].

O critério (2.56) só depende do conhecimento da variância de $a(n)$ e da informação de que seu cumulante de ordem (p, q) é não nulo para algum p e q tais que $p + q > 2$. Assim, o critério proposto é universal no sentido de não impor nenhuma outra restrição à distribuição de probabilidade do sinal de entrada e portanto pode ser usado em uma grande quantidade de problemas de desconvolução.

Em [Shalvi e Weinstein 94] é mostrado que, para o caso do equalizador duplamente infinito, a maximização com restrição mostrada em (2.56) não apresenta máximos locais, tendo somente um máximo global e outros pontos de mínimo ou de sela. Assim, independente da inicialização, um algoritmo de busca pelo gradiente convergiria para a solução desejada. Entretanto, esta condição não pode ser satisfeita na prática. Assim, com um equalizador finito, uma quantidade finita de dados, ruído aditivo e não-linearidades que podem estar presentes no sistema desconhecido (canal), estes pontos de sela podem vir a se tornar máximos locais. No próximo capítulo discutiremos com mais detalhes esta questão.

2.5.4 Critérios Normalizados

A restrição no critério (2.56) também pode ser substituída por uma normalização. Temos então a seguinte família de critérios normalizados [Shalvi e Weinstein 94]:

$$\max \left\{ \frac{|C_{p,q}^y|}{(C_{1,1}^y)^{\frac{p+q}{2}}} \right\} \quad (2.57)$$

Supondo $q = p$, o critério acima pode ser escrito em termos dos símbolos transmitidos $a(n)$, da seguinte forma:

$$\frac{|C_{2p}^y|}{(C_2^y)^p} = \frac{|C_{2p}^a|}{(C_2^a)^p} \underbrace{\left(\frac{\sum_l |s_l|^{2p}}{(\sum_l |s_l|^2)^p} \right)}_{f_{2p}(s)} \quad (2.58)$$

É conhecido que $0 \leq f_{2p}(s) \leq 1$ para qualquer \mathbf{s} sendo que o limite superior é alcançado se e somente se \mathbf{s} possuir um único elemento não nulo (solução ZF). Além disso, vale a igualdade $f_{2p}(s) = f_{2p}(\alpha s)$ para qualquer constante α não nula.

Como as estatísticas do sinal transmitido $a(n)$ em (2.58) podem ser consideradas conhecidas, a maximização do critério normalizado é equivalente à maximização de $f_{2p}(s)$, que por sua vez é o próprio critério proposto por Donoho [Donoho81] no contexto de sinais reais, do qual um caso particular com $p = 4$ é o critério de desconvolução proposto inicialmente por Wiggins [Wiggins78].

É interessante observar que, apesar da normalização, não existe em (2.57) nenhuma condição que relacione a potência do sinal na saída do equalizador com a do sinal de entrada. É fácil observar que, apenas quando se ajusta $C_{1,1}^y$ para ser igual a $C_{1,1}^a$, o critério (2.57) coincide com (2.56) [Tugnait92]. Entretanto, se esta igualdade não for considerada e lembrando que a relação $f_{2p}(s) = f_{2p}(\alpha s)$ é válida, o critério apresentará outros pontos de máximo (pontos estacionários). Esta questão será detalhadamente discutida no próximo capítulo.

Em [Shalvi e Weinstein 94] é provado que, no caso de um equalizador duplamente infinito e sem ruído, o único ponto de máximo do critério (2.57) ocorre se $s(n) = g\delta(n-\tau)$, onde τ é um atraso constante e arbitrário e g é um ganho complexo, resultante do fato do critério ser invariante sob a multiplicação da resposta combinada por uma constante.

É interessante observar que uma ambigüidade de ganho também ocorrerá em (2.56) se $C_{1,1}^a$ não for conhecido *a priori* e se a restrição indicada for substituída por igualar $C_{1,1}^y$ a uma constante positiva arbitrária.

2.5.5 Critérios sem Restrição

Ao invés da restrição utilizada no critério (2.56), pode-se adicionar ao critério um funcional da potência do sinal de saída do equalizador $y(n)$ como termo de penalidade. Tem-se então a seguinte família de critérios sem restrição:

$$\max \{ |C_{p,q}^y| + f(C_{1,1}^y) \} \quad (2.59)$$

onde $f : [0, \infty) \rightarrow R^1$ é uma função real e contínua. Em [Shalvi e Weinstein 94] ela é escolhida de forma que a função:

$$g(x) = |C_{p,q}^a| x^{p+q} + f(C_{1,1}^a x^2) \quad (2.60)$$

não tenha máximos em $x = 0$ ou em $x = \infty$, o que garante que o máximo de (2.59) corresponde à solução desejada a menos de um ganho, ou seja, $s(n) = g\delta(n - \tau)$. Considere então a seguinte escolha para $f(\cdot)$:

$$f(C_{1,1}^y) = \sum_{i=0}^M k_i (C_{1,1}^y)^i \quad (2.61)$$

onde k_0, k_1, \dots, k_M são constantes reais. Fazendo $k_0 = r^2$, $k_1 = 2r$, $k_2 = -2$ e $k_i = 0$ para $i = 3, \dots, M$ temos um caso particular de (2.61), resultando no critério:

$$\max \{ |C_{2,2}^y| - 2(C_{1,1}^y)^2 + 2rC_{1,1}^y + r^2 \} \quad (2.62)$$

onde r também é uma constante real.

Pode ser verificado que a função $g(x)$ não possui máximo em $x = 0$ ou em $x = \infty$ sempre que $r > 0$. Se $a(n)$ for uma variável aleatória complexa e de média nula tal que

$$C_{2,0}^a = E[a^2(n)] = 0, \quad C_{2,2}^a = E[|a(n)|^4] - 2E^2[|a(n)|^2] < 0, \quad (2.63)$$

condições que são satisfeitas por muitas constelações utilizadas em comunicações digitais, então, através de manipulações diretas [Shalvi e Weinstein 90], encontra-se que (2.62) é equivalente a

$$\min E \left[(|y(n)|^2 - r)^2 \right] \quad (2.64)$$

Se escolhermos então $r = E [|a(n)|^4] / E [|a(n)|^2]$, chega-se ao critério proposto por Godard [Godard80] e Treichler [Treichler e Agee 83], mostrado na equação (2.40) para $p=2$.

Vistos os critérios propostos por Shalvi e Weinstein, partimos então para a obtenção dos algoritmos adaptativos, que buscam otimizar recursivamente as funções custos citadas.

2.5.6 O Algoritmo Super-Exponencial

O algoritmo de Shalvi-Weinstein, calculado a partir do critério com restrição, é também conhecido como algoritmo super-exponencial (ASE). Este algoritmo foi proposto primeiramente em [Shalvi e Weinstein 93], mas apesar de ter seu bom desempenho comprovado através de simulações, o artigo não é muito rigoroso em sua dedução, realizando algumas aproximações válidas somente em um contexto bastante restrito. Uma dedução mais rigorosa foi realizada por P. Regalia e M. Mboup em [Regalia e Mboup 99]. Este artigo será utilizado como base na dedução a seguir, que envolve também considerações de [Shalvi e Weinstein 93] e de [Shalvi e Weinstein 94]. Além disso, discutiremos as aproximações realizadas por [Shalvi e Weinstein 93] no momento oportuno.

Primeiramente, iremos deduzir o algoritmo no espaço da resposta combinada, canal-equalizador. Em seguida, explicaremos a origem do termo *super-exponencial*. Por último, encontraremos o algoritmo no espaço de parâmetros do equalizador, que é o que realmente nos interessa, por ser o algoritmo implementável, já que o canal não é conhecido.

Já partindo-se do critério sem restrição e lembrando que este critério pode ser reduzido ao critério de Godard, também mostraremos que, utilizando o método de apro-

ximação estocástica, é possível se chegar ao próprio CMA. Alguns conceitos e a notação a ser utilizada na dedução do algoritmo serão definidos a seguir.

A resposta combinada canal-equalizador, dada por (2.47), pode ser escrita como:

$$\mathbf{s} = \mathbf{H}\mathbf{w} \quad (2.65)$$

onde $\mathbf{s} = [s_0 \ s_1 \ \dots \ s_{M+L-1}]^T$ e \mathbf{H} é a matriz de convolução do canal, com $(M + L - 1)$ linhas, e cujos elementos são dados por: $H_{ij} = h_{i-j}$, $1 \leq j \leq L$

$$\mathbf{H}^T = \begin{bmatrix} h_0 & h_1 & \dots & h_{M-1} & 0 & \dots & 0 \\ 0 & h_0 & h_1 & \dots & h_{M-1} & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & h_0 & h_1 & \dots & h_{M-1} \end{bmatrix}$$

Além das hipóteses já comentadas na seção 2.5.2, se $a(n)$ for complexo, também iremos supor que possui simetria circular, ou seja, que $E[a^2(n)] = 0$.

O conjunto T , chamado *attainable set*, nada mais é do que o conjunto de todas as respostas combinadas \mathbf{s} que podem ser obtidas utilizando-se um equalizador finito. Matematicamente temos:

$$T \triangleq \left\{ \mathbf{s} : s_n = \sum_{i=0}^L w_i h_{n-i}, w_i \in \mathbb{R} \right\} \quad (2.66)$$

Se tivermos, por exemplo, um canal MA e um equalizador FIR, sabemos que a resposta combinada \mathbf{s} nunca corresponderá a uma solução ZF, o que representaria uma equalização perfeita do canal. Neste caso, então, esta solução não pertence a T . Portanto, as soluções \mathbf{s} possíveis estão intimamente ligadas aos valores que os coeficientes do equalizador \mathbf{w} podem assumir.

Considerando o conjunto de todos os vetores \mathbf{s} possíveis, chamamos o operador projeção deste conjunto em T de P_T . Este pode ser escrito matematicamente como: $P_T = \mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H$.

É interessante também definir neste momento o conceito de equalizador sub-parametrizado, primeiramente estudado em [Regalia e Mboup 99]. Caso $P_T = I$, o *attainable set* abrange todo o conjunto de respostas \mathbf{s} possíveis, o que inclui a solução ZF. Diz-se então que o equalizador possui ordem suficiente. Neste caso, o espaço das linhas da matriz \mathbf{H} inclui todo o espaço complexo \mathcal{C}^{M+L-1} . Entretanto, se $P_T \neq I$, o equalizador é chamado sub-parametrizado e somente um sub-espaço de \mathcal{C}^{M+L-1} pode ser atingido. Por exemplo, se temos um canal com resposta ao impulso infinita ($M \rightarrow \infty$) e o equalizador for finito, então somente um sub-espaço de dimensão finita poderá ser atingido variando-se os coeficientes do equalizador, o que não é suficiente para a equalização do canal, resultando em $P_T \neq I$. No entanto, P_T poderá ser igual à matriz identidade I se e somente se $L \rightarrow \infty$ e a resposta impulsiva do canal $\mathbf{h}(z)$ for limitada e não nula para todo $|z| \geq 1$. Dependendo do caso, $P_T = I$ pode ser atingido trabalhando-se com equalizadores fracionários (ou multisensor), tendo um canal com resposta impulsiva não nula para qualquer z e L suficiente. Caso contrário, sempre teremos $P_T \neq I$, o que inclui o equalizador *baud-rate* estudado aqui.

Prosseguiremos então com a dedução do algoritmo no espaço da resposta combinada canal-equalizador.

Algoritmo no Espaço da Resposta Combinada Canal-Equalizador

Para continuarmos esta dedução, definimos a m -ésima potência de Hadamard, denotada por $\mathbf{s}^{\odot m}$, como:

$$\mathbf{s}^{\odot m} = \underbrace{\mathbf{s} \odot \dots \odot \mathbf{s}}_m \quad (2.67)$$

cuja k -ésima componente é s_k^m , denotaremos por $\mathbf{s}^{\odot \langle 2p-1 \rangle}$ o vetor $\mathbf{s}^{\odot p} \odot (\mathbf{s}^*)^{\odot (p-1)}$ cuja componente pode ser escrita como $[\mathbf{s}^{\odot \langle 2p-1 \rangle}]_k = |s_k|^{2p-2} s_k$.

Precisamos portanto encontrar quais são os pontos de máximo de $f_{2p}(s)$, dada por (2.58), o que resolve o problema de equalização em questão. Em [Regalia e Mboup 99], os autores mostram que, $\mathbf{s} \in T$ é um ponto estacionário de $f_{2p}(s)$ se e somente se

$$P_T(\mathbf{s}^{\odot \langle 2p-1 \rangle}) = \alpha \mathbf{s} \quad (2.68)$$

para algum escalar α . Se a norma ℓ_2 de \mathbf{s} for unitária, então $\sqrt[p]{\alpha} = f_{2p}(s)$, que é o

valor obtido no ponto estacionário. A norma ℓ_2 de \mathbf{s} é definida como:

$$\|\mathbf{s}\| = \left(\sum_i |s_i|^2 \right)^{1/2}. \quad (2.69)$$

A demonstração de (2.68) e das afirmações acima pode ser encontrada, de maneira resumida, no Apêndice deste trabalho.

Considerando o caso do equalizador com ordem suficiente, ou seja, $P_T = I$, o problema já está resolvido pois (2.68) se reduz a $\mathbf{s}^{\odot(2p-1)} - \alpha\mathbf{s} = \mathbf{0}$ que, em termos de componentes pode ser escrito como:

$$s_k(|s_k|^{2p-2} - \alpha) = 0, \quad \text{para todo } k \quad (2.70)$$

Isto significa que todos os termos não nulos possuem mesma amplitude, dada por $^{2p-2}\sqrt{\alpha}$. Godard já havia chegado nesta conclusão em [Godard80], mas de uma forma que não incluía o caso do equalizador sub-parametrizado (a semelhança entre os critérios de Godard e de SW é maior do que parece à primeira vista, questão que será estudada no próximo capítulo). Se dois ou mais termos forem nulos, o ponto estacionário será um ponto de mínimo ou de sela [Regalia e Mboup 99].

Já para o caso do equalizador sub-parametrizado, (2.70) não é mais válida pois $P_T \neq I$. Assim, $\mathbf{s}^{\odot(2p-1)} - \alpha\mathbf{s} \neq \mathbf{0}$ (para maiores detalhes, consultar o Apêndice) não sendo portanto possível encontrar uma solução com forma fechada para \mathbf{s} . Uma saída então é encontrar os pontos de máximo de $f_{2p}(s)$, de maneira iterativa. O seguinte mapeamento não-linear $D_{2p}(\mathbf{s})$ possui, como pontos fixos, os mesmos pontos estacionários de $f_{2p}(s)$ [Regalia e Mboup 99]:

- 1) Considere $\mathbf{s} \in T$, com norma unitária ($\|\mathbf{s}\| = 1$). A escolha da norma é arbitrária por enquanto.
- 2) Projete sua potência de Hadamard em T :

$$v = P_T(\mathbf{s}^{\odot(2p-1)}) \quad (2.71)$$

3) Faça com que o resultado possua norma unitária: $\mathbf{q} = v/\|v\|$ (mesma norma utilizada no passo 1)

É possível observar que os pontos fixos deste mapeamento, ou seja, pontos em que $\mathbf{q} = D_{2p}(\mathbf{s}) = \mathbf{s}$, com $\mathbf{s} \in T$, são exatamente os mesmos pontos estacionários que satisfazem (2.68). Caso o ponto estacionário (ou um ponto fixo do mapeamento) ainda não tenha sido atingido, vale a desigualdade:

$$f_{2p}(\mathbf{q}) > f_{2p}(\mathbf{s}) \quad \text{para } \mathbf{s} \in T \text{ com norma unitária} \quad (2.72)$$

ou seja, a busca continuará até que um ponto de máximo seja atingido o que garante a convergência do algoritmo. Portanto, o processo iterativo

$$\mathbf{s}_{n+1} = D_{2p}(\mathbf{s}_n) \quad (2.73)$$

onde n denota o número de iterações, se aproximará do máximo de $f_{2p}(\mathbf{s})$, exceto num conjunto de inicializações que podem levar a máximos locais indesejados.

A demonstração de que (2.72) é válida pode ser encontrada em [Regalia e Mboup 99] e consiste basicamente em, assumindo que o algoritmo utilize norma ℓ_2 , mostrar que

$$\sum_k |q_k|^{2p} > \sum_k |s_k|^{2p}, \quad \text{sempre que } \mathbf{q} = D_{2p}(\mathbf{s}) \neq \mathbf{s} \quad (2.74)$$

Para o caso do equalizador com ordem suficiente, (2.73) se reduz ao algoritmo super-exponencial (ASE) proposto em [Shalvi e Weinstein 93]. O termo *super-exponencial* surgiu pois o algoritmo converge para a solução ZF $\mathbf{z}_j = (\dots, 0, 1(\text{posição } j), 0, \dots)$ a uma taxa super-exponencial. Considerando a norma ℓ_∞ nos passos 1) e 3), temos:

$$\|\mathbf{z}_j - \mathbf{s}_{n+1}\|_\infty = \|\mathbf{z}_j - \mathbf{s}_n\|_\infty^{2p-1} \quad (2.75)$$

Por indução,

$$\|\mathbf{z}_j - \mathbf{s}_n\|_\infty = (\|\mathbf{z}_j - \mathbf{s}_0\|_\infty)^{(2p-1)^n} \quad (2.76)$$

que se contrai “super-exponencialmente” pois, se a condição inicial $\mathbf{s}_0 \in T$ tem $\|\mathbf{s}_0\|_\infty = 1$, então $\|\mathbf{z}_j - \mathbf{s}_0\|_\infty < 1$.

Já para o caso do equalizador sub-parametrizado, na dedução feita em [Shalvi e Weinstein 93], Shalvi-Weinstein argumentam que se $P_T \left(\mathbf{s}_n^{\odot(2p-1)} \right) \approx \mathbf{s}_n^{\odot(2p-1)}$ para todo n , então o algoritmo continuaria convergindo para uma solução próxima de Wiener. Entretanto, não fica claro até que ponto a aproximação $P_T \left(\mathbf{s}_n^{\odot(2p-1)} \right) \approx \mathbf{s}_n^{\odot(2p-1)}$ é válida. Já o algoritmo deduzido acima, com (2.72), converge sempre, sem aproximações.

Algoritmo no Espaço de Parâmetros do Equalizador

Resta ainda encontrar o algoritmo no espaço de parâmetros do equalizador \mathbf{w} . Definindo \mathbf{w} , \mathbf{w}' e \mathbf{w}'' como o estado do equalizador antes, durante e depois de cada iteração, e tendo por base o algoritmo dado por (2.71), gostaríamos que o vetor de coeficientes do equalizador \mathbf{w}' fosse tal que $\mathbf{s} = \mathbf{H}\mathbf{w}'$ seja igual a $\mathbf{s}^{\odot(2p-1)}$. Entretanto, o número de coeficientes é finito e portanto nem todos os vetores \mathbf{s} podem ser atingidos. Assim, deseja-se somente que \mathbf{w}' minimize a distância entre $\mathbf{H}\mathbf{w}'$ e $\mathbf{s}^{\odot(2p-1)}$, o que será obtido pela projeção deste último no *attainable set* [Shalvi e Weinstein 93]:

$$\min_{\mathbf{w}'} \|\mathbf{H}\mathbf{w}' - \mathbf{s}^{\odot(2p-1)}\|^2 \Rightarrow \mathbf{w}' = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H \mathbf{s}^{\odot(2p-1)} \quad (2.77)$$

A operação de normalização equivalente ao passo 3) do algoritmo deduzido anteriormente é equivalente a:

$$\mathbf{w}'' = \frac{1}{\sqrt{\mathbf{w}'^H \mathbf{H}^H \mathbf{H} \mathbf{w}'}} \mathbf{w}' \quad (2.78)$$

Projetando o algoritmo dado por (2.77) e (2.78) de volta ao espaço da resposta combinada \mathbf{s} , obtemos exatamente (2.71) e o passo de normalização 3).

Mesmo estando no domínio dos parâmetros do equalizador, o algoritmo dado por (2.77) e (2.78) continua em função de \mathbf{H} e \mathbf{s} , que não são conhecidos. O próximo passo então é escrevê-lo em função de cumulantes dos sinais de entrada e de saída do equalizador. Utilizando as propriedades de cumulantes mostradas na seção 2.5.1, temos:

$$\begin{aligned} cum(u(n-j); u^*(n-i)) &= cum \left(\sum_{k_1} h_{k_1-j} a(n-k_1); \sum_{k_2} h_{k_2-i}^* a^*(n-k_2) \right) \\ &= \sum_{k_1} \sum_{k_2} h_{k_1-j} h_{k_2-i}^* cum(a(n-k_1); a^*(n-k_2)) \end{aligned}$$

$$= C_{1,1}^a \sum_k h_{k-j} h_{k-i}^* = C_{1,1}^a (\mathbf{H}^H \mathbf{H})_{ij} \quad (2.79)$$

Temos também que:

$$\begin{aligned} cum(y(n) : p; y^*(n) : p-1; u^*(n-i)) &= \\ &= cum\left(\sum_{l_1} s_{l_1} a_{n-l_1}; \dots \sum_{l_p} s_{l_p} a_{n-l_p}; \sum_{m_1} s_{m_1}^* a_{n-m_1}^*; \dots \sum_{m_{p-1}} s_{m_{p-1}}^* a_{n-m_{p-1}}^*; \sum_k h_{k-i}^* a_{n-k}^*\right) \\ &= \sum_k h_{k-i}^* \sum_{l_1} \dots \sum_{l_p} \sum_{m_1} \dots \sum_{m_{p-1}} s_{l_1} \dots s_{l_p} s_{m_1}^* \dots s_{m_{p-1}}^* cum(a_n : p; a_n^* : p-1; a_{n-k}^*) \end{aligned} \quad (2.80)$$

Utilizando (2.51) para simplificar a expressão (2.80) acima, resulta:

$$\begin{aligned} cum(y(n) : p; y^*(n) : p-1; u^*(n-i)) &= cum(a(n) : p; a^*(n) : p) \sum_k h_{k-n}^* s_k^p (s_k^*)^{p-1} \\ &= cum(a(n) : p; a^*(n) : p) (\mathbf{H}^H \mathbf{s}^{\odot(2p-1)}) \end{aligned} \quad (2.81)$$

Finalmente, substituindo (2.79) e (2.81) em (2.77) e (2.78) obtém-se:

$$\begin{aligned} \mathbf{w}' &= \mathbf{R}^{-1} \mathbf{d} \\ \mathbf{w}'' &= \frac{1}{\sqrt{\mathbf{w}'^H \mathbf{R} \mathbf{w}'}} \mathbf{w}' \end{aligned} \quad (2.82)$$

onde \mathbf{R} é uma matriz $L \times L$ cujos elementos são dados por:

$$\mathbf{R}_{ij} = \frac{cum(u(n-j); u^*(n-i))}{cum(a(n); a^*(n))} = \frac{E[u(n-j)u^*(n-i)]}{E[|a(n)|^2]} \quad (2.83)$$

e \mathbf{d} é um vetor $L \times 1$ cujos elementos são dados por:

$$d_n = \frac{cum(y(n) : p; y^*(n) : p-1; u^*(n-i))}{C_{p,p}^a} \quad (2.84)$$

A multiplicação por \mathbf{R}^{-1} em (2.82) corresponde a uma operação de branqueamento espectral sugerida em [Benveniste et al. 80] e [Shalvi e Weinstein 90]. O termo \mathbf{w}' possui uma estrutura bastante similar à predição linear. A grande diferença está no vetor \mathbf{d} que nesta última é formado por cumulantes de segunda ordem, enquanto que,

em (2.82), é dado por cumulantes de ordem superior.

Um novo problema que surge aqui é o fato do algoritmo (2.82) estar expresso em função dos cumulantes do sinal recebido $u(n)$ e do sinal na saída do equalizador $y(n)$. Os valores exatos destes cumulantes não estão disponíveis. Uma solução seria substituí-los por estimativas (médias temporais, por exemplo) baseadas nos valores observados. Existe também a possibilidade de se obter um algoritmo recursivo, ou através da implementação sequencial do algoritmo super-exponencial visto, ou através do método de aproximação estocástica. Estas duas abordagens serão discutidas em seguida.

Algoritmo Obtido Através de Implementação Sequencial

Primeiramente, para implementar sequencialmente o algoritmo dado por (2.82), precisamos substituir os cumulantes por médias temporais. O algoritmo então pode ser escrito da seguinte forma:

$$\mathbf{w}' = \frac{1}{\alpha_p} \mathbf{Q}_n \hat{\mathbf{d}}_n \quad (2.85)$$

onde

$$\mathbf{Q}_n^{-1} = (1 - \beta_n) \mathbf{Q}_{n-1}^{-1} + \beta_n \mathbf{u}_n^* \mathbf{u}_n^T \quad (2.86)$$

sendo portanto proporcional à matriz \mathbf{R}^{-1} , com β um escalar constante. Devido à operação de normalização, vale a seguinte relação $E[y_n^2] = E[a_n^2]$ que também pode ser confirmada através da equação (2.52). Podemos então escrever α_p somente em função dos cumulantes do sinal transmitido, $a(n)$, eliminando a dependência de cumulantes de $y(n)$. Além disso, α_p depende da ordem p sendo considerada. Para $p = 2$ teremos, usando (2.34) que é equivalente ao $C_{2,2}^a$:

$$\alpha_2 = \delta = \frac{C_{2,2}^a}{C_{1,1}^a} = \frac{E[|a_n|^4] - \kappa E^2[|a_n|^2]}{E[|a_n|^2]} \quad (2.87)$$

com κ igual a 3 se o sinal for real e 2 se o sinal for complexo com $E[a^2] = 0$. Considerando cumulantes de terceira ordem, α_1 será:

$$\alpha_1 = \gamma = \frac{C_{1,2}^a}{C_{1,1}^a} = \frac{E[|a_n|^2 a_n^*]}{E[|a_n|^2]} \quad (2.88)$$

Já \mathbf{d}_n também depende da ordem dos cumulantes a ser considerada. Com $p = 2$, ou

seja, cumulantes de quarta ordem temos que

$$\begin{aligned} \mathbf{d}_n &= cum(y_n : 2; y_n^*; u_{n-i}^*) = E [|y_n|^2 y_n u_{n-i}^*] - \kappa E [|y_n|^2] E [y_n u_{n-i}^*] \\ &= E [|y_n|^2 y_n u_{n-i}^*] - \kappa E [|a_n|^2] E [y_n u_{n-i}^*] \end{aligned} \quad (2.89)$$

onde κ é igual a 3 caso os sinais sejam reais e igual a 2 caso os sinais sejam complexos com simetria circular. Portanto a estimativa $\hat{\mathbf{d}}_n$ pode ser escrita como:

$$\hat{\mathbf{d}}_n = (1 - \beta)\hat{\mathbf{d}}_{n-1} + \beta \mathbf{u}_n^* (|y_n|^2 - \kappa E [|a_n|^2]) y_n \quad (2.90)$$

Com cumulantes de terceira ordem teremos:

$$\mathbf{d}_n = cum(y_n; y_n^*; u_{n-i}^*) = E [|y_n|^2 u_{n-i}^*] \quad (2.91)$$

e portanto a estimativa será dada por

$$\hat{\mathbf{d}}_n = (1 - \beta)\hat{\mathbf{d}}_{n-1} + \beta \mathbf{u}_n^* |y_n|^2 \quad (2.92)$$

Substituindo (2.86) e (2.90) ou (2.92), dependendo da ordem dos cumulantes sendo considerada, em (2.85), chegaremos às equações mostradas abaixo para cumulantes de terceira e quarta ordem respectivamente:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \frac{\beta}{\gamma} \mathbf{Q}_n \mathbf{u}_n^* y_n \left(y_n^* - \frac{E [|a_n|^2 a_n^*]}{E [|a_n|^2]} \right) \quad (2.93)$$

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \frac{\beta}{\delta} \mathbf{Q}_n \mathbf{u}_n^* y_n \left(|y_n|^2 - \frac{E [|a_n|^4]}{E [|a_n|^2]} \right) \quad (2.94)$$

Observe que a equação para atualização da matriz \mathbf{Q} , (2.86), ainda é bastante complexa devido à inversão de matriz necessária. Esta operação pode ser evitada. Invertendo-se (2.86), tem-se:

$$\mathbf{Q}_{n+1} = \frac{1}{1 - \beta} \left(\mathbf{Q}_n - \frac{\beta \mathbf{Q}_n \mathbf{u}_n^* \mathbf{u}_n^T \mathbf{Q}_n}{1 - \beta + \beta \mathbf{u}_n^T \mathbf{Q}_n \mathbf{u}_n^*} \right) \quad (2.95)$$

Algoritmo Obtido Pelo Método de Aproximação Estocástica

Já o algoritmo obtido utilizando-se o método de aproximação estocástica, tem como ponto de partida o critério SW sem restrição. Considerando o uso de cumulantes de quarta ordem, $p = 2$ e $q = 2$, a função custo mostrada em (2.59) pode ser escrita como:

$$\Psi = C_{2,2}^y + f(E[|y_n|^2]) \quad (2.96)$$

O algoritmo de busca pelo gradiente, lembrando que estamos procurando o ponto de máximo da função custo Ψ , é dado por:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu \frac{\partial \Psi}{\partial \mathbf{w}} \quad (2.97)$$

Calculando o gradiente de (2.96) e substituindo em (2.97) teremos [Shalvi e Weinstein 94]:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu \left(\text{sgn} \left(C_{2,2}^y \right) \left(E[|y_n|^2 y_n \mathbf{u}_n^*] - \kappa E[|y_n|^2] E[y_n \mathbf{u}_n^*] \right) + \frac{1}{2} f' \left(E[y_n^2] \right) E[y_n \mathbf{u}_n^*] \right) \quad (2.98)$$

Os operadores esperança podem ser substituídos por estimativas baseadas em médias temporais, ou pelo valor instantâneo de seu argumento. Este último é o método conhecido por aproximação estocástica. Teremos então:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu \mathbf{u}_n^* y_n \left(\text{sgn} \left(C_{2,2}^a \right) \left(|y_n|^2 - \kappa \hat{E}[|y_n|^2] \right) + \frac{1}{2} f' \left(\hat{E}[|y_n|^2] \right) \right) \quad (2.99)$$

onde foi usada a igualdade $\text{sgn} \left(C_{2,2}^y \right) = \text{sgn} \left(C_{2,2}^a \right)$ (resultante de (2.52)).

Finalmente, se escolhermos $f(\cdot)$ como em (2.62) e com $r = E[|a_n|^4] / E[|a_n|^2]$, então, caso $C_{2,2}^a < 0$ (sub-gaussiano), (2.99) será

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \mathbf{u}_n^* y_n \left(|y_n|^2 - \frac{E[|a_n|^4]}{E[|a_n|^2]} \right) \quad (2.100)$$

que é o próprio algoritmo de Godard, dado por (2.42), para $p = 2$.

É possível observar que as estruturas de (2.94) e (2.100) são muito parecidas, tendo como diferença apenas a multiplicação pela matriz \mathbf{Q}_n . Como já comentado ante-

riormente, esta multiplicação corresponde a uma operação de branqueamento espectral e pode melhorar significativamente a velocidade de convergência do algoritmo. Além disso, como o algoritmo (2.94) procura realizar uma média cumulativa, é esperado que seja estatisticamente mais estável do que o algoritmo de Godard, que é, por sua vez, baseado em aproximações estocásticas de primeira ordem. Uma comparação mais detalhada entre estes algoritmos será realizada no capítulo 4.

2.5.7 Revisitando a Literatura

O algoritmo super-exponencial foi objeto de uma quantidade de trabalhos significativamente menor se comparado aos dedicados a técnicas como a do CMA. Entretanto, existem artigos muito interessantes que merecem ser discutidos. A própria equivalência entre os critérios de SW e CM, que será detalhadamente estudada no próximo capítulo, acaba trazendo muitas informações sobre questões de convergência e existência de mínimos locais, já que análises realizadas para o critério CM podem ser estendidas para o critério SW. É lógico que o contrário também é verdadeiro. Discutiremos em seguida alguns artigos que fornecem maiores detalhes sobre o algoritmo super-exponencial.

Primeiramente, Mboup e Regalia mostram, em [Mboup e Regalia 99] e mais tarde em [Mboup e Regalia 00], que o ASE é equivalente a um algoritmo de busca pelo gradiente, cuja função custo pertence à família de funções de Donoho ($f_{2p}(\mathbf{s})$). Assumindo que o canal seja real, estável, causal e que não exista adição de ruído, o algoritmo super-exponencial (2.71) é equivalente ao seguinte algoritmo de busca do gradiente:

$$\begin{aligned} v &= \mathbf{s}_n + \frac{1}{2pf_{2p}(\mathbf{s}_n)} P_T \nabla f_{2p}(\mathbf{s}_n) \\ \mathbf{s}_{n+1} &= \frac{v}{\|v\|} \end{aligned} \quad (2.101)$$

É fácil notar que a função custo associada ao algoritmo dado por (2.101) pertence à família de funções de Donoho. A análise de convergência então se reduz a uma descrição analítica da superfície de erro. Com (2.101), a convergência do algoritmo pode ser demonstrada de uma forma mais simples do que aquela baseada em (2.72).

Reescrevendo (2.101) como

$$\alpha_n \mathbf{s}_{n+1} = \lambda_n \mathbf{s}_n + \mathbf{x}_n \quad (2.102)$$

onde

$$\lambda_n \triangleq f_{2p}(\mathbf{s}_n) \quad \alpha_n \triangleq \|P_T(\mathbf{s}_n^{\odot(2p-2)})\|^2 \quad (2.103)$$

e \mathbf{x}_n é

$$\mathbf{x}_n = \frac{1}{2p} P_T \nabla f_{2p}(\mathbf{s}_n) \quad (2.104)$$

basta mostrar que as seguintes desigualdades são válidas:

$$0 < \lambda_n < \alpha_n < \lambda_{n+1} < 1 \quad (2.105)$$

e

$$\lambda_{n+1} - \lambda_n > 2p(\alpha_n - \lambda_n) \quad (2.106)$$

exceto quando \mathbf{s}_n é um ponto estacionário, caso em que valem as igualdades $\lambda_n = \alpha_n = \lambda_{n+1} = \alpha_{n+1}$. Assim a função custo no instante $n+1$ deve ser maior do que a operação de normalização no instante n , que por sua vez deve ser maior do que a função custo no instante n . Observando (2.102), pode-se concluir que, mostrando a validade da relação acima, demonstra-se que o algoritmo converge.

Além disso, os autores mostraram de maneira bastante simples, que o passo de adaptação do algoritmo é ótimo no sentido da velocidade de convergência, ou seja, é o passo que permite ao algoritmo convergir mais rapidamente. Este resultado é anunciado em forma de teorema afirmando que:

- 1) A cada iteração n , se o passo de adaptação μ_n for escolhido no intervalo

$$0 < \mu_n < \frac{2\lambda_n}{2\lambda_n^2 - \alpha_n^2} \quad (2.107)$$

então o algoritmo permanece estável.

- 2) A escolha

$$\mu = \frac{1}{\lambda_n}, \quad (2.108)$$

que dá origem ao algoritmo super-exponencial, é ótima com relação à velocidade de convergência.

A demonstração é feita considerando que não há adição de ruído e que o vetor gradiente é exato. Note que nenhuma destas hipóteses é satisfeita na prática. Um estudo do algoritmo baseado em simulações será realizado no capítulo 4.

Li e Ding, em [Li e Ding 95], realizam um estudo de convergência de equalizadores FIR partindo do critério de SW, em termos do efeito da inicialização no estado final do equalizador. Os resultados também podem ser estendidos para o critério CM, lembrando que existe uma relação muito próxima entre os dois. A análise se baseia em um novo conceito, chamado *Cones de Mínimo Global Único* S_n^\pm e que pode ser definido da seguinte forma:

Os cones de mínimo global único são definidos de forma que cada um possua somente um mínimo global, dado por uma solução ZF. A notação utilizada é tal que o cone S_n^+ possua, como mínimo global, a solução $\mathbf{s} = [\dots 0 \ 1(\text{posição } n) \ 0 \ \dots]$ enquanto que S_n^- teria como mínimo global $\mathbf{s} = [\dots 0 \ -1(\text{posição } n) \ 0 \ \dots]$. Matematicamente podem ser definidos como:

$$S_n^+ \triangleq \{\mathbf{s} \in \ell^1(\mathbb{R}) : s_n > 0 \text{ e } s_n > |s_k|, \text{ todo } k \neq n\} \quad (2.109)$$

$$S_n^- \triangleq \{\mathbf{s} \in \ell^1(\mathbb{R}) : s_n < 0 \text{ e } -s_n > |s_k|, \text{ todo } k \neq n\} \quad (2.110)$$

Outras propriedades importantes de S_n^\pm são que este é um conjunto convexo e que, se \mathbf{s} pertence a S_n^\pm , então $\alpha\mathbf{s}$ também pertence, para qualquer $\alpha > 0$

Além disso, os autores definem a chamada *kurtosis normalizada* como sendo:

$$K_{nm}(x) \triangleq \frac{K(x)}{\sigma_x^4} \quad (2.111)$$

onde $K(x)$ é a kurtosis dada por (2.34) e σ_x^2 é a variância de x .

Assim, se a saída do equalizador satisfizer a seguinte condição

$$\frac{K_{nm}(y_n)}{K_{nm}(a_n)} > 0.5 \quad (2.112)$$

no instante inicial e o vetor resposta combinada $\mathbf{s}_{inicial}$ estiver contido em um cone de mínimo global único S_n^\pm , utilizando um passo de adaptação suficientemente pequeno, é

demonstrado que \mathbf{s} irá convergir para um ponto de mínimo contido neste mesmo cone. Mais especificamente, podemos escrever que $\mathbf{s} \in T \cap S_n^\pm$, lembrando que o *attainable set* T é definido como sendo o conjunto de todos os vetores \mathbf{s} possíveis de serem obtidos.

A demonstração consiste basicamente em mostrar que, caso (2.112) seja satisfeita e $\mathbf{s}_{inicial} \in S_n^\pm$, o custo inicial é maior do que o custo em qualquer valor da fronteira do cone. Utilizando um algoritmo que busca o menor custo, \mathbf{s} não poderá nunca cruzar a fronteira, convergindo então para um ponto de mínimo que pertença a S_n^\pm .

O artigo também trata da questão de localização dos pontos de mínimo, colocando que se existe a possibilidade de inversão perfeita do canal, ou seja, se for possível atingir uma solução ZF contida em $T \in S_n^\pm$, então não existem pontos de mínimo na fronteira do cone. Entretanto, se esta configuração não for possível, o que acontece com muita frequência, e alguma solução ZF estiver próxima a, mas não contida em, $T \in S_n^\pm$, então existirá apenas um mínimo em $T \in S_n^\pm$, próximo a esta solução ZF, enquanto que todos os outros possíveis pontos de mínimo estarão próximos à fronteira do cone.

Este artigo também possui outros resultados muito interessantes, com relação à equivalência entre os critérios CM e SW, e que serão discutidos no próximo capítulo. Aqui foram tratados artigos que procuram elucidar melhor características tanto do critério SW como do algoritmo super-exponencial.

2.6 Critério do Módulo Constante Generalizado

O critério e o algoritmo mostrados a seguir foram propostos primeiramente por R. Cavalcanti em [Cavalcanti et al. 98, Cavalcanti99]. A inspiração para este novo algoritmo surgiu de uma medida de qualidade de sinal, a razão sinal-variação de envelope (SVR - *Signal to Envelope Variation Ratio*), proposta em [Brandão95], que é utilizada para a seleção ótima de antenas em sistemas de diversidade espacial. A SVR busca medir a qualidade de recepção de um sinal originalmente de módulo constante com base na variação de seu envelope em relação a sua média.

2.6.1 O critério

O algoritmo CMA (seção 2.4.8) procura restaurar a propriedade de módulo constante do sinal inicialmente transmitido, penalizando desvios do sinal na saída do equalizador com relação a um módulo constante, determinado por R_2 . Entretanto, em sinais de módulo constante como no caso das modulações PSK (*Phase Shift Keying*), não existe razão prática para especificar o valor do módulo já que qualquer sinal de módulo constante na saída do equalizador representa uma equalização bem sucedida. O critério proposto se baseia nesta observação.

O valor instantâneo do parâmetro SVR é proporcional a:

$$SVR \propto \frac{E [|y(n)|^2]}{(|y(n)|^2 - E [|y(n)|^2])^2} \quad (2.113)$$

Baseado então nesta idéia, foi proposto o critério do módulo constante generalizado que minimiza a seguinte função custo:

$$J_{GCM} = \frac{E \left[(|y(n)|^2 - E [|y(n)|^2])^2 \right]}{(E [|y(n)|^2])^2} \quad (2.114)$$

Assim, da mesma forma que o critério de Godard, este também penaliza desvios em relação a um módulo constante mas agora o valor exato deste módulo não é especificado. Portanto, qualquer sinal de módulo constante na saída do equalizador representa uma equalização bem sucedida, implicando em um número virtualmente infinito de soluções que satisfazem o critério proposto. Neste sentido, o critério foi chamado de *critério do módulo constante generalizado* (GCM).

2.6.2 O algoritmo: Questão da aproximação estocástica

Baseado no fato da SVR ser uma medida instantânea, o algoritmo GCMA foi deduzido em [Cavalcanti et al. 98] através de um método diferente da aproximação estocástica usual, usada na obtenção de algoritmos como o LMS ou o CMA. Ao invés

de derivar o gradiente de (2.114) e em seguida substituir os valores médios pelos valores instantâneos dos sinais envolvidos, retirando o operador esperança, foi calculado o gradiente do que poderia-se chamar função custo instantânea:

$$J'_{GCM} = \frac{(|y(n)|^2 - E[|y(n)|^2])^2}{(E[|y(n)|^2])^2} \quad (2.115)$$

A fim de facilitar a notação, utilizaremos $\bar{z} = E[|y(n)|^2]$ e $\mathbf{b} = E[y(n)\mathbf{u}^*(n)]$. Derivando então (2.115) em relação a \mathbf{w} , vetor de coeficientes do equalizador, temos:

$$\frac{\partial J'_{GCM}}{\partial \mathbf{w}} = \frac{4}{\bar{z}^2} \left(|y(n)|^2 y(n) \mathbf{u}^*(n) - \mathbf{b} \frac{|y(n)|^4}{E[|y(n)|^2]} + |y(n)|^2 \mathbf{b} - y(n) \mathbf{u}^*(n) \bar{z} \right) \quad (2.116)$$

A equação de atualização dos coeficientes do equalizador, dada por (2.117) abaixo, utiliza (2.116) diretamente lembrando que a aproximação estocástica já foi realizada antes do cálculo do gradiente :

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \frac{\partial J'_{GCM}}{\partial \mathbf{w}} \quad (2.117)$$

Em (2.117), μ é o passo de adaptação e \bar{z} e \mathbf{b} podem ser calculados através de médias temporais:

$$\begin{aligned} \bar{z}(n+1) &= \lambda \bar{z}(n) + (1-\lambda) |y(n)|^2 \\ \mathbf{b}(n+1) &= \lambda \mathbf{b}(n) + (1-\lambda) y(n) \mathbf{u}^*(n) \end{aligned} \quad (2.118)$$

onde λ é o fator de esquecimento.

Deduzindo o algoritmo novamente utilizando o método de aproximação estocástica clássico, calculou-se o gradiente de (2.114) obtendo:

$$\frac{\partial J_{GCM}}{\partial \mathbf{w}} = \frac{4}{\bar{z}^2} \left(E[|y(n)|^2 y(n) \mathbf{u}^*(n)] - \mathbf{b} \frac{E[|y(n)|^4]}{E[|y(n)|^2]} \right) \quad (2.119)$$

Assim a equação de adaptação dos coeficientes do equalizador é dada por:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \frac{y(n)\mathbf{u}^*(n)}{\bar{z}^2} \left(|y(n)|^2 - \frac{E[|y(n)|^4]}{E[|y(n)|^2]} \right) \quad (2.120)$$

Este algoritmo será chamado de GCMA2.

Comparando (2.117) e (2.120) observa-se que os dois algoritmos são diferentes. Entretanto, isto não acontece com critérios como o de Godard por exemplo, onde, se utilizarmos o valor instantâneo da função custo para o cálculo do gradiente, o algoritmo resultante será exatamente o mesmo do que se realizarmos a aproximação estocástica ao final do cálculo do gradiente. Uma análise detalhada dos dois algoritmos e seus desempenhos será realizada nos capítulos seguintes.

2.7 Conclusão

Neste capítulo, vimos os principais critérios e algoritmos autodidatas que serão utilizados como referência no decorrer deste trabalho. Iniciamos com uma breve revisão em equalização supervisionada, por ser a forma mais tradicional de se realizar equalização.

Em seguida, vimos os algoritmos de Bussgang e mostramos como o algoritmo de decisão direta, Sato e principalmente o CMA, que será um dos algoritmos estudados neste trabalho, podem ser incluídos nesta família de algoritmos. Discutimos como o CMA busca otimizar uma função custo que penaliza desvios do sinal na saída do equalizador com relação a um módulo constante, definido pelos momentos do sinal inicialmente transmitido e como o critério de Godard possui mínimos locais que podem levar à convergência do algoritmo para soluções indesejadas.

O fato do critério de Godard especificar, através de R_2 , o módulo do sinal de saída do equalizador não deixa de ser uma restrição à potência deste sinal. Veremos, no próximo capítulo, que a relação entre este critério e o critério de Shalvi-Weinstein com restrição é maior do que pode parecer num primeiro momento.

Quanto à família de critérios decorrente do trabalho de Shalvi-Weinstein, vimos

como fica mais simples dividi-la em três classes distintas: os critérios com restrição que restringem explicitamente a potência do sinal de saída do equalizador; os critérios normalizados, que podem ou não estar sujeitos a alguma restrição de potência e que, se estiverem, equivalem ao critério com restrição; e os critérios sem restrição, que possuem uma função custo com um termo de penalidade a mais. Dependendo de como este termo é escolhido, o critério se reduz ao próprio critério de Godard. O algoritmo super-exponencial é proveniente do critério normalizado, com a restrição de potência.

Finalmente, vimos o critério GCM, que tem como principal característica a ausência de algum tipo de restrição ao sinal de saída do equalizador. No próximo capítulo, faremos a comparação entre este e o critério de Shalvi-Weinstein normalizado que também pode não ter restrição. O algoritmo resultante deste critério, o GCMA, foi obtido utilizando-se um método de aproximação estocástica pouco usual, o que resulta em elementos interessantes a serem investigados, inclusive quanto ao seu desempenho quando comparado aos outros algoritmos. Este estudo será realizado mais detalhadamente no capítulo 4.

Capítulo 3

Comparação Entre os Critérios

Neste capítulo, iremos tratar da comparação entre os diversos critérios descritos no capítulo 2. Para isto, revisitamos alguns resultados obtidos na literatura incluindo, em seguida, o critério do módulo constante generalizado (GCM) na análise. Mostramos também as equivalências entre os critérios citados sob outros pontos de vista, diferentes dos encontrados na literatura e mais simples do que estes.

3.1 Revisitando Resultados da Literatura

3.1.1 Primeira Abordagem

A questão da comparação entre os critérios do Módulo Constante e de Shalvi-Weinstein foi primeiramente abordada na literatura pelo trabalho de Li e Ding, em 1995 [Li e Ding 95]. Para o caso do equalizador duplamente infinito, Foschini [Foschini85] e Shalvi e Weinstein [Shalvi e Weinstein 90] já haviam mostrado que os critérios CM e SW com restrição, respectivamente, não possuíam mínimos locais, sendo as soluções ZF as únicas possíveis. Assim, comparando as análises feitas pelos dois, Li e Ding mostram que existe uma correspondência única entre os pontos de mínimo dos dois critérios, ou seja, o número de mínimos globais será o mesmo para os dois.

Já para o caso do equalizador finito, Ding et al. [Ding et al. 91, Ding e Kennedy 92, Ding et al. 92, Ding e Johnson 93] já haviam mostrado que o critério CM apresenta mínimos locais indesejáveis, o que torna a inicialização um fator crucial para a convergência ao mínimo global desejado. Entretanto esta análise ainda não havia sido feita para o critério SW com restrição. Assim, uma comparação entre os dois critérios poderia trazer informações interessantes sobre este último, ainda não devidamente estudado. É necessário notar que, para o caso do critério SW sem restrição, Shalvi e Weinstein [Shalvi e Weinstein 90] mesmo já haviam demonstrado que, dependendo das escolhas dos parâmetros necessários, acabava-se encontrando o critério CM e portanto, obviamente, neste caso, um possui exatamente os mesmos pontos críticos do outro (seção 2.5.5).

Realizando uma análise basicamente geométrica, Li e Ding mostram que existe uma correspondência única e colinearidade entre os pontos de mínimo dos critérios CM e SW com restrição também para o caso de um equalizador finito. Para isto, os autores primeiramente definem três conceitos:

Attainable Set T: já definido na seção 2.5.6. Por conveniência, sua representação matemática é repetida abaixo:

$$T \triangleq \{ \mathbf{s} : s_n = \sum_{i=0}^L w_i h_{n-i}, w_i \in \mathbb{R} \} \quad (3.1)$$

Cones de Mínimo Global Único S_n^\pm : já definidos na seção 2.5.7. Podem ser escritos matematicamente como:

$$S_n^+ \triangleq \{ \mathbf{s} \in \ell^1(\mathbb{R}) : s_n > 0 \text{ e } s_n > |s_k|, \text{ todo } k \neq n \} \quad (3.2)$$

$$S_n^- \triangleq \{ \mathbf{s} \in \ell^1(\mathbb{R}) : s_n < 0 \text{ e } -s_n > |s_k|, \text{ todo } k \neq n \} \quad (3.3)$$

Super Esfera $\Phi(r), r > 0$: Nada mais é do que uma esfera de raio r . O termo *super* estende a definição para espaços com dimensão maior do que três. Matematicamente temos:

$$\Phi(r) = \mathbf{s} \in \ell^1(\mathbb{R}) / \|\mathbf{s}\| = r \quad (3.4)$$

A demonstração consiste portanto em, assumindo que \mathbf{s}_0 seja um ponto de equilíbrio estável do critério de SW em $T \cap S_n^\pm \cap \Phi(1)$, mostrar que ele também será um ponto de mínimo do critério CM. O fato de \mathbf{s}_0 pertencer a $T \cap S_n^\pm \cap \Phi(1)$ não surpreende pois \mathbf{s} deve ser uma solução possível e portanto pertencente ao *attainable set* T , e deve satisfazer a restrição de potência do critério de SW, pertencendo, portanto, também a $\Phi(1)$. Em seguida considera-se o caminho inverso, ou seja, demonstra-se que se \mathbf{s}_0 é um equilíbrio estável do critério CM em $T \cap S_n^\pm$ então ele também é um ponto de mínimo do critério SW em $T \cap S_n^\pm \cap \Phi(1)$.

É interessante discutir aqui o porque de se escrever “ponto de mínimo do critério de SW” se, como visto na seção 2.5, o critério de SW busca maximizar a função custo. Relembrando, o critério de SW com restrição é dado pela maximização de $|C_{2,2}^y|$, sujeito a $E[|y_n|^2] = E[|a_n|^2]$. Além disso, tem-se a seguinte relação: $|C_{2,2}^y| = |C_{2,2}^a \sum_l |s_l|^4|$. Como o critério CM só é válido caso o sinal transmitido tenha distribuição sub-Gaussiana, ou seja, $C_{2,2}^a < 0$, o critério de SW pode ser reescrito como sendo a minimização de $C_{2,2}^y = C_{2,2}^a \sum_l |s_l|^4$ sujeito à mesma restrição já citada.

Baseado então na equivalência entre os dois critérios, temos que o critério de Shalvi-Weinstein, para o caso do equalizador finito, apresenta os mesmos problemas de convergência do critério CM, possuindo o mesmo número de mínimos locais que este. Portanto, depende da inicialização para atingir a solução desejada.

Toda a análise realizada acima foi feita tendo como hipóteses que o sinal transmitido $a(n)$ era i.i.d, sub-gaussiano e real, e que o canal introduzia distorções lineares. Na próxima seção, discutiremos uma abordagem diferente, cujas hipóteses restringem menos a validade das conclusões obtidas.

3.1.2 Resultados Obtidos por P. Regalia

Phillip Regalia voltou a abordar a questão da comparação entre os critérios alguns anos mais tarde, em 1999 [Regalia99]. Sua análise, entretanto, foi realizada de uma

maneira bastante distinta da realizada anteriormente por Li e Ding. Aqui foram tratados os critérios do Módulo Constante e de Shalvi-Weinstein normalizado, sendo que suas equivalências são mostradas de forma simples, como poderá ser visto em seguida.

A fim de facilitar a análise, precisamos primeiramente escrever o critério de Shalvi-Weinstein normalizado, dado pela equação (2.57), em termos de momentos. Considerando o caso particular $p = 2$ e $q = 2$, e sabendo que o cumulante de quarta ordem é igual ao kurtosis, dado pela equação (2.34), a função custo do critério pode ser escrita como:

$$J_{SW} = \frac{C_{2,2}^y}{(C_{1,1}^y)^2} = \frac{E[|y(n)|^4] - 2E^2[|y(n)|^2] - |E[y^2(n)]|^2}{(E[|y(n)|^2])^2} = \frac{E[|y(n)|^4]}{(E[|y(n)|^2])^2} - \kappa \quad (3.5)$$

onde κ é igual a 3 para o caso dos sinais serem reais e igual a 2 caso sejam complexos com simetria circular, situação na qual $E[y^2(n)] = 0$.

Mais uma vez, como o sinal transmitido $a(n)$ é sub-gaussiano, a função custo do critério de SW, dada por (3.5), deverá ser minimizada.

Regalia [Regalia99] propõe escrever as funções custo J_G e J_{SW} em coordenadas polares, ou seja, em termos de vetores com norma unitária, um ganho ρ e ângulos de rotação θ . No caso de sinais complexos, estes últimos devem incluir a fase do sinal. Para isto, então, temos que o vetor de coeficientes do equalizador pode ser expresso em função de um vetor de norma unitária:

$$\mathbf{w} = \rho \bar{\mathbf{w}} \quad (3.6)$$

onde $\bar{\mathbf{w}}$ possui norma Euclidiana unitária ($\|\bar{\mathbf{w}}\| = 1$) e $\rho = \|\mathbf{w}\|$, um ganho que ajusta a norma de \mathbf{w} . Neste caso, a saída do equalizador será

$$\bar{y}(n) = \mathbf{x}^T(n) \bar{\mathbf{w}}(n), \quad \text{onde } y(n) = \rho \bar{y}(n) \quad (3.7)$$

Note que, ao utilizar vetores de norma unitária, o autor está embutindo na análise, a restrição de potência ($E[|y_n|^2] = E[|a_n|^2]$) existente no critério de SW com restrição, garantindo que a potência do sinal na saída do equalizador seja igual à potência do sinal

transmitido. A equivalência entre os critérios só é válida considerando esta restrição.

O critério de Shalvi-Weinstein, dado por (3.5) pode ser facilmente escrito em função destes novos vetores de norma unitária da seguinte forma:

$$J_{SW}(\theta) = \frac{E[|\bar{y}(n)|^4]}{(E[|\bar{y}(n)|^2])^2} - \kappa \quad (3.8)$$

o que significa que não existe dependência com relação à componente radial ρ e somente com os ângulos θ .

Já o critério de Godard, dado por (2.40), para $p = 2$ (o que resulta no CMA) pode ser escrito da seguinte forma, em função destes novos vetores de norma unitária:

$$J_G(\rho, \theta) = E[(|y(n)|^2 - R_2)^2] = E[(\rho^2|\bar{y}(n)|^2 - R_2)^2] = \rho^4 E[|\bar{y}(n)|^4] - 2\rho^2 R_2 E[|\bar{y}(n)|^2] + R_2^2 \quad (3.9)$$

Claramente, a função acima depende tanto da componente radial ρ como dos ângulos θ . Fica então difícil comparar as duas funções (3.8) e (3.9), pois dependem de variáveis diferentes. O ideal seria que J_G dependesse somente de θ , como J_{SW} . No sentido de procurar “eliminar” a variável ρ , são realizados os seguintes cálculos:

Encontrando o gradiente da função J_G em relação ao parâmetro ρ obtém-se:

$$\nabla_{\rho} J_G = \frac{\partial J_G}{\partial \rho} = 4\rho (\rho^2 E[|\bar{y}(n)|^4] - R_2 E[|\bar{y}(n)|^2]) \quad (3.10)$$

O valor ótimo para ρ é aquele que anula o gradiente, ou seja, aquele que fornece $\nabla_{\rho} J_G = 0$. Esta condição é satisfeita por $\rho = 0$ e $\rho = \pm \sqrt{\frac{R_2 E[|\bar{y}(n)|^2]}{E[|\bar{y}(n)|^4]}}$. O primeiro, no entanto, resulta em um máximo de J_G (mostrado em [Treichler e Agee 83]), enquanto que as duas outras soluções representam pontos de mínimo (mostrado por [Ding e Kennedy 92]). Assim, o valor ótimo para ρ^2 que nos interessa é dado por:

$$\rho_{opt}^2 = R_2 \frac{E[|\bar{y}(n)|^2]}{E[|\bar{y}(n)|^4]} \quad (3.11)$$

Substituindo (3.11) em (3.9), a função custo J_G se reduz a:

$$J_{G_{red}}(\theta) = R_2^2 \left(1 - \frac{(E[|\bar{y}(n)|^2])^2}{E[|\bar{y}(n)|^4]} \right) \quad (3.12)$$

A função acima é chamada por Regalia de *superfície de erro reduzida*. Agora sim, a função depende somente de θ e a comparação se torna possível. Considerando primeiramente o caso real e utilizando (3.8) tem-se:

$$J_{G_{red}}(\theta) = R_2^2 \left(1 - \frac{1}{\frac{E[|\bar{y}(n)|^4]}{(E[|\bar{y}(n)|^2])^2}} \right) = R_2^2 \left(1 - \frac{1}{J_{SW} + 3} \right) \quad (3.13)$$

A relação (3.13) mostra que a superfície de erro reduzida $J_{G_{red}}(\theta)$ é uma deformação de J_{SW} , tendo sido obtida a partir da minimização de J_G em relação a ρ . Esta deformação é monotônica já que

$$\frac{\partial J_{G_{red}}(\theta)}{\partial J_{SW}(\theta)} = \frac{R_2^2}{(J_{SW}(\theta) + 3)^2} > 0 \quad (3.14)$$

e preserva a ordem pois

$$J_{G_{red}}(\theta_1) - J_{G_{red}}(\theta_2) = R_2^2 \frac{J_{SW}(\theta_1) - J_{SW}(\theta_2)}{(J_{SW}(\theta_1) + 3)(J_{SW}(\theta_2) + 3)} \quad (3.15)$$

o que significa que $J_{G_{red}}(\theta_1) > J_{G_{red}}(\theta_2) \Leftrightarrow J_{SW}(\theta_1) > J_{SW}(\theta_2)$. Note que $J_{G_{red}}(\theta) < R_2^2$ para todo θ implicando, de (3.13), que o termo $J_{SW}(\theta_1) + 3$ é sempre positivo.

Para o caso complexo, considerando $\bar{y}(n)$ circular, a equação (3.13) será:

$$J_G(\theta) = R_2^2 \left(1 - \frac{1}{J_{SW} + 2} \right) \quad (3.16)$$

para a qual continuam valendo os argumentos que mostram a monotonicidade e preservação de ordem da deformação.

Com (3.14) e (3.15) é fácil observar que, no espaço de parâmetros θ , os pontos estacionários de $J_G(\theta)$ e $J_{SW}(\theta)$ coincidem, assim como suas classificações (mínimo local

ou global, máximo, pontos de cela). Portanto, para ρ como mostrado em (3.11), θ que minimiza J_G também minimiza J_{SW} . É muito importante ficar claro que, apesar dos pontos críticos serem os mesmos, as duas superfícies de erro não são iguais. Aproveitando a discussão feita pelo próprio autor, além do que já foi comentado, a equivalência demonstrada significa que os dois critérios possuem a mesma capacidade de reduzir IIS e a mesma degradação de desempenho no caso da presença de ruído ou de correlação da fonte. Entretanto, os algoritmos de busca do gradiente resultantes de cada função custo possivelmente serão diferentes em termos de velocidade de convergência e desajuste.

É interessante notar que aqui, diferentemente da análise de Li e Ding, válida somente para o caso real, assume-se pouco mais do que a estacionariedade até quarta ordem do sinal de entrada do equalizador. No caso complexo as hipóteses são um pouco mais restritivas devido às condições de circularidade que devem ser obedecidas. Caso não sejam, os critérios CM e SW podem ter pontos estacionários diferentes.

Como curiosidade é interessante mencionar que, com base nos resultados discutidos acima, o autor propõe ainda um algoritmo de Godard, como o CMA, válido também para fontes super-gaussianas.

Finalmente, como já havia sido comentado na seção 2.5.3, o critério de Shalvi-Weinstein normalizado remete ao critério proposto inicialmente por Donoho. Após a análise mostrada aqui, podemos concluir que não só o algoritmo super-exponencial mas também o CMA, procura os pontos extremos da mesma função custo, $f_{2p}(s)$, dada por (2.58). Assim, os algoritmos pertencem a uma mesma família, apesar de considerarem hipóteses diferentes sobre as estatísticas do sinal transmitido: critério CM supõe sinal com distribuição sub-gaussiana enquanto que o critério de SW não.

3.1.3 Regiões de Atração de Pontos Estacionários no Critério de Shalvi-Weinstein

O fato dos critérios de Godard e Shalvi-Weinstein serem equivalentes mas resultarem em algoritmos diferentes, que possivelmente realizam trajetórias diferentes antes da convergência, pode ficar mais claro com a análise do artigo [Gu e Tong 01], publicado

em 2001, onde M. Gu e L. Tong fazem uma análise das regiões de atração dos pontos estacionários do critério de Shalvi-Weinstein. A motivação maior do artigo é trazer novas informações sobre a questão da inicialização: como inicializar o algoritmo para que a convergência ocorra para a solução desejada? Isto está diretamente relacionado à região de atração do ponto crítico desejado. Em toda a análise, é suposto que a fonte é sub-gaussiana, os símbolos transmitidos e a resposta impulsiva do canal são reais, a matriz de convolução do canal possui posto completo e o ruído é branco e gaussiano.

Região de atração Φ nada mais é do que a região que contém o ponto estacionário x_{opt} , de forma que qualquer condição inicial x_0 pertencente à Φ , levará o algoritmo de busca do gradiente, com passo de adaptação suficientemente pequeno, a gerar uma sequência x_k satisfazendo $J(x_{k+1}) < J(x_k)$ para todo k , permanecer em Φ e convergir para x_{opt} . A definição dada está considerando pontos de mínimo da função custo. O mesmo vale para regiões de atração de pontos de máximo, onde $J(x_{k+1}) > J(x_k)$ deverá ser satisfeita e o algoritmo deve ser considerado no mesmo sentido do vetor gradiente.

O artigo possui vários resultados interessantes. Primeiramente, é mostrado, utilizando-se receptores de Shalvi-Weinstein (com restrição) que, no caso de fontes i.i.d. com símbolos de mesma potência e canais ortonormais, Φ não se altera com a redução ou aumento da relação sinal-ruído (SNR), ou seja, a dificuldade para se inicializar o algoritmo na região de atração de um mínimo global será a mesma para um caso pouco ou muito ruidoso. Por canal ortonormal entende-se um canal ortogonal com norma unitária. O canal ortogonal é aquele no qual as linhas da matriz de convolução do canal, \mathbf{H} , são ortogonais. Relembrando as hipóteses assumidas, como \mathbf{H} também possui posto completo, o operador de projeção será $P_T = I$, ou seja, estamos trabalhando com equalizadores com ordem suficiente, o que só pode ser satisfeito sob condições bastante restritas: equalizadores fracionários, resposta impulsiva $h(z)$ do canal não nula para todo z e equalizador com ordem $(L - 1)$ suficiente. A hipótese de canais ortogonais é importante no caso de aplicações em sistemas CDMA (*Code Division Multiple Access*) e OFDM (*Orthogonal Frequency Division Multiplexing*).

Já para canais ortogonais e fontes não identicamente distribuídas, o tamanho de Φ depende da constelação, da potência do sinal transmitido, do nível de ruído e do canal. Considerando a variância do ruído fixa, o tamanho de Φ irá diminuir conforme se diminui a potência do sinal de entrada ou o valor absoluto do seu cumulante. As-

sim, quanto mais sub-gaussiana for a fonte, maior a região de atração e, portanto, a convergência para a solução desejada é mais fácil de ser obtida, já que a possibilidade de se inicializar o algoritmo na sua região de atração é maior. Usualmente, sinais de constelações mais densas apresentam cumulantes maiores. Por exemplo, a constelação 8-PAM possui cumulante -1.238 enquanto que o do BPSK é igual a -2. Existe portanto um compromisso entre eficiência de largura de banda e capacidade de convergência [Gu e Tong 01].

Entretanto, o resultado que mais nos interessa na análise deste trabalho, é a comparação entre as regiões de atração dos critérios Shalvi-Weinstein e Módulo Constante. Os autores mostram, para o caso de fonte i.i.d. e resposta combinada canal-equalizador bi-dimensional, que, não havendo ruído, as regiões de atração para os dois receptores são iguais. O mesmo continua valendo caso exista ruído e o canal seja ortogonal. Já se o canal for ortogonal, as regiões de atração se tornam diferentes, fazendo com que os algoritmos possuam transientes diferentes em seus processos de convergência.

Observe que as hipóteses assumidas para a validade dos resultados mostrados são bastante restritivas. Assim, no geral, pode-se concluir que as regiões de atração dos pontos estacionários dos critérios de SW e CM serão diferentes e portanto os algoritmos terão desempenhos diferentes.

3.1.4 Relação com Wiener e ZF

O primeiro artigo tratando da relação entre equalizadores obtidos com o critério de Wiener (supervisionado) e os obtidos utilizando-se o critério CM foi escrito por H. Zeng, L. Tong e C. Johnson, em 1998 [Zeng et al. 98]. O artigo procura responder a três perguntas:

- 1) O critério CM possui mínimos locais? Se sim, aonde estão?
- 2) Qual o desempenho em termos de erro quadrático médio, dos equalizadores obtidos através do critério CM?
- 3) Qual a relação entre os equalizadores obtidos por Wiener e os obtidos pelo

critério CM?

Note que a primeira pergunta só faz sentido pois os autores tratam o caso do equalizador fracionário. Considerando o caso do equalizador cujos coeficientes estão espaçados de um período de símbolo (*baud rate*), como considerado até aqui nesta dissertação, tal questão já havia sido respondida de maneira afirmativa por Ding et al. [Ding et al. 91]. Além disso os autores consideram sinais binários, i.i.d. e ruído aditivo, branco e gaussiano.

Utilizando também uma abordagem geométrica, baseada no chamado teorema de Weierstrass [Luenberger69], os autores procuram encontrar uma região compacta B com uma fronteira ∂B e uma referência interna \mathbf{w}_{ref} , de tal forma que o custo (critério CM) na fronteira seja maior do que o do ponto de referência. Assim, tem-se pelo menos um ponto de mínimo em B .

Desta forma é mostrado que o critério CM apresenta sim mínimos locais para o caso em que existe ruído, sempre considerando regiões próximas das soluções de Wiener. Se o mínimo erro quadrático médio for menor do que um certo valor, garante-se a existência de uma solução do critério CM na vizinhança de uma solução de Wiener. É interessante notar que a idéia de erro quadrático médio está associada exclusivamente a Wiener e, no entanto, aqui ela é usada para analisar um equalizador cego. Estes são chamados de mínimos locais “bons” pois resultam em um baixo erro quadrático médio (EQM). Um limitante superior para o EQM destes últimos também é encontrado, dependendo das características da solução de Wiener associada.

Quanto à questão que realmente mais nos interessa neste trabalho, foi concluído que os equalizadores obtidos pelo critério CM e por Wiener são aproximadamente colineares. É calculado também o valor do EQM em excesso do equalizador CM em relação ao de Wiener.

Um segundo trabalho, escrito pelos mesmos autores e publicado em 1999 [Zeng et al. 99], pode ser visto como uma extensão do trabalho anterior, generalizando os resultados para o caso de sinais reais e não mais binários, e trabalhando com equalizadores *baud rate*. A análise realizada segue a mesma idéia do artigo anterior, baseando-se no teorema de Weierstrass. Os resultados obtidos são bastante similares aos já comentados, a saber:

estabelecimento de um limitante superior, em termos de EQM, para o desempenho de equalizadores obtidos através do critério CM; considerações sobre o erro em excesso deste em relação ao de Wiener; constatação da colinearidade dos dois equalizadores.

Pode-se concluir portanto que equalizadores obtidos pelo critério CM podem ter desempenhos semelhantes a equalizadores de Wiener (supervisionados) para altas SNRs, em termos de taxa de erro de bits. Entretanto, como o critério CM é cego, seu equalizador também pode ter um desempenho bastante ruim, se representar um mínimo local indesejável, podendo ser evitado através de uma inicialização correta.

Finalmente, em 1999, M. Gu e L. Tong escreveram outro artigo sobre o assunto [Gu e Tong 99], incluindo os equalizadores obtidos pelo critério de SW em suas análises. Pela equivalência já mostrada nas seções anteriores entre este e o critério CM, muitos dos resultados já discutidos são válidos também para o critério de SW. Neste artigo, os autores consideram uma situação bastante restritiva, utilizando canais ortogonais e respostas combinadas canal-equalizador bidimensionais. Além disso são considerados sinais i.i.d. e ruído aditivo, branco e gaussiano.

Os autores concluem, assim, que os equalizadores obtidos pelos critérios CM, SW, Wiener e ZF são colineares e unicamente correspondentes se e somente se não existir ruído ou o canal for ortogonal. Além disso, para respostas combinadas bidimensionais, é mostrado que para todo equalizador CM/SW existe uma solução de Wiener próxima, ou seja, não existem soluções espúreas (sem associação com alguma solução de Wiener). Desta forma, os três critérios são semelhantes e levam a desempenhos próximos. Outro resultado bastante interessante é que a diminuição da SNR aproxima a relação sinal-interferência (SIR) do receptor CM/SW da apresentada pelo receptor de Wiener, tornando mais próximos os desempenhos dos dois receptores.

3.2 GCM neste contexto

A função custo do critério GCM, dada por (2.114), através de algumas manipulações algébricas, também pode ser escrita da seguinte forma:

$$J_{GCM} = \frac{E \left[(|y(n)|^2 - \bar{z})^2 \right]}{(E [|y(n)|^2])^2} = \frac{E [|y(n)|^4]}{(E [|y(n)|^2])^2} - 1 \quad (3.17)$$

Com (3.5) temos então que:

$$J_{GCM} = J_{SW} + \kappa - 1 \quad (3.18)$$

Assim vê-se que a função custo GCM é equivalente à de Shalvi-Weinstein a menos de uma constante. Note que, diferentemente da equivalência entre os critérios de Godard e Shalvi-Weinstein mostrada por Regalia [Regalia99], agora não só os pontos críticos das duas funções são os mesmos. As superfícies de erro são realmente iguais a menos de um *offset*. Portanto, como já foi visto no capítulo 2, a grande diferença entre os algoritmos resultantes destes dois critérios está na forma como foram obtidos: o ASE é obtido através do método super-exponencial e o GCMA é obtido através de uma aproximação estocástica diferente da usual. A comparação entre estes últimos será feita detalhadamente no capítulo 4.

Comparando com a análise realizada por Regalia, tanto o critério GCM como o SW normalizado possuem vales de soluções correspondentes a mínimos globais ou locais, o que pode ser entendido como uma flexibilização do raio da constelação dos símbolos transmitidos. Não existe portanto a necessidade de se fixar o parâmetro ρ como havia sido necessário no caso da comparação com o critério de Godard. O que podemos concluir portanto é que, enquanto o critério de Godard possui um ponto de mínimo, definido por um ρ_{opt} e um θ_{opt} , os critérios GCM e SW normalizado possuem um vale de soluções, definido simplesmente por θ_{opt} e tendo como uma solução possível (particular), o mesmo ρ_{opt} que leva ao ponto de mínimo do critério de Godard. Assim, podemos dizer que, agora, não existe a restrição de potência garantindo que o sinal na saída do equalizador tenha a mesma potência do sinal transmitido. Na próxima seção, podem

ser encontradas figuras (3.1, 3.3 e 3.5) ilustrando as superfícies de erro dos critérios CM, SW normalizado e GCM respectivamente, facilitando a visualização do que acaba de ser dito.

Com base nesta análise é importante observar que o critério GCM, no caso do equalizador finito, também possui os mesmos mínimos locais que os critérios CM e SW, fato que não havia sido percebido no trabalho original [Cavalcanti et al. 98, Cavalcanti99]. Portanto, a inicialização é muito importante para que os algoritmos convirjam para a solução desejada (mínimo global).

Incluindo o critério GCM nas análises realizadas pelos artigos discutidos na seção 3.1, temos que, no caso do artigo [Gu e Tong 01], esta não é direta, já que toda a análise é feita com critérios que possuem restrição sobre o módulo do sinal da saída do equalizador (CM e SW com restrição). Fica realmente difícil pensarmos em regiões de atração de um vale de soluções, ao invés de pontos estacionários. Além disso, é esperado que o algoritmo GCMA tenha realmente alguma diferença em termos de desempenho quando comparado com os algoritmos ASE e CMA, devido à forma como foi deduzido.

Quanto à análise feita na seção 3.1.4, onde são incluídos os receptores de Wiener e ZF, em [Gu e Tong 99] os autores concluíram que equalizadores obtidos através dos critérios CM, SW, Wiener e ZF são colineares e unicamente correspondentes se e somente se não existir ruído ou o canal for ortogonal. Neste caso, esta conclusão também pode ser estendida para o critério GCM. Isto porque, como foi comentado acima, o critério possui vales de soluções que incluem os pontos estacionários do critério CM e portanto também os do critério SW com restrição. Assim, o vale de soluções também incluirá a solução de Wiener. A correspondência única poderia ser entendida como a existência de um vale de soluções para cada ponto estacionário dos critérios CM, SW com restrição e Wiener. Através de simulações simples, utilizando canais e equalizadores com dois coeficientes, sem ruído e comparando os resultados com as soluções de Wiener, este resultado foi confirmado.

3.3 Equivalências do Ponto de Vista do Vetor Gradiente

Ao invés de comparar os diversos critérios a partir de suas funções custo, como havia sido feito até agora, podemos também comparar seus vetores gradiente. Analisando o caso real (a extensão para o caso complexo é direta), o vetor gradiente para o critério de Shalvi-Weinstein normalizado, (2.57), fazendo $p = 4$ e $q = 0$ será:

$$\frac{\partial J_{SW}}{\partial \mathbf{w}} = \frac{4}{(E[y(n)^2])^2} (E[y^3(n)\mathbf{u}(n)] - R^y E[y(n)\mathbf{u}(n)]) \quad (3.19)$$

onde $R^y = \frac{E[y^4(n)]}{E[y^2(n)]}$

Como a função custo do critério GCM é igual à de SW utilizada no cálculo acima a menos de uma constante, seu gradiente também será o mesmo, dado por (3.19).

Já calculando-se o gradiente do critério de Godard, partimos de (2.40) e, fazendo $p = 2$ e derivando em relação a $\mathbf{w}(n)$, temos:

$$\frac{\partial J_G}{\partial \mathbf{w}} = 4 (E[y^3(n)\mathbf{u}(n)] - R_2 E[y(n)\mathbf{u}(n)]) \quad (3.20)$$

onde $R_2 = \frac{E[a^4(n)]}{E[a^2(n)]}$

Comparando então os vetores gradiente (3.19) e (3.20) vê-se que são muito parecidos, possuindo basicamente a mesma estrutura. Caso R^y seja igual a R_2 , os dois vetores só se diferenciam pelo fator $1/(E[y^2(n)])^2$, que multiplica (3.19). Esta condição, entretanto, só é satisfeita no caso de equalização perfeita (solução ZF) pois só neste caso, quando o vetor resposta combinada do sistema possuir a forma $\mathbf{s} = [0 \dots 0 \ 1 \ 0 \dots 0]$, vale a igualdade abaixo:

$$\frac{E[y^4(n)]}{E[y^2(n)]} = \frac{E[a^4(n)] \sum_l s_l^4 + 3(E[a^2(n)])^2 [(\sum_l s_l^2)^2 - \sum_l s_l^4]}{E[a^2(n)] \sum_l s_l^2} = \frac{E[a^4(n)]}{E[a^2(n)]} \quad (3.21)$$

onde foi utilizada a equação (2.49), considerando-se que $a(n)$ seja i.i.d.

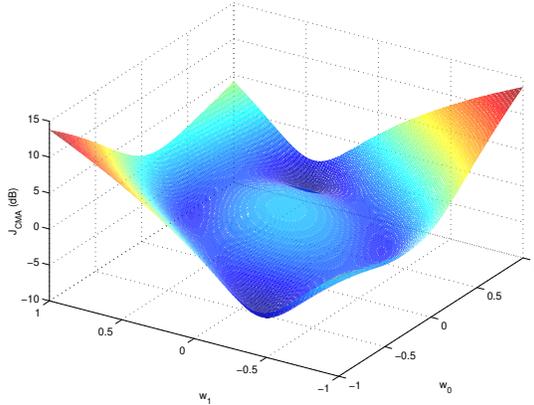


Figura 3.1: Superfície de Erro de Godard; canal: $h(z) = 1 - 0.6z^{-1}$

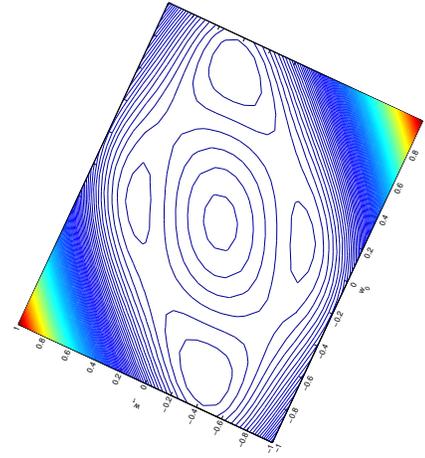


Figura 3.2: Curvas de Nível da Superfície de Erro de Godard; canal: $h(z) = 1 - 0.6z^{-1}$

Portanto, caso (3.21) seja satisfeita e como o fator $1/(E[y^2(n)])^2$ não interfere na obtenção dos pontos críticos, (3.19) e (3.20) possuirão os mesmos pontos estacionários. Note que aqui, como no caso da análise feita por Regalia, a restrição de potência acaba aparecendo, agora na relação (3.21) que deve ser satisfeita.

Entretanto, no que resulta ter $R^y \neq R_2$? Observando as duas superfícies de erro, mostradas nas Figuras 3.1 e 3.3, vê-se que, enquanto Godard possui pontos de mínimo (local ou global), Shalvi-Weinstein, assim como o critério GCM (figura 3.5), possui um vale de soluções que levam tanto ao mínimo global como ao mínimo local da função custo. Os pontos críticos do critério de Godard fazem parte destes vales.

As Figuras 3.1, 3.3 e 3.5 ilustram as superfícies de erro para o critério de Godard, SW normalizado e GCM, respectivamente. O canal utilizado foi $h = 1 - 0.6z^{-1}$, que é o canal também utilizado nas análises realizadas por Ding em [Ding et al. 91]. Ele é um canal bastante simples, permitindo o uso de um equalizador com dois coeficientes, o que facilita a visualização da superfície de erro, introduzindo alguma IIS, o que torna a análise mais interessante. As curvas de nível também são mostradas para a superfície de erro de Godard (figuras 3.2) pois seus mínimos locais são um pouco rasos, dificultando a visualização. A figura 3.4 ilustra a mesma superfície da figura 3.3, mas sob um outro ponto de vista, mostrando que qualquer ponto pertencente ao vale de soluções leva a um

mesmo valor de erro. Para este canal, os mínimos globais, obtidos através da solução de Wiener, correspondem a $w=[0.91 \ 0.4]$ e seu simétrico e os mínimos locais a $w=[-0.14 \ 0.67]$ e seu simétrico.

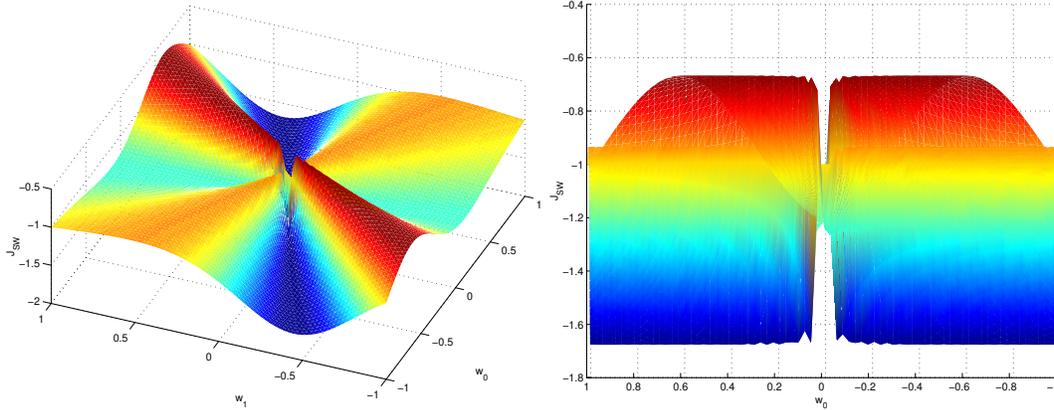


Figura 3.3: Superfície de Erro de Shalvi-Weinstein; canal: $h(z) = 1 - 0.6z^{-1}$ Figura 3.4: Superfície de Erro de Shalvi-Weinstein; canal: $h(z) = 1 - 0.6z^{-1}$

É interessante notar que, apesar do gradiente do critério de Shalvi-Weinstein depender de R^y , o algoritmo super-exponencial, deduzido na seção 2.5.6, não utiliza R^y e sim R_2 . Lembre-se que ao final da dedução do ASE, a restrição de potência $E[|y_n|^2] = E[|a_n|^2]$ é explicitamente utilizada. Assim o algoritmo garante a igualdade $R^y = R_2$ e portanto irá convergir para os mesmos pontos estacionários que o algoritmo CMA e não para os vales mostrados na figura 3.3 acima. Note que isto já não acontece com o GCMA. Este algoritmo não possui nenhuma restrição de potência e portanto convergirá para os vales de soluções mostrados na figura 3.5. Uma análise mais detalhada destes algoritmos será realizada no capítulo 4.

3.4 Conclusão

Neste capítulo, primeiramente foi estudada a equivalência entre os critérios CM e SW, partindo-se de resultados já existentes na literatura, de onde se conclui que os dois critérios possuem os mesmos pontos estacionários, sempre considerando a validade da restrição de potência do sinal de saída do equalizador. Entretanto, suas superfícies de erro são diferentes e os algoritmos resultantes em cada caso também serão. Lembre-se

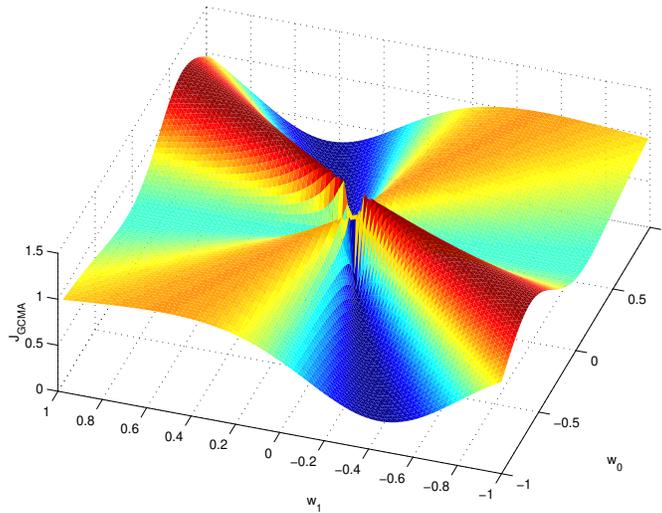


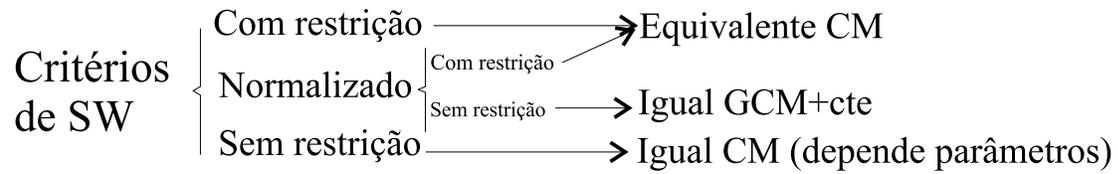
Figura 3.5: Superfície de Erro do critério GCM; canal: $h(z) = 1 - 0.6z^{-1}$

que o ASE é naturalmente deduzido com a presença da matriz \mathbf{Q} , diferentemente do CMA e portanto é esperado que tenha um melhor desempenho do que este último.

Incluindo o critério GCM nesta análise, mostramos que este equivale ao critério SW normalizado a menos de uma constante e que, por não possuir nenhuma restrição quanto ao módulo do sinal na saída do equalizador, possui vales de soluções correspondentes ao mínimo global e local, o que resulta, virtualmente, na existência de infinitas soluções. Estes vales possuem os pontos estacionários do critério CM e mesmo os do critério de Wiener, para o caso em que não existe ruído.

Por último, estas equivalências foram também mostradas por meio de um ponto de vista alternativo, partindo do vetor gradiente dos critérios. Vimos que, satisfazendo a relação $R^y = R_2$, que na verdade nada mais é do que a própria restrição de potência do sinal de saída do equalizador, os critérios CM e SW realmente terão os mesmos pontos estacionários. Caso contrário, a diferença entre estes dois termos leva o critério de SW normalizado a também possuir vales de soluções, enquanto que o critério de Godard converge para pontos estacionários.

Finalmente, a título de sintetizar as relações de equivalência existentes entre os critérios estudados no capítulo, propomos o esquema abaixo:



Capítulo 4

Comparação Entre os Algoritmos

Neste capítulo, é feita uma comparação entre os algoritmos CMA, ASE, GCMA e GCMA2 (obtido através da aproximação estocástica clássica), através de simulações. Primeiramente, os algoritmos são comparados em termos de convergência, utilizando-se um canal simples e um equalizador com dois coeficientes. São observadas inicializações que levam a mínimos locais, e como se dá a convergência do GCMA e do GCMA2 para vales de soluções. Em seguida, buscamos utilizar canais mais complexos, de fase não mínima e adicionar ruído. Por último, é feita uma análise mais detalhada do algoritmo GCMA, à luz da questão da aproximação estocástica, já que este algoritmo apresenta um bom desempenho e ainda não havia sido estudado em [Cavalcanti et al. 98, Cavalcanti99].

4.1 Formas de Análise de Desempenho

A fim de podermos analisar e comparar o desempenho dos diversos algoritmos em questão, utilizaremos como medida de desempenho, a IIS residual dada por:

$$IIS(\mathbf{s}) = \frac{\sum_i |s_i|^2 - |s|_{max}^2}{|s|_{max}^2} \quad (4.1)$$

Para que a comparação em termos de velocidade de convergência possa ser feita de maneira justa, precisamos determinar um parâmetro que deve ser atendido por todos os algoritmos. Uma forma de se fazer isso é considerar que todos convergem o mais rapidamente possível e compararmos o valor da IIS residual ao final do processo de convergência, ou determinarmos que todos os algoritmos devem convergir para um mesmo patamar de IIS residual e compararmos qual o atinge primeiramente.

Além disso, precisamos determinar a partir de que valor de IIS residual, pode-se considerar que o sistema foi equalizado, ou seja, que o olho na saída do equalizador está aberto. Para isto realizamos algumas simulações, chegando à conclusão de que o sistema pode ser considerado equalizado, a partir de uma IIS residual de -20dB . A título de exemplo, as figuras 4.1 e 4.2 mostram os símbolos à saída de um canal dado por $h(z) = 0.3482 - 0.8704z^{-1} + 0.3482z^{-2}$, e à saída do equalizador, para modulação BPSK. É possível notar, em 4.2, que os símbolos da saída do equalizador possuem uma variação de no máximo ± 0.5 em torno dos símbolos da constelação ($\{+1, -1\}$), resultando em uma estimação correta por parte do decisor. O mesmo resultado pode ser observado nas figuras 4.3 e 4.4 que ilustram o caso da modulação QPSK. Neste caso, os símbolos da constelação são dados por $\{+1 + j, +1 - j, -1 + j, -1 - j\}$ e a variação é de no máximo ± 0.5 tanto no sentido do eixo real como no do eixo imaginário.

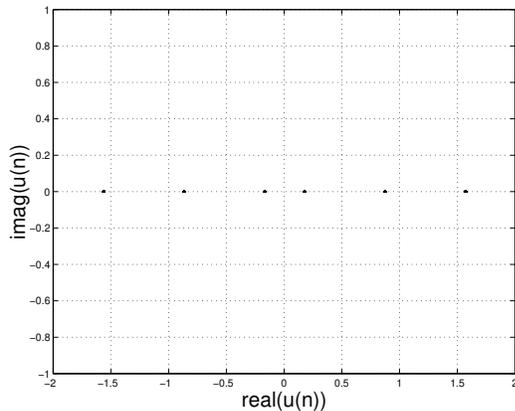


Figura 4.1: Símbolos na saída do canal, sem ruído, IIS= -20dB , modulação BPSK

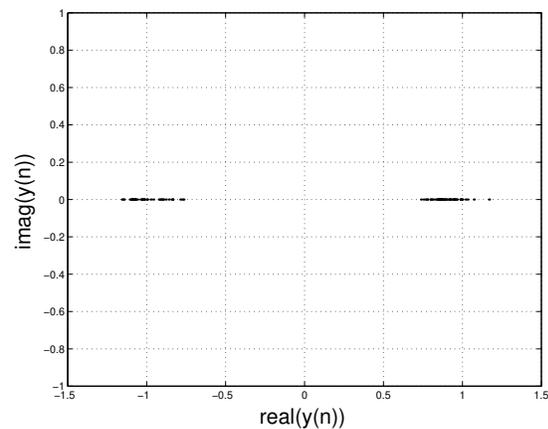


Figura 4.2: Símbolos na saída do equalizador, sem ruído, IIS= -20dB , modulação BPSK

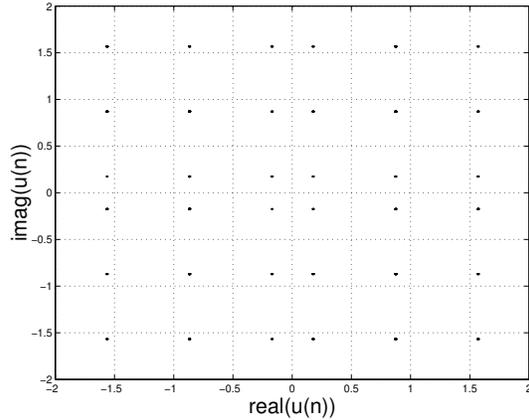


Figura 4.3: Símbolos na saída do canal, sem ruído, IIS= $-20dB$, modulação QPSK

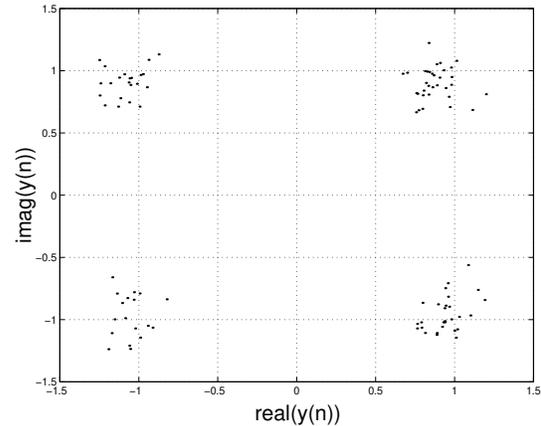


Figura 4.4: Símbolos na saída do equalizador, sem ruído, IIS= $-20dB$, modulação QPSK

Já adicionando-se ruído ($SNR = 20dB$), a variação dos símbolos da saída do equalizador com relação aos símbolos da constelação aumenta um pouco, mas continua sendo bastante aceitável, chegando a ± 0.6 . As figuras 4.5 e 4.6 ilustram esta última situação para as modulações BPSK e QPSK respectivamente. É importante deixar claro que o ruído, adicionado à saída do canal, é modelado como branco e gaussiano. Sua variância, para efeitos de simulação, é calculada a partir da SNR (*Signal to Noise Ratio*) da seguinte forma:

$$\sigma_{ruído}^2 = \sqrt{\frac{1}{10^{\frac{SNR}{10}}}} \quad (4.2)$$

Com estes parâmetros, partimos então para a análise dos algoritmos em termos de convergência, ou seja, analisamos como, dependendo da inicialização, os algoritmos convergem para mínimos globais ou locais.

4.2 Análise de Convergência dos Algoritmos

Primeiramente, relembremos os algoritmos mais importantes para nossa análise, todos já deduzidos no capítulo 2:

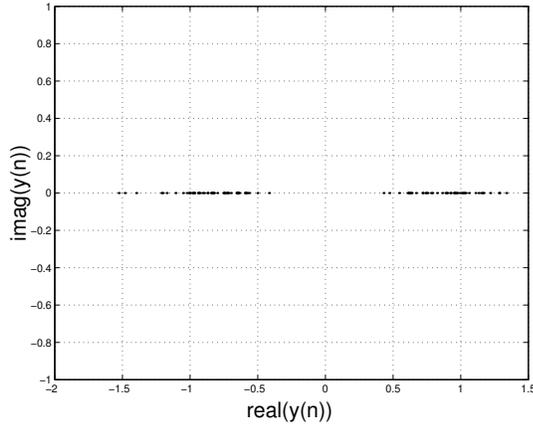


Figura 4.5: Símbolos na saída do canal, $SNR = 20dB$, $IIS = -20dB$, modulação BPSK

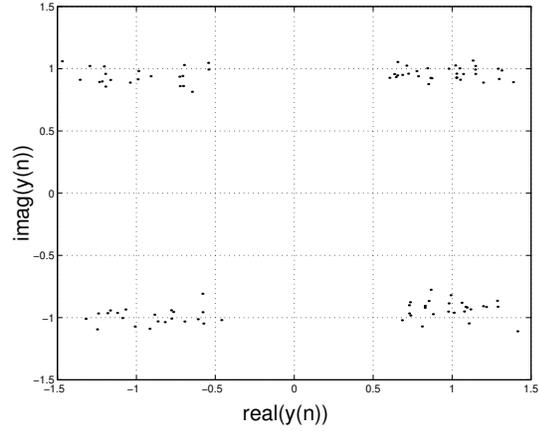


Figura 4.6: Símbolos na saída do equalizador, $SNR = 20dB$, $IIS = -20dB$, modulação QPSK

CMA:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \mathbf{u}_n^* y(n) (|y_n|^2 - R_2) \quad (4.3)$$

$$\text{onde } R_2 = E[|a_n|^4] / E[|a_n|^2]$$

ASE (cumulantes de quarta ordem):

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \frac{\beta}{\delta} \mathbf{Q}_n \mathbf{u}_n^* y(n) (|y_n|^2 - R_2) \quad (4.4)$$

onde $\mathbf{Q} \propto \mathbf{R}_u^{-1}$, $\mathbf{R}_u = E[\mathbf{u}_n^* \mathbf{u}_n^T]$, $\delta = \frac{E[|a_n|^4] - \lambda E^2[|a_n|^2]}{E[|a_n|^2]}$ e β é o passo de adaptação, também utilizado como fator de esquecimento para o cálculo iterativo de \mathbf{Q}

GCMA:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{\mu}{\bar{z}^2} \left(|y_n|^2 y(n) \mathbf{u}_n^* - R^{y'} \mathbf{b} + |y_n|^2 \mathbf{b} - y_n \mathbf{u}_n^* \bar{z} \right) \quad (4.5)$$

onde $\bar{z} = E[|y_n|^2]$, $\mathbf{b} = E[y_n \mathbf{u}_n^2]$ e $R^{y'} = |y_n|^4 / E[|y_n|^2]$. As médias temporais necessárias podem ser calculadas da seguinte forma:

$$\bar{z}(n+1) = \lambda \bar{z}(n) + (1 - \lambda) |y(n)|^2$$

$$\mathbf{b}(n+1) = \lambda \mathbf{b}(n) + (1-\lambda)y(n)\mathbf{u}^*(n) \quad (4.6)$$

GCMA2 (obtido a partir da aproximação estocástica clássica):

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{\mu}{\bar{z}^2} (|y_n|^2 y_n \mathbf{u}_n^* - R^y y_n \mathbf{u}_n^*) \quad (4.7)$$

onde $R^y = E[|y_n|^4] / E[|y_n|^2]$

Nos algoritmos citados acima, o CMA e o ASE convergem para pontos estacionários, enquanto o GCMA e o GCMA2 convergem para vales de soluções que contém os pontos estacionários do CMA e do ASE. Isto já foi mostrado no capítulo anterior e pode ser visto facilmente através de simulações.

Utilizando o canal $h(z) = 1 - 0.6z^{-1}$, que permite uma análise mais detalhada por ser simples e permitir o uso de um equalizador com dois coeficientes, sabemos, das análises já realizadas por Ding [Ding91], que o critério CM possuirá um mínimo global e seu simétrico e um mínimo local e seu simétrico.

O critério de Shalvi-Weinstein, no caso da família de critérios normalizados, possui vales de soluções pois a restrição de potência $E[y_n^2] = E[a_n^2]$ não é considerada. Entretanto, lembrando a dedução do ASE, feita no capítulo 2, temos que a restrição de potência é considerada, e portanto o ASE utiliza R_2 ao invés de R^y . Assim, o algoritmo satisfaz a restrição discutida na seção 3.3, convergindo para os mesmos pontos estacionários que o critério de Godard e não possuindo mais vales de soluções. A figura 4.7 ilustra a convergência destes dois algoritmos sendo inicializados em vários pontos distintos, no plano formado pelos dois coeficientes do equalizador. Os passos de adaptação utilizados foram suficientemente pequenos para garantir que os algoritmos ficassem presos em todos os mínimos locais existentes, mesmo que rasos.

O critério GCM, por ser equivalente ao critério de SW normalizado, também possui vales de soluções correspondentes a mínimos globais e locais. Entretanto, diferentemente do ASE, o GCMA é deduzido sem a utilização de restrições de potência. Assim, este convergirá para uma das soluções pertencentes aos vales, apresentando, virtualmente, um número infinito de valores possíveis. O valor exato desta solução depende

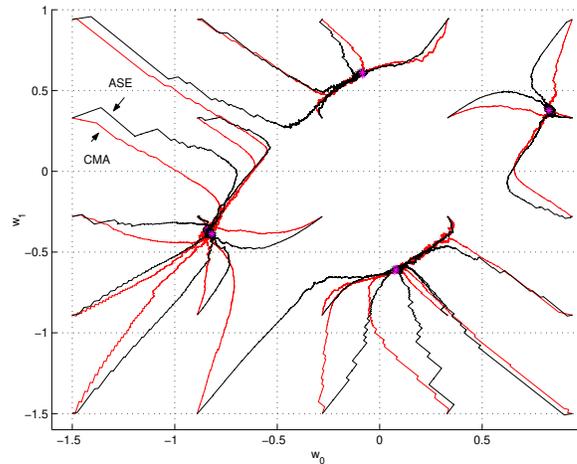


Figura 4.7: Convergência dos algoritmos CMA e ASE, canal: $h(z) = 1 - 0.6z^{-1}$, CMA: $\mu = 0.002$, ASE: $\beta = 0.008$

da inicialização do algoritmo. É interessante lembrar que estes vales contém os pontos estacionários do critério CM, para os quais o ASE também converge. A figura 4.8 a seguir mostra um mapa da convergência destes algoritmos. É importante comentar que o fator de esquecimento, λ , no algoritmo GCMA, utilizado para calcular as médias temporais necessárias no algoritmo, deve possuir um valor próximo de 1. Quanto mais baixo for seu valor, mais os vales de soluções (principalmente o correspondente aos mínimos locais) do algoritmo se distanciam dos pontos críticos para os quais o CMA converge, pois maior será o ruído de gradiente envolvido no processo de adaptação.

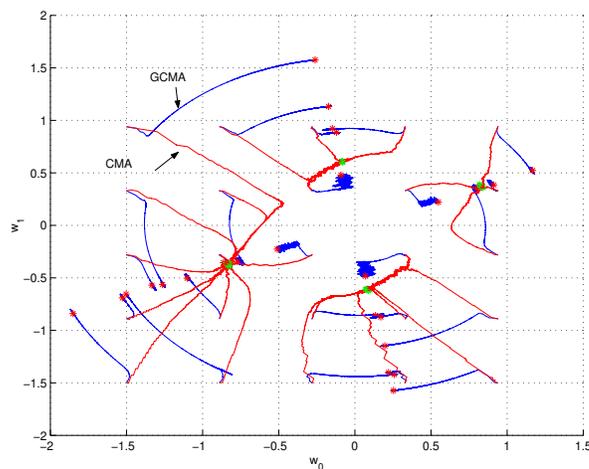


Figura 4.8: Convergência dos algoritmos GCMA e CMA, canal: $h(z) = 1 - 0.6z^{-1}$, GCMA: $\mu = 0.0025, \lambda = 0.97$; CMA: $\mu = 0.002$

Comparando-se também os algoritmos GCMA e GCMA2, é possível observar, através da figura 4.9 que os dois possuem os mesmos vales de convergência, o que já era esperado já que os dois algoritmos provém da mesma função custo. Além disso, o segundo possui uma característica que pode causar problemas: os coeficientes do equalizador convergem para os vales de soluções mas, uma vez atingindo-os, continuam o processo de adaptação, se distanciando cada vez mais da origem. Imagine então que queremos transmitir uma sequência muito longa de dados. Os coeficientes do equalizador podem atingir valores extremamente altos, o que não é, de forma alguma, desejável. Através da figura 4.8, é possível observar que os coeficientes do equalizador no caso do GCMA, principalmente quando seus valores iniciais são baixos (próximos do ponto $(0,0)$), também se distanciam um pouco da origem, mas é fácil notar que este efeito é muito menor neste último.

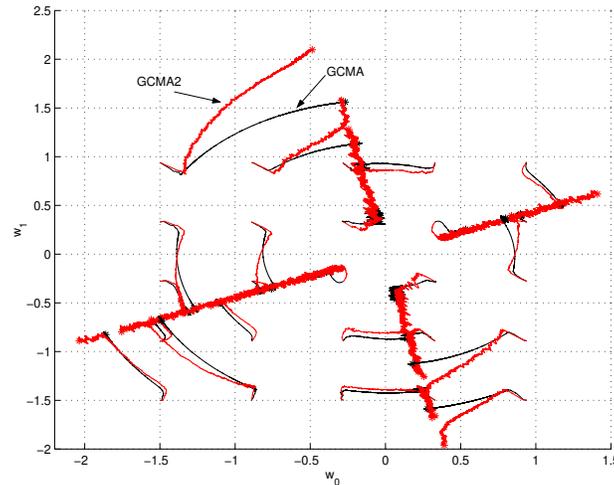


Figura 4.9: Convergência dos algoritmos GCMA e GCMA2, canal: $h(z) = 1 - 0.6z^{-1}$, no dois algoritmos foram utilizados os mesmos parâmetros: $\mu = 0.0025, \lambda = 0.97$

Em seguida veremos os desempenhos destes algoritmos na tarefa de equalizar vários canais distintos.

4.3 Desempenho dos Algoritmos

Estamos interessados aqui no desempenho dos diversos algoritmos citados na seção anterior. Antes de iniciar a discussão dos resultados, entretanto, é necessário comentar

que, nos casos em que os algoritmos utilizados foram o GCMA e o GCMA2, na saída do equalizador, foi colocado um controlador automático de ganho (CAG). Este foi implementado de maneira bastante simples. Chamando de $v(n)$ os símbolos na saída do CAG e lembrando que a seguinte relação é válida:

$$E [|y_n|^2] = E [|a_n|^2] \|s_n\|^2 \quad (4.8)$$

podemos escrever:

$$v_n = \frac{y_n}{\|s_n\|} \Rightarrow E [|v_n|^2] = \frac{E [|y_n|^2]}{\|s_n\|^2} = \frac{E [|y_n|^2]}{E [|y_n|^2]} E [|a_n|^2] = E [|a_n|^2] \quad (4.9)$$

garantindo assim a igualdade de potência entre o sinal transmitido e o sinal na saída do CAG.

Basta então dividirmos a saída do equalizador por $\|s_n\|$ que, utilizando (4.8), é equivalente à $\sqrt{\frac{E [|y_n|^2]}{E [|a_n|^2]}}$. Caso a modulação utilizada seja BPSK, $E [|a_n|^2] = 1$ e caso seja QPSK, $E [|a_n|^2] = 2$. O termo $E [|y_n|^2]$ foi calculado como uma média temporal, da mesma forma que o termo \bar{z} no algoritmo GCMA. Note que o CAG não irá influenciar no cálculo da IIS residual, pois esta é normalizada pelo termo $|s|_{max}^2$, como pode ser observado na equação (4.1). Entretanto, é importante que seja citado já que representa um certo aumento no custo computacional, mesmo que pequeno.

A fim de que os algoritmos pudessem ser comparados de maneira justa, utilizaram-se passos de adaptação de forma que todos convergissem para um mesmo patamar de IIS residual. Isto, no entanto, só é possível mantendo-se uma certa relação entre os passos de adaptação dos diversos algoritmos considerados. Comparando (4.3) e (4.4) observa-se que a única diferença está no passo (ASE possui o termo $1/\delta$) e na presença da matriz \mathbf{Q} . Quanto ao passo, podemos simplesmente alterar o passo do CMA para $\frac{\mu}{\delta}$. Já com relação à matriz, seu cálculo recursivo pode ser escrito como $\mathbf{Q}_n^{-1} = (1 - \beta_Q)\mathbf{Q}_{n-1}^{-1} + \beta_Q\mathbf{u}_n^*\mathbf{u}_n^T$. Portanto, existe um termo β_Q que multiplica $\mathbf{u}_n^*\mathbf{u}_n^T$ e que não existe no algoritmo CMA. Este termo no entanto pode ser incorporado ao próprio passo deste último algoritmo. Considerando $\beta = 1$ em (4.4), temos que os dois algoritmos convergirão para um mesmo patamar de IIS residual se o passo de adaptação do CMA for $\mu = \beta_Q/\delta$. É importante ressaltar que a norma da resposta impulsiva do canal é unitária.

O mesmo raciocínio pode ser usado com relação ao GCMA. Entretanto, este possui o termo $1/\bar{z}^2$ que multiplica o resto do gradiente. Assim, seu passo de adaptação deveria ser $\mu = \frac{\bar{z}^2 \beta_Q}{\delta}$ o que resultaria em um passo variável, já que o valor exato de \bar{z} não é conhecido. Desejamos, no entanto, manter um passo de adaptação fixo para todos os algoritmos. Podemos então realizar uma aproximação, fazendo $\bar{z} = E[|a_n|^2]$ já que este último termo é conhecido. Note que aqui, esta relação não representa uma restrição de potência, já que estamos trabalhando somente com o passo de adaptação do algoritmo e não com seu vetor gradiente.

Observe ainda que β_Q é um fator de esquecimento no cálculo de \mathbf{Q} . Para que seja compatível com as médias temporais calculadas no algoritmo GCMA, podemos fazer $\beta_Q = (1 - \lambda)$, onde λ é o fator de esquecimento utilizado no cálculo de \bar{z} e \mathbf{b} . O mesmo é feito para o GCMA2. Teremos então os seguintes algoritmos:

CMA:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{(1 - \lambda)}{\delta} \mathbf{u}_n^* y(n) (|y_n|^2 - R_2) \quad (4.10)$$

ASE:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \frac{1}{\delta} \mathbf{Q}_n \mathbf{u}_n^* y(n) (|y_n|^2 - R_2) \quad (4.11)$$

GCMA:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{(1 - \lambda)}{\delta} \frac{1}{\bar{z}^2} \left(|y_n|^2 y(n) \mathbf{u}_n^* - R^y \mathbf{b} + |y_n|^2 \mathbf{b} - y_n \mathbf{u}_n^* \bar{z} \right) \quad (4.12)$$

GCMA2:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{(1 - \lambda)}{\delta} \frac{1}{\bar{z}^2} (|y_n|^2 y_n \mathbf{u}_n^* - R^y y_n \mathbf{u}_n^*) \quad (4.13)$$

Visto isso, as próximas seções ilustram os resultados obtidos através de simulações.

4.3.1 Modulação BPSK

Comparando então o desempenho dos diversos algoritmos citados, simulamos cada um para canais de fase não mínima (resposta impulsiva com dois zeros, um no interior do círculo unitário e um no exterior), considerando os zeros interiores ao círculo unitário, no plano z , próximos à origem e próximos ao círculo unitário. Os resultados são mostrados nas figuras 4.10-4.12, respectivamente para canais com zeros, no plano z , em 0.2 e 5, 0.5 e 2, 0.8 e 1.25. Foram utilizados equalizadores sempre com 15 coeficientes. A inicialização dos algoritmos foi feita com *center spike*, onde o oitavo coeficiente, já que os canais utilizados eram sempre simétricos, possuía valor -1. A modulação utilizada para os símbolos transmitidos foi sempre BPSK. Primeiramente, não foi adicionado ruído. A tabela 4.1 resume a inicialização utilizada.

<i>Parâmetros de simulação para obtenção das Figuras 4.10-4.12</i>	
Todos	Modulação BPSK, sem ruído Equalizador com 15 coeficientes Inicialização: $[0...0 \ -1 \ (8^a \text{ posição}) \ 0...0]^T$
GCMA	$\bar{z}_0 = 0$ $\bar{b}_0 = 0$
ASE	$\mathbf{Q}_0 =$ média entre os primeiros 50 vetores de entrada \mathbf{u}

Tabela 4.1: Parâmetros de inicialização para as simulações cujos resultados são mostrados nas figuras 4.10-4.12

Iniciando a discussão dos resultados obtidos pelos dois algoritmos mais clássicos e mais estudados na literatura, o CMA e o ASE, é facilmente observado através das simulações que o segundo possui uma velocidade de convergência muito maior que o primeiro. Comparando suas equações, (4.10) e (4.11) respectivamente, podemos observar que a grande diferença está no segundo possuir a matriz \mathbf{Q} . Como já foi brevemente comentado, esta matriz é proporcional à matriz de autocorrelação do sinal de entrada do equalizador \mathbf{u}_n e, portanto, é equivalente a uma operação de pré-branqueamento deste sinal. Observe que seu efeito é percebido mais facilmente quanto mais correlacionado for o sinal de entrada do equalizador, ou seja, quanto mais próximos os zeros do canal estiverem do círculo unitário no plano z , maior será a diferença de desempenho entre o ASE e o CMA. Isto pode ser confirmado através das figuras 4.10 e 4.12. Além disso,

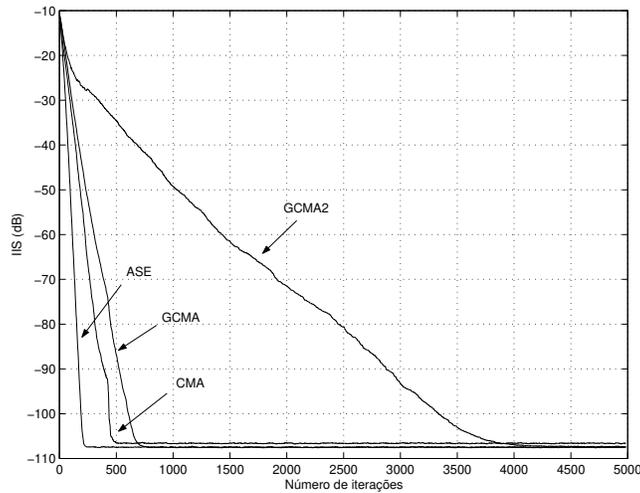


Figura 4.10: Comparação entre os algoritmos CMA: $\mu = 0.025$, GCMA: $\mu = 0.03$ e $\lambda = 0.65$, GCMA2: $\mu = 0.03$ e $\lambda = 0.65$, e ASE: $\beta_Q = 0.06$; canal: $h(z) = 0.1856 - 0.9650z^{-1} + 0.1856z^{-2}$, média 200 simulações ($\lambda = 0.94$)

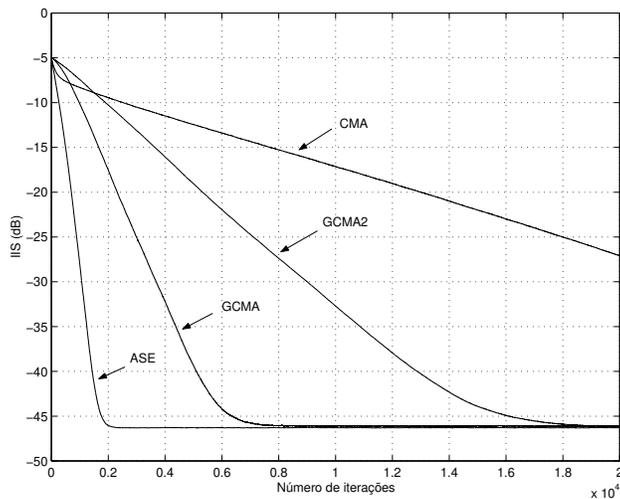


Figura 4.11: Comparação entre os algoritmos CMA: $\mu = 0.0018$, GCMA: $\mu = 0.0018$ e $\lambda = 0.8$, GCMA2: $\mu = 0.0018$ e $\lambda = 0.8$, e ASE: $\beta_Q = 0.0036$; canal: $h(z) = 0.3482 - 0.8704z^{-1} + 0.3482z^{-2}$, média 100 simulações ($\lambda = 0.9964$)

vimos na seção 2.5.7, que o passo de adaptação do algoritmo ASE é ótimo no sentido da velocidade de convergência. Apesar da demonstração da validade desta afirmação considerar um cálculo exato do vetor gradiente, o bom desempenho do algoritmo é facilmente observado.

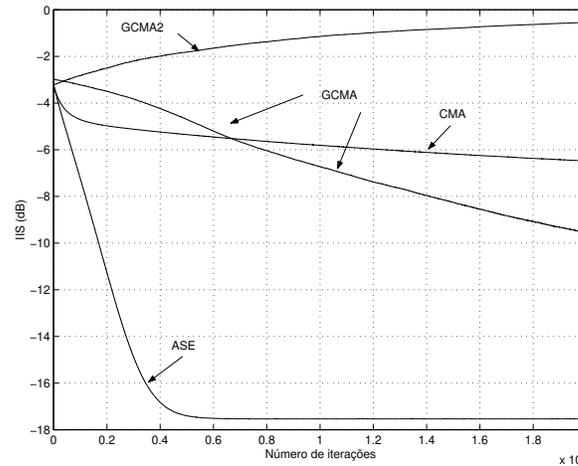


Figura 4.12: Comparação entre os algoritmos CMA: $\mu = 0.0005$, GCMA: $\mu = 0.0005$ e $\lambda = 0.8$, GCMA2: $\mu = 0.0005$ e $\lambda = 0.8$, e ASE: $\beta_Q = 0.001$; canal: $h(z) = 0.4015 - 0.8231z^{-1} + 0.4015z^{-2}$, média 100 simulações ($\lambda = 0.999$)

Considerando agora o algoritmo GCMA2, dado por (4.13), vê-se que converge mais lentamente do que o CMA na figura 4.10, mais rapidamente na figura 4.11 e não chega a convergir na figura 4.12. Comparando suas equações, vemos que a grande diferença está em um utilizar R^y enquanto o outro utiliza R_2 . Isto já foi discutido na seção 4.2: o primeiro caso resulta em uma convergência para vales de soluções enquanto que o segundo, para pontos estacionários. Entretanto, em termos de desempenho, vemos que o GCMA2 é muito mais instável do que o CMA e mesmo que o GCMA. Para os casos em que houve convergência (note que foi considerada uma média entre 200 simulações para garantir que o algoritmo realmente converge), alterando-se um pouco os parâmetros utilizados, o algoritmo já diverge. As figuras 4.13 e 4.14 ilustram estas situações. Em relação às simulações já realizadas, só foi alterado o valor do fator de esquecimento λ .

Devido à grande instabilidade apresentada pelo GCMA2, além do problema do valor dos coeficientes do equalizador irem sempre aumentando depois da convergência, como mostrado na seção anterior, este não será considerado nas próximas análises.

Quanto ao algoritmo GCMA, vê-se que possui um desempenho bastante próximo do CMA (figura 4.10) para um canal com zeros mais próximos da origem no plano z , o que já era esperado pois quanto mais próximos os zeros estão da origem, mais os desempenhos dos algoritmos se assemelham, já que o sinal de entrada do equalizador

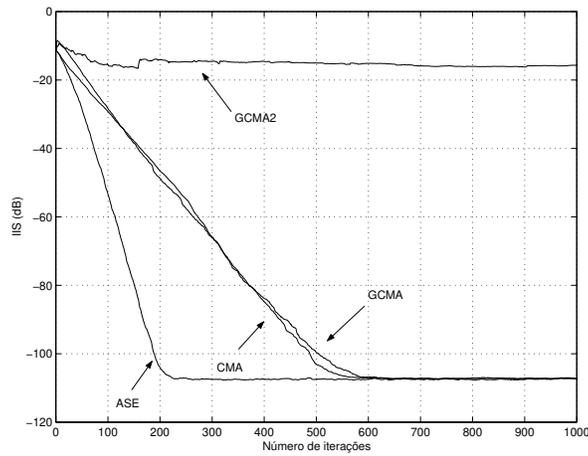


Figura 4.13: Comparação entre os algoritmos CMA: $\mu = 0.025$, GCMA: $\mu = 0.03$ e $\lambda = 0.75$, GCMA2: $\mu = 0.03$ e $\lambda = 0.75$, e ASE: $\beta_Q = 0.06$; canal: $h(z) = 0.1856 - 0.9650z^{-1} + 0.1856z^{-2}$, média 100 simulações ($\lambda = 0.94$)

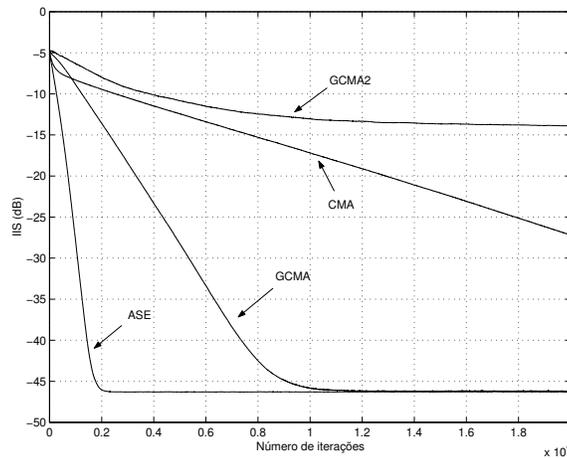


Figura 4.14: Comparação entre os algoritmos CMA: $\mu = 0.0018$, GCMA: $\mu = 0.0018$ e $\lambda = 0.65$, GCMA2: $\mu = 0.0018$ e $\lambda = 0.65$, e ASE: $\beta_Q = 0.0036$; canal: $h(z) = 0.3482 - 0.8704z^{-1} + 0.3482z^{-2}$, média 100 simulações ($\lambda = 0.9964$)

será mais decorrelacionado. Já nas figuras 4.11 e 4.12, vemos que o GCMA converge mais rapidamente do que o CMA, com a vantagem de possuir um custo computacional bem mais baixo do que ASE, já que não possui a necessidade da inversão de matriz. Um estudo mais detalhado deste algoritmo será realizado na próxima seção, por ser um algoritmo ainda pouco conhecido e apresentar um bom desempenho quando comparado ao CMA.

Mantendo os mesmos canais já utilizados e adicionando ruído, podemos ver através das figuras 4.15-4.17, que as discussões feitas para o caso sem ruído continuam valendo. O ASE continua tendo o melhor desempenho em termos de velocidade de convergência, seguido pelo GCMA, enquanto o CMA mostrou ser, novamente, o mais lento. Os parâmetros utilizados para inicialização dos algoritmos são mostrados na tabela 4.1, tendo como única diferença a adição de ruído. Foi considerada uma SNR=20dB.

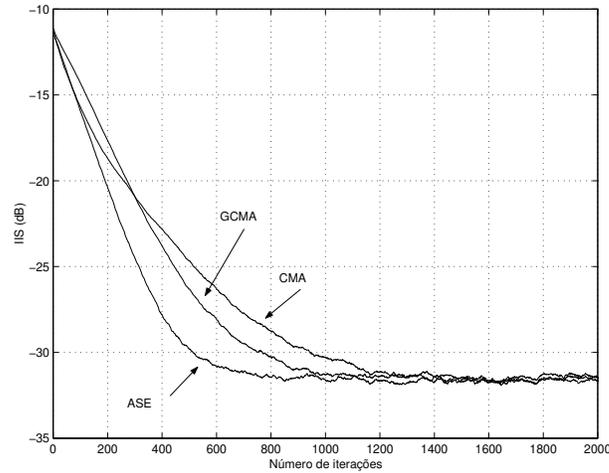


Figura 4.15: Comparação entre os algoritmos CMA: $\mu = 0.004$, GCMA: $\mu = 0.004$ e $\lambda = 0.8$, e ASE: $\beta_Q = 0.007$; canal: $h(z) = 0.1856 - 0.9650z^{-1} + 0.1856z^{-2}$, SNR=20dB, média 100 simulações

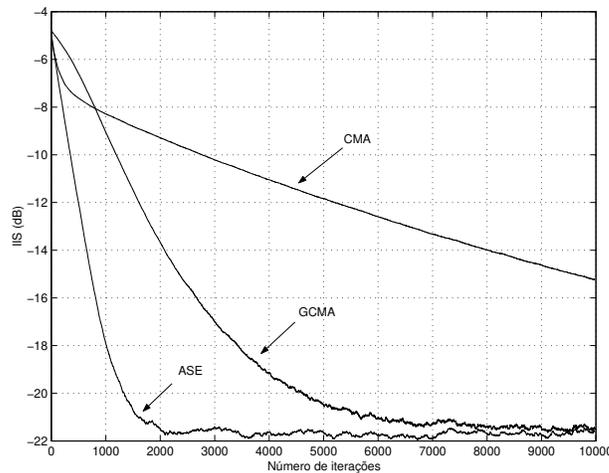


Figura 4.16: Comparação entre os algoritmos CMA: $\mu = 0.0018$, GCMA: $\mu = 0.0018$ e $\lambda = 0.8$, e ASE: $\beta_Q = 0.0036$; canal: $h(z) = 0.3482 - 0.8704z^{-1} + 0.3482z^{-2}$, SNR=20dB, média 100 simulações ($\lambda = 0.9964$)

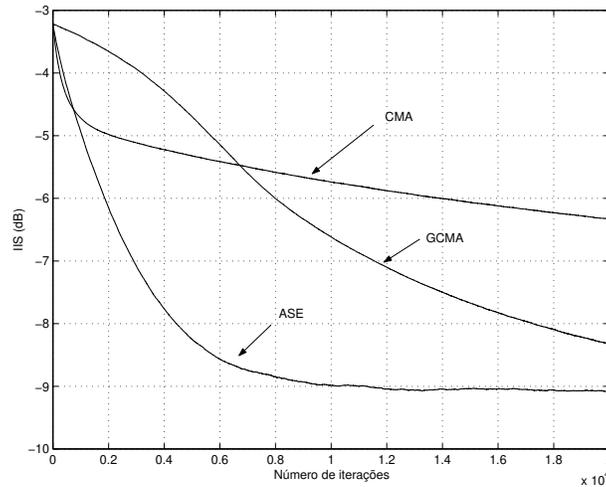


Figura 4.17: Comparação entre os algoritmos CMA: $\mu = 0.0005$, GCMA: $\mu = 0.0005$ e $\lambda = 0.8$, e ASE: $\beta_Q = 0.001$; canal: $h(z) = 0.4015 - 0.8231z^{-1} + 0.4015z^{-2}$, SNR=20dB, média 100 simulações ($\lambda = 0.999$)

É interessante observar que os passos de adaptação dos algoritmos, no caso da figura 4.15, tiveram que ser reduzidos para que houvesse convergência. Principalmente o CMA se mostrou bastante instável, divergindo para valores de passo de adaptação maiores.

4.3.2 Modulação QPSK

Já para a modulação QPSK, primeiramente foi utilizado o canal de fase não-mínima sugerido no artigo [Shalvi e Weinstein 90]. Este canal possui a seguinte resposta impulsiva:

$$\mathbf{h}(z) = -0.4 + 0.84z^{-1} + 0.336z^{-2} + 0.1344z^{-3} + 0.0538z^{-4}$$

As figuras 4.18 e 4.19 ilustram respectivamente o diagrama de zeros e a resposta em frequência do canal citado. É possível observar, portanto, que este é um canal fácil de ser equalizado já que apresenta distorções pequenas tanto em amplitude quanto em fase.

A tabela 4.2 apresenta os parâmetros utilizados para inicialização dos algoritmos simulados. Note que a inicialização dos coeficientes do equalizador foi feita colocando-

se o coeficiente não nulo na oitava posição. Isto foi feito seguindo-se a posição do termo de maior valor absoluto da resposta impulsiva do canal.

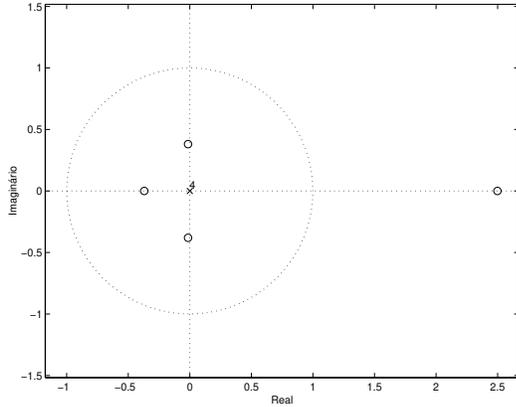


Figura 4.18: Diagrama de zeros do canal com resposta impulsiva $\mathbf{h}(z) = -0.4 + 0.84z^{-1} + 0.336z^{-2} + 0.1344z^{-3} + 0.0538z^{-4}$

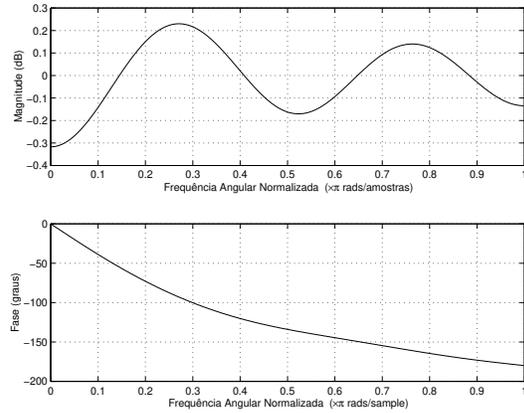


Figura 4.19: Resposta em frequência do canal com resposta impulsiva $\mathbf{h}(z) = -0.4 + 0.84z^{-1} + 0.336z^{-2} + 0.1344z^{-3} + 0.0538z^{-4}$

<i>Parâmetros de simulação para obtenção das Figuras 4.20 e 4.21</i>	
Todos	Modulação QPSK, caso com ruído: SNR= 20dB Equalizador com 20 coeficientes Inicialização:[0...0 1 (8ª posição) 0...0] ^T
CMA	sem ruído: $\mu = 0.0025$ com ruído: $\mu = 0.0005$
GCMA	sem ruído: $\mu = 0.0165$ e $\lambda = 0.7$ com ruído: $\mu = 0.0025$ e $\lambda = 0.9$ $\bar{z}_0 = 0$ $\bar{b}_0 = 0$
ASE	sem ruído: $\beta = 0.01$ com ruído: $\beta = 0.002$ \mathbf{Q}_0 = média entre os primeiros 50 vetores de entrada \mathbf{u}

Tabela 4.2: Parâmetros de inicialização para as simulações cujos resultados são mostrados nas figuras 4.10-4.12

As figuras 4.20 e 4.21 ilustram os resultados obtidos respectivamente não adicionando-se ruído e utilizando-se uma SNR=20dB. É possível observar que todos os algoritmos

mos apresentam desempenhos próximos nas duas situações consideradas. Apesar de possuírem, inicialmente, velocidades de convergência distintas, os algoritmos acabam convergindo com um mesmo número de iterações. Como os zeros do canal (figura 4.18), no interior do círculo unitário, estão próximos à origem do plano z , este resultado já era esperado.

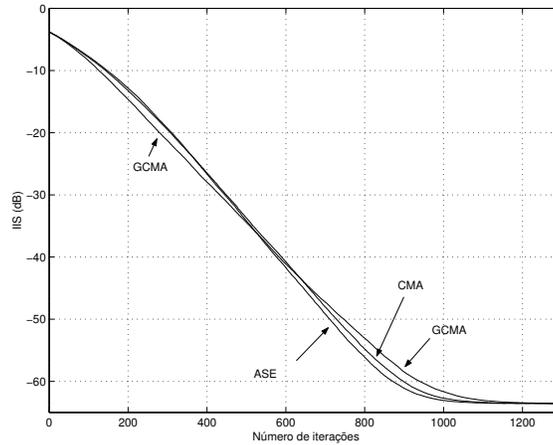


Figura 4.20: Comparação entre os algoritmos CMA: $\mu = 0.0025$, GCMA: $\mu = 0.0165$ e $\lambda = 0.7$, e ASE: $\beta_Q = 0.01$; canal: $\mathbf{h}(z) = -0.4 + 0.84z^{-1} + 0.336z^{-2} + 0.1344z^{-3} + 0.0538z^{-4}$, média 100 simulações

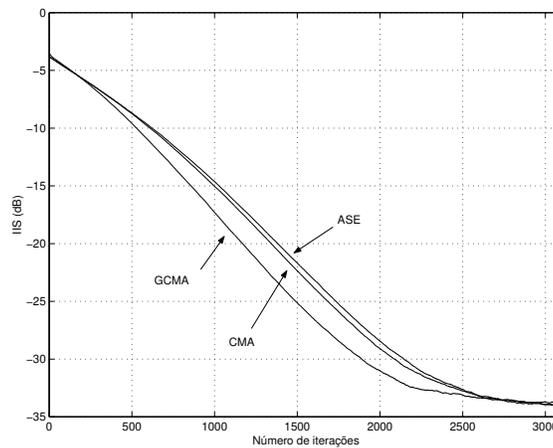


Figura 4.21: Comparação entre os algoritmos CMA: $\mu = 0.0005$, GCMA: $\mu = 0.0025$ e $\lambda = 0.9$, e ASE: $\beta_Q = 0.002$; canal: $\mathbf{h}(z) = -0.4 + 0.84z^{-1} + 0.336z^{-2} + 0.1344z^{-3} + 0.0538z^{-4}$, SNR=20dB média 100 simulações

No caso com ruído (figura 4.21), observe que o GCMA apresenta uma velocidade de convergência inicial maior do que os outros algoritmos, sendo que o ASE converge praticamente junto com o CMA. Entretanto os três chegam à convergência ao mesmo tempo. Na verdade, quanto mais próximos os zeros da resposta impulsiva do canal estão do círculo unitário, maior será a diferença de desempenho entre os algoritmos. Isto pode ser facilmente observado no próximo caso a ser tratado.

Em seguida, utilizou-se um dos canais propostos em [Proakis95], que já não apresenta a facilidade de equalização observada no canal proposto por Shalvi e Weinstein, estudado anteriormente. O canal escolhido possui a seguinte resposta impulsiva:

$$\mathbf{h}(z) = 0.407 + 0.815z^{-1} + 0.407z^{-2}$$

Seu diagrama de zeros e sua resposta em frequência são mostrados respectivamente nas figuras 4.22 e 4.23. Neste caso é possível observar que os zeros estão bem próximos do círculo unitário e portanto o canal é mais difícil de ser equalizado.

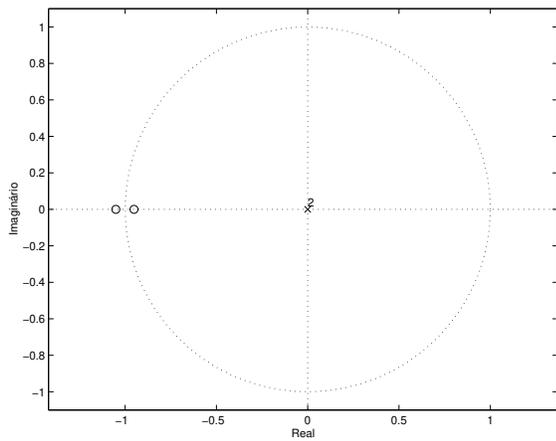


Figura 4.22: Diagrama de zeros do canal com resposta impulsiva $\mathbf{h}(z) = 0.407 + 0.815z^{-1} + 0.407z^{-2}$

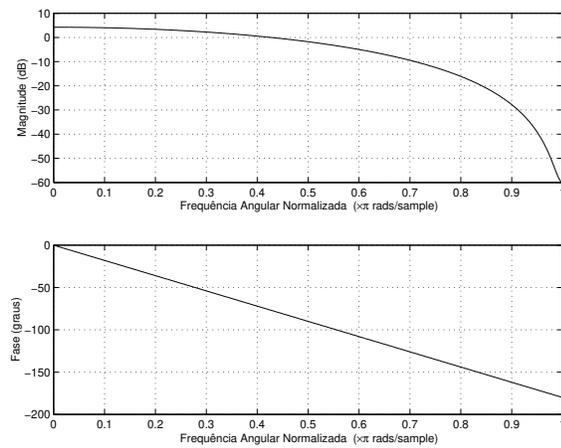


Figura 4.23: Resposta em frequência do canal com resposta impulsiva $\mathbf{h}(z) = 0.407 + 0.815z^{-1} + 0.407z^{-2}$

A tabela 4.3 apresenta os parâmetros utilizados nas inicializações dos algoritmos.

Parâmetros de simulação para obtenção das Figuras 4.20 e 4.21	
Todos	Modulação QPSK, sem ruído Equalizador com 30 coeficientes Inicialização: $[0 \dots 0 \ 1 \ (15^{\text{a}} \text{ posição}) \ 0 \dots 0]^T$
CMA	$\mu = 0.002$
GCMA	sem ruído: $\mu = 0.002$ e $\lambda = 0.9$ $\bar{z}_0 = 0$ $\bar{b}_0 = 0$
ASE	sem ruído: $\beta = 0.004$ $\mathbf{Q}_0 =$ média entre os primeiros 50 vetores de entrada \mathbf{u}

Tabela 4.3: Parâmetros de inicialização para as simulações cujos resultados são mostrados nas figuras 4.10-4.12

Os resultados obtidos são mostrados na figura 4.24. Observe que neste caso, o ASE apresenta um desempenho significativamente superior ao dos outros algoritmos, o que se deve à presença da matriz \mathbf{Q} . Note que o sinal de entrada do equalizador aqui está bastante correlacionado, o que prejudica o desempenho do CMA e do GCMA.

Como já havia sido observado, e pode ser confirmado aqui, conforme os zeros da resposta impulsiva do canal se aproximam do círculo unitário, mais o desempenho do GCMA se aproxima do CMA. Alterando-se um pouco o fator de esquecimento ou o passo de adaptação, foi observado que o algoritmo pode apresentar inclusive um desempenho pior do que o CMA.

Podemos concluir portanto que a situação na qual o GCMA apresenta melhor desempenho, se comparado ao CMA, acontece quando os zeros da resposta impulsiva do canal não estão nem muito perto da origem nem muito perto do círculo unitário no plano z . Conforme estes zeros se aproximam da origem, o sinal de entrada do equalizador se torna mais descorrelacionado e, portanto, é esperado que o desempenho de todos os algoritmos seja semelhante. Entretanto, conforme este sinal se torna mais correlacionado, ou seja, os zeros se aproximam do círculo unitário, o GCMA perde desempenho, se aproximando cada vez mais do CMA e podendo até ficar pior do que este. Na próxima seção estudaremos mais detalhadamente o desempenho do GCMA.

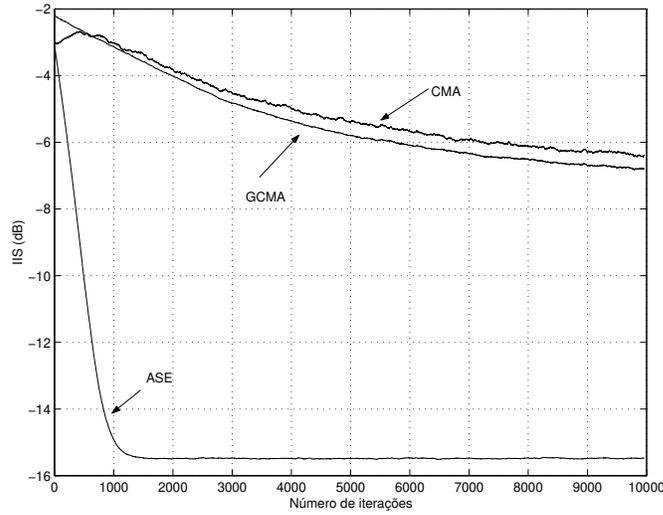


Figura 4.24: Comparação entre os algoritmos CMA: $\mu = 0.002$, GCMA: $\mu = 0.002$ e $\lambda = 0.9$, e ASE: $\beta_Q = 0.004$; canal: $\mathbf{h}(z) = 0.407 + 0.815z^{-1} + 0.407z^{-2}$, média 100 simulações

4.4 Estudo do GCMA

Para analisar o comportamento do GCMA, partimos novamente da equação de adaptação (4.5), deduzida, de forma alternativa, a partir da função custo “instantânea”, dada em (2.115). Em (4.5), o segundo termo do lado direito da equação poderia ser separado em dois termos, como mostrado abaixo:

$$\left(\underbrace{|y_n|^2 y(n) \mathbf{u}_n^* - R^{y'} \mathbf{b}}_{\text{termo I}} + \underbrace{|y_n|^2 \mathbf{b} - y_n \mathbf{u}_n^* \bar{z}}_{\text{termo II}} \right) \quad (4.14)$$

O termo I é semelhante aos gradientes do CMA, ASE e GCMA2, tendo como principais diferenças:

-a multiplicação por $R^{y'}$, enquanto que nos outros algoritmos temos R_2 e R^y

-a existência de um operador esperança em \mathbf{b} , enquanto que os outros gradientes utilizam o valor instantâneo do termo $y_n \mathbf{u}_n^*$

Na verdade, em média, este termo é equivalente ao vetor gradiente da função custo do método super-exponencial, ou seja, é equivalente ao gradiente exato do GCMA2 (antes da aproximação estocástica):

$$\begin{aligned}
 \text{Termo } I = TI &= |y_n|^2 y_n \mathbf{u}_n^* - E[y_n \mathbf{u}_n^*] \frac{|y_n|^4}{E[|y_n|^2]} \\
 E[TI] &= E[|y_n|^2 y_n \mathbf{u}_n^*] - E[y_n \mathbf{u}_n^*] \frac{E[|y_n|^4]}{E[|y_n|^2]} \\
 E[TI] &= E[|y_n|^2 y_n \mathbf{u}_n^*] - R^y E[y_n \mathbf{u}_n^*]
 \end{aligned} \tag{4.15}$$

Apesar da semelhança, através de simulações foi possível constatar que, na verdade, os gradientes dos algoritmos citados acima são bastante diferentes.

As figuras 4.25 e 4.26 ilustram o que foi dito. Considerando agora um algoritmo expresso por:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu \left(|y_n|^2 y_n \mathbf{u}_n^* - R^{y'} \mathbf{b} \right) \tag{4.16}$$

que será chamado $GCMA_I$, a ser comparado com o CMA e o GCMA2. Para o caso do canal $h(z) = 1 - 0.6z^{-1}$, um equalizador com dois coeficientes e sem ruído, a figura 4.25 mostra como, utilizando-se os melhores parâmetros possíveis em termos de velocidade de convergência, ou seja, os parâmetros para os quais os algoritmos convergem mais rapidamente, os três algoritmos acabam tendo desempenhos bastante semelhantes. Entretanto, o $GCMA_I$ já se mostra um pouco mais lento do que os outros dois.

Consideramos então um canal menos amigável, de fase não-mínima e com resposta impulsiva dada por $h(z) = 0.3482 - 0.8704z^{-1} + 0.3482z^{-2}$. Temos que com um equalizador de 15 coeficientes, sem ruído e parâmetros escolhidos de forma a que o algoritmo GCMA2 convergisse (como foi comentado anteriormente, ele é bastante instável), a situação se altera um pouco. A figura 4.26 ilustra os resultados obtidos. Observe que o algoritmo em questão, $GCMA_I$, acaba tendo um desempenho bastante ruim, com uma IIS residual muito alta. Este mesmo comportamento foi observado para os outros canais, com zeros em 0.2 e 5 e com zeros em 0.8 e 1.25.

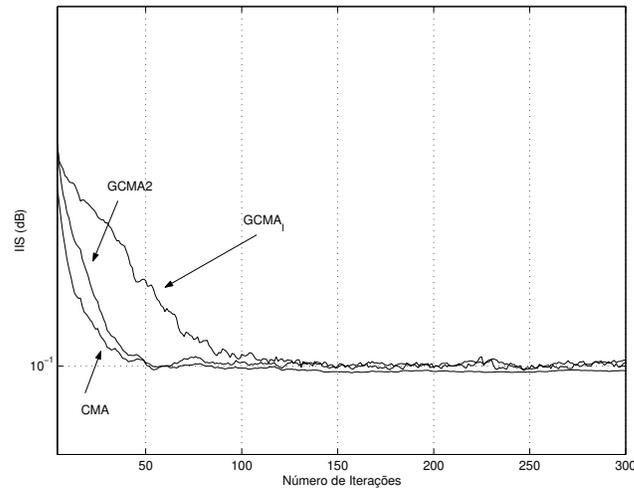


Figura 4.25: Comparação entre os algoritmos CMA, $\mu = 0.035$, GCMA2, $\mu = 0.035$ e $\lambda = 0.7$ e $GCMA_I$, $\mu = 0.035$ e $\lambda = 0.7$; canal: $h(z) = 1 + 0.6z^{-1}$; média 20 simulações

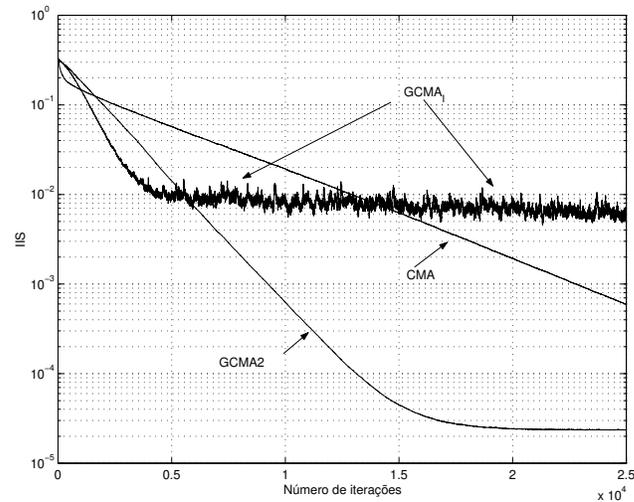


Figura 4.26: Comparação entre os algoritmos CMA, $\mu = 0.0018$, GCMA2, $\mu = 0.0018$ e $\lambda = 0.8$ e $GCMA_I$, $\mu = 0.0018$ e $\lambda = 0.8$; canal: $h(z) = 0.3482 - 0.8704z^{-1} + 0.3482z^{-2}$; média 20 simulações

Já o termo II não é semelhante a nenhum outro termo existente nos algoritmos vistos. A soma deste termo faz com que o algoritmo melhore muito de desempenho, como já foi visto nas simulações anteriores e como pode ser concluído do estudo do termo I acima, que sozinho resulta em um desempenho bastante pobre. Note que o termo é composto por dois elementos que diferem com relação ao operador esperança.

Assim, é facilmente visto que, em média, este termo acaba se anulando:

$$\begin{aligned} \text{Termo II} = TII &= |y_n|^2 E[y_n \mathbf{u}_n^*] - y_n \mathbf{u}_n^* E[|y_n|^2] \\ E[TII] &= E[|y_n|^2] E[y_n \mathbf{u}_n^*] - E[y_n \mathbf{u}_n^*] E[|y_n|^2] = 0 \end{aligned} \quad (4.17)$$

Simulando a evolução deste termo conforme o algoritmo converge, obtivemos os resultados mostrados nas figuras 4.27-4.29, onde foram utilizados respectivamente os mesmos canais que resultaram nas figuras 4.10-4.12. Note que este termo II, na verdade, é um vetor com o mesmo número de elementos que o número de coeficientes do equalizador. Nas figuras citadas, é mostrada a evolução do primeiro elemento deste vetor conforme o número de iterações aumenta.

Na figura 4.10, vimos que o GCMA possui um desempenho bastante próximo do CMA, o que poderia levar a conclusão equivocada de que o termo II não tivesse muita influência na convergência do algoritmo, já que, como mostra (4.17), este se anula em média. Através da figura 4.27, vê-se que o termo II fica praticamente oscilando em torno de um mesmo valor, próximo de zero. Entretanto, sem este termo ($GCMA_I$), o algoritmo nem chega a convergir para os parâmetros considerados.

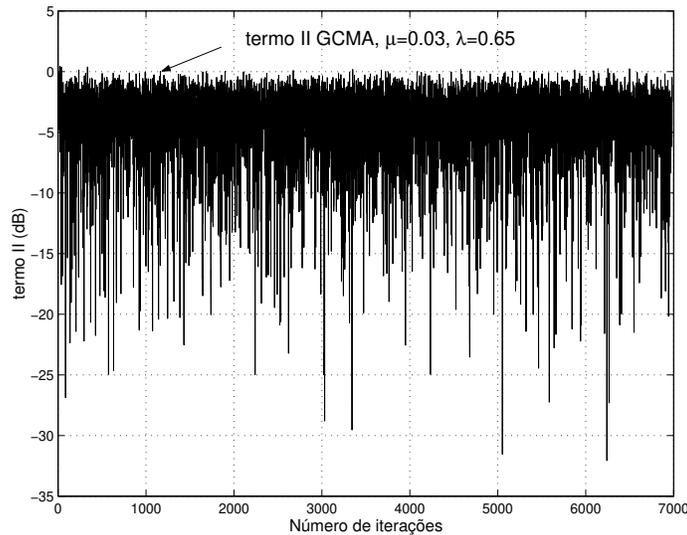


Figura 4.27: Evolução do termo II em função do número de iterações, $\mu = 0.03$, $\lambda = 0.65$; canal: $h(z) = 0.1856 - 0.9650z^{-1} + 0.1856z^{-2}$

Conforme alteramos o canal, a evolução do termo II também se altera. Observando as figuras 4.11 e 4.12, vemos que o GCMA possui um desempenho bem melhor que o CMA. É esperado portanto que o termo II tenha uma influência maior na convergência do algoritmo, o que é confirmado através dos resultados mostrados nas figuras 4.28 e 4.29.

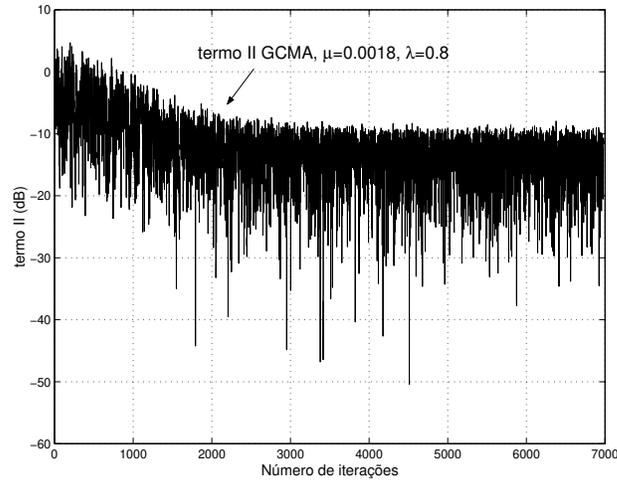


Figura 4.28: Evolução do termo II em função do número de iterações, $\mu = 0.0018$, $\lambda = 0.8$; canal: $h(z) = 0.3482 - 0.8704z^{-1} + 0.3482z^{-2}$

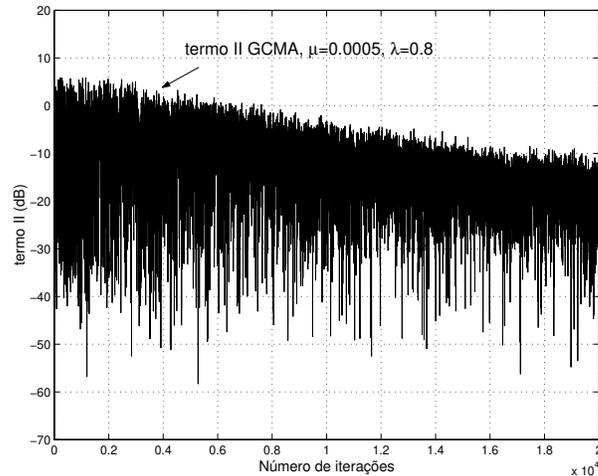


Figura 4.29: Evolução do termo II em função do número de iterações, $\mu = 0.0005$, $\lambda = 0.8$; canal: $h(z) = 0.4015 - 0.8231z^{-1} + 0.4015z^{-2}$

Comparando ainda o gradiente do GCMA com o do CMA, podemos escrever a seguinte variável:

$$\Delta_{grad} = \nabla CMA - \nabla GCMA \quad (4.18)$$

onde ∇CMA é o gradiente do algoritmo CMA que pode ser visto em (4.3) e $\nabla GCMA$ é o gradiente do GCMA visto em (4.5). Temos então que

$$\Delta_{grad} = \mathbf{b}R^{y'} - |y_n|^2\mathbf{b} + y_n\mathbf{u}_n^*\bar{z} - R_2y_n\mathbf{u}_n^* \quad (4.19)$$

ou ainda,

$$\Delta_{grad} = y_n\mathbf{u}_n^*(\bar{z} - R_2) + \mathbf{b} \left(R^{y'} - |y_n|^2 \right) \quad (4.20)$$

Calculando a esperança de Δ_{grad} teremos

$$\begin{aligned} E[\Delta_{grad}] &= E[y_n\mathbf{u}_n^*(\bar{z} - R_2) + \mathbf{b} \left(\underbrace{E[R^{y'}]}_{R^y} - E[|y_n|^2] \right)] \\ &= \mathbf{b}(\bar{z} - R_2) + \mathbf{b}(R^y - \bar{z}) \\ &= \mathbf{b}(R^y - R_2) \end{aligned} \quad (4.21)$$

Podemos concluir então que a média da diferença entre o vetor gradiente do CMA e do GCMA só tende a zero quando valer a relação $R^y = R_2$, que é exatamente o que diferencia o vetor gradiente da função custo de Shalvi-Weinstein do vetor gradiente da função custo de Godard. Assim mostra-se que, na média, os vetores gradiente do CMA e do GCMA são diferentes. Mais uma vez, a diferença entre estes dois termos leva o GCMA a convergir para vales de soluções enquanto o CMA converge para pontos estacionários. Da mesma forma como o ASE, que por possuir a constante R_2 , converge para os mesmos pontos críticos que o CMA, se a igualdade $R^y = R_2$ fosse válida, o GCMA também convergiria para estes mesmos pontos estacionários, não possuindo mais vales de soluções.

4.4.1 Considerações sobre o efeito do fator de esquecimento λ

O fator de esquecimento λ é utilizado para o cálculo das médias temporais necessárias no GCMA. É esperado que, quanto menor seja o valor deste parâmetro, mais imprecisos sejam os cálculos de $\bar{z} = E[|y(n)|^2]$ e de $\mathbf{b} = E[y^*(n)\mathbf{u}(n)]$, ou seja, maior será o ruído de gradiente envolvido no processo de adaptação. Mas no que isto influi exatamente no desempenho do algoritmo?

O primeiro resultado, e talvez também o mais preocupante, é que o valor do parâmetro λ influi no número de mínimos locais apresentado pelo algoritmo. Utilizando novamente o canal com resposta impulsiva $h(z) = 1 - 0.6z^{-1}$, modulação BPSK, sem adição de ruído e um equalizador com dois coeficientes, foi simulada a convergência do algoritmo quando inicializado em diversos pontos, com $\lambda = 0.95$ e com $\lambda = 0.6$. Os resultados são mostrados respectivamente nas figuras 4.30 e 4.31.

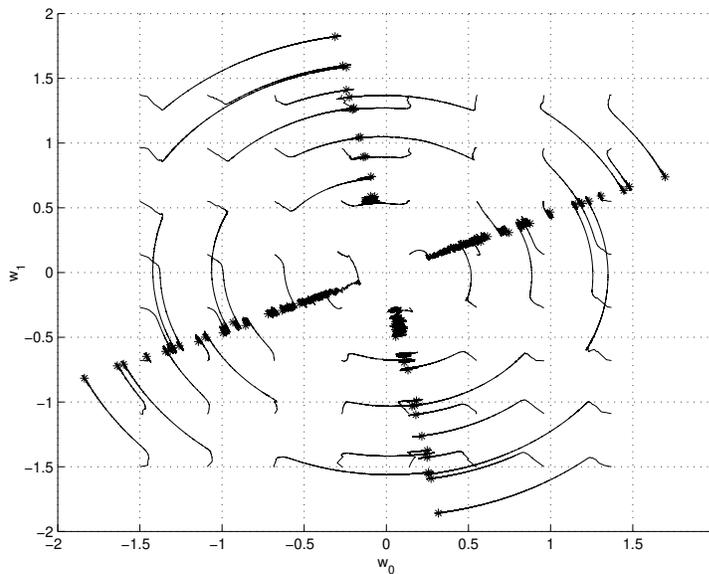


Figura 4.30: Convergência do algoritmo GCMA, canal $h(z) = 1 - 0.6z^{-1}$, parâmetros: $\mu = 0.0025$ e $\lambda = 0.97$

Observe então como, no caso em que $\lambda = 0.6$, surge um vale de mínimos locais espúrio. Chamamos de “espúrio” pois este não é um mínimo local proveniente direta-

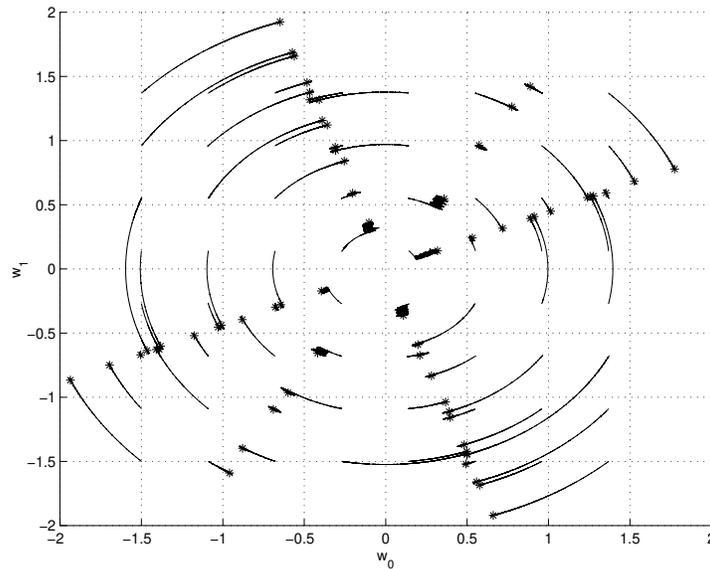


Figura 4.31: Convergência do algoritmo GCMA, canal $h(z) = 1 - 0.6z^{-1}$, parâmetros: $\mu = 0.0025$ e $\lambda = 0.6$

mente da função custo. Note que ele não aparece na superfície de erro mostrada na figura 3.5. Entretanto, neste ponto, é interessante relembrarmos como o GCMA foi deduzido. Na verdade, o algoritmo é proveniente do que podemos chamar de uma função custo “instantânea”. Esta função pode realmente possuir um vale de mínimos espúrios dependente do parâmetro λ .

Este vale de mínimos espúrio é altamente indesejável já que aumenta o risco do algoritmo ficar preso em mínimos locais. Vale observar que, para o canal $h(z) = 1 - 0.6z^{-1}$ utilizado aqui, o olho na saída do equalizador só pode ser considerado aberto caso haja convergência para o mínimo global. Quando existe convergência para algum mínimo local, a IIS residual fica muito alta, sendo que no mínimo local espúrio, esta é ainda mais alta do que no mínimo local proveniente do critério utilizado.

Já para canais de fase não mínima e equalizadores com um número maior de coeficientes, tal análise se torna bastante complexa. Nas simulações realizadas até agora, com inicialização *center spike*, não foi observada esta convergência para mínimos locais espúrios em nenhum momento, já que o CMA e o ASE não possuem estes mínimos e todos os algoritmos convergiram sempre para a mesma IIS residual.

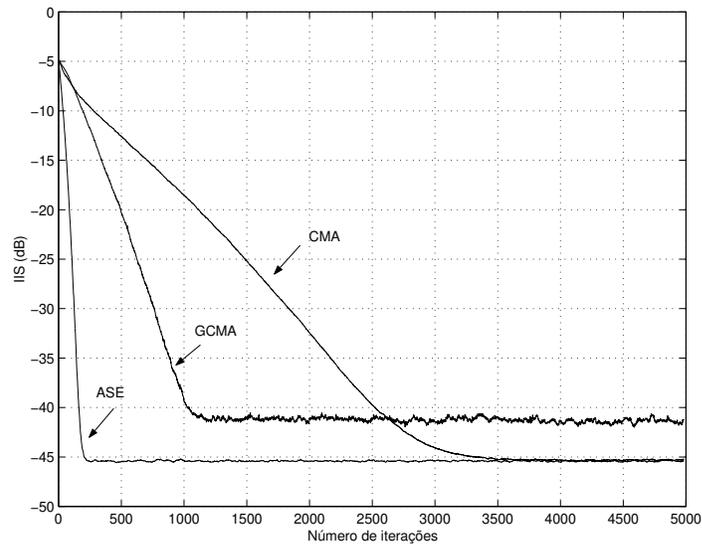


Figura 4.32: Comparação entre os algoritmos CMA, $\mu = 0.02$, GCMA, $\mu = 0.02$ e $\lambda = 0.55$, ASE, $\beta_Q = 0.04$ canal: $h(z) = 0.3482 - 0.8704z^{-1} + 0.3482z^{-2}$; média 100 simulações

A figura 4.32 ilustra um caso em que, como o valor de λ utilizado é muito baixo, o algoritmo GCMA acaba convergindo para um valor de IIS residual mais alto do que o ASE ou o CMA. Entretanto, comparando o valor do conjunto de coeficientes do equalizador após a convergência, observou-se que, a menos de um ganho constante (que pode ser facilmente corrigido utilizando-se um CAG), os valores obtidos foram muito próximos para os três algoritmos. A diferença existente era da ordem de 10^{-3} , perfeitamente factível já que estamos considerando algoritmos recursivos obtidos através de aproximações estocásticas. Assim, provavelmente este comportamento se deve mesmo ao aumento do ruído de gradiente, por causa do baixo valor de λ .

Outro fator indesejável observado durante as simulações foi que, dependendo dos valores de passo de adaptação μ e fator de esquecimento λ , o algoritmo GCMA apresenta um comportamento oscilatório. A figura 4.33 ilustra este caso. Foi notado, entretanto, que este desempenho indesejável se deve mais ao passo de adaptação, que já está muito próximo do valor limite para o qual o algoritmo ainda converge, do que ao fator de esquecimento. Aumentando-se o valor deste último, o algoritmo deixa de convergir. Escolhendo-se corretamente o valor de μ , como na figura 4.11 por exemplo, este comportamento não foi observado, mesmo utilizando-se valores baixos para λ .

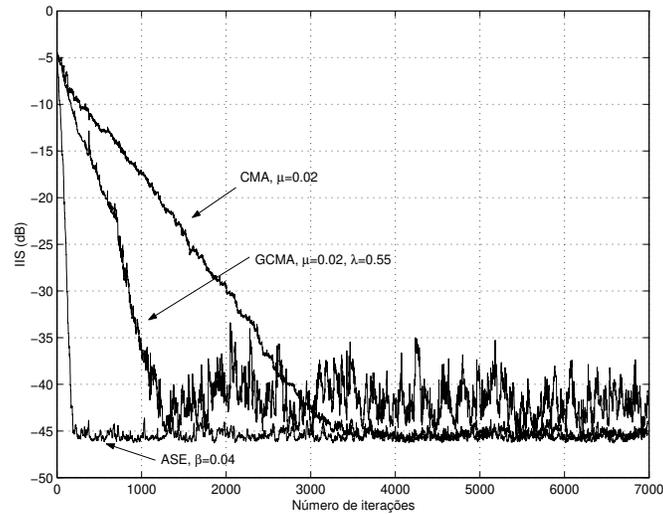


Figura 4.33: Comparação entre os algoritmos CMA, $\mu = 0.02$, GCMA, $\mu = 0.02$ e $\lambda = 0.55$, ASE, $\beta_Q = 0.04$ canal: $h(z) = 0.3482 - 0.8704z^{-1} + 0.3482z^{-2}$

4.4.2 Alterando a inicialização

Até agora, a inicialização de todos os algoritmos foi feita utilizando-se o método *center spike*, com o coeficiente não nulo sempre com módulo igual a 1. Segundo a literatura, esta é a melhor inicialização para o CMA e, portanto, também para o ASE, devido à equivalência entre os dois, já discutida anteriormente.

No entanto, observando-se a forma como o GCMA converge, podemos esperar que, inicializando os coeficientes do equalizador mais próximos da reta determinada pelo vale de soluções, o algoritmo convirja mais rápido para a solução desejada, pois a distância que ele precisa percorrer para atingir a solução será menor. Da mesma forma, se inicializássemos o CMA ou o ASE próximos de seus pontos críticos, estes também convergiriam mais rapidamente. Entretanto, a localização destes pontos não é conhecida. Observando-se mais uma vez a figura 4.30, vê-se que os vales de soluções cruzam a origem do plano determinado pelos coeficientes do equalizador. Assim, quanto mais próximo da origem o algoritmo GCMA for inicializado, mais próximo ele estará dos vales de soluções e portanto, podemos esperar que ele convirja mais rapidamente. Note que a função custo do GCMA não é definida quando todos os coeficientes do equalizador forem nulos pois, neste ponto, seu denominador $E^2 [|y|^2]$ também será nulo. Podemos, no entanto, inicializar os coeficientes com valores menores do que 1, mas significativos

o suficiente para que o algoritmo funcione corretamente.

Assim, mantendo a inicialização como *center spike* mas fazendo o coeficiente não nulo igual a -0.3, em vez de -1, simulamos o algoritmo GCMA e comparamos com o ASE inicializado com -1. Foram utilizados os canais com zeros em 0.2 e 5 e em 0.5 e 2, modulação BPSK, sem ruído. Os resultados obtidos são mostrados respectivamente nas figuras 4.34 e 4.35.

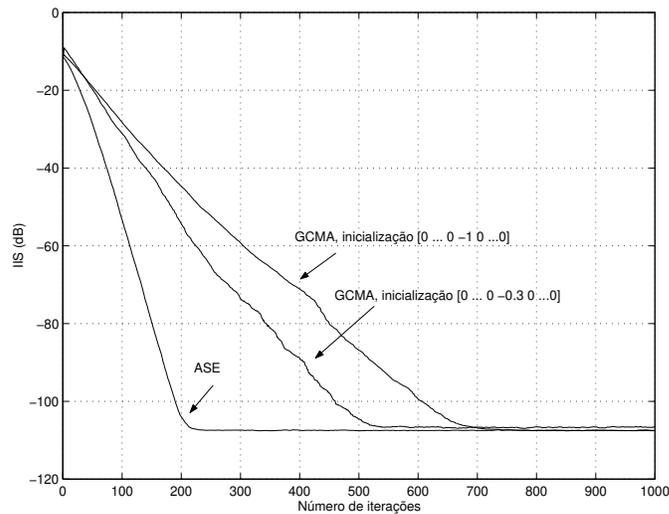


Figura 4.34: Comparação entre os algoritmos GCMA inicializado com $[0 \dots 0 - 1 0 \dots 0]$: $\mu = 0.03$ e $\lambda = 0.65$, inicializando com $[0 \dots 0 - 0.3 0 \dots 0]$: $\mu = 0.0035$ e $\lambda = 0.7$ ASE, $\beta_Q = 0.06$ canal: $h(z) = 0.1856 - 0.9650z^{-1} + 0.1856z^{-2}$; média 100 simulações

É possível observar, portanto, que existe um ganho em termos de velocidade de convergência, apesar do algoritmo continuar não atingindo o desempenho do ASE. No caso da figura 4.34 vemos um ganho de um pouco menos de 200 iterações e, no caso do canal mostrado em 4.35, houve um ganho de 2000 iterações. Para o canal com zeros em 0.8 e 1.25, não houve ganho nem perda de desempenho.

A desvantagem, entretanto, deste método, é que o algoritmo se torna um pouco mais instável, sendo necessário um valor de passo de adaptação bem menor (da ordem de 0.1) do que o utilizado quando a inicialização é feita com o coeficiente não nulo igual a 1. Utilizando-se valores menores do que 0.3 para o coeficiente não nulo, não foi observada nenhuma melhora maior no desempenho, além do fato de que quanto menor for este valor, menor deve ser o passo de adaptação utilizado.

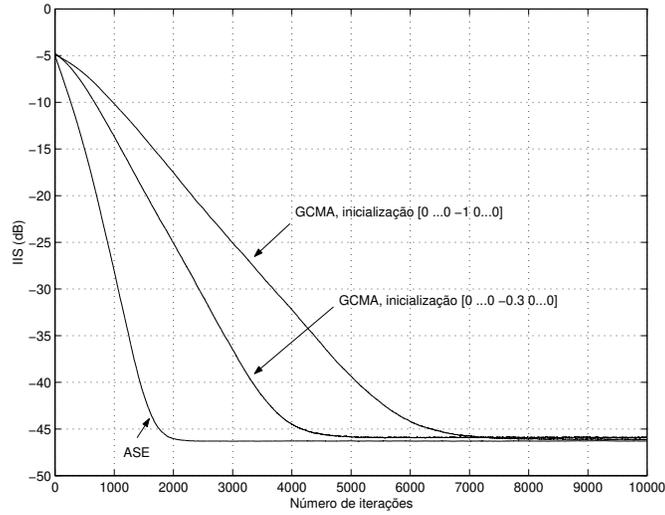


Figura 4.35: Comparação entre os algoritmos GCMA inicializado com $[0 \dots 0 - 1 \ 0 \dots 0]$: $\mu = 0.0018$ e $\lambda = 0.8$, inicializando com $[0 \dots 0 - 0.3 \ 0 \dots 0]$: $\mu = 0.00025$ e $\lambda = 0.8$, ASE, $\beta_Q = 0.0036$ canal: $h(z) = 0.3482 - 0.8704z^{-1} + 0.3482z^{-2}$; média 100 simulações

4.4.3 Alterando o passo de adaptação

Como vimos, o melhor desempenho do algoritmo ASE deve-se à multiplicação do termo correspondente ao gradiente pela matriz \mathbf{Q} . Este comportamento é facilmente explicado já que a matriz representa uma operação de pré-branqueamento do sinal de entrada do equalizador. Incluindo esta matriz no passo de adaptação do algoritmo CMA, é fácil observar por (4.3) e (4.4) que este novo algoritmo será igual ao ASE, já que o fator δ , existente neste último, pode ser incorporado ao passo do algoritmo, β . Entretanto, qual seria o efeito de realizar esta alteração no algoritmo GCMA?

A equação deste novo algoritmo, considerando o caso real, será:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{\mu}{\bar{z}^2} \mathbf{Q}_n \left(y_n^3 \mathbf{u}_n - R^{y'} \mathbf{b} + y_n^2 \mathbf{b} - y_n \mathbf{u}_n \bar{z} \right) \quad (4.22)$$

onde a matriz \mathbf{Q}_n será calculada da mesma forma que no algoritmo ASE.

As figuras 4.36 e 4.37 mostram os resultados obtidos simulando-se este novo algoritmo e comparando com o ASE. Foram simulados os canais com zeros em 0.5 e 2, e 0.8 e 1.25 respectivamente, sem ruído e a modulação utilizada foi BPSK. Como era es-

perado, o algoritmo converge bem mais rapidamente do que no caso em que não existe a matriz, chegando a ter um desempenho bastante próximo daquele obtido pelo ASE. No entanto, vale ressaltar que seu custo computacional também aumenta bastante.

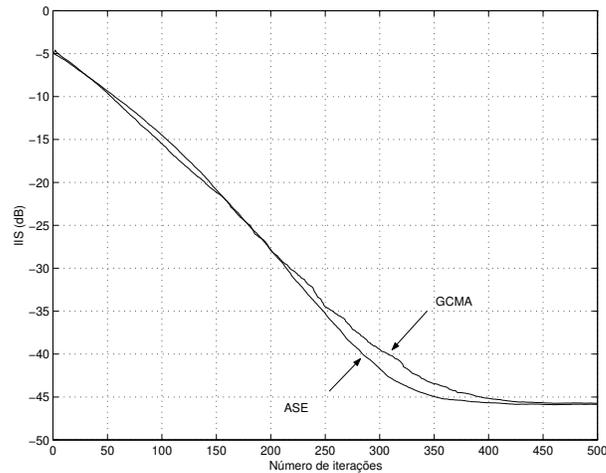


Figura 4.36: Comparação entre os algoritmos ASE, $\beta = 0.02$ e GCMA com a matriz \mathbf{Q} , $\mu = 0.02$ e $\lambda = 0.7$, canal: $h(z) = 0.3482 - 0.8704z^{-1} + 0.3482z^{-2}$; média 100 simulações

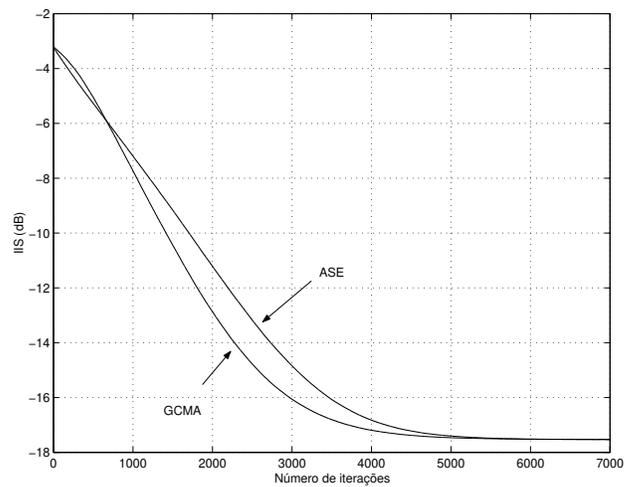


Figura 4.37: Comparação entre os algoritmos ASE, $\mu = 0.001$ e GCMA com a matriz \mathbf{Q} , $\mu = 0.001$ e $\lambda = 0.65$, canal: $h(z) = 0.4015 - 0.8231z^{-1} + 0.4015z^{-2}$; média 100 simulações

4.5 Conclusão

Neste capítulo, analisamos o desempenho dos algoritmos ASE, CMA, GCMA e GCMA2, através de simulações para diversos canais, modulações e mesmo com a adição de ruído. Vimos que o algoritmo ASE possui um desempenho muito melhor que os outros para a grande maioria dos casos estudados, principalmente quando os zeros da resposta impulsiva do canal estão próximos ao círculo unitário no plano z . Tal comportamento se deve à presença da matriz \mathbf{Q} , que pode ser vista como uma operação de pré-branqueamento do sinal de entrada do equalizador.

Já com relação ao GCMA, vimos que seu desempenho pode ser bem melhor do que o obtido pelo CMA, possuindo um custo computacional bem menor do que o ASE. Entretanto, ele perde desempenho conforme os zeros do canal se aproximam do círculo unitário, situação na qual seu desempenho se aproxima do obtido pelo CMA. Analisando mais detalhadamente seu algoritmo, vimos que ele pode ser visto como a soma de dois termos, um semelhante ao gradiente dos demais algoritmos e outro que surge da forma como foi deduzido, utilizando-se uma aproximação estocástica pouco usual. Apesar deste segundo termo se anular em média, sua presença é essencial para o bom desempenho do algoritmo.

Através de simulações, vimos o efeito do fator de esquecimento λ no desempenho do algoritmo, concluindo que este sempre deve estar o mais próximo possível de 1. Foi mostrado também que a inicialização com *center-spike* tendo o coeficiente não nulo com módulo igual a 1 talvez não seja a melhor, já que inicializando o algoritmo mais próximo da configuração trivial (todos os coeficientes nulos), ele também estará mais próximo dos vales de soluções e poderá convergir mais rapidamente.

Poderíamos dizer portanto que o bom desempenho do algoritmo não se deve exatamente à flexibilização do raio da constelação mas sim à forma como este foi obtido. A possibilidade de convergência para um vale de soluções, pode trazer algumas vantagens como o fato do algoritmo convergir mais rapidamente quando inicializado mais próximo da origem no espaço de parâmetros do equalizador, mas vimos que por exemplo o GCMA2, obtido através do método de aproximação estocástica clássico e mesmo o $GCMA_I$, apesar de também convergirem para vales de soluções, apresentaram desem-

penhos ruins, se mostrando instáveis e divergindo facilmente. Assim, a estabilidade do algoritmo e seu bom desempenho se devem realmente à forma como foi deduzido.

Capítulo 5

Conclusões e Perspectivas

A análise de critérios e algoritmos autodidatas deve ser sempre muito cuidadosa no sentido de observar semelhanças e diferenças com relação a critérios e algoritmos já existentes. Nos últimos anos temos visto um número muito grande de publicações sugerindo novos critérios e algoritmos, e alguns, se analisados mais detalhadamente, revelam-se equivalentes, ou mesmo iguais, aos já existentes.

A equivalência entre os critérios do módulo constante e Shalvi-Weinstein com restrição de potência, garantindo que o sinal na saída do equalizador terá a mesma potência do sinal inicialmente transmitido, já é um resultado consolidado na literatura. Mostramos aqui que o critério do módulo constante generalizado é igual ao de Shalvi-Weinstein sem restrição de potência, a menos de uma constante. Concluímos então que, em termos de critério, não há nele nenhuma característica propriamente nova. Observou-se, entretanto, que critérios sem a restrição de potência ainda não haviam sido estudados mais detalhadamente na literatura. Em sua comparação com o critério CM, viu-se que o critério GCM pode ser entendido como considerando uma flexibilização do raio da constelação.

Por outro lado, a forma como o algoritmo GCMA foi deduzido é bastante interessante e traz novos elementos à discussão. Basicamente, embora decorrentes de um mesmo critério, a diferença entre os algoritmos super-exponencial e GCMA está na forma como foram deduzidos, considerando-se também que o primeiro utiliza a restrição de potência,

enquanto o segundo não utiliza esta restrição. Vimos, através de simulações, que o GCMA apresenta um bom desempenho, com velocidade de convergência superior ao CMA, apresentando um custo computacional menor que o ASE. Intuitivamente, podemos dizer que, enquanto a direção do gradiente no algoritmo ASE é rotacionada pela matriz \mathbf{Q} , de forma a apontar para a solução procurada, no GCMA temos a adição de um termo, que também ajuda a direcionar o vetor gradiente na direção desejada. Utilizando-se diversos canais e modulações, observou-se que o GCMA converge para um intervalo maior de valores de passos de adaptação do que o CMA. Entretanto, ele perde desempenho conforme os zeros da resposta impulsiva do canal se aproximam do círculo unitário no plano z . Foi visto também, através da comparação do GCMA com o GCMA2, que seu bom desempenho não é exatamente proveniente da flexibilização do raio da constelação, já que esta é uma característica dos dois algoritmos citados, mas sim da forma como foi deduzido, utilizando-se uma aproximação estocástica pouco usual.

Tendo sido vistos então os principais resultados deste trabalho, podemos resumir as contribuições desta tese da seguinte forma:

- Esclarecimento de algumas questões não muito claras na literatura, como a função da restrição de potência do critério de Shalvi-Weinstein na demonstração de sua equivalência com o critério CM.
- Proposta alternativa de se enxergar a relação entre o critério CM e Shalvi-Weinstein, através dos vetores gradiente.
- Inclusão do critério GCM, que não possui restrição de potência, na análise das equivalências entre os critérios.
- Estudo da consequência do método pouco usual de aproximação estocástica e sua consequência na dedução do algoritmo GCMA.
- Comparação dos algoritmos em termos de velocidade de convergência e desajuste
- Considerações sobre os efeitos dos parâmetros envolvidos no algoritmo GCMA sobre o seu desempenho

Da análise realizada para o algoritmo GCMA, podemos propor, como continuação deste trabalho, uma análise mais detalhada quanto ao intervalo de valores que o passo de adaptação pode assumir e mesmo uma análise quanto ao passo de adaptação ótimo para o algoritmo.

Uma continuação natural deste trabalho seria a extensão do estudo de equivalências entre critérios e algoritmos para o caso multi-usuário, que tem sido bastante estudado ultimamente e está diretamente relacionado à busca de uma solução para duas desvantagens dos algoritmos baseados em estatísticas de ordem superior estudados aqui: convergência lenta e a existência de mínimos locais indesejáveis. Esta nova forma de equalização vem, senão para resolver, pelo menos para procurar diminuir estes problemas.

Apêndice

A demonstração da veracidade de (2.68) é baseada na utilização de uma outra variável \mathbf{q} definida como

$$\mathbf{q}(t) = \mathbf{s} + t\mathbf{r} \quad (\text{A.1})$$

onde t é um escalar real qualquer, de forma que \mathbf{q} varrerá T conforme \mathbf{r} varra T . Introdz-se então esta nova função:

$$D_{2p} = [f_{2p}(\mathbf{q}(t))]^{2p} = \frac{\sum_{k=0}^{\infty} |q_k(t)|^{2p}}{(\sum_{k=0}^{\infty} |q_k(t)|^2)^p} \quad (\text{A.2})$$

e supõe-se que \mathbf{s} tenha norma ℓ_2 unitária.

Temos então que $\mathbf{s}_0 \in T$ é um ponto estacionário de D_{2p} se e somente se a derivada direcional se anular em \mathbf{s}_0 para todos os vetores direcionais $\mathbf{r} \in T$. Matematicamente, podemos escrever como:

$$\left. \frac{dD_{2p}(\mathbf{q}(t))}{dt} \right|_{t=0} = 0, \quad \text{para todo } \mathbf{r} \in T \quad (\text{A.3})$$

Calculando-se a derivada direcional e igualando a zero, conclui-se que a seguinte expressão deve ser satisfeita:

$$\mathbf{s}^{\odot(2p-1)} \perp \mathbf{r} - \langle \mathbf{s}, \mathbf{r} \rangle \mathbf{s}, \quad \text{para todo } \mathbf{r} \in T \quad (\text{A.4})$$

onde \perp denota ortogonalidade e $\langle \cdot, \cdot \rangle$ denota o produto interno em ℓ_2 , ou seja,

$$\langle \mathbf{s}, \mathbf{r} \rangle = \sum_k s_k^* r_k \quad (\text{A.5})$$

Introduz-se, portanto, o complemento ortogonal do *attainable set*, denominado T^\perp . Podemos também decompor T em um subespaço unidimensional, colinear com $\mathbf{s}_0 \in T$ e seu complemento ortogonal:

$$\begin{aligned} T_s &\triangleq \{ \mathbf{x} \in T : \mathbf{x} = \alpha \mathbf{s} \text{ para alguma escalar } \alpha \} \\ T_s^\perp &\triangleq \{ \mathbf{x} \in T : \mathbf{x} \perp \mathbf{s} \} \end{aligned} \quad (\text{A.6})$$

Como $\mathbf{r} \in T$, o termo $\mathbf{r} - \langle \mathbf{s}, \mathbf{r} \rangle \mathbf{s}$ também pertence a T mas é ortogonal a \mathbf{s} pois, lembrando que este possui norma ℓ_2 unitária,

$$\langle \mathbf{s}, \mathbf{r} - \langle \mathbf{s}, \mathbf{r} \rangle \mathbf{s} \rangle = \sum_k s_k^* r_k - \langle \mathbf{s}, \mathbf{r} \rangle |\mathbf{s}|^2 = 0 \quad (\text{A.7})$$

Portanto, $\mathbf{r} - \langle \mathbf{s}, \mathbf{r} \rangle \mathbf{s} \in T_s^\perp$ para todo $\mathbf{r} \in T$. Assim, (A.4) é válida se e somente se a projeção de $\mathbf{s}^{\odot(2p-1)}$ em T_s^\perp for nula, o que, segundo (A.6) significa que a projeção de $\mathbf{s}^{\odot(2p-1)}$ ao longo de T se reduz à sua projeção ao longo de T_s , ou seja,

$$P_T(\mathbf{s}^{\odot(2p-1)}) = \alpha \mathbf{s}, \text{ para alguma escalar } \alpha \quad (\text{A.8})$$

Para completar a demonstração, com \mathbf{s} tendo norma unitária, é mostrado que $\alpha = \langle \mathbf{s}, \alpha \mathbf{s} \rangle = (f_{2p}(\mathbf{s}))^{2p}$.

Desta forma, para o caso do equalizador sub-parametrizado onde $P_T \neq I$, (A.8) pode ser escrita como $\mathbf{s}^{\odot(2p-1)} - \alpha \mathbf{s} = \mathbf{b}$ para algum $\mathbf{b} \in T^\perp$. Como \mathbf{b} não é conhecido *a priori*, não é possível se encontrar uma solução em forma fechada para \mathbf{s} . Busca-se portanto, resolver o problema de forma iterativa, como mostrado em (2.71).

Bibliografia

R. R. F. Attux, “Sobre Dinâmica Caótica e Convergência em Algoritmos de Equalização Autodidata”, Tese de Mestrado, UNICAMP, **2001**

S. Bellini, “Busgang Techniques for Blind Deconvolution and Equalization”, in *Blind Deconvolution*, Simon Haykin (Ed.), Prentice Hall, **1994**

A. Benveniste, M. Goursat, G. Ruget, “Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications”, *IEEE Transactions on Automatic Control*, Vol. AC-25, No 3, pp. 385-399, Junho **1980**

A. L. Brandão, “Quality Assessment for Pre-Detection Diversity Switching”, *Anais IEEE/ Personal Indoor Mobile Radio Communications (PIMRC)*, pp. 577-581, **1995**

F. R. P. Cavalcanti, A. L. Brandão, J. M. T. Romano, “A Generalized Constant Modulus Algorithm for Blind Equalization”, *Anais Global Telecommunication Conference (GLOBECOM)*, **1998**

F. R. Cavalcanti, “Antenas Inteligentes & Processamento Espaço-Temporal para Sistemas de Comunicação sem Fio”, Tese de Doutorado, UNICAMP, **1999**

R. Cusani, Laurenti, “A Convergence analysis of the CMA blind equalizer”, *IEEE Transactions on Communications*, Vol. 43, No 2-4 Parte 3, pp. 1304-1307, Fev. - Abril **1995**

Z. Ding, Ye Li, “Blind Equalization and Identification”, 1ª edição, Marcel Dekker, Inc., Signal Processing and Communications Series, **2001**

- Z. Ding, R. A. Kennedy, B. D. Anderson, C. R. Jonhson, "Ill- Convergence of Godard Blind Equalizers in Data Communication Systems", *IEEE Transactions on Communications*, Vol. 39, No 9, pp. 1313-1327, Set. **1991**
- Z. Ding, R. A. Kennedy, B. D. Anderson, C. R. Jonhson, "Local Convergence of Sato Blind Equalizer and Generalizations Under Practical Constraints", *IEEE Transactions on Information Theory*, Vol. 39, No 1, pp-129-144, Jan. **1993**
- Z. Ding, C. R. Jonhson, R. A. Kennedy, "On the (Non)Existence of Undesirable Equilibria of Godard Blind Equalizers", *IEEE Transactions on Signal Processing*, Vol. 40, No 10, pp. 2425-2432, Out. **1992**
- Z. Ding, C. R. Johnson, "On the Non-vanishing Stability of Undesirable Equilibria for FIR Godard Blind Equalizers", *IEEE Transactions on Signal Processing*, Vol.41, pp. 1940-1944, Maio **1993**.
- Z. Ding, R. A. Kennedy, "On the Whereabouts of Local Minima for Blind Adaptive Equalizers", *IEEE Trans. Circuits and Systems II*, Vo. 39, pp 119-123, Fev. **1992**
- D. Donoho, "On Minimum Entropy Deconvolution", in *Applied Time Series Analysis II*, D. F. Findley, Ed. New York: Academic, **1981**
- G. Foschini, "Equalization Without Altering or Detecting Data", *Bell Syst. Tech. J.*, Vol. 64, No 8, pp. 1885-1911, Out. **1985**
- D. N. Godard, "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems", *IEEE Transactions on Communications*, Vol. COM-28, No 11, pp. 1867-1875, Nov. **1980**,
- R. Godfrey, F. Rocca, "Zero Memory Non-linear Deconvolution", *Geophysical Prospecting*, Vol. 29, pp. 189-228, **1981**
- M. Gu, L. Tong, "Domains of Attraction of Shalvi-Weisntein Receivers", *IEEE Trans. on Signal Processing*, Vol. 49, No 7, pp. 1397-1408, Julho **2001**
- M. Gu, L. Tong, "Geometrical Characterizations of Constant Modulus Receivers", *IEEE Trans. Signal Processing*, Vol. 47, No 10, pp. 2745-2756, Out. **1999**
- S. Haykin, "Adaptive Filter Theory", 3rd edition, Prentice Hall, **1996**

- C. R. Johnson, P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, R. A. Casas, "Blind Equalization Using the Constant Modulus Criterion: A Review", *Proceedings of the IEEE*, Vol. 86, No 10, pp. 1927-1950, Out. **1998**
- Ye Li, Z. Ding, "Convergence Analysis of Finite Length Blind Adaptive Equalizers", *IEEE Transactions on Signal Processing*, Vol. 43, No 9, pp. 2121-2129, Set. **1995**
- Ye Li, K.J. Ray, Z. Ding, "Length- and Cost-Dependent Local Minima of Unconstrained Blind Channel Equalizers", *IEEE Trans. Signal Processing*, Vol. 44, No 11, pp. 2726-2735, Nov. **1996**
- R. Lucky, J. Salz, E. Weldon, "Principles of Data Communication", MacGraw-Hill, Nova York, **1968**
- D.G. Luenberger, "Optimization by Vector Space Methods", Wiley, **1969**
- O. Macchi, E. Eweda, "Convergence Analysis of Self-Adaptive Equalizers", *IEEE Transactions on Information Theory*, Vol. IT-30, No 2, Special Issue on Linear Adaptive Filtering, pp. 161-176, Mar. **1984**
- J. E. Mazo, "Analysis of Decision Directed Equalizer Convergence", *Bell Syst. Tech. J.*, Vol 59, pp. 1557-1876, Dez. **1980**
- M. Mboup, P. Regalia, "A Gradient Search Interpretation of the Super-Exponential Algorithm", *IEEE Trans. on Information Theory*, Vol. 46, No 7, pp. 2731-2734, Nov. **2000**
- M. Mboup, P. Regalia, "On The Equivalence Between the Super-Exponential Algorithm and a Gradient Search Method", *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, Vol. 5, pp. 2643 -2646, Phoenix, **1999**
- C. L. Nikias, A. P. Petropulu, "Higher-order Spectra Analysis: A Nonlinear Signal Processing Framework", Prentice Hall, **1993**
- J. G. Proakis, "Digital Communications", McGraw-Hill, 3^a edição, **1995**
- P. A. Regalia, "On the Equivalence Between the Godard and Shalvi Weinstein Schemes of Blind Equalization", *Signal Processing*, Vol. 73, No 1-2, pp. 185-190, Jan. **1999**

- P. A. Regalia, M. Mboup, "Undermodeled Equalization: A Characterization of Stationary Points for a Family of Blind Criteria", *IEEE Transactions on Signal Processing*, Vol. 47, No 3, pp. 760-770, Mar. **1999**
- H. Robbins, S. Monro, "A Stochastic Approximation Method", *The Annals of Mathematical Statistics*, Vol. 22, pp. 400-407, **1951**
- C. A. F. Rocha, J. M. T. Romano, "Equalização Autodidata: Fundamentos, Novas Propostas e Perspectivas", *Revista Brasileira de Telecomunicações*, Vol. 10, No 1, pp 25-36, Dez. **1995**
- Y. Sato, "A method for self recovering equalization", *IEEE Transactions on Communications*, Vol. COM-23, No 6, pp. 679-682, Junho **1975**
- O. Shalvi, E. Weinstein, "New Criteria for Blind Deconvolution of Nonminimum Phase Systems (Channels)", *IEEE Transactions on Information Theory*, Vol. 36, No 2, pp. 312-321, Mar. **1990**
- O. Shalvi, E. Weinstein, "Super-Exponential Methods for Blind Deconvolution", *IEEE Transactions on Information Theory*, Vol. 39, No 2, pp. 504-519, Mar. **1993**
- O. Shalvi, E. Weinstein, "Universal Methods for Blind Deconvolution", in *Blind Deconvolution*, Simon Haykin (Ed.), Prentice Hall, **1994**
- R. A. Soni, T. A. Schirtzinger, W. K. Jenkins, "Learning Strategies for Adaptive Equalizers Using the Constant Modulus Error Criterion", *International Journal of Adaptive Control and Signal Processing*, Vol. 12, pp. 97-116, **1998**
- J. Treichler, B. Agee, "New Approach to Multipath Correction of Constant Modulus Signals", *IEEE Trans. Acoust., Speech and Signal Processing*, Vol. ASSP-31, No 2, pp. 459-472, Abril **1983**
- J.K. Tugnait, "Comments on 'New Criteria for Blind Deconvolution of Nonminimum Phase Systems (Channels)", *IEEE Trans. Information Theory*, Vol. 38, pp. 210-213, Jan. **1992**
- R. A. Wiggins, "Minimum Entropy Deconvolution", *Geoeexplorations*, Vol. 16, pp. 21-35, **1978**

H. Zeng, L. Tong, C.R. Johnson, "An Analysis of Constant Modulus Receivers", *IEEE Trans. Signal Processing*, Vol. 47, No 11, pp. 2990-2999, **1999**

H. Zeng, L. Tong, C.R. Johnson, "Relationships Between the Constant Modulus and Wiener Receivers", *IEEE Trans. Information Theory*, Vol. 44, No 4, pp. 1523-1538, Julho **1998**